



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO  
DEPARTAMENTO DE MATEMÁTICA  
Mestrado Profissional em Matemática em Rede Nacional



Bruno Ramos Sales Mendes de Barros

**A TEORIA DE RESPOSTA AO ITEM COMO INSTRUMENTO DE  
ELABORAÇÃO DE TESTES PARA PROFESSORES DE MATEMÁTICA**

Recife

2022





UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO  
DEPARTAMENTO DE MATEMÁTICA  
Mestrado Profissional em Matemática em Rede Nacional



Bruno Ramos Sales Mendes de Barros

**A TEORIA DE RESPOSTA AO ITEM COMO INSTRUMENTO DE  
ELABORAÇÃO DE TESTES PARA PROFESSORES DE MATEMÁTICA**

Dissertação de mestrado apresentada ao Departamento de Matemática da Universidade Federal Rural de Pernambuco como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador: Prof. Dr. Fabiano Barbosa Mendes da Silva

Recife  
2022

Dados Internacionais de Catalogação na Publicação  
Universidade Federal Rural de Pernambuco  
Sistema Integrado de Bibliotecas  
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

---

B277t

Barros, Bruno Ramos Sales Mendes de

A Teoria de Resposta ao Item como instrumento de elaboração de testes para professores de matemática / Bruno Ramos Sales Mendes de Barros. - 2022.

150 f. : il.

Orientador: Fabiano Barbosa Mendes da Silva.

Inclui referências e apêndice(s).

Dissertação (Mestrado) - Universidade Federal Rural de Pernambuco, Programa de Mestrado Profissional em Matemática (PROFMAT), Recife, 2022.

1. Teoria de Resposta ao Item. 2. EIRT. 3. Escalas de habilidade. 4. Modelos logísticos unidimensionais. 5. Banco de itens. I. Silva, Fabiano Barbosa Mendes da, orient. II. Título

CDD 510

---

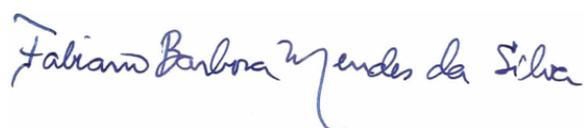
BRUNO RAMOS SALES MENDES DE BARROS

**A Teoria de Resposta ao Item como Instrumento de Elaboração de Testes para Professores de Matemática.**

*Trabalho apresentado ao Programa de Mestrado Profissional em Matemática – PROFMAT do Departamento de Matemática da UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO, como requisito parcial para obtenção do grau de Mestre em Matemática.*

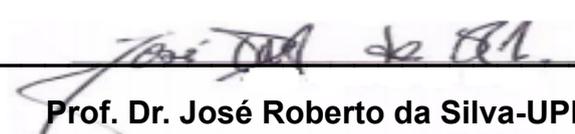
Aprovado em 04/03/2022

BANCA EXAMINADORA



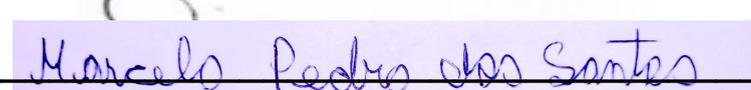
---

**Prof. Dr. Fabiano Barbosa Mendes da Silva** (Orientador) – UFRPE



---

**Prof. Dr. José Roberto da Silva-UPE**



---

**Prof. Dr. Marcelo Pedro dos Santos** – PROFMAT/UFRPE



*Dedicado à estrela que mais brilha no universo: minha filha Júlia.*



## **AGRADECIMENTOS**

A minha filha Júlia, por todo amor e compreensão, principalmente nos momentos da minha ausência por causa da dedicação a esse mestrado.

Aos meus pais, irmãos e a Maria José, por todo apoio durante essa jornada de estudos e ao longo da vida.

Aos meus amigos de turma do Profmat/UFRPE, pela parceria e companheirismo. Em especial, ao meu amigo Cícero José, companheiro das longas viagens semanais até a universidade e que sempre esteve presente nos momentos que eu precisei durante o curso.

Ao meu orientador, professor Fabiano Mendes, pela ajuda, confiança, compreensão, ensinamentos e solicitude na condução deste trabalho.

A todos os professores do Profmat/UFRPE, em especial ao apoio dado pelos professores Ross Nascimento e Anete Soares.



*“Quando a educação não é libertadora, o sonho do oprimido é ser o opressor.”*

Paulo Freire



## DECLARAÇÃO

Eu, **Bruno Ramos Sales Mendes de Barros** declaro, para devidos fins e efeitos, que a dissertação sob título **A Teoria de Resposta ao Item como instrumento de elaboração de testes para professores de matemática**, entregue como Trabalho de Conclusão de curso para obtenção do título de mestre, com exceção das citações diretas e indiretas claramente indicadas e referenciadas, é um trabalho original. Eu estou consciente que a utilização de material de terceiros incluindo uso de paráfrase sem a devida indicação das fontes será considerado plágio, e estará sujeito à processos administrativos da Universidade Federal Rural de Pernambuco e sanções legais. Declaro ainda que respeitei todos os requisitos dos direitos de autor e isento a Pós-graduação PROFMAT/UFRPE, bem como o professor orientador **Fabiano Barbosa Mendes da Silva**, de qualquer ônus ou responsabilidade sobre a sua autoria.

Recife, 28 de abril de 2022.

*Bruno Ramos S.M. de Barros*

---

Bruno Ramos Sales Mendes de Barros

## RESUMO

Este estudo tem por objetivo apresentar a Teoria da Resposta ao Item (TRI) aos professores que lecionam matemática no ensino básico, abordando os conceitos fundamentais que a embasam. Para tanto, foi realizada uma pesquisa a partir de um questionário abordando alguns aspectos sobre a TRI, na qual foi constatado que esses professores possuíam um baixo nível de conhecimento acerca do tema e que isso poderia contribuir negativamente na sua prática pedagógica e em ações para melhoria da educação. Este material apresenta as principais definições, conceitos e aspectos gerais sobre a TRI, tais como a abordagem histórica, os modelos logísticos unidimensionais mais utilizados em testes e ainda construção da curva de probabilidade de acerto em um item, caracterizando o que cada parâmetro significa. São também abordados temas como a estimação das habilidades e dos parâmetros dos itens, a construção e a interpretação das escalas de habilidades, e ainda o conceito de banco de itens. O *software* EIRT, apresentado em detalhes neste estudo, possibilita a geração e a análise dos dados estatísticos de um teste baseado na TRI, e, juntamente aos elementos teóricos aqui abordados, pode servir como ponto de partida para que os professores revisem suas práticas pedagógicas e proponham ações que visem à melhoria do ensino e aprendizagem em suas escolas ou redes de ensino.

**Palavras-chave:** Teoria de Resposta ao Item, EIRT, escalas de habilidade, modelos logísticos unidimensionais, banco de itens.

## ABSTRACT

This study aims to present the Item Response Theory (IRT) to teachers who teach mathematics in basic education, addressing the fundamental concepts that underlie it. Therefore, a survey was carried out from a questionnaire addressing some aspects about IRT, in which it was found that these teachers had a low level of knowledge about the subject and that this could negatively contribute to their pedagogical practice and actions to improve the education. This material presents the main definitions, concepts and general aspects of the IRT, such as the historical approach, the one-dimensional logistic models most used in tests and also the construction of the probability curve of success in an item, characterizing what each parameter means. Topics such as the estimation of skills and item parameters, the construction and interpretation of skill scales, and the concept of an item bank are also addressed. The EIRT software, presented in detail in this study, enables the generation and analysis of statistical data from a test based on the IRT, and, along with the theoretical elements discussed here, can serve as a starting point for teachers to review their pedagogical practices and propose actions aimed at improving teaching and learning in their schools or teaching networks.

**Keywords:** Item Response Theory, EIRT, skill scales, one-dimensional logistic models, item bank.



## LISTA DE FIGURAS

Figura 1 - Curva Característica do Item (CCI) .....	37
Figura 2 - Equivalências entre as escalas (0,1), Enem (500,100) e Saeb (250,50).....	40
Figura 3 – Equivalência da Curva Característica do Item (CCI) nas escalas (0,1) e (500, 100).....	42
Figura 4 - Curva Característica do Item (CCI) para o modelo 3PL.....	43
Figura 5 – Variação do parâmetro $c$ da curva característica de três itens.....	45
Figura 6 - Parâmetro da dificuldade de três itens para diferentes valores de $b$ .....	46
Figura 7 – CCI de um item com índice de discriminação (parâmetro $a$ ) negativo.....	48
Figura 8 -Curvas características dos itens para diferentes valores do parâmetro de discriminação (parâmetro $a$ ).....	49
Figura 9 - Curva característica do item 5 conforme o modelo 3PL.....	52
Figura 10 - Curva características dos 5 itens do teste hipotético conforme o modelo 3PL.....	53
Figura 11 - Curvas de Informação dos itens A e B.....	55
Figura 12 – Função de Informação de cinco itens e Função de Informação do teste formado por esses itens.....	56
Figura 13 - Matriz 5 x 10 das respostas de cinco sujeitos em um teste de dez itens.....	57
Figura 14 – Trecho da matriz de respostas dos 100 sujeitos submetidos a 20 itens.....	65
Figura 15 – Trecho da matriz de respostas dos 100 sujeitos submetidos a 20 itens na forma dicotômica após a correção.....	65
Figura 16 – Curvas características do item 8 e 9 gerados no <i>software</i> EIRT.....	72
Figura 17 – Curvas características dos itens 5 e 16 geradas no <i>software</i> EIRT.....	74
Figura 18 – Curva característica do item 1 gerado pelo <i>software</i> EIRT.....	75
Figura 19 - Curvas características dos 20 itens gerados pelo <i>software</i> EIRT.....	75
Figura 20 - Curva de Informação dos itens 18 e 20 gerado pelo EIRT.....	76
Figura 21 - Curvas de Informação dos 20 itens do teste gerado pelo EIRT.....	76
Figura 22 - Relações entre posições de itens e proficiências em uma escala hipotética.....	83
Figura 23 – Trecho do mapa de itens do Enem na área de Matemática e suas tecnologias.....	87
Figura 24 - Curvas Características (CCIs) dos itens âncoras 13 e 16.....	88
Figura 25 – Posicionamento dos itens âncoras 13 e 16 e dos sujeitos 85 e 86 na escala de habilidade.....	90
Figura 26 – Evolução das proficiências médias dos estudantes em matemática no SAEB de 1995 a 2017.....	91
Figura 27 - Etapas do processo de elaboração de um item.....	94
Figura 28 – Estatísticas de um item da prova do 5º ano no SARESP 2011, e ancorada no nível Básico (175 a < 225).....	99
Figura 29 - Distribuição de professores por sexo (esquerda) e por sexo X rede de ensino (direita).....	103
Figura 30 - Quantidade e percentual de professores de matemática por rede de ensino.....	104
Figura 31 - Etapas do ensino básico (à esquerda) e tempo de ensino (à direita) dos professores de matemática.....	105
Figura 32 - Formação dos professores que ensinam matemática.....	106
Figura 33 - Proporção sobre o conhecimento dos professores de matemática em relação a TRI.....	107
Figura 34 - Professores x Estudo da TRI durante a formação.....	109
Figura 35 - Opinião sobre a melhor forma de mensurar o desempenho dos estudantes em avaliações externas em larga escala.....	110
Figura 36 - Conhecimento do professor sobre a função probabilidade de acerto no item da TRI.....	113
Figura 37 - Conhecimento do professor sobre a variável ( $\theta$ ) e parâmetros dos itens ( $a$ , $b$ e $c$ ).....	114
Figura 38 - Interesse do professor de matemática em entender como se calcula a probabilidade de acerto em um item, incluindo a compreensão do significado da variável e dos parâmetros.....	115
Figura 39 - Conhecimento do professor de matemática sobre a Curva Característica do Item.....	116
Figura 40 - Conhecimento do professor de matemática sobre as estimativas da habilidade e parâmetros dos itens na TRI.....	117

Figura 41 - Conhecimento do professor de matemática sobre a interpretação dos resultados das avaliações com uso da TRI. ....	118
Figura 42 - Coerência pedagógica nas respostas dos participantes em um teste. ....	119
Figura 43 - Percentual de professores em relação a sua compreensão da mudança de escala na TRI. ....	120
Figura 44 - Conhecimento sobre os <i>softwares</i> utilizados na TRI. ....	121
Figura 45 - Percentual dos professores que responderam que entender sobre a TRI contribui para melhoria da sua prática profissional na TRI. ....	122
Figura 46 – Escala de habilidade de Matemática e suas tecnologias de níveis entre 742 e 767 para uma certa prova do Enem. ....	131
Figura 47 – Comparação do padrão de respostas e notas/proficiência de dois examinandos na escala do Enem - (500,100). ....	132
Figura 48 - Relação entre o número de acertos e a nota no Enem 2011 em Matemática e suas tecnologias. ....	133
Figura 49 – Proficiências médias dos estudantes em Matemática e suas tecnologias (2009-2018). ....	136

## LISTA DE TABELAS

Tabela 1 - Classificação, valores e percentual esperado para o parâmetro de dificuldade na TRI. ....	47
Tabela 2 - Classificação para o parâmetro de discriminação (parâmetro $a$ ) de um item pela TRI. ....	49
Tabela 3 - Parâmetros dos cinco itens do teste hipotético estimados a partir do modelo logístico 3PL. ....	50
Tabela 4 - Probabilidade de acerto em cada um dos cinco itens para diferentes níveis de habilidades utilizando o modelo logístico 3PL ( $D=1$ ). ....	51
Tabela 5 - Respostas de cinco sujeitos em um teste de dez itens. ....	57
Tabela 6 - Parâmetros dos 20 itens conforme a TRI estimados pelo EIRT. ....	71
Tabela 7 - Classificação dos 20 itens conforme o parâmetro de discriminação (parâmetro $a$ ). ....	73
Tabela 8 - Classificação dos 20 itens conforme o índice de dificuldade (parâmetro $b$ ) na TRI. ....	73
Tabela 9 - Comparação dos escores (TCT) e habilidades - $\theta$ (TRI) estimados pelo <i>software</i> EIRT. ....	77
Tabela 10 - Habilidades descritas pela TRI nas escalas (0,1), Enem e Saeb, conforme a posição do examinando. ....	79
Tabela 11 - Parâmetros dos itens 13 e 16. ....	88
Tabela 12 - Notas mínimas e máximas em Matemática e suas Tecnologias no Enem de 2009 a 2020. ....	134
Tabela 13 – Quantidade de participantes, no Enem 2015, em Matemática e suas Tecnologias, por faixa de proficiência. ....	135

## LISTA DE QUADROS

Quadro 1 - Comparação entre a TCT e TRI.....	28
Quadro 2- Métodos de estimação para os parâmetros dos itens e habilidades. ....	59
Quadro 3 - Alguns programas de computador para estimação na TRI.....	62
Quadro 4 – Diretrizes para o relatório das estatísticas da TRI e TCT com EIRT.....	66
Quadro 5 - Trecho das descrições dos níveis da escala de matemática Saeb – 5º ano do ensino fundamental .....	85
Quadro 6 – Níveis e descrição das habilidades de uma escala hipotética - (0,1).....	90
Quadro 7 – Diretrizes para construção e interpretação de uma escala de habilidade. ....	92
Quadro 8 - Trecho da matriz de referência de Matemática do Saeb: Temas e descritores para o 5º ano do ensino fundamental. .....	95
Quadro 9 - Trecho da matriz de referência de Matemática e suas tecnologias do Enem.....	96

## LISTAS DE SIGLAS

**2PL** - Modelo Logístico Unidimensional de Três Parâmetros

**3PL** - Modelo Logístico Unidimensional de Três Parâmetros

**ANEB** - Avaliação Nacional da Educação Básica

**ANA** - Avaliação Nacional da Alfabetização

**ANRESC** - Avaliação Nacional do Rendimento Escolar

**BIB** – Bloco Incompletos Balanceados

**BNI** – Banco Nacional de Itens

**CCI** – Curva Característica do Item

**CII** – Curva de Informação do Item

**EAP** – Esperado a Posteriori (em inglês: *Expected a Posteriori*)

**ENEM** - Exame Nacional do Ensino Médio

**ENCCEJA** - Exame Nacional para Certificação de Competências de Jovens e Adultos

**INEP** - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

**LDB** - Lei de Diretrizes e Bases da Educação Nacional

**MAP** - Máximo a Posteriori.

**MV** – Máxima Verossimilhança

**MVC** - Máxima Verossimilhança Conjunta

**MVM** - Máxima Verossimilhança Marginal

**PISA** - Programa Internacional de Avaliação de Estudantes

**SAEB** - Sistema de Avaliação da Educação Básica

**SAEP** - Sistema de Avaliação Educacional de Pernambuco

**SARESP** - Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo

**SAVEAL** - Sistema de Avaliação Educacional de Alagoas

**TCT** – Teoria Clássica dos Testes

**TOEFL** - *Test of English as a Foreign Language*

**TRI** – Teoria de Resposta ao Item

**WML** - Máxima Verossimilhança Ponderada (em inglês: *Weighted Maximum Likelihood*)

## SUMÁRIO

1	INTRODUÇÃO.....	22
1.1	Justificativa.....	25
1.2	Objetivos .....	25
2	A TEORIA DE RESPOSTA AO ITEM – TRI .....	27
2.1	Uma visão geral sobre a psicometria aplicada a avaliações .....	27
2.2	Um pouco de história sobre a TRI.....	30
2.3	Características da TRI .....	34
2.3.1	A unidimensionalidade.....	34
2.3.2	A independência local .....	35
2.4	Modelos logísticos unidimensionais da Teoria de Resposta ao Item .....	36
2.5	Transformações de escalas .....	39
2.6	Modelo logístico de 3 parâmetros (3PL).....	43
2.6.1	Características gerais dos parâmetros dos itens.....	44
2.6.1.1	Acerto ao acaso (parâmetro $c$ ).....	44
2.6.1.2	Dificuldade de um item (parâmetro $b$ ) .....	45
2.6.1.3	Discriminação de um item (parâmetro $a$ ) .....	47
2.7	Probabilidade de acerto em um item e a construção da Curva Característica do Item (CCI).....	50
2.8	Função de Informação .....	53
3	ESTIMAÇÃO DOS PARÂMETROS DOS ITENS E HABILIDADES NA TRI .....	57
3.1	Método de Máxima Verossimilhança para estimação dos parâmetros dos itens e das habilidades.....	59
3.2	Recursos computacionais na TRI .....	61
3.2.1	O <i>software</i> EIRT.....	64
4	ALGUMAS APLICAÇÕES DA TRI NO DESENVOLVIMENTO DE TESTES .....	82
4.1	Interpretação e construção das escalas de habilidades .....	82
4.2	Banco de itens para avaliações educacionais .....	93
5	O CONHECIMENTO DO PROFESSOR DE MATEMÁTICA ACERCA DA TEORIA DE RESPOSTA AO ITEM .....	102
5.1	Metodologia.....	102
5.2	Resultados e discussões.....	103
5.2.1	Análise das informações gerais sobre o professor de matemática.....	103
5.2.2	Análise das respostas dos professores sobre o seu conhecimento acerca da Teoria de Resposta ao Item (TRI).....	106
6	CONCLUSÃO E TRABALHOS FUTUROS .....	123
	REFERÊNCIAS BIBLIOGRÁFICAS .....	126
	APÊNDICE A - Coerência pedagógica das respostas no Enem .....	131

APÊNDICE B – Instalação do software EIRT.....	137
APÊNDICE C – Matriz das Respostas dos 100 sujeitos aplicadas ao teste hipotético.....	138
APÊNDICE D – Matriz dicotômicas contendo as respostas dos 100 sujeitos ao teste.....	141
APÊNDICE E – Questionário: O conhecimento do professor de matemática acerca da Teoria de Resposta ao Item .....	144

## 1 INTRODUÇÃO

As avaliações educacionais, em geral, utilizam testes como instrumentos de medição a fim de obter informações de modo a construir diagnósticos e subsídios para a realização ou manutenção de ações na área educacional. Segundo Klein e Fontanive (1995), essas avaliações devem ser realizadas também para fornecer um contínuo monitoramento do sistema educacional com o objetivo de detectar os efeitos positivos ou negativos de políticas adotadas.

Luckesi (2008, p.76) afirma que o ato de avaliar

implica coleta, análise e síntese dos dados que configuram o objeto da avaliação, acrescido de uma atribuição de valor ou qualidade, que se processa a partir da comparação da configuração do objeto avaliado com um determinado padrão de qualidade previamente estabelecido para aquele tipo de objeto. O valor ou qualidade atribuídos ao objeto conduzem a uma tomada de posição a seu favor ou contra o objeto, ato ou curso de ação, a partir do valor ou qualidade atribuídos, conduz a uma decisão nova: manter o objeto como está ou atuar sobre ele.

Nesse sentido, uma das questões fundamentais no contexto educacional é determinar como realizar essas comparações que, em geral, são alcançadas a partir de testes. Um teste, segundo Cohen, Swerdlik e Sturman (2014, p.6), pode ser definido simplesmente como “um dispositivo ou procedimento que visa medir uma variável [latente]”. Porém, é mais complexo medir quando essa variável expressa fenômenos que não podem ser observados diretamente, como, por exemplo, autoestima, personalidade e depressão. Assim, como seria possível medir a inteligência ou alguma habilidade de uma pessoa?

Questões como essa têm ocupado gerações de pesquisadores no campo da mensuração psicológica - incluindo a área educacional- denominada Psicometria. A psicometria pode ser definida em termos gerais “como o conjunto de métodos, técnicas e teorias envolvidas na medição das variáveis psicológicas” (MUÑIZ, 2018, p.15, tradução nossa). Desse modo, pode-se afirmar que é nessa área que se concentram os estudos sobre a construção e interpretação de testes com qualidade estatisticamente comprovada.

Urbina (2007, p.14-15) afirma que os testes, como os realizados na área educacional, são ferramentas criadas para ajudar na realização de inferências a respeito de indivíduos ou grupos, e, quando usados de forma adequada, podem ser indispensáveis no processo avaliativo. Essa autora enfatiza que os testes “sempre são um meio para alcançar um fim, e nunca um fim

em si mesmos”. Ademais, os resultados desses testes podem ter grande relevância na vida das pessoas; portanto, é fundamental que os testes utilizados atendam requisitos de qualidade psicométrica demonstráveis.

Normalmente quando o instrumento utilizado na avaliação é o teste, tradicionalmente os resultados obtidos apenas levam em conta a soma das pontuações das questões acertadas, cuja análise central está no teste como um todo. Essa abordagem está relacionada com a Teoria Clássica dos Testes (TCT), que ainda é muito utilizada em provas escolares e em alguns vestibulares. No entanto, nessa perspectiva, os resultados dependem tanto do nível das questões aplicadas quanto da amostra dos respondentes a qual o teste foi aplicado, e isso constitui um problema dentro dessa teoria, pois inviabiliza, por exemplo, a comparação entre os sujeitos que realizarem diferentes testes em uma mesma avaliação.

Por outro lado, a Teoria de Resposta ao Item (TRI) surgiu para preencher as lacunas observada na TCT. Na TRI, a análise central é focada no item e não no teste como um todo, como observado na TCT. Nesse sentido, a TRI analisa a probabilidade de um sujeito acertar cada item de um teste em função da sua habilidade e das características desse item; e é a partir dessas análises -realizadas conforme os postulados dessa teoria e apoiada em alguns métodos estatísticos - que se pode encontrar o valor mais verossímil para a habilidade de um sujeito que realizou um teste, bem como pode-se estimar os parâmetros que caracterizam cada item.

No Brasil, desde a década de 1990, as avaliações externas em larga escala têm se mostrado importantes por proporcionarem um *feedback* necessário para tomada de decisões educacionais que conduzem ao acerto de ações e direcionamento de políticas públicas (VIANNA, 2003). Assim, os resultados nessas avaliações podem fornecer indícios importantes para que se reflita sobre o desenvolvimento do trabalho no âmbito escolar, pois “informam sobre os resultados educacionais de escolas e redes de ensino a partir do desempenho dos alunos em testes ou provas padronizadas que verificam se estes aprenderam o que deveriam ter aprendido, permitindo inferências sobre o trabalho educativo” (BLASIS; FALSARELLA; ALAVARSE, 2013, p. 12).

Nesse sentido, mesmo que nas últimas décadas os resultados nas avaliações em larga escala tenham se mostrado de grande relevância para a educação, muitos profissionais de escolas e de secretarias de educação desconhecem sua importância ou não compreendem sobre o seu funcionamento e, desse modo, os dados gerados por essas avaliações não são utilizados

para propor ações com intuito de melhorar o ensino-aprendizagem nas escolas e redes de ensino, o que pode configurar um problema no progresso da educação.

Para tanto, esse estudo objetiva trazer à luz os conceitos fundamentais sobre a TRI que fazem parte dessas avaliações - sem ter a pretensão de esgotar o tema - e apresentar a importância dessa teoria como instrumento na elaboração de testes, sejam eles de menor ou maior escala. Além disso, este trabalho também tem como objetivo investigar o nível de conhecimento dos professores de matemática, das redes de ensino em Alagoas, acerca da Teoria de Resposta ao Item (TRI).

Dessa forma, o trabalho foi estruturado em capítulos da seguinte maneira:

**Capítulo 2 – A TEORIA DE RESPOSTA AO ITEM – TRI** – explora as principais diferenças entre a TRI e TCT, traz aspectos históricos sobre a TRI e apresenta as características e suposições dessa teoria. São ainda apresentados os modelos logísticos de 1, 2 e 3 parâmetros, enfatizando as características de cada parâmetro, o conceito de escalas de habilidades e a transformação entre escalas distintas. Também é explorada a ideia do cálculo da probabilidade de acerto no item e a construção da curva característica do item. Por fim, esse capítulo versa ainda sobre a importância das funções e curvas de informação para os testes.

**Capítulo 3 - ESTIMAÇÃO DOS PARÂMETROS DOS ITENS E HABILIDADES NA TRI** – trata de como as estimativas dos parâmetros dos itens e das habilidades são calculados. Para tanto, o método apresentado nesse capítulo para o cálculo dessas estimativas foi o de Máxima Verossimilhança. Por fim, é apresentada uma subseção sobre a utilização de recursos computacionais para a TRI, onde é explorado o *software* EIRT, que é um suplemento (*add-in*) gratuito do Excel e utilizado para realizar as estatísticas da TRI (e também da TCT), indicado as etapas (passo-a-passo) de como utilizá-lo quando se quer analisar os seus próprios testes.

**Capítulo 4 - ALGUMAS APLICAÇÕES DA TRI NO DESENVOLVIMENTO DE TESTES** – apresenta o uso da TRI em diversas etapas da avaliação, como na construção e interpretação das escalas de habilidades, que configura uma etapa importante quando se quer compreender os resultados de uma avaliação de forma qualitativa ou ainda construir uma escala de habilidades para que se possa interpretar os resultados do seu próprio teste. Ademais, nesse

capítulo é mostrada a importância da utilização dessa teoria para avaliar itens de qualidade com intuito de utilizá-los para compor um banco de itens em avaliações educacionais.

**Capítulo 5 - O CONHECIMENTO DO PROFESSOR DE MATEMÁTICA ACERCA DA TEORIA DE RESPOSTA AO ITEM** – trata da pesquisa sobre o conhecimento dos professores de matemática, que atuam nas redes de ensino do estado de Alagoas, acerca da TRI. A pesquisa é exploratória-quantitativa e foi realizada a partir de um questionário contendo questões sobre informações gerais do professor pesquisado e também sobre o seu conhecimento acerca de tópicos básicos da TRI. A partir das respostas dadas nesse questionário são apresentados os resultados e as análises para cada questão da pesquisa.

## **1.1 Justificativa**

A realização deste trabalho ancora no fato de que a Teoria de Resposta ao Item (TRI) é um tema essencial na educação escolar, posto que as avaliações externas em larga escala, como Saeb e Enem, utilizam-se da TRI para estimar as proficiências dos estudantes e os parâmetros que caracterizam os itens do teste. Nesse sentido, é necessário que os professores que atuam no ensino básico possuam certo conhecimento sobre os conceitos fundamentais, seja para interpretar resultados de uma avaliação ou mesmo construir testes utilizando essa teoria.

Com o objetivo de apontar o nível de conhecimento dos professores sobre a TRI, foi realizada uma pesquisa com os professores de matemática da educação básica, que atuam nas redes de ensino de Alagoas, cujos resultados indicaram a necessidade de aprofundamento sobre o tema. Isto serviu como elemento motivador para a realização deste estudo, cuja centralidade reside na apresentação e discussão dos principais conceitos e elementos da TRI e na utilização do *software* EIRT para determinação dos dados estatísticos relativos a essa teoria em testes.

## **1.2 Objetivos**

### **Objetivos gerais**

- Apresentar os conceitos fundamentais sobre a TRI, bem como exemplificar a sua aplicação em algumas avaliações.

### Objetivos específicos

- Investigar o conhecimento que os professores de matemática, das redes de ensino de Alagoas, têm acerca da Teoria de Resposta ao Item -TRI.
- Destacar as principais diferenças entre a Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI).
- Utilizar o *software* EIRT (suplemento do Excel) para calcular as estatísticas básicas relativas a TRI.

## 2 A TEORIA DE RESPOSTA AO ITEM – TRI

Construtos psicológicos e educacionais como inteligência, habilidade, depressão, personalidade, autoestima, capacidade matemática e compreensão da leitura não podem ser observados ou medidos no mesmo sentido físico em que (digamos) se pode medir a altura e peso de uma pessoa (BAKER, 2001). Como então os construtos psicológicos e educacionais são medidos pelos testes? Diferentes modelos e métodos foram propostos em resposta a esta importante pergunta e essa questão ocupa a área de estudo da Psicometria, que representa um ramo da Psicologia, e se concentra nos problemas de medição utilizando a Estatística como pilar básico para a elaboração de teorias, e para o desenvolvimento de métodos e técnicas específicas para medição. Nesse sentido este capítulo terá como objetivo apresentar a Teoria de Resposta ao Item, que revela ser uma importante teoria psicométrica quando se trata de avaliações na área educacional.

### 2.1 Uma visão geral sobre a psicometria aplicada a avaliações

A utilização de instrumentos de medida, como os testes psicométricos, representa uma importante forma de avaliar os constructos psicológicos relacionados à educação. De acordo com Pasquali (2011) a Psicometria procura justificar o sentido das respostas dadas pelos indivíduos aos itens de um teste. O modelo psicométrico mais tradicional e comum, para estabelecer esse contexto, é a Teoria Clássica dos Testes (TCT), cujo objetivo principal é explicar o resultado obtido através da soma das pontuações de uma série de itens acertados, no que é denominado *escore total*.

Em geral, quando se trata de processo de seleção, utilizando esse modelo, quanto maior a quantidade de itens acertados maior a sua classificação. Assim, por exemplo, se em um teste X com 10 itens, em que um acerto vale 1 ponto e um erro vale 0, um sujeito A acerta os itens 1, 2, 4, 5 e 7 e o sujeito B acerta os itens 3, 6, 8, 9 e 10 significa que ambos tiraram 5 pontos. Para a TCT, tanto o sujeito A quanto o sujeito B têm a mesma aptidão, pois essa teoria preocupa-se apenas com o escore total. Pode-se então perguntar: será que esses dois sujeitos, que obtiveram a mesma pontuação, mas acertaram itens distintos, têm as mesmas habilidades? Certamente não, mas a TCT não diferencia esses sujeitos em relação às suas habilidades.

Por outro lado, era inevitável que outra teoria surgisse para preencher algumas lacunas deixadas pela TCT. Nesse sentido, a Teoria de Resposta ao Item (TRI) apresentou-se como uma alternativa trazendo “suposições adicionais que permitirão responder problemas que a TCT não conseguiu” (MUÑIZ, 2018, p. 187, tradução nossa). Desse modo, a análise pela TRI no teste X, descrito anteriormente, não se dá pelo o escore total, mas o seu interesse, nesse caso, concentra-se em avaliar as características de cada um dos 10 itens desse teste e também descobrir o traço latente (habilidade/proficiência) do respondente que melhor explica a probabilidade do conjunto de itens desse teste ser acertado ou errado, a partir das suas respostas.

Como explica Urbina (2007), a TCT já está bem estabelecida, é relativamente simples e amplamente utilizada. Já a TRI ainda está em evolução, é considerada mais complexa do ponto de vista matemático e ainda desconhecida por muitos profissionais. Além disso, Pasquali (2011) afirma que a TCT tem interesse em produzir testes de qualidade; já a TRI preocupa-se em produzir itens de qualidade. Algumas das principais diferenças entre essas teorias ficam evidentes no quadro a seguir:

Quadro 1 - Comparação entre a TCT e TRI.

<b>Teoria Clássica dos Testes (TCT)</b>	<b>Teoria de Resposta ao Item (TRI)</b>
Análise e interpretações estão associadas à prova como um todo.	Elementos centrais são os itens e não a prova como o todo.
Resultado: expresso pelo escore bruto ou padronizado.	Resultado: proficiência estimada pelo modelo estatístico.
Base numérica da comparação: o escore pode ser comparado com o escore de outro indivíduo submetido à mesma prova.	Base numérica da comparação: indivíduos e itens são colocados em uma escala comum, mesmo se submetidos a provas diferentes.

Fonte: Rabelo (2013).

O Quadro 1 apresenta algumas diferenças entre a TCT e a TRI. Nele, Rabelo (2013) destaca que as principais diferenças estão no fato que na TCT as interpretações da prova são feitas como um todo, posto que o resultado é calculado simplesmente pelo o escore total, não

levando em conta as características de cada item do teste. Já na TRI essa análise está centrada em cada item e o resultado do examinando (habilidade/proficiência) é estimado por modelos estatísticos, em que as respostas dadas em cada item medem uma faceta dessa habilidade. O mesmo autor afirma ainda que na TCT a comparação dos escores dos indivíduos se dá, normalmente, na mesma prova; por outro lado, na TRI, os indivíduos podem ser comparados em provas distintas, seja em uma mesma edição de um exame ou ao longo de uma série temporal. Essa característica da TRI revela-se extremamente importante principalmente para as avaliações externas de larga escala, por propiciar a comparação dos resultados dos examinandos ao longo do tempo.

Conforme Cohen, Swerdlik e Sturman (2014, p.166):

[Uma] vantagem da TCT sobre qualquer outro modelo de mensuração tem a ver com sua compatibilidade e facilidade de uso com técnicas estatísticas de ampla utilização (bem como com a maioria dos programas de análise de dados disponíveis) [...]. Mas especialistas em mensuração também listaram muitos problemas com a TCT. Para os iniciantes, um problema com essa teoria diz respeito a seu pressuposto relativo à equivalência de todos os itens em um teste; ou seja, todos os itens presumivelmente contribuem de igual modo para o escore total.

Vale destacar que, segundo os autores acima mencionados, um problema da TCT refere-se à equivalência de todos os itens em um teste, pois, mesmo que apresentem dificuldades distintas, segundo essa teoria, eles representam o mesmo valor na pontuação final de um examinando, o que parece não ser adequado. Ademais, apesar de grande utilização nos dias atuais, a TCT possui algumas limitações teóricas graves. Hambleton, Swaminathan e Rogers (1991, apud PASQUALI, 2011, p. 80-81) salientam quatro dessas limitações:

- 1) Os parâmetros clássicos dos itens (dificuldade e discriminação) dependem diretamente da amostra de sujeitos utilizada para estabelecê-los [...].
- 2) A avaliação das aptidões dos testandos também depende do teste utilizado [...].
- 3) A definição do conceito de fidedignidade ou precisão na teoria clássica dos testes constitui também uma fonte de dificuldades [...].
- 4) [A] teoria clássica dos testes consiste em que ela é orientada para o teste total e não para o item individual [...].

Como visto acima, na Teoria Clássica dos Testes existem diversos problemas e limitações que precisam ser observados para que possam ser feitas análises estatísticas com qualidade. Nesse sentido, Valentini e Laros (2011) afirmam que se um mesmo teste for aplicado a diferentes grupos de indivíduos com características distintas, e que um dos grupos seja bem

mais habilidoso do que um outro, tem-se que o parâmetro da dificuldade dos itens do teste receberia diferentes indicadores para cada um desses grupos. Entretanto, o parâmetro de dificuldade, observado na TCT, deveria ser uma característica dos itens do teste e não dos indivíduos examinados.

O problema supracitado do parâmetro da dificuldade em variar de um teste para um outro quando utilizado a TCT tem relevância,

mas ela se refere a um problema de amostragem e não tanto a análise da dificuldade do item em termos de número de acertos; assim se a amostra for aleatória e representativa, esse problema não existe na Psicometria Clássica. No entanto uma grande vantagem da TRI é que mesmo em amostras não representativas os parâmetros poderão ser estimados corretamente (EMBRETSON & REISE, 2000 apud PASQUALI, 2018, p. 12-13).

Nesse sentido, as estatísticas clássicas da TCT são “restritas a determinado momento, a determinado teste e ao grupo que respondeu à prova, dificultado o acompanhamento ao longo dos anos” (RABELO, 2013, p. 126). Assim, as dificuldades vistas acima levaram os psicometristas à procura de teorias alternativas que pudessem estabelecer (HAMBLETON; SWAMINATHAN; ROGERS, 1991, p. 5, tradução nossa):

- (a) Características do item que não são dependentes do grupo.
- (b) Pontuações que descrevem a proficiência do examinando que não dependem do teste.
- (c) Um modelo que é expresso no nível do item, em vez do que no nível de teste.
- (d) Um modelo que não requer rigorosamente testes paralelos para avaliar a confiabilidade.
- (e) Um modelo que fornece uma medida de precisão para cada nível de habilidade.

Essas características, destacadas acima, são oferecidas pela Teoria de Resposta ao Item (TRI), que constituiu um avanço para as aplicações em testes educacionais ou psicológicos e que será abordada com mais detalhes no decorrer deste estudo.

## **2.2 Um pouco de história sobre a TRI**

A Teoria de Resposta ao Item (TRI) tem sido estudada por diversos teóricos desde a década 1950, com foco em análises a partir das respostas dadas a um item, embora na sua

essência Pasquali (2018, p.14) diz que essa teoria já tinha sido estudada há mais de uma década, e destaca

os trabalhos de Richardson (1936), comparando os parâmetros dos itens obtidos pela teoria clássica da Psicometria com os moldes que hoje usam a TRI; os trabalhos de Lawley (1943, 1944), indicando alguns métodos para estimar os parâmetros dos itens, os quais se afastavam da teoria clássica e os trabalhos de Tucker (1946), que parece ter sido o primeiro a utilizar a expressão curva característica do item - Characteristic Curve, (ICC) - que constitui um conceito chave na TRI. Também deve ser mencionada a contribuição de Lazarsfeld (1950), que introduziu o conceito de traço latente, ainda que no contexto da medida das atitudes, conceito novamente que se constituiu num parâmetro chave da nova TRI.

Observa-se que foi por volta do início da década de 1940 que esses estudiosos iniciaram o processo de aperfeiçoamento e divulgação de uma nova teoria que trouxesse mais detalhes para a análise e interpretação dos itens de um teste para as medidas de aptidão. Mas, segundo Muñiz (2018), a formulação mais sistemática da TRI, que resultou nos trabalhos subsequentes dessa teoria, foi realizado em 1952 e 1953 pelo psicometrista Frederic Lord -nos Estados Unidos-, que elaborou um modelo teórico aplicado a testes para dados dicotômicos (certo ou errado) a partir do modelo da ogiva normal.

Anos depois, em 1957, Birnbaum dá um grande impulso nesse campo, substituindo os modelos ogiva normal de Lord pelos logísticos, matematicamente mais convenientes, gerando os desenvolvimentos matemáticos necessários para sua futura aplicação (MUÑIZ, 2018). Birnbaum foi quem formalizou e integrou muitos dos principais elementos teóricos da TRI utilizados atualmente, conforme suas contribuições apresentadas em alguns capítulos no destacado livro *Statistical Theories of Mental Test Scores*<sup>1</sup>.

Outro eminente psicometrista foi Georg Rasch, que trabalhava - na Dinamarca- nos mesmos tipos de problemas de medição que Lord, como os de medidas de traço latente aplicados a testes de desempenho e aptidão - baseados em respostas dicotômicas- e que em 1960 publicou um livro (*Probabilistic Models for Some Intelligence and Attainment tests*) que expõe o modelo logístico de um parâmetro (LUECHT; HAMBLETON, 2021, tradução nossa). Em vista disso, esse modelo ficou conhecido como *Modelo de Rasch*.

A modelagem da TRI se expandiu indo além dos estudos para dados dicotômicos. Nesse sentido, Samejima, em 1969, iniciou uma pesquisa independente sobre modelos alternativos para dados de desempenho educacional classificados em categorias ordenadas, como os

---

<sup>1</sup> Ver (LORD; NOVICK, 1968).

formatos de respostas do tipo *Likert*<sup>2</sup>. Esse modelo é conhecido como Modelo de Resposta Gradual (LUECHT; HAMBLETON, 2021, tradução nossa).

Diversas contribuições acerca da TRI tem sido oferecidas por outros estudiosos<sup>3</sup>, mas, segundo Fletcher (2010), foi na década de 1980 que essa teoria se tornou centro das atenções da psicometria para as medidas em testes educacionais. Esse autor destaca que foi apenas na terceira edição do prestigiado livro sobre medições em testes educacionais: *Educational Measurement*<sup>4</sup>, que

a teoria da resposta ao item é citada como uma de duas inovações que atualmente exercessem [sic] o maior impacto sobre a prática da aplicação de testes e medidas (a segunda inovação seria a tecnologia de informática). Na nova edição, os princípios e aplicações da nova teoria ganharam um enorme relevo em capítulo a parte e ainda em várias seções de outros capítulos que tratam de escalas, normas, procedimentos para tornar resultados de provas equivalentes, investigações do viés de conteúdo nas provas e ainda sobre tecnologias computacionais introduzidas na produção e aplicação de testes (FLETCHER, 2010, p.21).

Essa mudança de perspectiva pela comunidade científica, colocando a TRI em destaque para as análises em testes, deu-se graças ao avanço da informática nos anos 80 e contribuiu para a popularização do uso dessa teoria psicométrica. Como os cálculos na TRI são complexos e demasiadamente trabalhosos, o progresso dos computadores e a criação de *softwares* especializados tornaram-nos viáveis. Segundo Pasquali (2018) o primeiro *software* para as análises envolvidas na TRI surgiu somente em 1979 com o BICAL (por Wright, Mead e Bell), seguido depois pelo LOGIST (por Wingersky, Barton e Lord, em 1982) e pelo BILOG (por Mislevy e Bock, em 1984).

As aplicações da TRI no Brasil são, em sua maioria, realizadas na área de avaliação educacional, tendo como objetivo a análise de itens e comparações dos resultados em avaliações externas em larga escala. Conforme Gatti (1996 apud MOREIRA JUNIOR, 2011), a primeira aplicação da TRI no país foi realizada pela Secretaria de Educação do Estado de São Paulo em 1993, porém ela não foi aplicada com todo o seu potencial. Mas, considerando as avaliações em larga escala a nível nacional, Andrade, Tavares e Valle (2000, p. 5) afirmam que a

---

<sup>2</sup> *Likert* é um tipo de escala de resposta psicométrica usada habitualmente em questionários e muito popular em pesquisas de opinião. Ela consiste em pedir ao sujeito, que está a ser avaliado, para manifestar o grau de concordância, que normalmente vai desde o discordo totalmente (nível 1, por exemplo) até ao concordo totalmente (nível 5, por exemplo).

<sup>3</sup> Ver MUÑIZ (2018); ANDRADE; TAVARES; VALLE (2000); LUECHT; HAMBLETON (2021).

<sup>4</sup> Esse livro foi lançado em conjunto pelo *National Council on Measurement in Education* (EUA) e o *American Council on Education*, e editado por Robert L. Linn (LINN, 1989 apud FLETCHER, 2010).

TRI foi usada pela primeira vez em 1995 na análise dos dados do Sistema Nacional de Ensino Básico – SAEB [sic]. A introdução da TRI permitiu que os desempenhos de alunos de 4ª e 8ª séries do Ensino Fundamental e de 3ª série do [Ensino Médio] [...] pudessem ser comparados e colocados em uma escala única de conhecimento.

Nas últimas décadas, a TRI tem se tornado a técnica psicométrica mais utilizada em vários países quando se trata de testes educacionais. Ela é aplicada em diversas avaliações em todo mundo, como no Toefl (*Test of English as a Foreign Language*) e o Pisa (Programa Internacional de Avaliação de Estudantes). No Brasil, além da sua aplicação no Saeb<sup>5</sup> (Sistema de Avaliação da Educação Básica), pode-se destacar também a sua utilização no Enem (Exame Nacional do Ensino Médio), Encceja (Exame Nacional para Certificação de Competências de Jovens e Adultos), Saresp (Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo), Saepe (Sistema de Avaliação da Educação Básica de Pernambuco), Saveal (Sistema de Avaliação Educacional de Alagoas), entre outros.

No Saeb, até 1993, as estatísticas dessa avaliação eram computadas segundo a TCT para construção dos instrumentos de medida, atribuição dos escores e análises dos resultados; assim os resultados de um ano para o outro não eram comparáveis e isso constitui uma limitação importante para a análise dos resultados na linha do tempo. Como visto acima, foi a partir de 1995 que se introduziu o uso da TRI com o objetivo de garantir a comparação das notas do exame daquele ano com outros anos, a partir da atribuição das proficiências médias dos estudantes e também da análise qualitativa e pedagógica dos resultados com a construção das escalas de proficiência (RABELO, 2013). Já no Enem a TCT foi utilizada de 1998 a 2008, e a partir de 2009 começou a ser utilizada a TRI com o objetivo de: “(1) permitir a comparabilidade dos resultados entre os anos e (2) permitir a aplicação do Exame várias vezes ao ano” (ANDRADE; KARINO, 2012, n.p).

Portanto, como mencionado nessa seção, o desenvolvimento da Teoria de Resposta ao Item se deu por diversos estudiosos através de décadas na história, mas as aplicações práticas dessa teoria eram extremamente limitadas e apenas foram se popularizando a partir da década de 1980 com o advento dos computadores e *softwares* especializados. Além disso, atualmente

---

<sup>5</sup> As avaliações do Saeb compreendem três avaliações externas em larga escala: Avaliação Nacional da Educação Básica – Aneb, Avaliação Nacional do Rendimento Escolar - Anresc (nacionalmente conhecida como Prova Brasil) e a Avaliação Nacional da Alfabetização – ANA. Destaca-se que, a partir de 2019, todas essas avaliações externas passaram a ser identificadas apenas por: Saeb, acompanhado da etapa correspondente.

essa teoria é utilizada em diversas avaliações ao redor do mundo, e no Brasil tornou-se tema relevante a partir da sua utilização nas avaliações externas em larga escala.

### 2.3 Características da TRI

Diversos modelos da TRI foram construídos ao longo dos tempos, de modo que, mesmo sendo diferentes em alguns aspectos, todos eles têm em comum algumas características básicas que, segundo Hambleton, Swaminathan e Rogers (1991, p. 7, tradução nossa), são:

- a) O desempenho de um examinando em um item de um teste pode ser previsto (ou explicado) por um conjunto de fatores chamados traços, traços latentes ou habilidades.
- b) A relação entre o desempenho de um examinando em um item e o conjunto de traços subjacentes ao desempenho no item podem ser descritos por uma função monotonicamente crescente chamada de função característica do item ou curva característica do item (CCI).

Além disso, todos os modelos da TRI especificam que a probabilidade de responder a um item corretamente depende da(s) habilidade(s) dos examinandos e das características dos itens. Como todos os modelos, eles incluem um conjunto de suposições sobre os dados aos quais se aplicam. As principais premissas são unidimensionalidade e independência local, apresentadas a seguir.

#### 2.3.1 A unidimensionalidade

Como foi visto acima, a TRI postula que o desempenho de qualquer sujeito para executar alguma tarefa está relacionado a um conjunto de traços latentes ou habilidades que ele tem. Assim, quando um sujeito realiza um teste, vários fatores influenciam o seu desempenho. Hambleton, Swaminathan e Rogers (1991) afirmam que esses fatores podem incluir o nível de motivação, ansiedade, capacidade de resolver rapidamente, tendência de adivinhar as respostas quando estiver em dúvida e também da(s) sua(s) habilidade(s) cognitiva(s).

Entretanto, quando um modelo da TRI pressupõe que há apenas uma habilidade dominante - representada por  $\theta$ -, que é responsável por responder um conjunto de itens de um

teste, dizemos que ele é *unidimensional*. Modelos desse tipo são os mais comuns que se apresentam nas avaliações educacionais e, portanto, serão tratados nesse trabalho.

### 2.3.2 A independência local

O postulado da independência local diz que um determinado desempenho do sujeito em um item de um teste não afeta o desempenho em um outro item desse teste, isto é, as probabilidades de um sujeito acertar ou errar a um conjunto de itens são independentes entre si (PASQUALI, 2018). Assim, a probabilidade de acerto (ou erro) da sequência das respostas dadas pelos sujeitos –  $P(U_1, U_2, \dots, U_n | \theta)$  – será o produto das probabilidades de acerto (ou erro) em cada item de forma independente<sup>6</sup>. Matematicamente, a independência local pode ser representada por:

$$P(U_1, U_2, \dots, U_n | \theta) = P(U_1 | \theta) \cdot P(U_2 | \theta) \cdot \dots \cdot P(U_n | \theta) = \prod_{i=1}^n P(U_i | \theta), \quad (1)$$

em que  $\theta$  é a habilidade dominante;  $U_i$  representa a resposta de um indivíduo ao item  $i$  ( $i = 1, 2, 3, \dots, n$ ), em que  $U_i = 1$  se o sujeito acerta ao item, caso contrário  $U_i = 0$ . Temos ainda que  $P(U_i | \theta)$  é a probabilidade de um sujeito com habilidade  $\theta$  acertar ou errar o item  $i$ ; daí se tem  $P(U_i = 1 | \theta)$  representa a probabilidade de uma resposta correta dada ao item  $i$  por um sujeito com habilidade  $\theta$ . Por outro lado,  $P(U_i = 0 | \theta)$  representa a probabilidade de resposta errada dada por esse sujeito. Nota-se que esses eventos são complementares, logo  $P(U_i = 0 | \theta) = 1 - P(U_i = 1 | \theta)$ .

Neste trabalho, quando não causar dúvidas, será usada a notação abreviada para a probabilidade de acerto no item  $i$  como  $P_i$  e o seu complementar  $Q_i = 1 - P_i$  (que significa a probabilidade de erro no item  $i$ ). Assim, é possível reescrever a Equação (1) como:

$$P(U_1, U_2, \dots, U_n | \theta) = \prod_{i=1}^n P(U_i = 1 | \theta)^{U_i} \cdot [1 - P(U_i = 1 | \theta)]^{1-U_i} = \prod_{i=1}^n P_i^{U_i} \cdot Q_i^{1-U_i} \quad (2)$$

---

<sup>6</sup> A saber, essa função probabilidade de acerto (ou erro) em cada item depende das suas características e do valor da habilidade ( $\theta$ ) do sujeito respondente ao teste, e será apresentada em detalhes na Seção 2.4.

Nesse sentido, quando um examinando responde a um conjunto de itens resulta em um *padrão de respostas*, formados por acertos (valor = 1) e erros (valor = 0). Assim, se em um teste um sujeito acertou os itens 1, 2, e 4, e errou o item 3, segue que o padrão de suas respostas é  $U_1=1, U_2=1, U_3=0, U_4=1$ , ou seja, 1101. Daí, pela independência local descrita em (2), isso implica que:

$$P(U_1 = 1, U_2 = 1, U_3 = 0, U_4 = 1 | \theta) = P_1 \cdot P_2 \cdot Q_3 \cdot P_4$$

É importante destacar essa característica, pois é a partir dessa suposição que pode ser encontrada a habilidade mais plausível para um sujeito a partir do seu padrão de respostas em um teste, bem como a sua utilização para estimação dos parâmetros dos itens, que será discutido na Seção 3.

## 2.4 Modelos logísticos unidimensionais da Teoria de Resposta ao Item

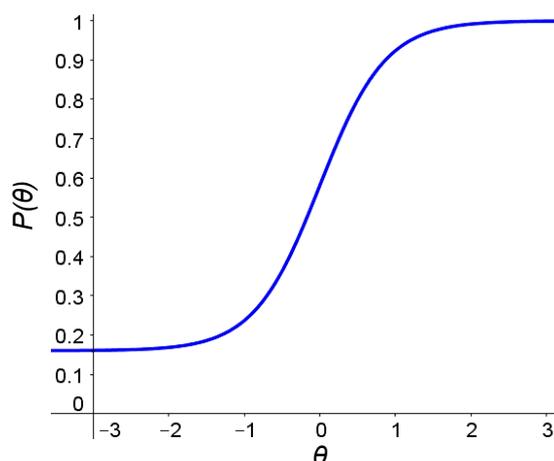
Os modelos logísticos da TRI, sejam eles dicotômicos ou dicotomizados (corrigidos como certo ou errado, como nos testes de múltiplas escolhas), constituem os modelos matemáticos que têm como objetivo representar a probabilidade de um indivíduo dar uma resposta correta a um item em função da sua habilidade e dos parâmetros desse item.

Assim, se a habilidade<sup>7</sup> de um sujeito é expressa como  $\theta$ , pode-se definir que a probabilidade de acerto em um item  $i$  é a função probabilidade  $P_i(\theta)$ , cujos valores funcionais variam em uma escala contínua de 0 a 1 (de 0% a 100%). Desse modo, é natural conjecturar que um sujeito que tem maior habilidade terá uma probabilidade maior de acertar um certo item do que um sujeito com habilidade inferior. Nesse sentido, à medida que a habilidade cresce, também cresce a  $P_i(\theta)$ , apresentando graficamente uma curva em uma forma de S, denominada Curva Característica do Item (CCI), como mostra a figura a seguir.

---

<sup>7</sup> Essa habilidade pode teoricamente assumir qualquer valor real entre  $-\infty$  e  $+\infty$ , mas na prática esses valores costumam figurar entre -4 e 4 desvios-padrão da média (NAKANO; PRIMI; NUNES, 2015).

Figura 1 - Curva Característica do Item (CCI).



Fonte: elaboração do autor.

No gráfico apresentado na Figura 1, o eixo horizontal representa a habilidade ( $\theta$ ). Já o eixo vertical denota a probabilidade de acerto no item. Como pode ser visto nessa figura, a probabilidade de acerto no item é pequena nos níveis mais baixos de habilidade e aproxima-se de 1 no mais altos. Além disso, em um teste cada item pode ser descrito por sua CCI e essa curva é a unidade conceitual básica dos modelos da TRI.

De acordo com Andrade, Tavares e Valle (2000, p. 7) os diversos modelos da TRI que encontramos na literatura dependem basicamente: “(1) da natureza do item (dicotômicos ou não-dicotômicos); (2) do número de populações envolvidas (apenas uma ou mais de uma); (3) da quantidade de traços latentes que está sendo medida (apenas um ou mais de um)”.

Embora existam vários modelos da TRI, os mais utilizados são os modelos logísticos unidimensionais. Dentre esses modelos logísticos destacam-se os de 1, 2 ou 3 parâmetros, que, respectivamente, avaliam apenas a dificuldade do item, ou a dificuldade e a discriminação, ou a dificuldade, discriminação e acerto ao acaso.

De acordo com a Teoria da Resposta ao Item, o modelo matemático padrão utilizado atualmente para representar a curva característica do item é a função logística, que “foi inicialmente descoberta em 1844 e tem sido amplamente utilizada nas ciências biológicas para modelar o crescimento de plantas e animais desde o nascimento até a maturidade” (BAKER, 2001, p. 21, tradução nossa). Além de modelar fenômenos biológicos, ela é amplamente utilizada em outras áreas como economia, sociologia, física e, segundo Muñiz (2018), também tem sido tradicionalmente utilizada na medicina para analisar problemas como a propagação de

doenças. Entretanto, foi no final dos anos 1950 que a função logística foi utilizada pela primeira vez para as análises de testes de aptidão e se tornou popular por sua simplicidade.

Nesse sentido, o primeiro modelo foi desenvolvido, em 1957, por Birnbaum que substituiu o modelo da ogiva normal -desenvolvido por Lord<sup>8</sup> (1952) - pela função logística de dois parâmetros. O modelo logístico de dois parâmetros (2PL), equivalente ao modelo desenvolvido por Birnbaum, é dado por:

$$P_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}}, \quad (3)$$

em que  $a_i$  é o parâmetro de discriminação do item  $i$ ,  $b_i$  é o parâmetro de dificuldade do item  $i$ ,  $\theta$  é a variável que representa a habilidade,  $e$  representa o número de Euler, cujo valor aproximado é 2,718 (com três casas decimais);  $D$  denota uma constante que serve como um fator de escala<sup>9</sup> igual a 1 e  $P_i(\theta)$  é a probabilidade de um sujeito com habilidade  $\theta$  responder corretamente um item  $i$ .

Na década de 1960, o matemático dinamarquês George Rasch sugeriu um modelo equivalente ao logístico de dois parâmetros, assumindo  $D = 1$  e também um valor comum para o parâmetro de discriminação dos itens, pondo  $a = 1$ , e com isso dependendo apenas do parâmetro da dificuldade - e da habilidade do sujeito. Por esse motivo é que o modelo logístico de um parâmetro é popularmente conhecido como *modelo de Rasch*, cuja função que representa esse modelo logístico de um parâmetro é dada por (BAKER, 2001):

$$P_i(\theta) = \frac{1}{1 + e^{-1(\theta - b_i)}} \quad (4)$$

---

<sup>8</sup> Frederic Lord foi o primeiro a desenvolver o modelo unidimensional de dois parâmetros aplicados à TRI, cujo modelo era baseado na distribuição normal acumulada (ogiva normal), que é dado por (LORD, 1952, p.5):

$$P_i = \int_{-\infty}^{\frac{c - a_i}{b_i}} N(y) dy, \text{ em que } c \text{ é a medida da habilidade, } a_i \text{ e } b_i \text{ são os valores que caracterizam o item, } y \text{ é a variável}$$

de integração e que  $N(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$  é a função de distribuição normal.

<sup>9</sup> Quando se deseja que a função logística forneça resultados semelhantes aos da função da ogiva normal utiliza-se  $D=1,7$  (ANDRADE; TAVARES; VALLE, 2000, p. 10).

Nos dois modelos logísticos acima descritos ocorre uma lacuna: eles não consideram que os sujeitos com baixa habilidade podem acertar itens ao acaso. Com base nessa observação, Birnbaum, em 1968, desenvolveu o modelo logístico de 3 parâmetros (3PL), modificando o modelo de dois parâmetros para incluir o parâmetro  $c$ , que representa a probabilidade de acerto ao acaso quando não se conhece a resposta. A função que representa o modelo logístico de três parâmetros (3PL) é dada por:

$$P_i(\theta) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta - b_i)}} \quad (5)$$

onde  $P_i(\theta)$ ,  $D$ ,  $e$ ,  $a_i$ ,  $b_i$  e  $\theta$  tem o mesmo significado que aquele já mencionado para o caso de dois parâmetros (2PL), e  $c$  é o valor de  $P(\theta)$  quando  $\theta \rightarrow -\infty$ .

Em algumas avaliações de larga escala, como o Enem e o Saeb, as estimativas dos parâmetros que caracterizam os itens e as estimativas das habilidades/proficiências dos estudantes é realizada através do modelo 3PL. Em vista disso, será dada uma maior ênfase a explicação desse modelo logístico neste estudo, e, a saber, os outros dois podem ser obtidos a partir dele.

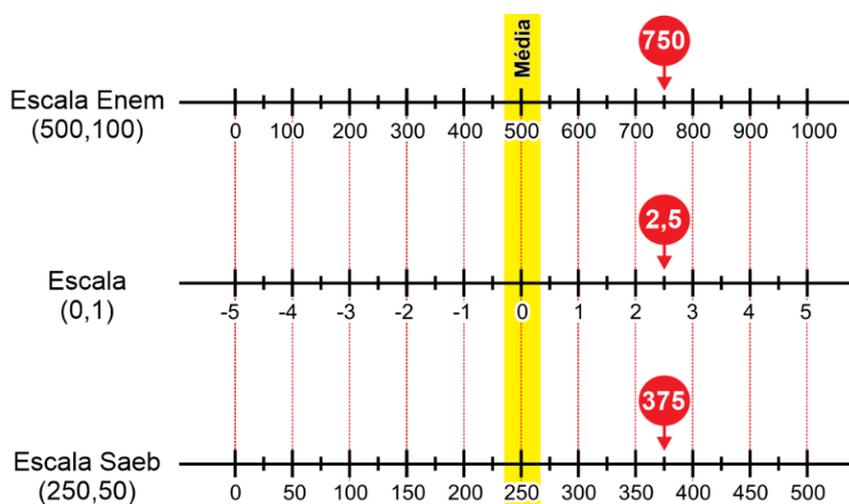
## 2.5 Transformações de escalas

O conceito de escala de habilidade (proficiência), de forma simplificada, constitui uma espécie de “régua”, graduada para qualquer valor real entre  $-\infty$  e  $+\infty$  e construída com base na Teoria de Resposta ao Item (TRI). Assim, inicialmente estabelece para essa escala uma origem e uma unidade de medida, de forma arbitrária, que representa, respectivamente, o valor médio e o desvio-padrão das habilidades dos indivíduos da população em estudo. Em geral, esses valores são representados através da notação  $(m, d)$ , em que  $m$  é a média (ou habilidade média) e  $d$  refere-se ao desvio padrão.

Normalmente a escala  $(0, 1)$  é a mais utilizada, mas não faz a menor diferença estabelecer média e desvio-padrão diferentes daqueles, pois a interpretação feita sobre duas escalas diferentes é a mesma, o importante é a relação de ordem entre seus pontos (RABELO, 2013).

Assim, por exemplo, se na escala (0,1) um sujeito com habilidade 2,5 está 2,5 desvios-padrão acima da habilidade média. Nesse sentido, na escala do Saeb, que foi arbitrada com média de 250 e desvio padrão de 50, isto é, (250,50), esse mesmo sujeito teria habilidade 375 e consequentemente estaria 125 pontos acima da média - isto equivale dizer que também ele estaria a 2,5 desvios-padrão acima da média. Caso fosse utilizada a escala do Enem (500, 100), esse indivíduo teria uma habilidade de 750 e estaria a 250 pontos acima da média (que da mesma forma equivale a 2,5 desvios-padrão). As equivalências acima destacadas podem ser vistas na figura a seguir:

Figura 2 - Equivalências entre as escalas (0,1), Enem (500,100) e Saeb (250,50).



Fonte: adaptado de Rabelo (2013).

Da mesma forma que foram analisados os pontos destacados na figura acima, pode-se realizar essa equivalência para qualquer ponto da escala. Assim, por exemplo, um sujeito com habilidade de -1,62 na escala (0,1) equivaleria a uma habilidade de 338 na escala (500, 100), do Enem. Entretanto, muitas vezes é comum recorrer ao uso de algumas transformações lineares para ajustar os valores da habilidade e dos parâmetros dos itens com o objetivo de facilitar a interpretação na escala, uma vez que a mudança procura transformar valores negativos ou com múltiplas casas decimais em números inteiros e positivos. Essa estratégia normalmente é realizada com o intuito de tornar mais natural a comunicação dos resultados para o público comum.

De modo geral, as mudanças dos parâmetros dos itens e da habilidade da escala (0,1) para outra escala qualquer  $(m,d)$  são obtidas a partir de transformações lineares, desde que os parâmetros dos itens também sejam transformados adequadamente, para manter a propriedade

de invariância (ARIAS; LLOREDA; LLOREDA, 2014). Para tal, seguem-se as expressões para essas transformações (MUÑIZ, 2018, p.223-224):

$$\text{i) } \theta^* = d \cdot \theta + m; \quad (6)$$

$$\text{ii) } b^* = d \cdot b + m; \quad (7)$$

$$\text{iii) } a^* = a / d, \quad (8)$$

em que  $\theta^*$ ,  $a^*$  e  $b^*$  representam, respectivamente, a habilidade e os parâmetros dos itens na escala  $(m, d)$ , e ainda  $a$  e  $b$  estão definidos como em (3).

Ademais, para o caso 3PL, o parâmetro  $c$ , de acerto ao acaso, em ambas as escalas, segue na mesma métrica, isto é:

$$\text{iv) } c^* = c. \quad (9)$$

Daí, observadas as transformações acima, tem-se que

$$\text{v) } P(U_i = 1 | \theta^*) = P(U_i = 1 | \theta), \quad (10)$$

que, para o caso 3PL, como em (5), pode ser verificado a seguir:

$$\begin{aligned} P(U_i = 1 | \theta^*) &= c_i^* + (1 - c_i^*) \frac{1}{1 + e^{-Da_i^*(\theta^* - b_i^*)}} \\ &= c_i + (1 - c_i) \frac{1}{1 + e^{-\frac{D}{d}[(d \cdot \theta + m) - (d \cdot b_i + m)]}} \\ &= c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta - b_i)}} = P(U_i = 1 | \theta). \end{aligned}$$

Como foi mencionado, a escolha da escala é arbitrária e depende apenas do desenvolvedor do teste. Por exemplo, suponha que um sujeito tenha proficiência  $\theta = -1,2$  e que os parâmetros de um item K ( $a=1,94$ ,  $b=-0,28$  e  $c=0,18$ ), estimados na escala  $(0,1)$ . Caso se queira transformar esses elementos para a escala do Enem  $(500,100)$ , é necessário utilizar as expressões (6), (7), (8) e também o fato de que  $c^* = c$ , descrito em (9). Daí, tem-se como a proficiência desse sujeito o valor  $\theta^* = 100(-1,2) + 500 = 380$  e como parâmetros de discriminação, dificuldade e acerto ao acaso, respectivamente,  $a^* = 1,94/100 = 0,0194$ ,  $b^* = 100(-0,28) + 500 = 472$  e  $c^* = c = 0,18$ .

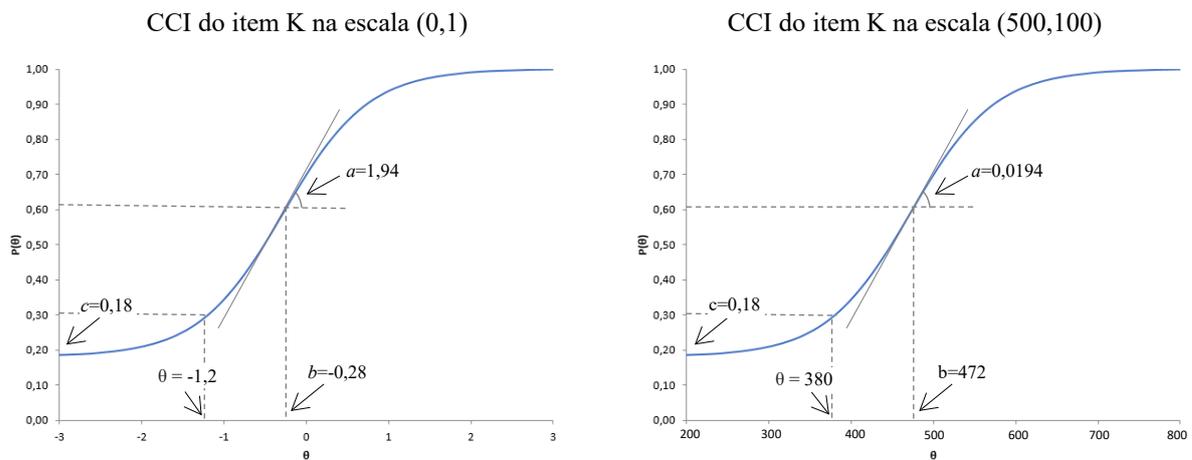
Ademais, a probabilidade de um sujeito responder corretamente a um certo item, independentemente da escala utilizada, é sempre a mesma, como em descrito em (10). Nesse caso, para o modelo 3PL e pondo  $D=1$ , tem-se:

$$P(U_i = 1 | \theta = -1,2) = 0,28 + (1 - 0,28) \frac{1}{1 + e^{-1,94 \cdot [-1,2 - (-0,28)]}}$$

$$= 0,28 + (1 - 0,28) \frac{1}{1 + e^{-1,0,0194 \cdot (380 - 472)}} = P(U_i = 1 | \theta^* = 380).$$

Assim, a partir dessas conversões e independentemente da escala escolhida serão produzidas interpretações congêneres tanto para as habilidades quanto para os parâmetros dos itens, e ainda para as representações gráficas da CCI, como na figura a seguir:

Figura 3 – Equivalência da Curva Característica do Item (CCI) nas escalas (0,1) e (500, 100).



Fonte: elaboração do autor.

Os gráficos apresentados na figura acima representam a curva característica do mesmo item K, mas em escalas distintas. O gráfico da esquerda refere-se ao item K que foi construído na escala (0,1); já o gráfico da direita exibe um comportamento equivalente para o item K na escala (500,100). No decorrer deste estudo será utilizada preferencialmente a escala (0,1), por ser mais universal e amplamente empregada em *softwares* computacionais.

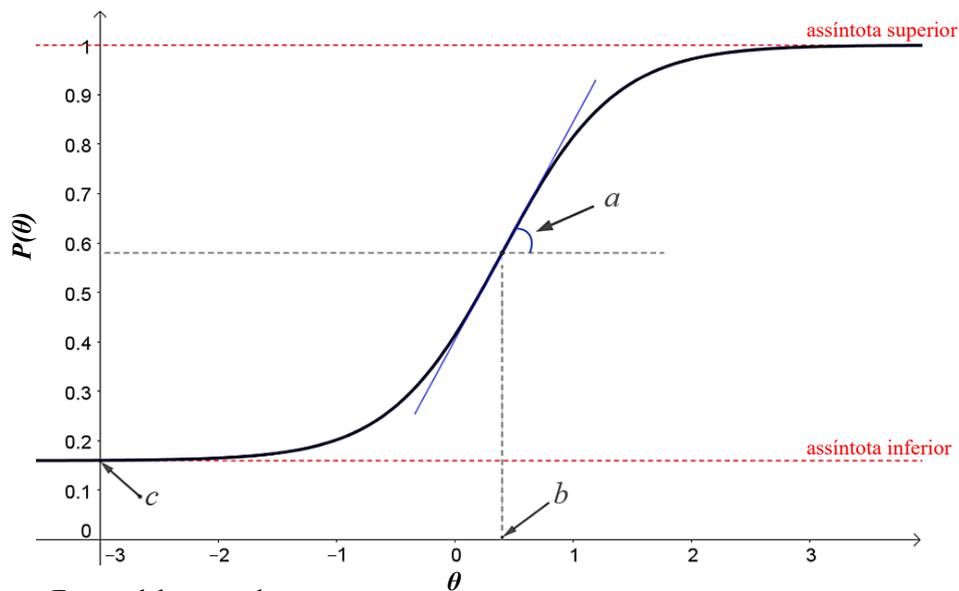
## 2.6 Modelo logístico de 3 parâmetros (3PL)

Entre os modelos sugeridos pela TRI, o modelo logístico unidimensional de 3 parâmetros é o mais utilizado em avaliações. Daí, com o intuito de destacar os itens e os sujeitos de um teste, podemos reescrever a função (5), que representa a probabilidade de um indivíduo  $j$  responder corretamente um item  $i$ , como<sup>10</sup>:

$$P(U_{ij}=1|\theta_j) = c_i + \frac{1-c_i}{1+e^{-Da_i(\theta_j-b_i)}}, \quad (11)$$

em que  $U_{ij}$  representa a resposta do sujeito  $j$  ao item  $i$  ( $U_{ij}=1$ , se o sujeito responde corretamente e  $U_{ij}=0$ , caso contrário) e que  $a_i$ ,  $b_i$ ,  $c_i$ ,  $e$ ,  $\theta_j$  e  $D$  tem o mesmo significado como em (5). De forma simplificada, pode-se usar a notação:  $P(U_{ij}=1|\theta_j) = P(\theta)$ . Ademais, o gráfico que representa a função (11) é a Curva Característica do Item (CCI) e está apresentado na Figura 4 a seguir.

Figura 4 - Curva Característica do Item (CCI) para o modelo 3PL.



Fonte: elaboração do autor.

<sup>10</sup> Uma notação mais completa para representar a probabilidade de um indivíduo  $j$  responder corretamente um item  $i$  é:  $P(U_{ij}=1|\theta_j, a_i, b_i, c_i) = c_i + \frac{1-c_i}{1+e^{-Da_i(\theta_j-b_i)}}$ , destacando, dessa forma, que a função  $P$  depende não apenas de  $\theta$ , mas também dos parâmetros  $a$ ,  $b$  e  $c$ .

O gráfico acima apresenta a relação existente entre  $\theta$  e  $P(\theta)$ . Nota-se que  $\theta \rightarrow +\infty$  implica que  $P(\theta) \rightarrow 1$ , e ainda  $\theta \rightarrow -\infty$  implica  $P(\theta) \rightarrow c$ . Isso significa que as assíntotas superior e inferior são, respectivamente, 1 e  $c$ . Além disso, no gráfico acima são destacados os parâmetros que caracterizam um item, cujos significados específicos e as respectivas características serão apresentados na Subseção 2.6.1 a seguir.

## 2.6.1 Características gerais dos parâmetros dos itens

No modelo logístico unidimensional de três parâmetros (3PL), a probabilidade que justifica o acerto ou o erro de um sujeito, quando submetido a um item, não depende apenas da habilidade dominante ( $\theta$ ) que ele possui, mas também das características dos itens. Essas características que cada item possui versa sobre a discriminação (parâmetro  $a$ ), a dificuldade (parâmetro  $b$ ) e o parâmetro de acerto ao acaso (parâmetro  $c$ ), que são medidos em uma escala contínua e têm interpretações específicas.

### 2.6.1.1 Acerto ao acaso (parâmetro $c$ )

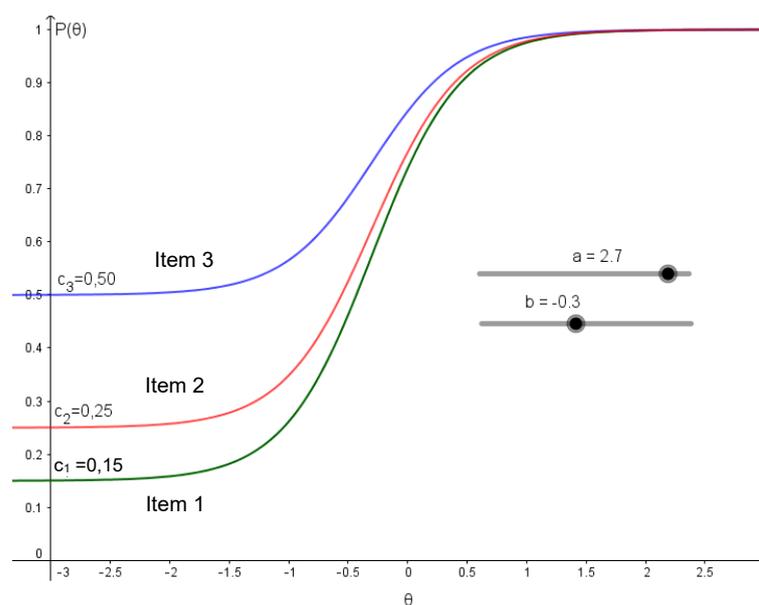
O parâmetro  $c$  representa a probabilidade de acerto das respostas dadas ao acaso, isto é, por “chute”. Normalmente isso ocorre quando um sujeito de baixa habilidade não conhece a resposta correta, mas arrisca uma resposta qualquer. Esse parâmetro assume teoricamente valores entre 0 e 1, pois trata de uma probabilidade. Segundo Baker e Kim (2004) o parâmetro  $c$ , geometricamente, corresponde ao limite inferior da CCI, ou seja, ele representa a assíntota quando  $\theta \rightarrow -\infty$  (ver Figura 4).

Neste modelo logístico, supondo que um teste de múltipla escolha possua  $n$  alternativas, é possível afirmar que cada alternativa tenha probabilidade  $1/n$  de ser escolhida. No Enem, por exemplo, como um item possui cinco alternativas de escolha, cada uma delas tem probabilidade de acerto igual a  $1/5 = 0,20$ . No caso do Saeb, em que cada item possui quatro alternativas, essa probabilidade é de  $1/4 = 0,25$ . Assim, se na análise estatística de um item, pela TRI, esse parâmetro  $c$  for muito superior a  $1/n$ , essa condição aponta que a opção correta se diferencia de algum modo das demais, atraindo sujeitos de baixa habilidade. Daí, quando isso acontece, é um

indicativo de que um item pode ter sido mal elaborado ou que o gabarito pode conter inconsistência.

Uma ilustração mostrando a variação do parâmetro  $c$  é apresentada na figura a seguir:

Figura 5 – Variação do parâmetro  $c$  da curva característica de três itens.



Fonte: elaboração do autor.

A Figura 5 mostra a curva característica de três itens, com variação apenas do parâmetro  $c$ . Daí, se for considerado que esses itens são de um teste com quatro alternativas cada, significa que os itens 1 e 2 são adequados em relação ao parâmetro  $c$ , pois apresentaram valores menores ou iguais a  $1/4 = 0,25$ . Por outro lado, é possível observar que o item 3 apresentou o valor  $c=0,50$ , que é muito superior ao valor esperado ( $1/4 = 0,25$ ), significando que pessoas com baixa habilidade tem alta probabilidade de acertar esse item. Dependendo do propósito da avaliação, como, por exemplo, quando se quer arquivar itens de qualidade em um banco de itens<sup>11</sup>, o item 3 poderá ser reformulado ou ainda excluído.

### 2.6.1.2 Dificuldade de um item (parâmetro $b$ )

Na TRI, a dificuldade de um item é representada pelo parâmetro  $b$  e pode ser definida como o nível mínimo de habilidade (proficiência) que um sujeito precisa possuir para ter

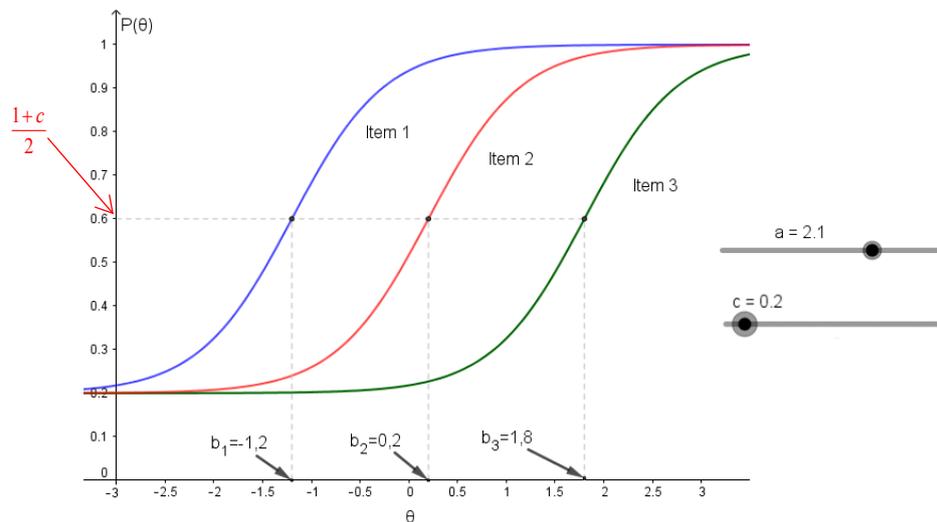
<sup>11</sup> Banco de itens é um conjunto de itens organizados, normalmente em computadores, cujas propriedades psicométricas são conhecidas (MUÑIZ, 2018). Normalmente é desse banco que são extraídos os itens que irão compor os testes em uma avaliação.

probabilidade relativamente alta de acertar um item. Esse parâmetro representa um número real e varia na escala de habilidade de  $-\infty$  a  $+\infty$ . Desse modo, como o parâmetro  $b$  e o  $\theta$  dos respondentes a um teste estão na mesma escala, eles podem ser comparados.

Nesse sentido, a “dificuldade do item descreve onde o item está localizado na escala de habilidade, ou seja, quanta habilidade o item requer para ser resolvido com sucesso, sendo, portanto, um índice da posição ou localização do item na escala de medição do traço ou habilidade ( $\theta$ )” (ARIAS; LLOREDA; LLOREDA, 2014, p. 135, tradução nossa). Então, considerando o item 2 da Figura 6 abaixo, cujo parâmetro  $b=0,2$ , tem-se que um sujeito com um  $\theta = 1$  tem alta probabilidade de acertar o item; já um sujeito com  $\theta = -1$  tem baixa probabilidade de acertar esse item.

Nota-se, observando a forma geométrica da CCI (Figura 4) - e de suas assíntotas: a superior e a inferior-, que o parâmetro  $b$  corresponde ao ponto na escala de habilidade ( $\theta$ ) para que o examinando tenha a probabilidade de acerto igual a  $\frac{1-c}{2} + c = \frac{1+c}{2}$  (em que  $c$  é o parâmetro de acerto ao acaso no item). Desse modo temos que o ponto de inflexão da CCI é dado por  $\left(b, \frac{1+c}{2}\right)$ .

Figura 6 - Parâmetro da dificuldade de três itens para diferentes valores de  $b$ .



Fonte: elaboração do autor.

O deslocamento para a direita da CCI, apresentado na Figura 6, significa que é necessário que um sujeito tenha uma maior habilidade para obtenção de determinada probabilidade de acerto no item. Desse modo, por exemplo, para o sujeito ter uma probabilidade de 60% de acertar um item é necessário, no primeiro caso (item 1), que ele tenha um  $\theta = -1,2$ ;

já na segunda curva (item 2), ele necessita ter um  $\theta = 0,2$  e, finalmente, para a última curva (item 3),  $\theta = 1,8$ . Assim, quanto mais à direita o valor do parâmetro  $b$ , mais difícil será o item e isso significa, nesse caso, que um sujeito precisa ter uma habilidade ( $\theta$ ) muito maior para acertar o item 3 do que o item 1.

Como foi dito, o parâmetro da dificuldade de um item representa um número real, mas na prática, conforme Rabelo (2013, p. 134), “seus valores situam-se tipicamente entre -3 (itens fáceis) e 3 (itens difíceis), pois entre esses extremos estão cerca de 99,7% dos casos”. Esse autor afirma ainda que a distribuição ideal de dificuldade dos itens em avaliações educacionais deve seguir uma distribuição conforme a Curva Normal, seguindo os percentuais apresentados na tabela a seguir:

Tabela 1 - Classificação, valores e percentual esperado para o parâmetro de dificuldade na TRI.

Classificação	Valores do parâmetro b	% esperado
Muito fáceis	até -1,28	10%
Fáceis	de -1,27 a -0,52	20%
Medianos	de -0,51 a 0,51	40%
Difíceis	de 0,52 a 1,27	20%
Muito difíceis	1,28 ou mais	10%

Fonte: Rabelo (2013).

A Tabela 1 mostra a classificação dos itens em cinco faixas, em que é esperado que quase a metade dos itens estejam na faixa de dificuldade mediana e que 30% sejam itens difíceis ou muito difíceis. Entretanto, segundo Rabelo (2013, p.135), nos testes de matemática “os itens apresentam índices de dificuldade elevada, o que torna muito árdua a tarefa de atender às faixas indicadas na tabela”.

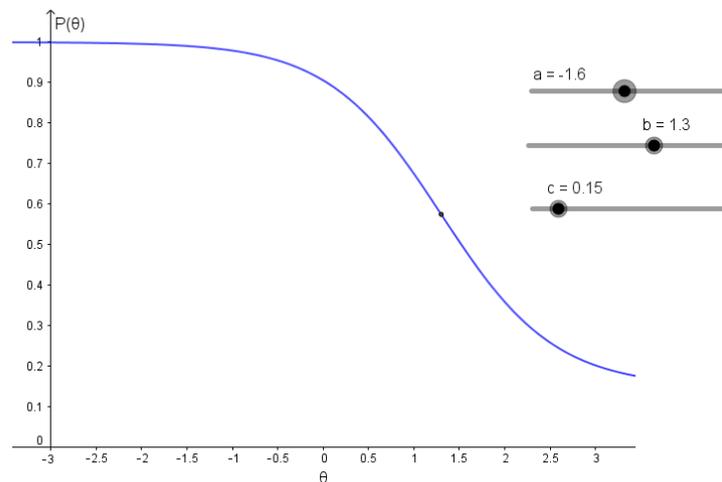
### 2.6.1.3 Discriminação de um item (parâmetro a)

Na TRI o parâmetro  $a$  é o índice de discriminação do item e “mede se o item consegue diferenciar alunos com baixa proficiência de alunos com alta proficiência em relação à dificuldade da questão, definindo assim a sinuosidade da curva [CCI]” (BORTOLUCCI, 2013,

p. 55). Geometricamente o parâmetro  $a$  representa a derivada da reta tangente à curva no ponto de sua inclinação máxima, ou seja, no ponto de inflexão da CCI. Isso ocorre no ponto da CCI de abscissa igual a  $b$ , como pode ser visto na Figura 4.

Esse parâmetro também varia de  $-\infty$  a  $+\infty$ , mas “na prática, a métrica desse parâmetro vai de 0 a 3, na qual 0 significa nenhuma discriminação e 3, a discriminação praticamente perfeita” (PASQUALI, 2011, p.27). Ademais, valores negativos para esse parâmetro ficam sem sentido prático para testes educacionais, pois significam que sujeitos de menor habilidade tendem a acertar o item enquanto o de maior habilidade tendem a errá-lo. Quando isso acontece pode indicar um possível erro no gabarito ou na formulação da questão, e dessa forma pode ser necessário realizar uma revisão. Gráficamente essa situação está apresentada na figura a seguir:

Figura 7 – CCI de um item com índice de discriminação (parâmetro  $a$ ) negativo.



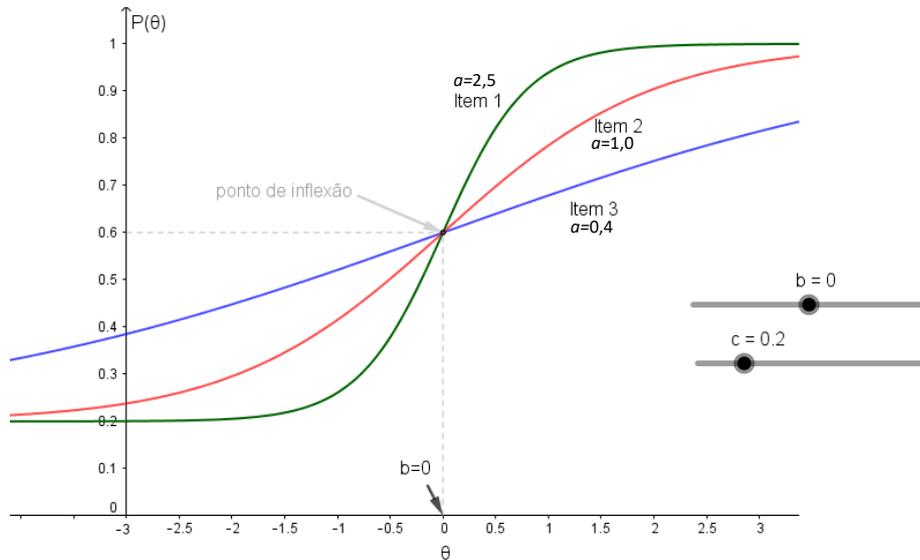
Fonte: elaboração do autor.

Andrade, Tavares e Valle (2000, p. 11), destacam que

[i] baixos valores de  $a$  indicam que o item tem pouco poder de discriminação (alunos com habilidades bastante diferentes têm aproximadamente a mesma probabilidade de responder corretamente ao item) e [ii] valores muito altos indicam itens com curvas características muito “íngremes”, que discriminam os alunos basicamente em dois grupos: os que possuem habilidades abaixo do valor do parâmetro  $b$  e os que possuem habilidades acima do valor do parâmetro  $b$ .

A situação apresentada em [i] pode ser notada pelo comportamento do gráfico do item 3 da figura a seguir. Já a situação [ii] está exemplificada pelo gráfico do item 1.

Figura 8 -Curvas características dos itens para diferentes valores do parâmetro de discriminação (parâmetro  $a$ ).



Fonte: elaboração do autor.

Na Figura 8, os itens apresentados possuem índices de discriminação distintos. Nota-se que a curva do item 1 é mais íngreme que a do item 3, e isto significa que o item 1 é mais discriminativo que o item 3, estando o item 2 com uma discriminação intermediária em relação aos outros dois.

Conforme Rabelo (2013), a interpretação a partir da classificação do parâmetro de discriminação pode ser realizada com base nas informações contidas na Tabela 2 a seguir, segundo o intervalo que ele está situado.

Tabela 2 - Classificação para o parâmetro de discriminação (parâmetro  $a$ ) de um item pela TRI.

Classificação	Valores do parâmetro de discriminação
Nenhuma	0
Muito baixa	(0 ; 0,35]
Baixa	(0,35 ; 0,65]
Moderada	(0,65 ; 1,35]
Alta	(1,35 ; 1,70]
Muito alta	(1,70 ; +∞)

Fonte: Rabelo (2013).

Assim, por exemplo, o item 1 que tem o parâmetro  $a = 2,5$  significa que ele possui uma discriminação muito alta; já o item 3 que apresenta  $a=0,4$  tem baixa discriminação, indicando

que o item não discrimina de forma significativa os sujeitos com baixa ou alta habilidade. Em geral, segundo Nakano, Primi e Nunes (2015, p. 102), “valores entre 0,6 e 1,8 são considerados adequados para a composição de uma boa medida, sendo que quanto maior a discriminação, melhor o item”.

Portanto, como visto nesta seção, compreender as características dos itens constitui um fundamento básico para interpretar ou construir testes educacionais quando se utiliza a TRI.

## 2.7 Probabilidade de acerto em um item e a construção da Curva Característica do Item (CCI)

Mesmo considerando o fato de que na prática os cálculos da TRI são realizados por *softwares* computacionais devido à sua complexidade, é importante entender, por exemplo, como se calcula a probabilidade de acerto a item em um teste, e como a Curva Característica do Item (CCI) é construída, pois isso constitui uma importante etapa para a compreensão dos fundamentos básicos da TRI.

Assim, pode-se perguntar: qual a probabilidade de um item ser acertado quando respondido por sujeitos com habilidades ( $\theta$ ) distintas? Tomando-se o modelo logístico unidimensional da TRI de 1, 2 ou 3 parâmetros, a probabilidade de um sujeito  $j$  com habilidade  $\theta$  acertar a um item  $i$  só pode ser calculada quando é conhecida a habilidade de cada sujeito, além dos parâmetros de cada item.

Nesse sentido, supondo que já tenha sido feita a estimação dos parâmetros dos itens, a construção da CCI, tomado o modelo logístico de 3 parâmetros (3PL), será ilustrada por meio de um exemplo. Para tanto, os valores desses parâmetros estão evidenciados na Tabela 3 abaixo, considerando  $\theta$  em um intervalo de -3 a 3 com incremento de 0,5 unidade.

Tabela 3 - Parâmetros dos cinco itens do teste hipotético estimados a partir do modelo logístico 3PL.

Item (i)	Parâmetro da discriminação $a_i$	Parâmetro da dificuldade $b_i$	Parâmetro do acerto ao acaso $c_i$
Item 1	1,29	0,69	0,13
Item 2	1,14	-0,23	0,15

Item (i)	Parâmetro da discriminação $a_i$	Parâmetro da dificuldade $b_i$	Parâmetro do acerto ao acaso $c_i$
Item 3	1,11	0,16	0,16
Item 4	0,67	0,85	0,18
Item 5	1,83	0,75	0,13

Fonte: elaboração do autor.

Para calcular a probabilidade de acerto nos itens para os diferentes valores de  $\theta$ , basta calcular as probabilidades para cada caso. Por exemplo: para um sujeito com  $\theta = -0,5$ , tem-se que a probabilidade de acerto desse sujeito no item 5, cujo parâmetro  $a$  é 1,83, o parâmetro  $b$  é 0,75 e o parâmetro  $c$  é 0,13 (ver Tabela 3), aplicada a expressão (11) e pondo  $D=1$ , será:

$$P(U_5 = 1 | \theta = -0,5) = c_5 + \frac{1 - c_5}{1 + e^{-1a_5(\theta - b_5)}} = 0,13 + \frac{1 - 0,13}{1 + 2,72^{-1,83[-0,5 - 0,75]}} \approx 0,210$$

Como pôde ser visto, o cálculo da probabilidade de uma resposta correta em um nível de habilidade é de simples resolução quando se usa o modelo logístico, desde que os parâmetros do item sejam conhecidos. Na Tabela 4 a seguir, são apresentados os resultados das probabilidades de acerto em cada item no intervalo de habilidade de -3 a 3, com incrementos de 0,5 unidade.

Tabela 4 - Probabilidade de acerto em cada um dos cinco itens para diferentes níveis de habilidades utilizando o modelo logístico 3PL (D=1).

$\theta$	$P(\theta)$				
	Item 1	Item2	Item 3	Item 4	Item 5
-3,0	0,137	0,185	0,184	0,238	0,131
-2,5	0,144	0,209	0,202	0,259	0,132
-2,0	0,156	0,250	0,230	0,286	0,136
-1,5	0,179	0,312	0,275	0,321	0,144
-1,0	0,218	0,400	0,342	0,364	0,164
-0,5	0,284	0,510	0,433	0,416	0,210
0,0	0,383	0,630	0,543	0,476	0,306
0,5	0,512	0,742	0,658	0,542	0,467
1,0	0,651	0,832	0,763	0,611	0,663

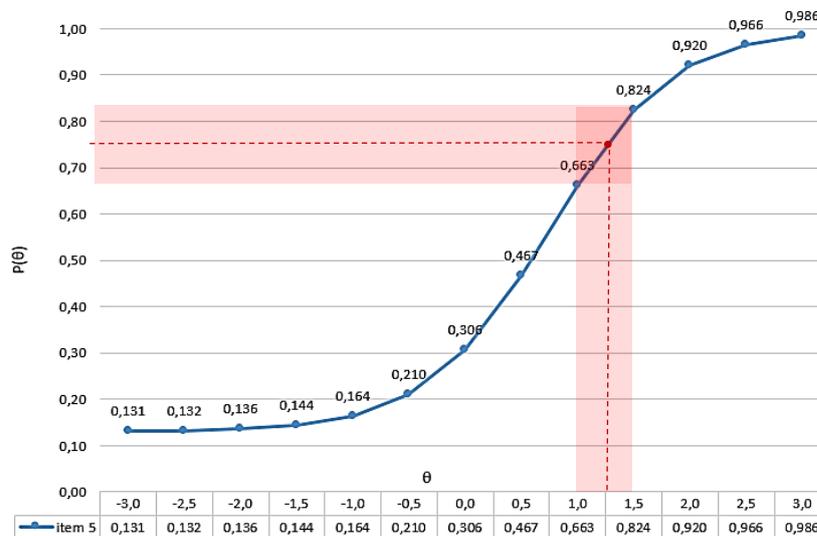
$\theta$	$P(\theta)$				
	Item 1	Item 2	Item 3	Item 4	Item 5
1,5	0,774	0,896	0,845	0,678	0,824
2,0	0,864	0,938	0,904	0,741	0,920
2,5	0,923	0,964	0,942	0,796	0,966
3,0	0,958	0,979	0,966	0,843	0,986

Fonte: elaboração do autor.

Nota:  $P(\theta)$  representa a probabilidade de acerto em um item para os diferentes níveis de habilidade  $\theta$ .

A partir dos dados da Tabela 4, foram construídos os gráficos das curvas características dos itens do exemplo discutido acima, utilizando-se a ferramenta Gráfico de Dispersão, do *software* Excel. A Figura 9, a seguir, evidencia uma boa aproximação da curva característica do item 5 e a Figura 10, mais abaixo, apresenta essas aproximações para as curvas características para todos os cinco itens.

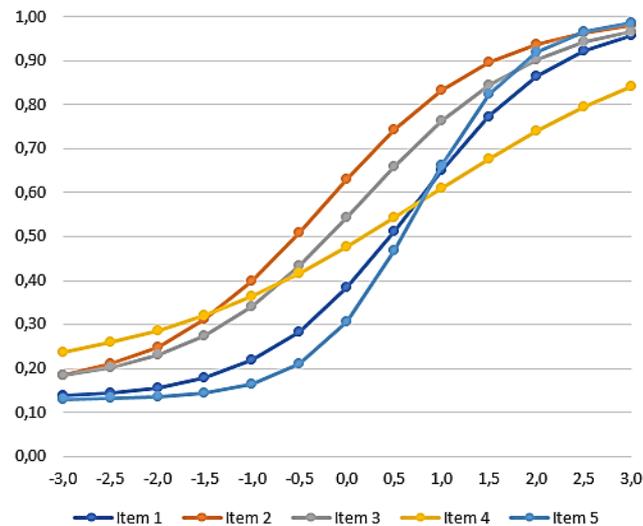
Figura 9 - Curva característica do item 5 conforme o modelo 3PL.



Fonte: elaboração do autor.

Observando o gráfico acima pode-se notar, por exemplo, que um sujeito que possui  $\theta=1,3$  tem aproximadamente 75% de probabilidade para acertar esse item. Além disso, o item 5 possui discriminação muito alta, conforme a classificação da Tabela 2, indicando possuir alta capacidade de diferenciar os sujeitos com baixa habilidade dos com alta habilidade em relação à sua dificuldade, o que não ocorre, por exemplo, com o item 4 apresentado na Figura 10 a seguir, que possui baixa capacidade de discriminação entre os sujeitos desse teste.

Figura 10 - Curva características dos 5 itens do teste hipotético conforme o modelo 3PL.



Fonte: elaboração do autor.

Como já dito na Seção 2.5, a escala da habilidade é contínua. Entretanto, no exemplo apresentado acima, foram tomados apenas 13 valores nessa escala com intuito de ilustrar a construção da CCI para cada item, a partir do gráfico de dispersão de dados. Portanto, conhecendo-se os parâmetros dos itens é possível construir boas aproximações das curvas de probabilidade de resposta correta para os diferentes níveis de habilidades ( $\theta$ ). Em geral, esses parâmetros não são conhecidos, mas podem ser estimados como apresentado na Seção 3. Ademais, um melhor ajuste dessa curva pode ser obtido a partir um processo de regressão logística, mas que não será tratado nesse trabalho.

## 2.8 Função de Informação

O conceito de informação psicométrica é um aspecto importante para análises dentro da Teoria de Resposta ao Item. Nesse sentido, segundo Pasquali (2018) a Função de Informação do Item é definida como o montante de informação psicométrica que um item contém ao longo da escala de habilidade, e é a partir dela que se permite analisar o quanto de informação um item tem para a medida da habilidade. A Função de Informação do Item é dada por (BAKER; KIM, 2004, p. 71):

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (12)$$

em que  $I_i(\theta)$  é a informação fornecida pelo item  $i$  no nível de habilidade  $\theta$ ;  $P_i(\theta)$  é a probabilidade de acerto no item  $i$  e  $Q_i(\theta) = 1 - P_i(\theta)$ . Além disso,  $P_i'(\theta) = \partial P_i(\theta) / \partial \theta$ .

Para o modelo logístico de três parâmetros (3PL) a equação (12) pode ser escrita como (ANDRADE; TAVARES; VALLE, 2000, p.12):

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)} \left[ \frac{P_i(\theta) - c_i}{1 - c_i} \right]^2 \quad (13)$$

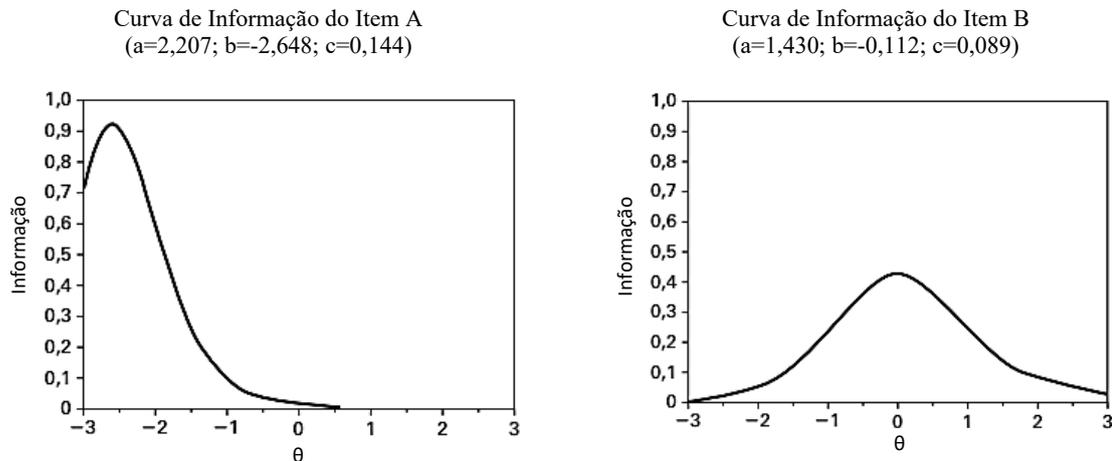
Essa equação mostra que os três parâmetros têm influência sobre o montante de informação do item. Desse modo, Andrade, Tavares e Valle (2000) dizem que essa informação do item é maior quando  $b_i$  se aproxima de  $\theta$ , quanto maior for o  $a_i$  e quanto mais  $c_i$  se aproximar de 0. Assim, quanto maior a quantidade de informações em um determinado nível de habilidade, mais próximas dos valores exatos serão as estimativas da habilidade e dessa forma pode-se observar uma estimativa mais precisa (BAKER; KIM, 2004).

Desse modo, Birnbaum (1968, p. 463) derivou a expressão (13) para encontrar o nível de habilidade quando o valor de informação do item,  $I(\theta)$ , é máximo:

$$\theta_{MAX} = b_i + \frac{1}{Da_i} \ln \left( \frac{1 + \sqrt{1 + 8c_i}}{2} \right) \quad (14)$$

O gráfico gerado a partir da função (12) é denominado Curva de Informação do Item (CII). A figura seguir apresenta a curva de informação dos itens A e B (hipotéticos) para ilustrar esses gráficos.

Figura 11 - Curvas de Informação dos itens A e B.



Fonte: Arias, Lloreda e Lloreda (2014).

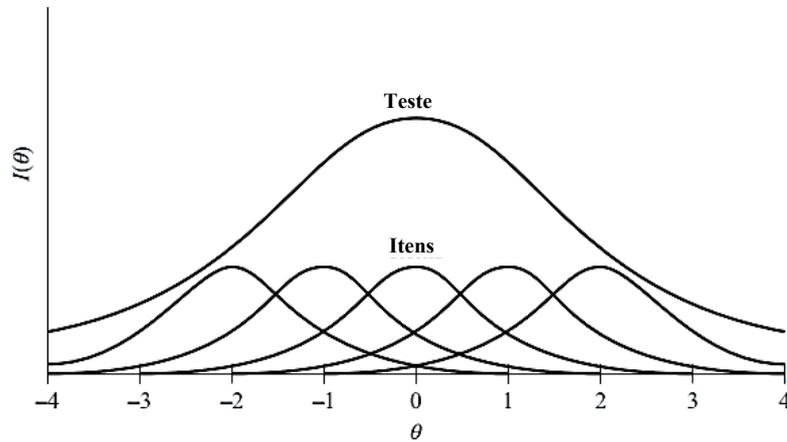
Na figura acima nota-se que o item A fornece o máximo de informações para sujeitos com  $\theta$  em torno de -2,6; já o item B em torno de  $\theta = 0$ . Além disso, o gráfico mostra que o item A fornece muita informação, mas segundo Arias, Lloreda, Lloreda (2014) por ser um item muito fácil a informação só é boa em níveis muito baixos de habilidades e só seria útil quando a análise se concentra em níveis de habilidade muito baixos. Já o item B, embora a quantidade de informações que possua seja menor, ele é representativo em uma ampla faixa de habilidade.

Por outro lado, um teste completo fornece mais informações do que cada um dos itens que o compõe. Nesse sentido, além da função de informação do item, outra medida também pode ser utilizada para fornecer informações ao instrumento de medida: a Função de Informação do Teste. Ela consiste na soma das informações fornecidas por cada item que compõe o teste e que foi definida por Birnbaum, em 1968, como (BAKER; KIM, 2004, p. 70):

$$I(\theta) = \sum_{i=1}^n \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (15)$$

Desse modo, a Função de Informação do Teste (15) descreve uma curva como a ilustrado na Figura 12 a seguir.

Figura 12 – Função de Informação de cinco itens e Função de Informação do teste formado por esses itens.



Fonte: MUÑIZ (2018).

Outra maneira de representar uma função de informação para o teste é através do erro-padrão de medida, chamado na TRI de erro-padrão de estimação, que é dado por (ANDRADE; TAVARES; VALLE, 2000, p.13):

$$EP(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (16)$$

Muñiz (2018) diz que faz sentido conceitual chamar a função (16) de uma "função de informação" uma vez que quanto mais alto o  $I(\theta)$  menor será o erro-padrão de estimativa e, portanto, maior será a informação que os estimadores conseguem explicar sobre a variável  $\theta$ . Nesse sentido, a função de informação pode ser considerada um indicador de precisão do teste.

Portanto, como foi visto nesta seção, a Função de Informação do Item (ou do teste), bem como as Curvas de Informação do Item (ou do teste), mostram-se como medidas, que segundo Muñiz (2018) traduzem confiabilidade, validade e eficiência para um teste. Ademais, essas medidas constituem uma poderosa ferramenta para o construtor do teste, pois essas informações possibilitam a construção de testes personalizados de forma mais facilitada, permitindo a escolha de itens para serem aplicados tanto à examinandos comuns, como em uma distribuição normal, quanto para examinandos com baixas ou altas habilidades.

### 3 ESTIMAÇÃO DOS PARÂMETROS DOS ITENS E HABILIDADES NA TRI

Um dos objetivos principais da TRI é a estimação dos parâmetros dos itens e das habilidades dos respondentes a um teste. Contudo, após a realização de um teste são apenas conhecidas as respostas dos sujeitos. Assim, há o interesse em estimar tanto os parâmetros dos itens quanto as habilidades desses sujeitos. Normalmente os dados empíricos de uma amostra, após uma aplicação de um teste para  $n$  sujeitos e  $I$  itens, são apresentados segundo uma matriz de  $n$  por  $I$ ; quando o teste é de habilidade, as respostas dos sujeitos serão da forma 1 (para itens acertados) ou 0 (para itens errados). A Tabela 5 e a Figura 13, apresentadas a seguir, ilustram esse fato considerando um teste com 10 itens realizado por 5 sujeitos.

Tabela 5 - Respostas de cinco sujeitos em um teste de dez itens.

Sujeito ( $n$ )	Item ( $I$ )										Total de acertos
	1	2	3	4	5	6	7	8	9	10	
1	1	0	0	0	0	0	0	0	0	0	1
2	1	1	0	0	1	0	0	0	0	0	3
3	1	0	1	1	0	1	0	1	1	1	7
4	1	1	1	1	0	0	0	1	1	1	7
5	1	1	1	1	1	1	1	1	1	1	10

Fonte: elaboração do autor.

Figura 13 - Matriz 5 x 10 das respostas de cinco sujeitos em um teste de dez itens.

1	0	0	0	0	0	0	0	0	0
1	1	0	0	1	0	0	0	0	0
1	0	1	1	0	1	0	1	1	1
1	1	1	1	0	0	0	1	1	1
1	1	1	1	1	1	1	1	1	1

Fonte: elaboração do autor.

O que se quer saber, a partir dos dados da tabela acima, é o quanto de habilidade que cada um desses cinco sujeitos possui e quais são os parâmetros que caracterizam esses itens.

Diante das informações apresentadas na Tabela 5, pode-se perguntar, por exemplo: será que os sujeitos 3 e o 4, que acertaram a mesma quantidade de itens, terão o mesmo nível de habilidade? E porque todos acertaram o item 1, mas apenas o sujeito 5 acertou o item 7?

Segundo Pasquali (2018) questões como essas são o objetivo da TRI; porém há um problema a ser resolvido, pois apenas as respostas dos sujeitos não produzem soluções para esses questionamentos. Portanto, *a priori*, não sabemos o nível de habilidade de nenhum sujeito nem as características psicométricas dos itens do teste. Assim, do ponto de vista teórico, conforme Andrade, Tavares e Valle (2000, p. 27), esse problema poderá ser resolvido quando observada uma das três situações:

- [1] já conhecemos os parâmetros dos itens, temos apenas que estimar as habilidades;
- [2] já conhecemos as habilidades dos respondentes, estaremos interessados apenas na estimação dos parâmetros dos itens;
- [3] desejamos estimar os parâmetros dos itens e as habilidades dos indivíduos simultaneamente.

Nas duas primeiras situações, quando são conhecidos os parâmetros ou as habilidades, estimar aquilo que falta é relativamente mais simples. Por outro lado, a estimação observada na situação [3], que é o caso mais complexo, é mais comum de ocorrer porque tanto os parâmetros dos itens quanto as habilidades, em geral, são desconhecidas. Avaliações como o Saesp, Enem e Saeb utilizam-se das situações [1] e [3]: primeiro os parâmetros dos itens são estimados previamente em uma pré-testagem - como em [3] - e os dados obtidos são analisados para que os itens possam ser arquivados em um banco de itens<sup>12</sup>. Daí, em um momento *a posteriori*, são construídos os testes com os itens, cujos parâmetros já foram estimados, com o intuito de estimar as habilidades/proficiências dos sujeitos -como na situação [1]. Na TRI, o processo de estimação dos parâmetros dos itens é conhecido como *calibração* dos itens.

Na verdade, as questões apresentadas nas situações [1], [2] e [3] são resolvidas a partir da utilização de alguns métodos estatísticos, como os de Máxima Verossimilhança e os Bayesianos. O quadro a seguir apresenta alguns desses métodos de estimação:

---

<sup>12</sup> Ver Seção 4.2- Banco de itens para avaliações educacionais, p. 62.

Quadro 2- Métodos de estimação para os parâmetros dos itens e habilidades.

Estimar os parâmetros dos itens	Estimação das habilidades	Estimação dos parâmetros dos itens e das habilidades
<ul style="list-style-type: none"> <li>- Máxima Verossimilhança (MV)</li> <li>-Máxima Verossimilhança Marginal (MVM).</li> <li>- Bayesiano.</li> </ul>	<ul style="list-style-type: none"> <li>- Máxima Verossimilhança (MV)</li> <li>- Máximo a Posteriori (MAP).</li> <li>- Esperança a Posteriori (EAP).</li> </ul>	<ul style="list-style-type: none"> <li>- Máxima Verossimilhança Conjunta (MVC).</li> <li>- Máxima Verossimilhança Marginal (MVM)</li> <li>-Bayesiano</li> </ul>

Fonte: VALLE (2000), ANDRADE; TAVARES; VALLE (2000).

Para tanto, a título de ilustração, será exemplificado aqui apenas o mais comum utilizado em avaliações: o método de Máxima Verossimilhança. Os demais métodos para as estimativas da TRI, descritos no Quadro 2, estão escritos em ampla literatura, como em Andrade, Tavares e Valle (2000); Valle (2000); e Baker e Kim (2004).

### 3.1 Método de Máxima Verossimilhança para estimação dos parâmetros dos itens e das habilidades

Inicialmente, partir das hipóteses da independência local, considere-se um teste composto por  $I$  itens  $i$  respondidos por  $n$  sujeitos  $j$  ( $i \in \{1, 2, \dots, I\}$  e  $j \in \{1, 2, \dots, n\}$ ), levando em conta a habilidade  $\theta_j$  e o vetor  $\zeta_i$ , que representa os parâmetros  $a_i, b_i$  e  $c_i$  do item  $i$  e simplesmente por  $\zeta$  o conjunto de todos  $\zeta_i$ . Ademais,  $u_j$  é vetor das respostas do sujeito  $j$ , em que em que  $u_{ij}$  representa a resposta do sujeito  $j$  ao item  $i$  ( $u_{ij}=1$ , se o sujeito responde corretamente e  $u_{ij}=0$ , caso contrário).

A equação (2), da Subseção 2.3.2, trata da probabilidade do padrão de resposta antes do fato (*a priori*); mas, quando o padrão de resposta já foi observado, após a coleta dos dados empíricos, essa interpretação da probabilidade (*a posteriori*, ou seja, probabilidade após o fato) é denominada *verossimilhança*, que para um sujeito  $j$ , condicionado a sua habilidade  $\theta_j$  e  $I$  itens  $i$  é denotada por (VALLE, 2000):

$$L(u_j | \theta_j, \zeta) = \prod_{i=1}^I P_{ji}^{u_{ij}} Q_{ji}^{1-u_{ij}} \quad (17)$$

No caso geral, a função de *verossimilhança conjunta* baseada nas respostas dos  $n$  sujeitos, pode ser escrita como (VALLE, 2000):

$$L(u | \theta, \zeta) = \prod_{j=1}^n L(u_j | \theta_j, \zeta) = \prod_{j=1}^n \prod_{i=1}^I P_{ji}^{u_{ij}} Q_{ji}^{1-u_{ij}} \quad (18)$$

Daí, com o intuito de simplificar os cálculos, a função (18) normalmente é transformada em uma função logarítmica aplicando  $\ln$  a ambos os membros, em que se obtém a função *ln-verossimilhança* (VALLE, 2000):

$$\ln L(u | \theta, \zeta) = \sum_{j=1}^n \sum_{i=1}^I [u_{ij} \cdot \ln P_{ji} + (1-u_{ij}) \cdot \ln Q_{ji}] \quad (19)$$

Assim, o objetivo é calcular o  $\zeta_i$  e  $\theta_j$  que maximiza a função de verossimilhança. Desse modo, os estimadores de máxima verossimilhança de  $\zeta_i$  e  $\theta_j$ , em que  $i = 1, \dots, I$  e  $j = 1, \dots, n$ , são os valores que maximizam a função (19) (ou a função (18)), ou seja, são as soluções da equação, conforme o caso (ANDRADE; TAVARES; VALLE, 2000):

- Estimar os parâmetros dos itens conhecidas as habilidades dos sujeitos:  $\frac{\partial \ln L(\zeta)}{\partial \zeta_i} = 0$
- Estimar as habilidades dos sujeitos conhecidos os parâmetros dos itens:  $\frac{\partial \ln L(\theta)}{\partial \theta_j} = 0$
- Estimar conjuntamente os parâmetros dos itens e as habilidades dos sujeitos:  $\frac{\partial \ln L(\theta, \zeta)}{\partial \theta_j \partial \zeta_i} = 0$

Assim, calculando-se as derivadas parciais de primeira ordem da função com intuito de maximizar cada expressão acima, pode-se estimar os parâmetros  $a_i$ ,  $b_i$ ,  $c_i$  e  $\theta_j$ . Para tanto, as

equações encontrada não possuem solução explícita, não podendo ser resolvida diretamente, e por isso é necessário algum método iterativo para a obtenção das estimativas de máxima verossimilhança dos parâmetros dos itens e para as habilidades dos sujeitos. Em geral, utiliza-se o algoritmo Newton-Raphson ou “Scoring” de Fischer<sup>13</sup> para encontrar essas estimativas.

Ressalta-se que o método de Máxima Verossimilhança é o mais comum e utilizado nos procedimentos de estimação na TRI. Entretanto, como a maioria dos métodos de estimação, o método de Máxima Verossimilhança apresenta alguns problemas, entre eles quando o indivíduo erra ou acerta todos os itens, que nesse caso produz um  $\theta$  infinito. Entretanto, Andrade, Tavares e Valle (2000) afirmam que os métodos bayesianos apresentam uma solução em que esses problemas são contornados.

Portanto nessa seção foi tratado sobre as estimativas para encontrar os parâmetros dos itens e as habilidades dos examinandos, em especial foi apresentado um dos métodos mais comuns de estimação na TRI: a Máxima Verossimilhança. Entretanto, independentemente do método estatístico selecionado para esse fim, em geral os cálculos são realizados por *softwares* computacionais por sua complexidade, como será ilustrado na Seção 3.2 a seguir.

### 3.2 Recursos computacionais na TRI

Atualmente a TRI é uma teoria bastante difundida e utilizada em diversas avaliações ao redor do mundo, mas o seu crescimento e popularização estiveram atrelados ao desenvolvimento dos computadores e *softwares* que viabilizassem a sua utilização, pois, na prática, os cálculos estatísticos envolvidos nessa teoria são demasiadamente trabalhosos, não sendo realizáveis manualmente ou por calculadora.

Embora alguns pesquisadores prefiram desenvolver os seus próprios *softwares* computacionais para esse fim, em geral os cálculos são realizados a partir da utilização *softwares* já consolidados, sejam eles gratuitos ou não. Os *softwares* mais utilizados para estatísticas da TRI são o BILOG, o BILOG-MG, o TESTFACT<sup>14</sup> e o PASCALE, que são pagos;

---

<sup>13</sup> Ver em Andrade, Tavares e Valle (2000).

<sup>14</sup> O TESTFACT produz várias estatísticas utilizando a TCT para os itens de um teste apresentados na forma dicotômica. Também tem recursos importantes para a TRI, usados na verificação da dimensionalidade dos testes, como técnicas de análise fatorial específicas para serem aplicadas em itens (ANDRADE; TAVARES; VALLE, 2000; ARIAS; LLOREDA; LLOREDA, 2014).

e outros, como o *software* R ou o EIRT, são gratuitos. Uma lista ampliada dos *softwares* aplicados à TRI pode ser observada no quadro a seguir:

Quadro 3 - Alguns programas de computador para estimação na TRI.

<b>Programa</b>	<b>Modelo</b>	<b>Desenvolvedor(es)/ano</b>
BIGSTEPS	Modelo de Rasch e derivados	Linacre e Wright (1998)
BILOG	1p, 2p, 3p	Mislevy e Bock (1984)
BILOG-MG	1p, 2p, 3p, DIF e equiparação	Zimowski, Muraki, Mislevy e Bock: <a href="http://www.ssicentral.com">http://www.ssicentral.com</a>
ConQuest	1p, multi-categoria, multidimensional	Adams, Wu e Wilson (2015)
DIMENSION	Multidimensional	Hattie e Krakowski (1994)
flexMIRT	Multidimensional	Cai (2013)
IRTPRO	1p, 2p, 3p, multi-categoria	Cai, Thissen e Du Toit (2011)
LOGIST	1p, 2p, 3p	Wingersky (1983). Wingersky et al. (1982)
MICROSCALE	Multicategoria (1p)	Medias Interactive Technologies (1986)
MIRTE	1p, 2p, 3p	Carlson (1987)
MULTILOG	Multi-categoria	Thissen, Chen e Bock: <a href="http://www.ssicentral.com">http://www.ssicentral.com</a>
NOHARM	1p, 2p, 3p, multidimensional	Fraser e McDonald (1988)
PARSCALE	1P, 2P, 3P, multi-categoria	Muraki y Bock (1991): <a href="http://www.ssicentral.com">http://www.ssicentral.com</a>
PML	1p	Gustafsson (1980)
RASCAL	1p	Assessment System Corporation: <a href="http://www.assess.com">http://www.assess.com</a>
RIDA	1p	Glas (1990)
RUMM2030	1p, 2p, 3p, multi-categoria	Andrich y Luo: <a href="http://www.rummlab.com.au/">http://www.rummlab.com.au/</a>
<i>Software</i> R	1p, 2p, 3p, multi-categoria	R Core Team (2014)
WinGen	Gerar dados TRI	Han (2007): <a href="http://www.hantest.net/wingen">http://www.hantest.net/wingen</a>
WinSteps	Modelo de Rasch e derivados	Linacre (2015): <a href="http://www.winsteps.com">http://www.winsteps.com</a>
XCALIBRE	1p, 2p, 3p	<a href="http://www.assess.com">http://www.assess.com</a>

Fonte: MUNIZ (2018, p. 211).

Nota: 1p: modelo de Rasch, 2p: modelo logístico de 2 parâmetros, 3p: modelo logístico de 3 parâmetros.

No Brasil os *softwares* mais utilizados para análises estatísticas da TRI é o BILOG e o BILOG-MG<sup>15</sup>. Esses dois programas permitem fazer as estimativas para os modelos com 1, 2 ou 3 parâmetros para itens dicotômicos ou dicotomizados, em métricas logísticas ou normais. Ademais, pode-se obter as curvas características e de informação de cada item e do teste, além de calcular alguns elementos da TCT, como, por exemplo, porcentagem de acerto em cada item, correlação bisserial e ponto-bisserial. Em geral, a diferença básica entre os dois programas é que o BILOG-MG permite a análise de mais de um grupo de respondentes, enquanto o BILOG permite apenas análises de respondentes considerados como proveniente de uma única população (ANDRADE; TAVARES; VALLE, 2000).

Por outro lado, Arias, Lloreda e Lloreda (2014) afirmam que os *softwares* livres e/ou gratuitos, para estimação utilizando os modelos da TRI, são muito escassos. Nesse sentido, o *software* R, que usa a linguagem de programação R, é uma boa opção para análises e manipulação de dados da TRI (e também da TCT). As análises estatísticas realizadas com o uso desse *software* dão-se a partir de pacotes, que contêm funções pré-programadas. Assim, na medida em que o usuário necessita realizar uma análise estatística específica, é necessária a instalação de um pacote que contém a função pretendida para essa análise. Ademais, o R também permite que o usuário desenvolva suas próprias funções de acordo com as suas necessidades, constituindo assim uma grande vantagem em relação aos *softwares* congêneres<sup>16</sup>.

Como alternativa aos programas computacionais mencionados, este estudo evidenciará o uso do *software* EIRT, que é um suplemento do Excel para o cálculo das estatísticas relativas a TRI em testes educacionais. Esta escolha se deu por ele ser livre, gratuito e de fácil utilização, não necessitando que o usuário tenha o domínio de linguagem de programação, como no caso do *software* R.

---

<sup>15</sup> O BILOG-MG é o *software* utilizado para produzir as estatísticas em diversas avaliações do Saeb, como na Avaliação Nacional do Rendimento Escolar (ANRESC) – também conhecida como Prova Brasil, que adota procedimentos de equalização sob a metodologia de múltiplos grupos, estimação via Máxima Verossimilhança Marginal (MVM) para calibração dos itens e estimação via valor Esperado a Posteriori (EAP) para cálculo das proficiências (BRASIL, 2015).

<sup>16</sup> Para mais detalhes sobre a utilização do *software* R para a TRI ver: DOS ANJOS; ANDRADE (2012) e/ou BAKER; KIM (2017).

### 3.2.1 O *software* EIRT

O *software* EIRT<sup>17</sup> é um suplemento (*add-in*) para o Excel que faz alguns cálculos da TRI e TCT aplicados aos modelos unidimensionais. Além disso, os itens para as análises podem ser do tipo binário, de múltipla escolha e graduado. Por fim, tem-se que os modelos suportados inclui os logísticos com um, dois e três parâmetros para os dados binários, o modelo de resposta nominal para os dados de múltipla escolha e o modelo de resposta gradual para os dados graduados.

Além disso, para o caso de estimadores paramétricos são utilizados os métodos tais como o Estimador de Máxima Verossimilhança Marginal e o Estimador Bayesiano Modal; já para o caso não-paramétrico são utilizados o método Kernel Smoothing e o Estimador de Máxima Verossimilhança Marginal Penalizada. Ademais, para as estimativas das habilidades são utilizados os métodos: Esperança a Posteriori (EAP - *Expected A Posteriori*) e Estimador de Máxima Verossimilhança Ponderada (WML - *Weighted Maximum Likelihood*) (GERMAIN; VALOIS; ABDOUS, 2007).

Assim, com o intuito de ilustrar como se obtém as estatísticas da TRI (e também da TCT) em um teste, utilizando o *software* EIRT, tomou-se como base de dados uma situação hipotética composta de uma amostra de 100 respondentes a um teste e submetidos a 20 itens - em que cada item possui uma única alternativa correta dentre quatro possibilidades. Ademais, questões sem marcação (em branco) foram corrigidas como erradas.

Para tanto, a matriz das respostas desses sujeitos encontra-se no **APÊNDICE C** (ou na forma dicotômica no **APÊNDICE D**), e estão apresentadas parcialmente nas matrizes da Figura 14 e Figura 15. Na Figura 14 são apresentadas as respostas dos sujeitos para cada item no teste; já na Figura 15 é apresentada uma matriz de forma dicotômica (binária), após a correção das respostas, onde 1 representa o acerto e 0 o erro no item. Desse modo, ressalta-se que o *software* EIRT pode ser utilizado tanto para os dados apresentados como os da Figura 15 quanto os da Figura 14. Destaca-se que nesta subseção os dados utilizados se apresentam como os da Figura 14.

---

<sup>17</sup> O EIRT (em inglês: *Item Response Theory applications using Excel add-in*) foi desenvolvido por Stéphane Germain, Pierre Valois e Belkacem Abdous. As informações sobre esse programa estão disponíveis em: <<https://libirt.psychometricon.net/>>. Acesso em: 20 mai. 2021.

Figura 14 – Trecho da matriz de respostas dos 100 sujeitos submetidos a 20 itens.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
2	<b>GABARITO</b>	<b>A</b>	<b>B</b>	<b>D</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>D</b>	<b>B</b>	<b>D</b>	<b>D</b>	<b>A</b>	<b>B</b>	<b>D</b>	<b>D</b>
3	SUJEITO 1	A	D	D	C	A	A	D	B	D	B	A	B	D	D	A	B	A	B	C	B
4	SUJEITO 2	D	B	A	C	A	A	C	D	A	B	A	C	D	B	B	A	A	B	D	D
5	SUJEITO 3	A	B	D	C	A	A	B	C	D	D	A	A	D	B	D	D	A	B	D	A
6	SUJEITO 4	D	C	D	C	A	C	A	B	A	D	C	C	C	C	A	D	C	B	C	A
7	SUJEITO 5	D	B	D	C	A	B	C	D	B	A	A	B	D	B	D	C	C	D	B	D

⋮

101	SUJEITO 99	A	B	D	C	A	A	B	C	D	C	A	A	D	B	D	D	A	B	D	D
102	SUJEITO 100	D	C	B	D	B	B	C	D	A	B	C	B	C	A	B	B	C	D	A	A

Fonte: elaboração do autor.

Figura 15 – Trecho da matriz de respostas dos 100 sujeitos submetidos a 20 itens na forma dicotômica após a correção.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
2																					
3	SUJEITO 1	1	0	1	1	1	1	0	0	1	0	1	0	1	0	0	0	1	1	0	0
4	SUJEITO 2	0	1	0	1	1	1	0	0	0	0	1	0	1	1	0	0	1	1	1	1
5	SUJEITO 3	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0
6	SUJEITO 4	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
7	SUJEITO 5	0	1	1	1	1	0	0	0	0	0	1	0	1	1	1	0	0	0	0	1

⋮

101	SUJEITO 99	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
102	SUJEITO 100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

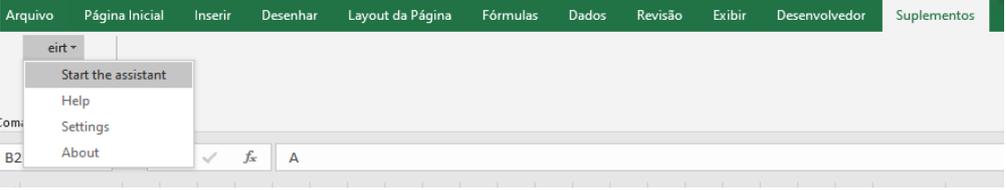
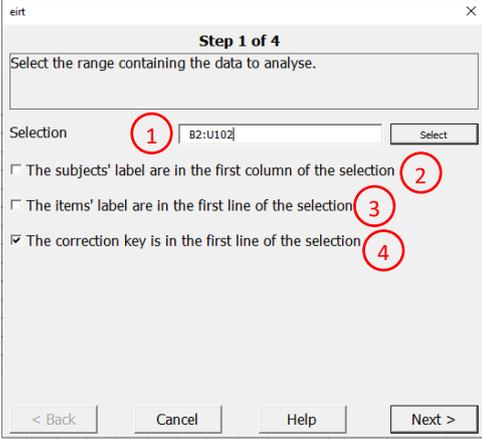
Fonte: elaboração do autor.

Nota: 1 = acerto; 0 = erro.

Assim, a partir das respostas dos sujeitos, dispostas em uma matriz como apresentado acima, e utilizando o EIRT, pode-se calcular as estatísticas relativas a TRI tais como as curvas características do item, parâmetros dos itens, curvas de informações dos itens, habilidade ( $\theta$ ) dos sujeitos, entre outros. Ademais, caso se queira, pode-se calcular também elementos da TCT como o escore total dos sujeitos, índice de dificuldade no item, correlação (bisserial) dos itens, entre outros.

Nesse sentido, com o intuito de fornecer ao professor uma noção de como as estatísticas da TRI (e/ou da TCT) são realizadas com o *software* EIRT, a partir dos dados empíricos

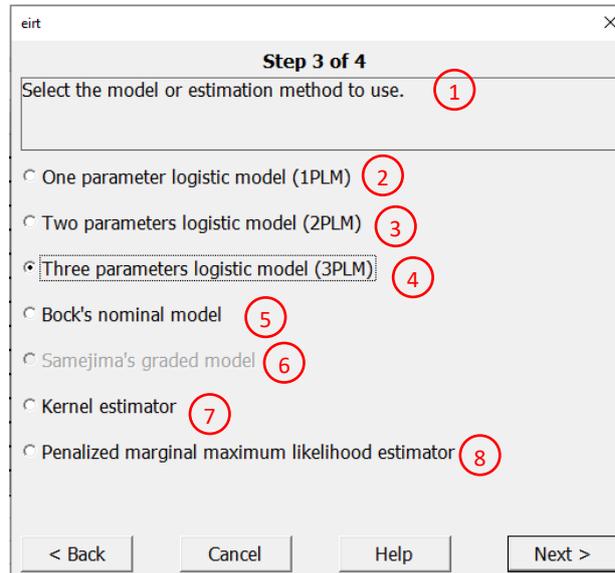


Passos	Descrição dos passos
3º passo:	<p>Selecionar o <i>software</i> EIRT na aba suplemento e, em seguida, clicar em “<i>Start the assistant</i>” para iniciar o assistente (<i>software</i>).</p> 
4º passo:	<p>Selecionar a opção 3 (<i>The correction key is in the frist line of the seleccion</i>) e clicar em “<i>Next&gt;</i>” para ir para o próximo passo.</p>  <ol style="list-style-type: none"> <li>1) Seleccione o intervalo contendo os dados para análise.</li> <li>2) ( ) Os rótulos dos sujeitos estão na primeira coluna da seleção.</li> <li>3) ( ) Os rótulos dos itens estão na primeira linha da seleção.</li> <li>4) (x) As chaves de correção (gabarito) estão na primeira linha da seleção.</li> </ol>
5º passo:	<p>Selecionar a opção 2 (<i>Multiple choice</i>). Assim, após selecionar essa opção, nota-se que a chave de correção (gabarito) já estará preenchida. Em seguida clicar em “<i>Next &gt;</i>” para ir para o próximo passo.</p>

Passos	Descrição dos passos
	<div data-bbox="660 309 1182 801" style="border: 1px solid black; padding: 5px;"> <p style="text-align: center;"><b>Step 2 of 4</b></p> <p>Select the type of data to analyse and specify the format. <span style="border: 1px solid red; border-radius: 50%; padding: 2px;">1</span></p> <p><input type="radio"/> Dichotomous <span style="border: 1px solid red; border-radius: 50%; padding: 2px;">2</span></p> <p><input checked="" type="radio"/> Multiple choice <span style="border: 1px solid red; border-radius: 50%; padding: 2px;">3</span></p> <p><input type="radio"/> Graded <span style="border: 1px solid red; border-radius: 50%; padding: 2px;">4</span></p> <p><input type="radio"/> Mixed <span style="border: 1px solid red; border-radius: 50%; padding: 2px;">5</span></p> <p>Success <input style="width: 100px;" type="text" value="1"/></p> <p>Missing value <input style="width: 100px;" type="text"/></p> <p>Answer key <input style="width: 150px;" type="text" value="A,B,D,C,A,A,B,C,D,C,A,A,D,B,D,D"/> 20 items <span style="border: 1px solid red; border-radius: 50%; padding: 2px;">7</span></p> <p style="text-align: center;"> <input style="width: 60px;" type="button" value=" &lt; Back "/> <input style="width: 60px;" type="button" value=" Cancel "/> <input style="width: 60px;" type="button" value=" Help "/> <input style="width: 60px;" type="button" value=" Next &gt; "/> </p> </div> <p>1) Seleccione o tipo de dados para analisar e especifique o formato.</p> <p>2) ( ) Dicotômicos</p> <p>3) (x) Múltipla escolha</p> <p>4) ( ) Graduado</p> <p>5) ( ) Misturado</p> <p>6) Valores ausentes (“questões em branco”)</p> <p>7) Chave de correção (“gabarito”)</p>

Selecionar “*Three parameters logistic model (3PLM)*”. Em seguida clicar em “*Next >*” para ir para o próximo passo.

6º passo:



- 1) Selecione o modelo ou método de estimação para usar.
- 2) ( ) Modelo logístico de um parâmetro (1PLM)
- 3) ( ) Modelo logístico de dois parâmetros (2PLM)
- 4) (x) Modelo logísticos de 3 parâmetros (3PLM)
- 5) ( ) Modelo nominal de Bock
- 6) ( ) Modelo [de resposta] graduado de Samejima
- 7) ( ) Estimador Kernel
- 8) ( ) Estimador de máxima verossimilhança marginal penalizada

Selecionar os elementos que serão estimados pela TRI e/ou TCT. Em seguida clique em “Next>” para gerar os relatórios relativos a TRI e também da TCT.

7º passo:

- 1) Selecione os elementos para incluir no relatório
- 2) ( ) Estatísticas da teoria clássica dos testes
- 3) ( ) Matriz de correlação
- 4) (x) Escore
- 5) ( ) Teste de ajuste
- 6) ( ) Teste de independência local
- 7) (x) Estimativas dos parâmetros
- 8) (x) Estimativas das variáveis latentes (habilidades)
- 9) (x) Curvas características
- 10) (x) Funções de informações (curvas de informação)
- 11) ( ) Erro padrão
- 12) ( ) Todos os pontos de dados para o gráfico

	13) ( ) Escore na escala da abscissa
--	--------------------------------------

Fonte: elaboração do autor.

O relatório gerado contém as informações conforme selecionadas nos passos descritos no Quadro 4. Nesse exemplo foi utilizado a opção para respostas de múltiplas escolhas, modelo logístico de três parâmetros e as estatísticas pretendidas foram: escore, estimativas dos parâmetros (TRI), estimativas das variáveis latentes (habilidades), curvas características (CCI) e de informação (CII) dos itens do teste.

Para tanto, os resultados são apresentados em um relatório contendo as informações selecionadas. Assim, por exemplo, pode-se destacar desse relatório os parâmetros dos itens, como observado na tabela a seguir:

Tabela 6 - Parâmetros dos 20 itens conforme a TRI estimados pelo EIRT.

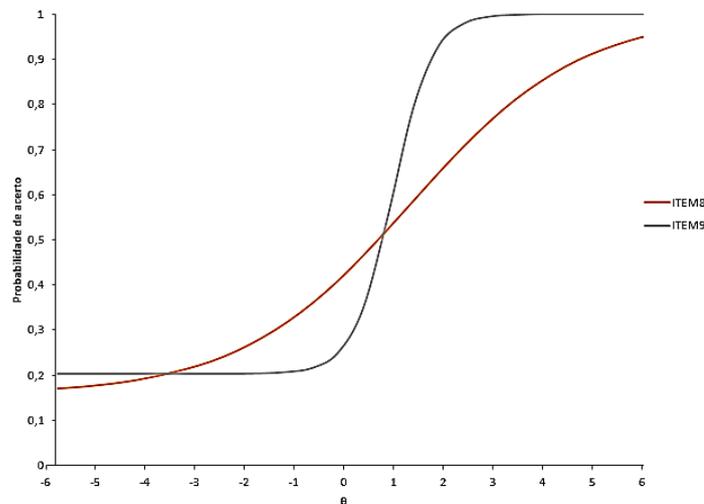
Item	Parâmetro a	Parâmetro b	Parâmetro c
1	0,697	0,648	0,167
2	1,662	0,194	0,156
3	0,896	1,032	0,159
4	0,956	-0,188	0,161
5	0,986	-2,382	0,165
6	1,454	0,031	0,144
7	0,779	0,726	0,177
8	0,586	1,329	0,156
9	2,522	0,978	0,203
10	0,700	1,124	0,177
11	1,603	0,407	0,190
12	0,761	0,314	0,168
13	2,131	-0,316	0,164
14	1,340	-0,202	0,161
15	1,208	1,346	0,147
16	1,691	1,706	0,170
17	1,102	-0,631	0,175
18	1,685	-1,203	0,165

Item	Parâmetro a	Parâmetro b	Parâmetro c
19	1,386	0,291	0,151
20	0,795	-1,385	0,166

Fonte: elaboração do autor

Analisando o parâmetro  $a$  da tabela acima, observa-se que o item mais discriminativo é o item 9 e o que menos discrimina os sujeitos do teste é o item 8. Isso significa que o item 9 consegue diferenciar alunos com baixa proficiência de alunos com alta proficiência em relação à dificuldade do item. Já o item 8 não consegue diferenciar significativamente os respondentes desse teste, conforme pode ser observado na Figura 16 a seguir, cuja CCI é mais íngreme no ponto de inflexão para o item 9 do que para o item 8.

Figura 16 – Curvas características do item 8 e 9 gerados no *software* EIRT.



Fonte: elaboração do autor.

Organizando os 20 itens desse teste conforme a classificação relativa ao parâmetro de discriminação do item (parâmetro  $a$ ), que foi descrita na Tabela 2, p.49, tem-se:

Tabela 7 - Classificação dos 20 itens conforme o parâmetro de discriminação (parâmetro  $a$ ).

Classificação	Item
Nenhuma	-
Muito baixa	-
Baixa	8
Moderada	1, 10, 12, 7, 20, 3, 4, 5, 17, 15, 14
Alta	19, 6, 11, 2, 18, 16
Muito alta	13, 9

Fonte: elaboração do autor.

De acordo com os dados da tabela acima nota-se que nas faixas “nenhuma” e “muito baixa” não foi observado nenhum item; já na faixa de classificação “baixa” apenas um item foi observado. Na faixa “moderada” concentraram-se 55% dos itens, e nas faixas “alta” juntamente com “muito alta” figuram 40% dos itens. Observa-se ainda que a maioria dos itens se encontra nas últimas faixas indicando que os itens desse teste discriminam bem os respondentes ao teste.

Por outro lado, organizando os itens como descritos na Tabela 6 e de acordo com a classificação dos parâmetros de dificuldade, conforme a classificação sugerida na Tabela 1, p.47, tem-se:

Tabela 8 - Classificação dos 20 itens conforme o índice de dificuldade (parâmetro  $b$ ) na TRI.

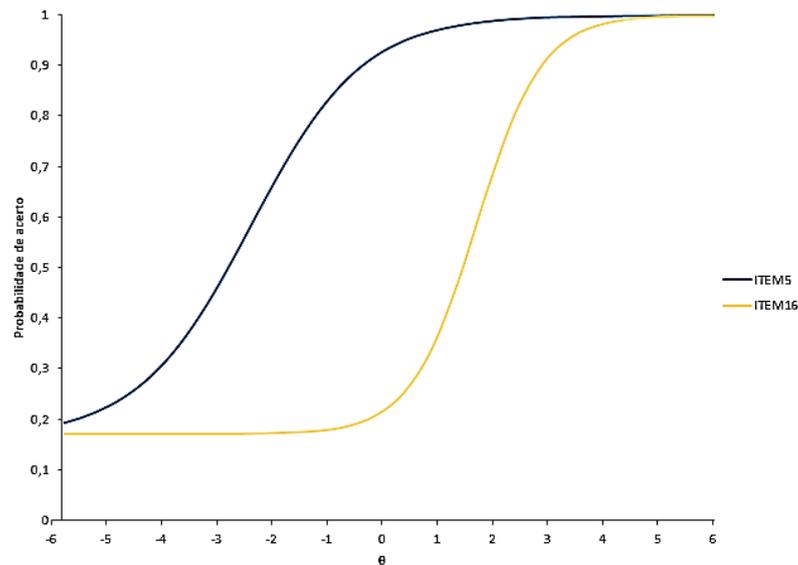
Classificação	Itens
Muito fáceis	5, 20
Fáceis	18, 17
Medianos	13, 14, 4, 6, 2, 19, 12, 11
Difíceis	1, 7, 9, 3, 10
Muito difíceis	8, 15, 16

Fonte: elaboração do autor.

Desse modo, observa-se na Tabela 8 acima que o item mais fácil é o 5 ( $b=-2,382$ ) e o item 16 ( $b=1,706$ ) é o mais difícil desse teste, segundo essa classificação da TRI. Além disso,

pode-se notar, a partir das curvas características desses itens, que o item 16 está mais à direita que o item 5, configurando assim que aquele item é mais difícil que este, conforme apresentado na Figura 17 a seguir.

Figura 17 – Curvas características dos itens 5 e 16 geradas no software EIRT.

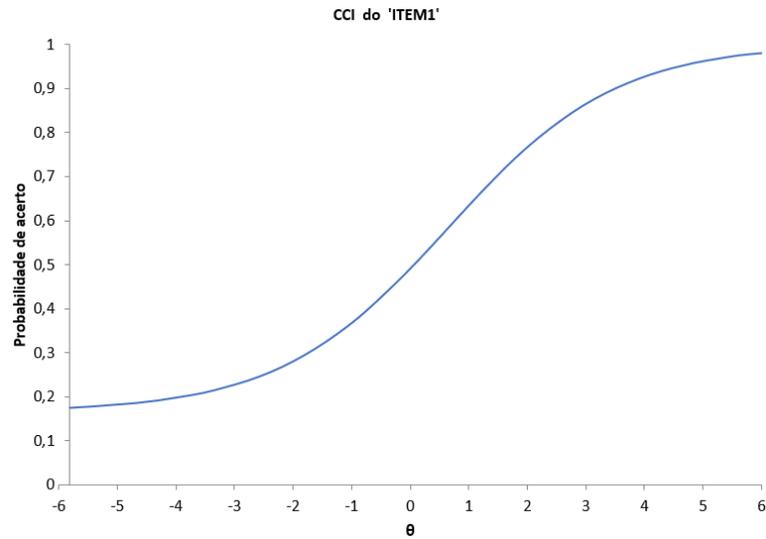


Fonte: elaboração do autor.

Ademais, como cada questão de múltipla escolha desse teste possui quatro alternativas, espera-se que o parâmetro  $c$  para cada item seja menor ou igual  $1/4=0,25$ . Nesse sentido, analisando esse parâmetro, descrito na Tabela 6, conclui-se que todos os valores estão dentro do esperado, indicado que nesse teste os sujeitos com baixa habilidade não teve probabilidade alta de acertar os itens no “chute”.

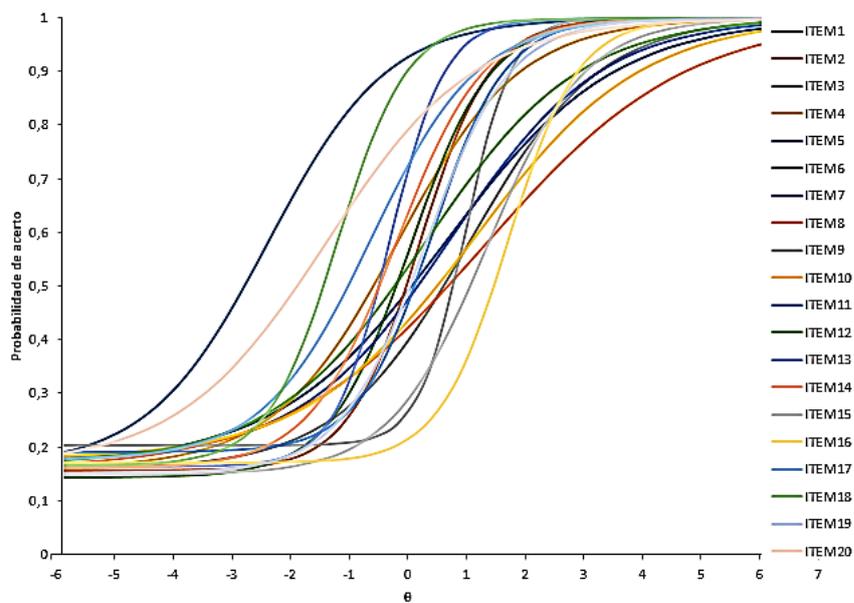
Outra representação destacada e que pode ser observada no relatório do *software* EIRT é a curva característica dos itens do teste, que são apresentadas individualmente ou em um plano único, como nas Figura 18 e Figura 19 abaixo:

Figura 18 – Curva característica do item 1 gerado pelo *software* EIRT.



Fonte: elaboração do autor.

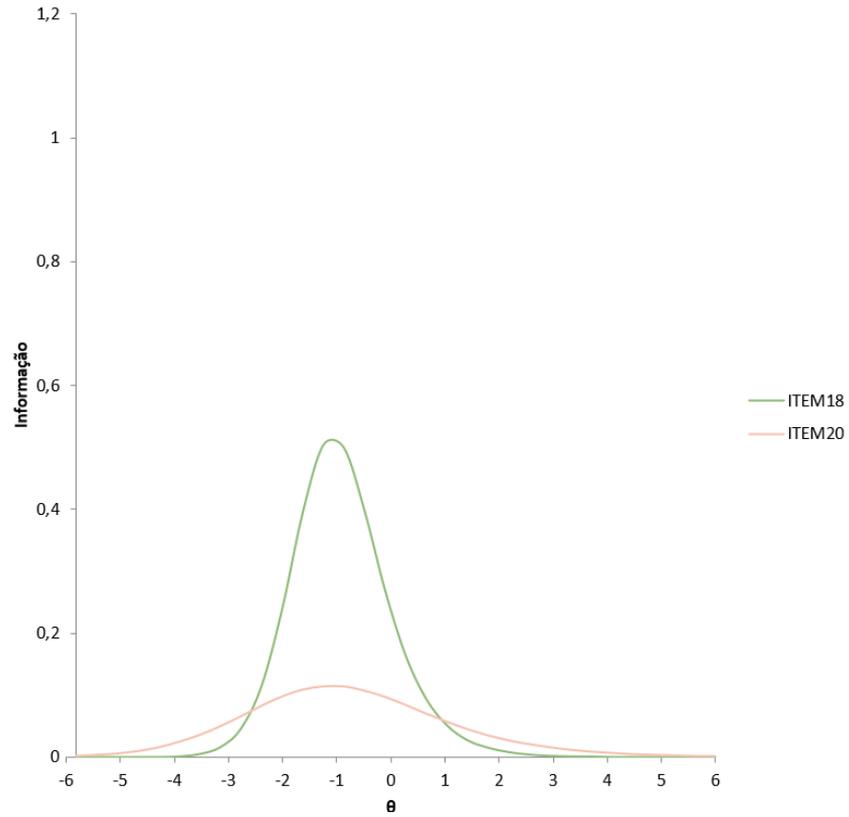
Figura 19 - Curvas características dos 20 itens gerados pelo *software* EIRT.



Fonte: elaboração do autor

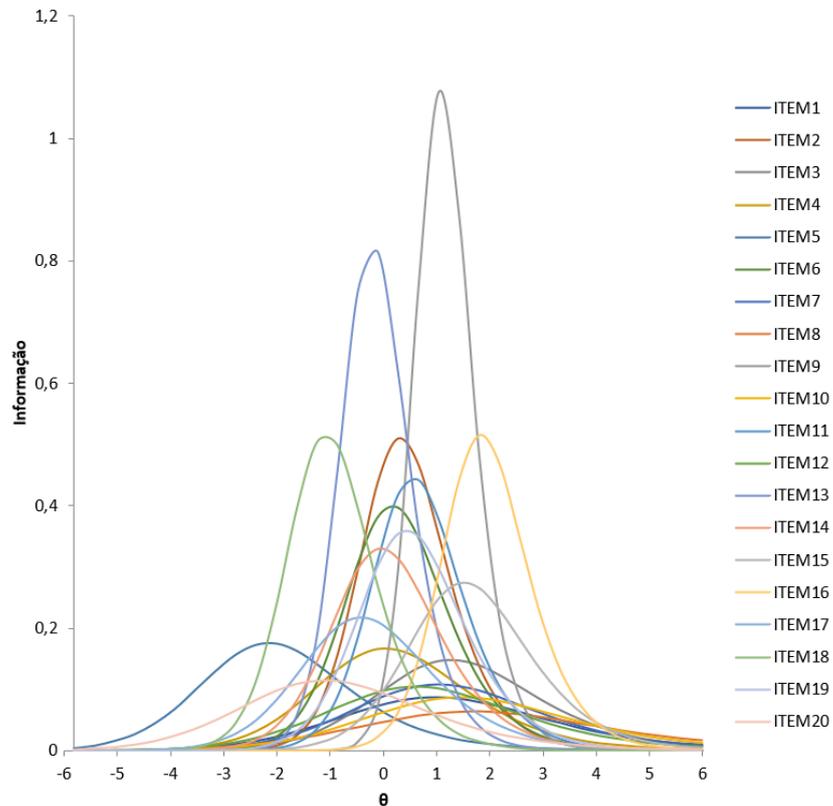
A partir das curvas características dos itens é que o construtor do teste pode fazer as análises da probabilidade de acerto no item ao longo da escala de habilidade ou ainda fazer comparações em relação as suas características. Além disso, o EIRT também gera a curva de informação para cada item (CII) do teste, como pode ser visto a seguir:

Figura 20 - Curva de Informação dos itens 18 e 20 gerado pelo EIRT.



Fonte: elaboração do autor.

Figura 21 - Curvas de Informação dos 20 itens do teste gerado pelo EIRT.



Fonte: elaboração do autor.

Algumas análises podem ser realizadas a partir das curvas de informação desses itens como apresentado na Seção 2.8. Por exemplo, nota-se na Figura 20 que o item 18 traz mais informação para esse teste que o item 20, quando observada certa faixa de habilidade. Daí se o propósito fosse uma análise para arquivá-los em um banco de itens o item 20 poderia ser descartado, pois, segundo Pasquali (2018), um item com esse comportamento pode indicar péssimos parâmetros psicométricos e não traz informação útil para esse teste.

Além das estatísticas e gráficos da TRI apresentados acima, encontrar o traço latente (habilidade/proficiências) dos sujeitos é o objetivo principal de muitas avaliações. Para tanto, o EIRT também fornece essas informações em seu relatório, em que podemos comparar o escore total de cada sujeito - descritos pela TCT- e as habilidades ( $\theta$ ) - descritas pela TRI-, conforme apresentado na tabela abaixo.

Tabela 9 - Comparação dos escores (TCT) e habilidades -  $\theta$  (TRI) estimados pelo *software* EIRT.

<b>Sujeito j</b>	<b>Escore</b>	<b>(<math>\theta</math>)</b>	<b>Posição</b>	<b>Sujeito j</b>	<b>Escore</b>	<b>(<math>\theta</math>)</b>	<b>Posição</b>
SUJEITO 99	20	2,261	1°	SUJEITO 24	11	-0,090	51°
SUJEITO 3	18	1,798	2°	SUJEITO 8	11	-0,097	52°
SUJEITO 80	18	1,653	3°	SUJEITO 37	12	-0,104	53°
SUJEITO 17	17	1,614	4°	SUJEITO 78	10	-0,110	54°
SUJEITO 19	17	1,496	5°	SUJEITO 64	11	-0,143	55°
SUJEITO 41	17	1,487	6°	SUJEITO 14	10	-0,194	56°
SUJEITO 28	17	1,450	7°	SUJEITO 23	10	-0,227	57°
SUJEITO 76	17	1,422	8°	SUJEITO 9	9	-0,230	58°
SUJEITO 98	17	1,358	9°	SUJEITO 67	10	-0,232	59°
SUJEITO 39	17	1,354	10°	SUJEITO 11	11	-0,241	60°
SUJEITO 74	17	1,332	11°	SUJEITO 92	10	-0,275	61°
SUJEITO 61	16	1,222	12°	SUJEITO 21	9	-0,293	62°
SUJEITO 51	16	1,112	13°	SUJEITO 6	9	-0,310	63°
SUJEITO 49	15	1,090	14°	SUJEITO 48	11	-0,313	64°
SUJEITO 35	15	1,056	15°	SUJEITO 7	8	-0,371	65°
SUJEITO 30	16	1,037	16°	SUJEITO 55	11	-0,372	66°
SUJEITO 32	15	0,970	17°	SUJEITO 5	9	-0,400	67°
SUJEITO 71	14	0,850	18°	SUJEITO 59	10	-0,438	68°
SUJEITO 73	14	0,806	19°	SUJEITO 83	7	-0,488	69°

Sujeito j	Escore	( $\theta$ )	Posição	Sujeito j	Escore	( $\theta$ )	Posição
SUJEITO 72	14	0,737	20°	SUJEITO 81	9	-0,501	70°
SUJEITO 65	15	0,674	21°	SUJEITO 42	7	-0,549	71°
SUJEITO 56	14	0,653	22°	SUJEITO 75	8	-0,586	72°
SUJEITO 18	14	0,625	23°	SUJEITO 40	9	-0,632	73°
SUJEITO 79	15	0,599	24°	SUJEITO 47	9	-0,641	74°
SUJEITO 45	14	0,596	25°	SUJEITO 77	9	-0,651	75°
SUJEITO 53	13	0,539	26°	SUJEITO 95	10	-0,668	76°
SUJEITO 44	13	0,527	27°	SUJEITO 68	8	-0,677	77°
SUJEITO 26	13	0,523	28°	SUJEITO 94	7	-0,721	78°
SUJEITO 85	14	0,512	29°	SUJEITO 89	9	-0,812	79°
SUJEITO 43	14	0,489	30°	SUJEITO 70	7	-0,840	80°
SUJEITO 93	13	0,479	31°	SUJEITO 96	7	-0,924	81°
SUJEITO 82	13	0,456	32°	SUJEITO 16	6	-0,927	82°
SUJEITO 20	13	0,441	33°	SUJEITO 88	6	-0,931	83°
SUJEITO 60	12	0,432	34°	SUJEITO 91	6	-0,952	84°
SUJEITO 50	13	0,424	35°	SUJEITO 25	6	-0,963	85°
SUJEITO 2	11	0,399	36°	SUJEITO 15	6	-1,031	86°
SUJEITO 31	12	0,379	37°	SUJEITO 63	5	-1,080	87°
SUJEITO 22	12	0,300	38°	SUJEITO 33	6	-1,127	88°
SUJEITO 10	13	0,290	39°	SUJEITO 52	6	-1,160	89°
SUJEITO 27	12	0,274	40°	SUJEITO 87	5	-1,186	90°
SUJEITO 46	12	0,263	41°	SUJEITO 34	6	-1,187	91°
SUJEITO 62	13	0,246	42°	SUJEITO 97	7	-1,236	92°
SUJEITO 69	13	0,170	43°	SUJEITO 29	5	-1,264	93°
SUJEITO 36	12	0,092	44°	SUJEITO 54	6	-1,278	94°
SUJEITO 13	10	0,042	45°	SUJEITO 58	8	-1,286	95°
SUJEITO 38	12	0,034	46°	SUJEITO 66	7	-1,302	96°
SUJEITO 84	12	0,002	47°	SUJEITO 4	5	-1,304	97°
SUJEITO 90	11	-0,005	48°	SUJEITO 86	5	-1,445	98°
SUJEITO 12	11	-0,056	49°	SUJEITO 57	6	-1,625	99°
SUJEITO 1	10	-0,081	50°	SUJEITO 100	0	-2,254	100°

Fonte: elaboração do autor.

Da tabela acima pode-se observar que a classificação dos sujeitos pela TRI - conforme a habilidade ( $\theta$ ) - difere do resultado na TCT (escore total). Na TCT alguns respondentes possuem a mesma pontuação, o que não ocorre na TRI e com isso, nota-se que a utilização dessa teoria faz com que seja evitado um grande número de empates para a classificação final em um teste. Além disso, pode-se observar que nem sempre quem acertou mais questões teve a maior habilidade ( $\theta$ ), como, por exemplo, o sujeito 49 (posição 14) que acertou 15 questões, mas ficou melhor classificado que o sujeito 30 (posição 16), que acertou 16 questões. Isso se deve aos métodos estatísticos empregados na TRI para estimar as habilidades dos sujeitos, em que não é apenas observado o número total de acertos, mas também o padrão das suas respostas.

Além disso, como a escala nas diversas avaliações que utilizam a TRI é arbitrária, o construtor do teste pode transformar a habilidade ( $\theta$ ) da escala (0,1), que é a escala padrão para o *software* EIRT, como apresentado na Tabela 9, para outra escala qualquer ( $m,d$ ). Como já foi mencionado, isso normalmente é feito para transformar valores negativos ou com muitas casas decimais com intuito de facilitar a comunicação dos resultados do teste. Para tanto, essas transformações são realizadas, como observado na Seção 2.5, utilizando a fórmula  $\theta^* = d \cdot \theta + m$ , em que o  $\theta$  é o valor da habilidade na escala (0,1),  $m$  representa a média e  $d$  o desvio-padrão na escala ( $m,d$ ).

Nesse sentido, por exemplo, após a transformação da escala (0,1) para a escala do Enem (500,100) e Saeb (250,50), tem-se os resultados:

Tabela 10 - Habilidades descritas pela TRI nas escalas (0,1), Enem e Saeb, conforme a posição do examinando.

Sujeito j	Escala (1,0)	Enem (500,100)	Saeb (250,50)	Posição	Sujeito j	Escala (1,0)	Enem (500,100)	Saeb (250,50)	Posição
SUJEITO 99	2,261	726,1	363,1	1°	SUJEITO 24	-0,090	491,0	245,5	51°
SUJEITO 3	1,798	679,8	339,9	2°	SUJEITO 8	-0,097	490,3	245,2	52°
SUJEITO 80	1,653	665,3	332,7	3°	SUJEITO 37	-0,104	489,6	244,8	53°
SUJEITO 17	1,614	661,4	330,7	4°	SUJEITO 78	-0,110	489,0	244,5	54°
SUJEITO 19	1,496	649,6	324,8	5°	SUJEITO 64	-0,143	485,7	242,8	55°
SUJEITO 41	1,487	648,7	324,3	6°	SUJEITO 14	-0,194	480,6	240,3	56°
SUJEITO 28	1,450	645,0	322,5	7°	SUJEITO 23	-0,227	477,3	238,7	57°
SUJEITO 76	1,422	642,2	321,1	8°	SUJEITO 9	-0,230	477,0	238,5	58°
SUJEITO 98	1,358	635,8	317,9	9°	SUJEITO 67	-0,232	476,8	238,4	59°
SUJEITO 39	1,354	635,4	317,7	10°	SUJEITO 11	-0,241	475,9	237,9	60°

Sujeito j	Escala (1,0)	Enem (500,100)	Saeb (250,50)	Posição	Sujeito j	Escala (1,0)	Enem (500,100)	Saeb (250,50)	Posição
SUJEITO 74	1,332	633,2	316,6	11°	SUJEITO 92	-0,275	472,5	236,2	61°
SUJEITO 61	1,222	622,2	311,1	12°	SUJEITO 21	-0,293	470,7	235,4	62°
SUJEITO 51	1,112	611,2	305,6	13°	SUJEITO 6	-0,310	469,0	234,5	63°
SUJEITO 49	1,090	609,0	304,5	14°	SUJEITO 48	-0,313	468,7	234,3	64°
SUJEITO 35	1,056	605,6	302,8	15°	SUJEITO 7	-0,371	462,9	231,4	65°
SUJEITO 30	1,037	603,7	301,8	16°	SUJEITO 55	-0,372	462,8	231,4	66°
SUJEITO 32	0,970	597,0	298,5	17°	SUJEITO 5	-0,400	460,0	230,0	67°
SUJEITO 71	0,850	585,0	292,5	18°	SUJEITO 59	-0,438	456,2	228,1	68°
SUJEITO 73	0,806	580,6	290,3	19°	SUJEITO 83	-0,488	451,2	225,6	69°
SUJEITO 72	0,737	573,7	286,9	20°	SUJEITO 81	-0,501	449,9	224,9	70°
SUJEITO 65	0,674	567,4	283,7	21°	SUJEITO 42	-0,549	445,1	222,5	71°
SUJEITO 56	0,653	565,3	282,7	22°	SUJEITO 75	-0,586	441,4	220,7	72°
SUJEITO 18	0,625	562,5	281,3	23°	SUJEITO 40	-0,632	436,8	218,4	73°
SUJEITO 79	0,599	559,9	279,9	24°	SUJEITO 47	-0,641	435,9	217,9	74°
SUJEITO 45	0,596	559,6	279,8	25°	SUJEITO 77	-0,651	434,9	217,4	75°
SUJEITO 53	0,539	553,9	277,0	26°	SUJEITO 95	-0,668	433,2	216,6	76°
SUJEITO 44	0,527	552,7	276,3	27°	SUJEITO 68	-0,677	432,3	216,1	77°
SUJEITO 26	0,523	552,3	276,1	28°	SUJEITO 94	-0,721	427,9	213,9	78°
SUJEITO 85	0,512	551,2	275,6	29°	SUJEITO 89	-0,812	418,8	209,4	79°
SUJEITO 43	0,489	548,9	274,4	30°	SUJEITO 70	-0,840	416,0	208,0	80°
SUJEITO 93	0,479	547,9	273,9	31°	SUJEITO 96	-0,924	407,6	203,8	81°
SUJEITO 82	0,456	545,6	272,8	32°	SUJEITO 16	-0,927	407,3	203,6	82°
SUJEITO 20	0,441	544,1	272,0	33°	SUJEITO 88	-0,931	406,9	203,5	83°
SUJEITO 60	0,432	543,2	271,6	34°	SUJEITO 91	-0,952	404,8	202,4	84°
SUJEITO 50	0,424	542,4	271,2	35°	SUJEITO 25	-0,963	403,7	201,8	85°
SUJEITO 2	0,399	539,9	270,0	36°	SUJEITO 15	-1,031	396,9	198,5	86°
SUJEITO 31	0,379	537,9	268,9	37°	SUJEITO 63	-1,080	392,0	196,0	87°
SUJEITO 22	0,300	530,0	265,0	38°	SUJEITO 33	-1,127	387,3	193,6	88°
SUJEITO 10	0,290	529,0	264,5	39°	SUJEITO 52	-1,160	384,0	192,0	89°
SUJEITO 27	0,274	527,4	263,7	40°	SUJEITO 87	-1,186	381,4	190,7	90°
SUJEITO 46	0,263	526,3	263,1	41°	SUJEITO 34	-1,187	381,3	190,6	91°
SUJEITO 62	0,246	524,6	262,3	42°	SUJEITO 97	-1,236	376,4	188,2	92°
SUJEITO 69	0,170	517,0	258,5	43°	SUJEITO 29	-1,264	373,6	186,8	93°
SUJEITO 36	0,092	509,2	254,6	44°	SUJEITO 54	-1,278	372,2	186,1	94°

Sujeito j	Escala (1,0)	Enem (500,100)	Saeb (250,50)	Posição	Sujeito j	Escala (1,0)	Enem (500,100)	Saeb (250,50)	Posição
SUJEITO 13	0,042	504,2	252,1	45°	SUJEITO 58	-1,286	371,4	185,7	95°
SUJEITO 38	0,034	503,4	251,7	46°	SUJEITO 66	-1,302	369,8	184,9	96°
SUJEITO 84	0,002	500,2	250,1	47°	SUJEITO 4	-1,304	369,6	184,8	97°
SUJEITO 90	-0,005	499,5	249,7	48°	SUJEITO 86	-1,445	355,5	177,7	98°
SUJEITO 12	-0,056	494,4	247,2	49°	SUJEITO 57	-1,625	337,5	168,8	99°
SUJEITO 1	-0,081	491,9	245,9	50°	SUJEITO 100	-2,254	274,6	137,3	100°

Fonte: elaboração do autor.

Da tabela acima pode-se notar o melhor classificado foi o sujeito 99 e o pior foi o sujeito 100, cujas proficiências variaram de -2,254 (errou todas as questões) a 2,261 (acertou todas as questões), que é equivalente, por exemplo, a proficiência de 274,6 e 726,1 na escala do Enem ou ainda 137,3 a 363,1 na escala do Saeb.

Além disso, para algumas avaliações tem-se como objetivo apenas classificar os sujeitos comparando os valores das suas proficiências. Entretanto, apenas fazendo essas comparações não se pode afirmar o que eles aprenderam em termos de conteúdo ou quais habilidades adquiriram e, desse modo, é necessário construir as escalas de habilidades para atingir esse propósito. A seção a seguir abordará em detalhes essa questão.

Portanto, esta seção evidenciou que a utilização de *softwares* computacionais possibilita encontrar as estatísticas relativas à TRI – e também à TCT – facilitando o trabalho dos construtores do teste, poupando tempo para as interpretações qualitativas de uma avaliação. Em especial, foi exibido o *software* EIRT, que, segundo os seus desenvolvedores, “torna a utilização dos modelos da TRI o mais fácil e acessível possível [...] [e] será de grande ajuda para todos os ‘avaliadores’ que necessitam de uma ferramenta estatística para determinar e compreender as propriedades psicométricas do seu questionário ou exame” (VALOIS et. al, 2011, tradução nossa).

#### 4. ALGUMAS APLICAÇÕES DA TRI NO DESENVOLVIMENTO DE TESTES

O desenvolvimento de uma avaliação é uma tarefa trabalhosa, pois envolvem diversas etapas que requer o conhecimento de especialistas em medição e também do construto a ser avaliado. Assim, “um bom instrumento de avaliação deve ser construído de forma rigorosa se queremos que seja adequado para os usos pretendidos” (ARIAS; LLOREDA; LLOREDA, 2014, p. 26, tradução nossa). Nesse sentido, as aplicações da Teoria de Resposta ao Item, nas diferentes etapas de uma avaliação, configuram relevante destaque dessa teoria para os testes educacionais. Para tanto, nesta seção, serão apresentadas algumas aplicações mais comuns, tais como: a interpretação e a construção das escalas de habilidade e o banco de itens<sup>19</sup>.

##### 4.1 Interpretação e construção das escalas de habilidades

A utilização da TRI nas avaliações vem possibilitando grandes avanços em termos de acompanhamento no âmbito educacional. Nesse sentido, uma escala de habilidade (ou proficiência), possibilita a interpretação qualitativa dos valores das habilidades estimadas. Mas, segundo Valle (2001), assim como existe uma teoria matemática que possibilita a obtenção das estimativas das habilidades e dos parâmetros dos itens, também existe uma metodologia e todo um processo para as interpretações pedagógicas e construção de uma escala de habilidade.

Em geral, em cada ciclo de uma avaliação, os itens de um teste são posicionados em pontos específicos na escala -segundo algum critério definido pelo construtor da avaliação- e, em seguida, o valor da habilidade de cada examinando, estimado a partir das respostas dadas aos itens, é posicionado nessa mesma escala, permitindo interpretar o nível de conhecimento que cada um possui. Essa interpretação é possível, pois quando se constrói um item de um teste, tem-se como primeiro passo a escolha da habilidade que ele está associado.

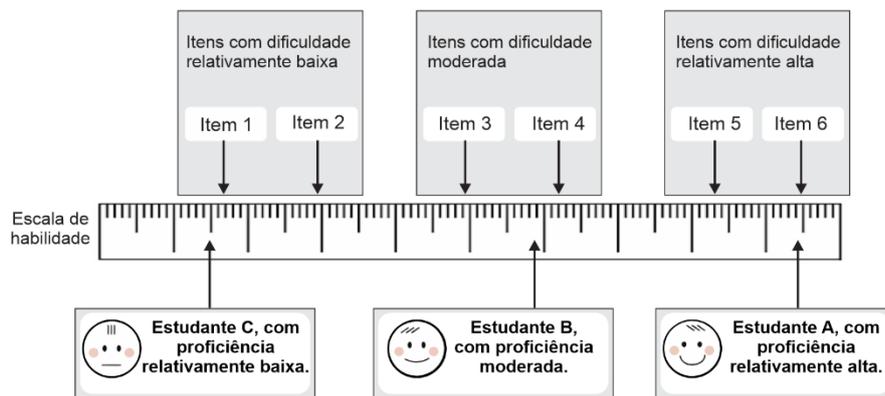
Destaca-se que quando um item é posicionado em um ponto da escala, ele determina uma marca que traduz que os examinandos, cujas proficiências (habilidades) estão acima deste ponto, têm maior probabilidade de acertá-lo, mas se estão abaixo significa que têm pouca probabilidade de acerto. A Figura 22 a seguir ilustra essas ideias, em que é possível observar que o Estudante C, com proficiência relativamente baixa, possui baixa probabilidade de acertar

---

<sup>19</sup> Outras aplicações podem ser encontradas em Pasquali (2018).

o Item 1 e possivelmente irá errar os itens de 2 a 6. Já o estudante B é esperado que seja capaz de acertar os itens de 1, 2 e 3 com sucesso, mas com pouca probabilidade de completar o item 4 com sucesso, e nenhuma probabilidade de completar os itens 5 e 6 com sucesso. O Estudante A, com proficiência relativamente alta, apresenta alta probabilidade de acertar os itens de 1 a 5, e baixa probabilidade de acertar o item 6.

Figura 22 - Relações entre posições de itens e proficiências em uma escala hipotética.



Fonte: adaptado de OECD (2005).

A TRI possibilita comparar as habilidades/proficiências dos respondentes a um teste, mas *a priori* não se pode afirmar o que cada sujeito com uma certa habilidade sabe a mais do que outro com uma habilidade inferior. Por esse motivo, os resultados por si só não possuem qualquer significado pedagógico, sendo então necessário que seja efetuada uma ligação desses valores com os conteúdos e as habilidades esperadas envolvidas na avaliação, com o intuito de “fornecer um sentido qualitativo e pedagógico às estimativas quantitativas, possibilitando ampliar o significado das proficiências e dos parâmetros de dificuldade dos itens” (BRASIL, 2019, p.47).

A utilização da TRI possibilita também realizar comparações entre diferentes edições de uma avaliação ou de diferentes provas em uma mesma edição dessa avaliação, cujos resultados estarão uma mesma escala. Assim, para muitas avaliações educacionais a comparação entre os resultados observados para cada ano/série é o objetivo principal a ser concretizado.

Entretanto, essa comparação para diferentes populações só faz sentido quando utilizado o processo denominado *equiparação* ou *equalização*, que, em linhas gerais, consiste em “colocar parâmetros de itens vindo de provas distintas ou habilidades de respondentes de

diferentes grupos na mesma métrica, isto é, em uma escala comum, tornando os itens e/ou as habilidades comparáveis” (VALLE, 2001, p. 73). Esse processo pode ser feito por duas maneiras: equalização via população ou equalização via itens comuns. De acordo com Andrade, Tavares e Valle (2000, p.81)

[na equalização via população] usamos o fato de que se um único grupo de respondentes é submetido a provas distintas, basta que todos os itens sejam calibrados conjuntamente para termos a garantia de que todos estarão na mesma métrica. Já na equalização via itens comuns, a garantia de que as populações envolvidas terão seus parâmetros em uma única escala será dada pelos itens comuns entre as populações, que servirão de ligação entre elas<sup>20</sup>.

Essas ideias apontam para o fato de que após a realização da equalização, os resultados tornam-se comparáveis, mas segundo Valle (2001), saber que em uma avaliação os estudantes de um ano/série para outro/a evoluíram sua habilidade média em matemática de 100 para 150, apenas indica que tiveram um aumento de 50%. Entretanto, apenas fazendo essas comparações não se pode afirmar o que eles aprenderam em termos de conteúdo ou quais habilidades adquiriram em cada um desses anos.

É com esse objetivo que são construídas as escalas de habilidades: interpretar de forma qualitativa as proficiências de cada sujeito ou ainda as proficiências médias de uma população.

Mas,

considerando que é praticamente impossível obter informações sobre todos os pontos da escala de proficiência, adota-se a divisão em intervalos, chamados **níveis de proficiência**, agrupando-se em cada nível [um ou mais] itens cujos parâmetros de dificuldade [, conforme a TRI,] possam ser incluídos naquele intervalo (BRASIL, 2019, p. 47, grifo nosso).

No quadro a seguir são apresentados alguns níveis de proficiência da escala de matemática do Saeb para o 5º ano do Ensino Fundamental, que associa cada nível da escala a uma habilidade (ou um conjunto de habilidades). Assim, por exemplo, se um sujeito tiver uma proficiência de 160 nessa escala significa que ele está no nível 2 e que ele possivelmente possui a maioria das habilidades descritas nesse nível; provavelmente possui ainda todas as habilidades

---

<sup>20</sup> Segundo Klein (2003, p. 127), no Saeb os testes são equiparados via itens comuns entre séries e entre anos e, desse modo, “permite que os alunos de todas as séries e de todos os anos sejam postos em uma mesma escala de proficiência de modo que seus desempenhos possam ser comparados”.

descritas no nível 1. Destaca-se que a quantidade de níveis e a descrição das habilidades em cada nível podem variar e dependem da escolha do construtor da escala em uma avaliação.

Quadro 5 - Trecho das descrições dos níveis da escala de matemática Saeb – 5º ano do ensino fundamental<sup>21</sup>

5º ano	
Nível*	Descrição das habilidades desenvolvidas
<b>Nível 1</b> - Desempenho maior ou igual a 125 e menor que 150	Os estudantes provavelmente são capazes de: <b>Grandezas e medidas</b> – determinar a área de figuras desenhadas em malhas quadriculadas por meio de contagem.
<b>Nível 2</b> - Desempenho maior ou igual a 150 e menor que 175	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes /de: <b>Números e operações; Álgebra e funções</b> – resolver problemas do cotidiano envolvendo adição de pequenas quantias de dinheiro. <b>Tratamento de informações</b> – localizar informações, relativas ao maior ou menor elemento, em tabelas ou gráficos.
⋮	⋮
<b>Nível 10</b> - Desempenho maior ou igual a 350	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <b>Espaço e forma</b> – reconhecer entre um conjunto de quadriláteros aquele que possui lados perpendiculares e com a mesma medida. <b>Grandezas e medidas</b> – converter uma medida de comprimento, expressando decímetros e centímetros, para milímetros.

Fonte: Brasil (2019, grifo nosso).

Nota: \* O Saeb não especifica as habilidades desenvolvidas no nível 0 da escala.

Essas descrições das habilidades, em cada nível da escala, favorecem a interpretação dos dados por parte da escola, professores, secretarias de educação, entre outros, e indo além do dado numérico da proficiência (ou proficiência média), observando o que os estudantes conseguem realizar em relação aos seus conhecimentos e habilidades quando posicionados em diferentes pontos da escala.

Destaca-se que os extremos desses intervalos (níveis de proficiência) são denominados *níveis âncoras*, que são pontos selecionados pelo construtor na escala de habilidade para serem interpretados pedagogicamente. Conforme Rabelo (2013) esses níveis âncoras normalmente

<sup>21</sup> O quadro completo com todos os níveis da escala de matemática pode ser encontrado em Brasil (2019, p. 54-58).

são apresentados por intervalos de um ou meio desvios padrão. Nesse sentido, por exemplo, tem-se que nas escalas do Saeb o intervalo que define cada nível é de 25 pontos (correspondente a meio desvio-padrão). Assim, os níveis âncoras escolhidos para o Saeb, arbitrariamente, são: 125, 150, 175, 200, 225, 250, 275, ..., 500.

Esses níveis âncoras são caracterizados por um conjunto de itens denominados *itens âncora*. Nesse sentido, segundo Andrade, Tavares e Valle (2000), um item pode ser considerado âncora, em um determinado nível Y da escala, se:

- i) a probabilidade de acerto do item no nível Y for de pelo menos 65%;
- ii) a probabilidade de acerto do item no nível imediatamente inferior for menor do que 50%;
- iii) a diferença entre as probabilidades de acerto do item nesses dois níveis consecutivos deve ser de pelo menos de 30%.

Contudo, alguns itens podem na prática não se enquadrar nos três critérios acima, e por esse motivo alguns construtores de testes às vezes flexibilizam ou utilizam outra definição, como é o caso do Inep, para o Enem, que apenas utiliza um critério para selecionar os itens âncoras, tal como descrito em Brasil (2014, p. 15):

a posição que cada item ocupa no mapa [de itens] representa a menor proficiência com a qual pelo menos 65% dos participantes o acertaram, ou seja, trata-se de um ponto que indica a dificuldade empírica do item, obtida após sua aplicação. Assim, podemos afirmar que pelo menos 65% dos participantes com determinada proficiência acertaram um item posicionado no mapa no ponto de dificuldade equivalente a essa proficiência.

Esse mapa de itens do Enem<sup>22</sup> – construído por área de conhecimento-, e mencionado acima, é um documento do Inep/Enem onde são apresentadas descrições das habilidades avaliadas nas questões desse exame, e tem por objetivo associar cada item a um ponto da escala, possibilitando visualizar o nível de dificuldade dos itens ao longo dela para que em outro momento as proficiências dos respondentes ao teste desse exame possam ser avaliadas de forma qualitativa.

Nesse caso, “geralmente, o posicionamento do item está um pouco acima do parâmetro de dificuldade, sendo uma combinação dos três parâmetros. A questão é sempre posicionada no valor da régua onde a probabilidade de acerto está próxima de 0,65” (BRASIL, 2012a, p.14).

---

<sup>22</sup> Ver em BRASIL, 2014.

Nesse sentido, pode-se inferir que esse posicionamento na escala pode ser encontrado calculando o seu ponto de ancoragem a partir da equação (11), isto é (CARLOS, 2016, p.46):

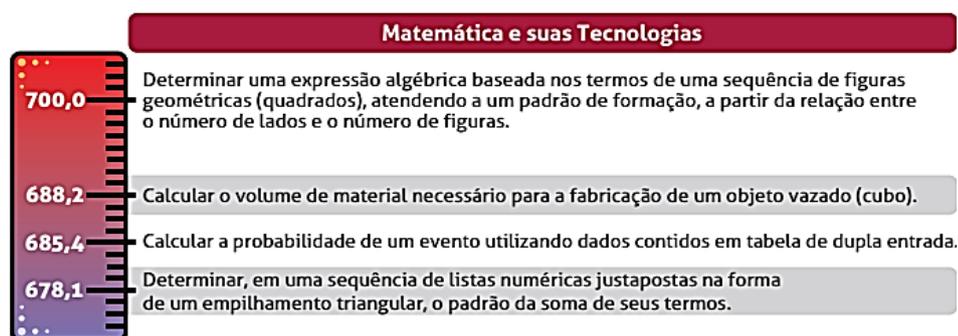
$$0,65 = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta - b_i)}} \Leftrightarrow \theta = b_i - \frac{\ln\left(\frac{1 - 0,65}{0,65 - c_i}\right)}{Da_i}$$

Portanto, o ponto de ancoragem para os itens do Enem, segundo os critérios observados acima e ressaltando que  $D=1$  para esse exame, é:

$$P_{\text{ancoragem}} = b_i - \frac{\ln\left(\frac{1 - 0,65}{0,65 - c_i}\right)}{a_i} \quad (20)$$

No Enem cada questão está associada a uma habilidade descrita no mapa de itens, cujo posicionamento na escala de habilidades localiza-se no ponto de ancoragem conforme a expressão (20). A figura a seguir apresenta o recorte de um mapa de itens do Enem, destacando os níveis entre 650 e 700 da escala de proficiência de matemática e suas tecnologias, bem como as suas respectivas habilidades.

Figura 23 – Trecho do mapa de itens do Enem na área de Matemática e suas tecnologias.



Fonte: Brasil (2012a).

A título de ilustração, é possível observar na Figura 23 que no nível 688,2, por exemplo, a habilidade que identifica uma questão ancorada nesse ponto é “Calcular o volume de material necessário para a fabricação de um objeto vazado (cubo)”. Além disso, pode-se dizer que questões ancoradas no ponto 700 são mais difíceis que essa questão; já as ancoradas em 685,4

são consideradas mais fáceis. Nota-se ainda que questões ancoradas em diferentes níveis há diferentes habilidades a elas associadas.

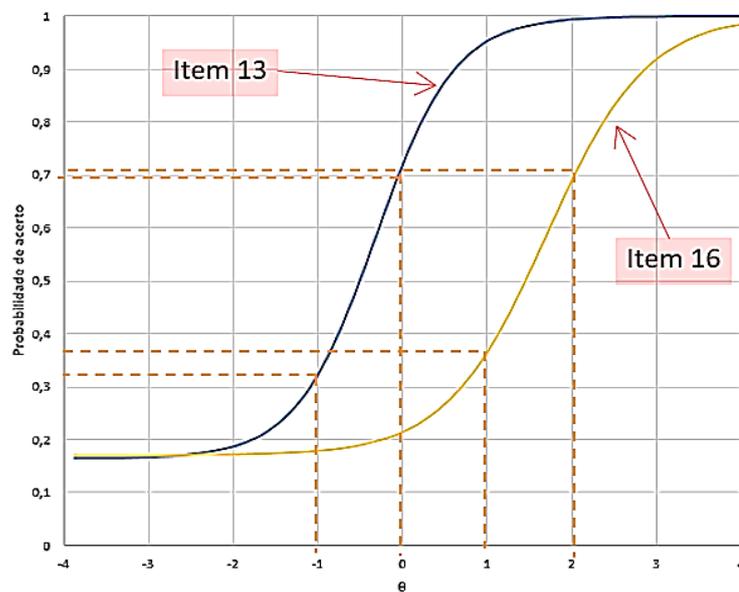
Para facilitar a compreensão sobre o processo de ancoragem, a partir da definição de Andrade, Tavares e Valle (2000), acima mencionada, apresenta-se a seguir uma tabela com dois itens<sup>23</sup> e seus respectivos parâmetros, analisados em uma escala de proficiência (0,1), cujos níveis âncoras escolhidos são -3, -2, -1, 0, 1, 2 e -3.

Tabela 11 - Parâmetros dos itens 13 e 16.

Item	Parâmetro $a$	Parâmetro $b$	Parâmetro $c$
Item 13	2,131	-0,316	0,164
Item 16	1,691	1,706	0,170

Fonte: elaboração do autor.

Figura 24 - Curvas Características (CCIs) dos itens âncoras 13 e 16.



Fonte: elaboração do autor.

Com base nos dados indicados na Tabela 11 e na Figura 24, é possível aplicar os critérios que definem um item âncora e verificar que:

<sup>23</sup> Estes itens compõem a Tabela 6, p. 50.

a) O item 13 é âncora do nível 0, pois

i) a probabilidade de acerto do item 13 no nível 0 é de pelo menos 65%:

$$P(U_{13} = 1 | \theta = 0) = 0,72 \geq 0,65.$$

ii) a probabilidade de acerto do item no nível imediatamente inferior é menor que 50%:

$$P(U_{13} = 1 | \theta = -1) = 0,32 < 0,50.$$

iii) a diferença entre as probabilidades de acerto do item nesses dois níveis consecutivos deve ser de pelo menos de 30%.

$$P(U_{13} = 1 | \theta = 0) - P(U_{13} = 1 | \theta = -1) = 0,72 - 0,32 = 0,40 \geq 0,30.$$

b) Segundo os mesmos critérios, o item 16 é âncora do nível 2, pois

i)  $P(U_{16} = 1 | \theta = 2) = 0,69 \geq 0,65$ .

ii)  $P(U_{16} = 1 | \theta = 1) = 0,36 < 0,50$ .

iii)  $P(U_{16} = 1 | \theta = 2) - P(U_{16} = 1 | \theta = 1) = 0,86 - 0,27 = 0,59 \geq 0,30$ .

Mesmo, *a priori*, não podendo se ter certeza se no teste existirão itens âncoras para cada nível, nem, quando existirem, quantos deles serão selecionados, Valle (2000, p.65) diz que é “fundamental que os níveis âncoras sejam escolhidos não muito próximos uns dos outros e também o número de itens deve ser bastante grande de modo a possibilitar a construção e interpretação da escala de habilidade”.

Uma vez que o item foi ancorado em um ponto ou nível, ele poderá ser utilizado para descrevê-lo de forma pedagógica na escala de proficiência/habilidade, que normalmente é apresentada na forma de um quadro estatístico que contém as descrições das habilidades em cada nível para os pontos ou intervalos da escala.

Desse modo, quando se quer representar a escala de habilidades em forma de intervalos, pode-se ter como exemplo a escala do Saeb – (250, 50), apresentada no Quadro 5. Por outro lado, quando se quer apresentar tal escala por pontos, com intuito de interpretar qualitativamente as proficiências dos sujeitos, pode-se construir como apresentado no Quadro 6 a seguir. Destaca-se que essa escolha dos níveis depende exclusivamente do construtor do teste. Neste caso, a escala utilizada foi a (0,1) e organizada em 7 níveis de proficiência, associados aos níveis âncoras conforme apresentados no quadro a seguir.

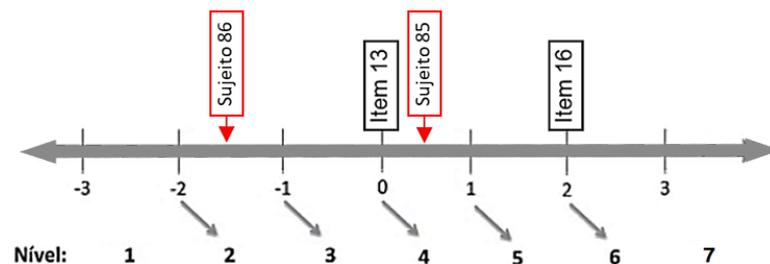
Quadro 6 – Níveis e descrição das habilidades de uma escala hipotética - (0,1).

Nível	Descrição das habilidades
Nível 1 ( $\theta < -2$ )	Neste nível os alunos: <ul style="list-style-type: none"> <li>• [descrição da habilidade A1]</li> <li>• [descrição da habilidade B1]</li> <li>• ...</li> </ul>
Nível 2 ( $\theta = -2$ )	E mais: <ul style="list-style-type: none"> <li>• [descrição da habilidade A2]</li> <li>• [descrição da habilidade B2]</li> <li>• ...</li> </ul>
Nível 3 ( $\theta = -1$ )	E mais: <ul style="list-style-type: none"> <li>• [descrição da habilidade A3]</li> <li>• [descrição da habilidade B3]</li> <li>• ...</li> </ul>
Nível 4 ( $\theta = 0$ )	E mais: <ul style="list-style-type: none"> <li>• [descrição da habilidade A4]</li> <li>• [descrição da habilidade B4]</li> <li>• ...</li> </ul>
Nível 5 ( $\theta = 1$ )	E mais: <ul style="list-style-type: none"> <li>• [descrição da habilidade A5]</li> <li>• [descrição da habilidade B5]</li> <li>• ...</li> </ul>
Nível 6 ( $\theta = 2$ )	E mais: <ul style="list-style-type: none"> <li>• [descrição da habilidade A6]</li> <li>• [descrição da habilidade B6]</li> <li>• ...</li> </ul>
Nível 7 ( $\theta > 2$ )	E mais: <ul style="list-style-type: none"> <li>• [descrição da habilidade A7]</li> <li>• [descrição da habilidade B7]</li> <li>• ...</li> </ul>

Fonte: elaboração do autor.

Assim, após construída a escala de habilidade explicitando em um quadro estatístico o que se espera que um respondente a um teste possua de conhecimento para cada nível, como apresentado acima, pode-se fazer as interpretações pedagógicas a partir da habilidade ( $\theta$ ) de cada respondente. Nesse sentido, se, por exemplo, forem posicionados os itens âncoras 13 e 16 na escala (0,1), como na Figura 25 abaixo e, em seguida, também posicionar nessa escala o sujeito 86 com  $\theta = -1,445$  e o sujeito 85 com  $\theta = 0,512$  (esses dados estão apresentados na Tabela 9), então, analisando de forma quantitativa, pode-se dizer que o sujeito 85 ficou melhor classificado no teste que o sujeito 86, como ilustra a figura abaixo.

Figura 25 – Posicionamento dos itens âncoras 13 e 16 e dos sujeitos 85 e 86 na escala de habilidade.



Fonte: elaboração do autor.

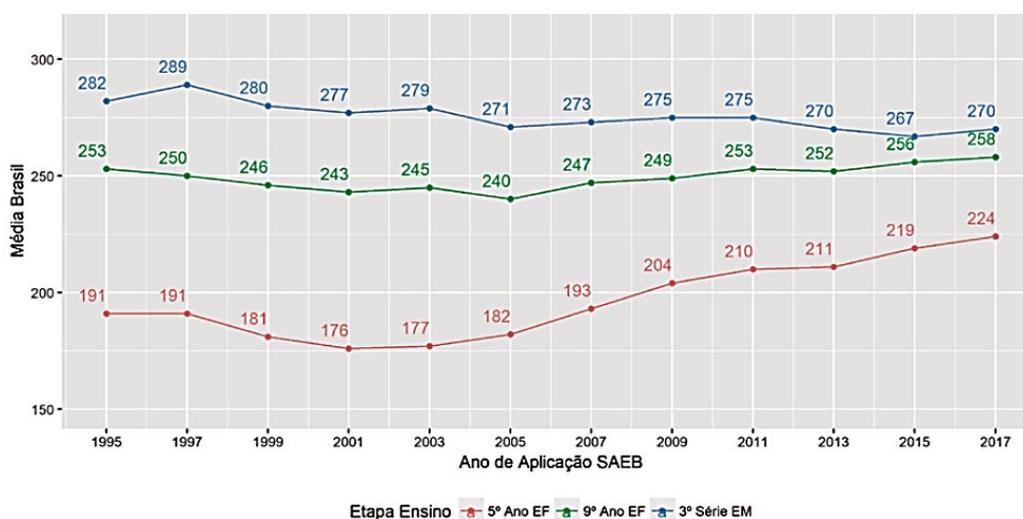
Vale ainda destacar que o sujeito 86 possivelmente não acertaria o item 13 nem o 16, pois o seu  $\theta$  é menor que os níveis que foram ancorados esses itens. Já o sujeito 85 possivelmente acertaria o item 13, mas provavelmente erraria o item 16. Ademais, usando as informações das habilidades descritas no Quadro 6, pode-se inferir que o sujeito 86 muito provavelmente adquiriu apenas as habilidades descritas nos níveis 1 e 2 - e não consolidou todas as exigidas no nível 3 em diante -, enquanto que o sujeito 85 provavelmente adquiriu as habilidades descritas nos níveis 1, 2, 3 e 4, e não consolidou todas as exigidas no nível 5 em diante.

Quando se utiliza a TRI, além das interpretações do que cada sujeito possui de conhecimento, pode-se ainda determinar as proficiências médias das turmas de uma escola, de um município, das unidades da federação ou ainda as proficiências médias de cada ano/série da avaliação da população em análise de um país. Nesse contexto, Valle (2001, p. 77) diz que é importante

posicionar as populações que foram estudadas e verificar em que pontos da escala elas se encontram, utilizando a sua habilidade média. Assim, a primeira informação que a escala nos fornecerá é a identificação do que os alunos sabem e do que não são capazes de fazer, ou seja, quais conteúdos que eles dominam e quais conteúdos ainda precisam melhorar.

Essa situação descrita acima pode ser ilustrada na figura a seguir, cujas proficiências médias dos estudantes foram estimadas na escala do Saeb, que produziu uma série estatística de 1995 a 2017 com o intuito de fazer comparações sobre a evolução do desempenho dos estudantes de três séries/anos do ensino básico.

Figura 26 – Evolução das proficiências médias dos estudantes em matemática no SAEB de 1995 a 2017.



Fonte: Inep/DAEB apud BRASIL (2018b).

Analisando os gráficos acima pode-se afirmar que houve uma discreta evolução da proficiência média de 1995 a 2017 para os estudantes do 9º ano e uma evolução um pouco melhor para os do 5º ano. Já para os estudantes do 3º ano do ensino médio houve uma queda na proficiência média nesse período. É a partir de dados como esses que as proficiências médias podem ser interpretadas para os diferentes níveis da escala de habilidade, que no exemplo apresentado na Figura 26 foram observados os dados do 5º e 6º ano do ensino fundamental e também o 3º ano do ensino médio. Por exemplo, analisando as proficiências médias em 2017, nota-se que os estudantes do 5º ano figuram no nível 4 (entre 200 e 225), os do 9º ano figuram no nível 3 (entre 175 e 200), e os do 3º ano do ensino médio figuram no nível 2 (entre 150 e 175), conforme o quadro das respectivas escalas de habilidade do Saeb<sup>24</sup>.

Além de destacar a importância de compreender como se interpreta os resultados de uma avaliação a partir do posicionamento dos itens e das proficiências dos examinados em uma escala, pode-se também ter como propósito a construção de uma escala de habilidade. Para tanto, quando se quer obter essa construção, pode-se seguir as etapas sugeridas por Valle (2001, p.74-77) e apresentadas a seguir:

Quadro 7 – Diretrizes para construção e interpretação de uma escala de habilidade.

<b>Etapa</b>	<b>Descrição</b>
1ª	Definir a(s) população(ões) e a(s) disciplina(as) que serão analisadas.
2ª	Elaborar e aplicar os instrumentos de medida (testes).
3ª	Equalizar o teste.
4ª	Definir a escala.
5ª	Escolher os níveis âncoras.
6ª	Identificar os itens âncora.
7ª	Interpretar cada ponto (ou nível) da escala.

Fonte: adaptado de Valle (2001).

<sup>24</sup>O quadro completo com as descrições das competências e habilidades em todos os níveis da escala de matemática para o 5º ano e 9º ano do ensino fundamental, bem como para o 3º ano do ensino médio pode ser encontrado em: BRASIL (2019).

Portanto, como foi dito, saber interpretar - ou ainda construir - uma escala de habilidade de uma avaliação constitui uma etapa importante para a consolidação plena dos conceitos da Teoria de Resposta ao Item. Ademais, pode-se afirmar que a proficiência média constitui um importante dado para as avaliações educacionais, pois é a partir desse tipo de informação que se pode propor ações ou políticas públicas para a melhoria da qualidade da educação.

#### 4.2 Banco de itens para avaliações educacionais

Em testes educacionais o item consiste na unidade básica de um instrumento de medida, e que normalmente é considerado sinônimo de questão, termo mais comum e utilizado frequentemente nas escolas (BRASIL, 2010). Eles podem ser de dois tipos de formato: (a) de *resposta livre* (resposta construída), que exige que o examinando forneça ou elabore uma resposta correta, ou (b) de *resposta objetiva* (resposta selecionada), que exige que o examinando selecione uma resposta de um conjunto de opções.

Nesse sentido, a construção dos itens constitui uma das etapas mais cruciais no processo de elaboração do instrumento de medida, pois eles “são as matérias-primas, os tijolos dos quais um instrumento de avaliação é formado, portanto uma construção deficiente dos mesmos afetará as propriedades métricas finais do instrumento de medida e a validade das inferências que são feitos a partir das pontuações” (HALADYNA; RODRIGUEZ, 2013; LANE, RAYMOND & HALADYNA, 2016; MUÑIZ et al., 2005; OSTERLIND, 1998; SCHMEISERY WELCH, 2006 apud MUÑIZ, 2018, p.268, tradução nossa).

Para elaborar bons itens<sup>25</sup> Fontanive (2005 apud BARRETO; MENDES; OLIVEIRA, 2009, p. 205) destaca como essenciais os seguintes aspectos:

- (a) o item deve ter exatidão de conteúdo, ou seja, não pode haver erros conceituais;
- (b) deve mensurar a habilidade que se deseja avaliar;
- (c) não deve conter ambiguidades em seu enunciado nem alternativas que gerem pistas que podem atrair respostas;
- (d) deve apresentar resultados estatísticos adequados.

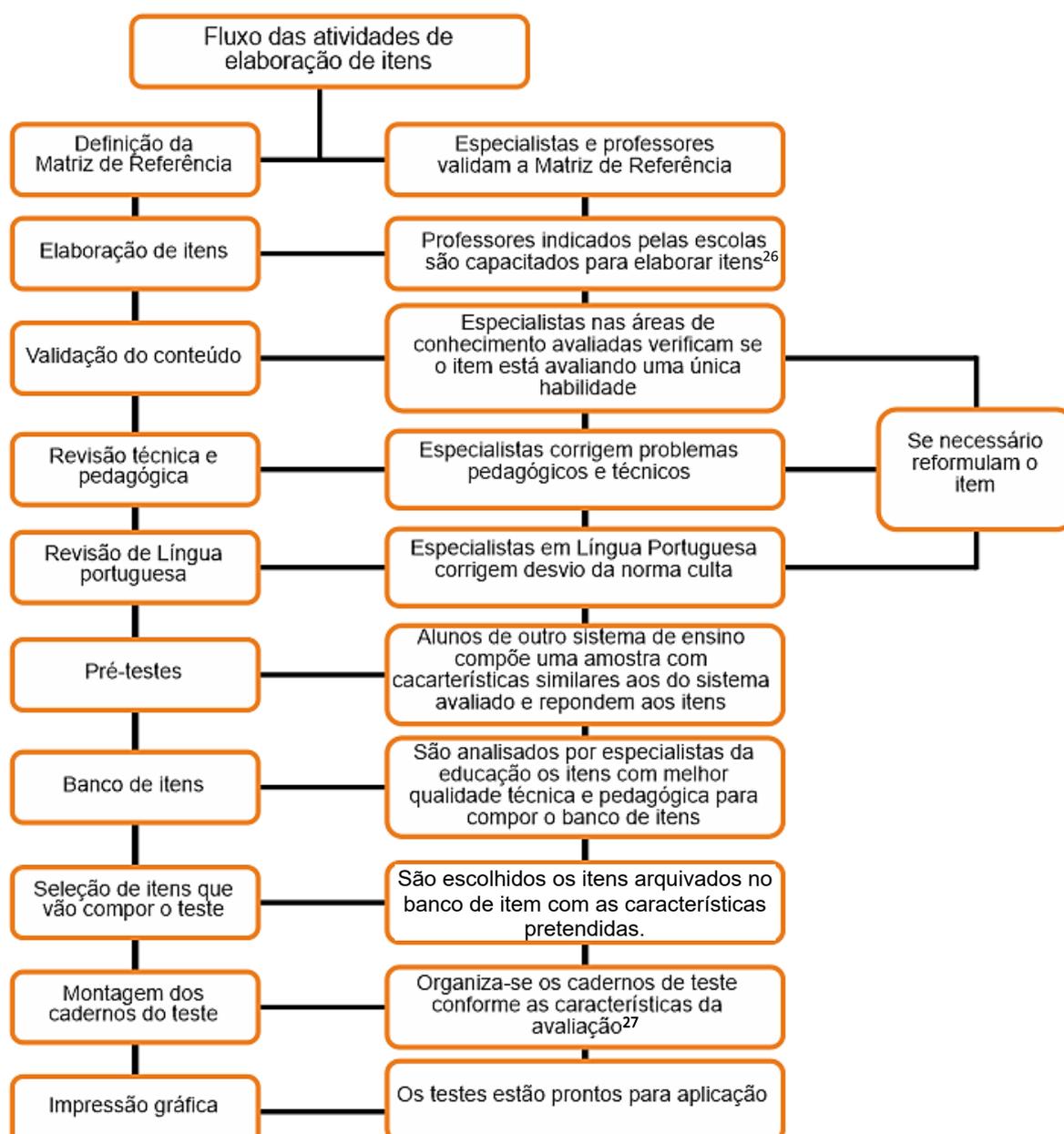
Assim, a construção de bons itens para compor os testes de habilidade/proficiência, que normalmente são utilizados em avaliações de larga escala, compreende diversas etapas

---

<sup>25</sup> Mais detalhes sobre o processo de elaboração de um item podem ser encontrados em: CAEd (2008); BRASIL (2010); RABELO (2013); MUÑIZ (2018); COHEN; SWERDLIK; STURMAN (2014).

envolvendo psicometristas, pedagogos e outros profissionais da educação. Como foi dito, itens de boa qualidade refletem na qualidade do instrumento de medida a ser produzido e consequentemente reflete na confiabilidade dos resultados de um processo de avaliação. Nesse sentido, a seguir é apresentado um fluxograma com o percurso para a construção de um item, bem como a sua aplicação nos testes.

Figura 27 - Etapas do processo de elaboração de um item.



Fonte: adaptada de CAEd (2008).

<sup>26</sup> A etapa de elaborações dos itens, para as avaliações do INEP, dar-se-á por meio de chamadas públicas, a partir de edital, com o intuito de convocar interessados em elaborar itens para a composição de provas.

<sup>27</sup> Em algumas avaliações, como as do Saeb, a montagem de cadernos é realizada utilizando a metodologia de Blocos Incompletos Balanceados – BIB, que permite que um grande número de itens seja aplicado ao conjunto de alunos

Como pode ser visualizado na Figura 27, uma das etapas que compõem uma avaliação, quando se quer avaliar pedagogicamente as habilidades que um sujeito tem, é a construção de uma Matriz de Referência, que é um instrumento norteador quando se constrói um item. Em geral, essas matrizes são estruturadas a partir de habilidades e/ou competências apresentando o que é esperado que um respondente a um teste tenha desenvolvido ao longo dos seus anos de estudos. A Matriz de Referência não é um currículo – que é mais amplo-, mas apenas uma referência para um determinado teste ao qual, como mencionado, figuram as competências e habilidades que pretende ser avaliadas e analisadas.

Em alguns testes, como as avaliações do Saeb, a matriz de referência apresenta tópicos ou temas subdivididos em descritores e estes, por sua vez, relacionam os conteúdos curriculares as competências e habilidades esperadas para cada ano/série de um componente curricular (ver Quadro 8 abaixo). É a partir dessa matriz que os descritores “indicam habilidades gerais que se esperam dos alunos [e também] constituem a referência para [a construção e] seleção dos itens que devem compor uma prova” (BRASIL, 2008, p.18). A seguir é apresentada a matriz de referência de matemática no Saeb, do 5º ano do ensino fundamental, para o tema “Espaço e forma”, que é composta por cinco descritores: D1, D2, D3, D4 e D5, conforme pode ser observado no quadro a seguir.

Quadro 8 - Trecho da matriz de referência de Matemática do Saeb: Temas e descritores para o 5º ano do ensino fundamental<sup>28</sup>.

<b>I. ESPAÇO E FORMA</b>	
D1	Identificar a localização/movimentação de objeto em mapas, croquis e outras representações gráficas.
D2	Identificar propriedades comuns e diferenças entre poliedros e corpos redondos, relacionando figuras tridimensionais com suas planificações.
D3	Identificar propriedades comuns e diferenças entre figuras bidimensionais pelo número de lados, pelos tipos de ângulos.
D4	Identificar quadriláteros observando as posições relativas entre seus lados (paralelos, concorrentes, perpendiculares).

avaliados, sem que cada aluno precise responder a todos eles. Para mais detalhes sobre a metodologia BIB ver: BRASIL (2019) e/ou BEKMAN (2001).

<sup>28</sup> A matriz de referência completa está disponível em BRASIL (2008, p.106-108).

<b>I. ESPAÇO E FORMA</b>	
D5	Reconhecer a conservação ou modificação de medidas dos lados, do perímetro, da área em ampliação e/ou redução de figuras poligonais usando malhas quadriculadas.

Fonte: Brasil (2008).

No Enem, com o objetivo de medir a proficiência/habilidade do participante em Matemática e suas tecnologias; Linguagens, códigos e suas tecnologias; Ciências Humanas e suas tecnologias; e Ciências da natureza e suas tecnologias, foram criadas, para cada uma dessas áreas, uma matriz de referência que toma como base os documentos oficiais que subsidiam o Ensino Médio (BRASIL, 2012a).

Para a matriz de referência de Matemática e suas tecnologias do Enem, por exemplo, tem-se que elas foram divididas em sete competências e estas, por sua vez, subdivididas em 30 habilidades a serem avaliadas, cuja intenção é saber quais delas foram desenvolvidas pelo respondente ao realizar o exame. O quadro a seguir apresenta a competência de área 1, subdividida em cinco habilidades: H1, H2, H3, H4 e H5.

Quadro 9 - Trecho da matriz de referência de Matemática e suas tecnologias do Enem<sup>29</sup>.

<b>Competência de área 1 - Construir significados para os números naturais, inteiros, racionais e reais</b>	
H1	Reconhecer, no contexto social, diferentes significados e representações dos números e operações - naturais, inteiros, racionais ou reais.
H2	Identificar padrões numéricos ou princípios de contagem.
H3	Resolver situação-problema envolvendo conhecimentos numéricos.
H4	Avaliar a razoabilidade de um resultado numérico na construção de argumentos sobre afirmações quantitativas.
H5	Avaliar propostas de intervenção na realidade utilizando conhecimentos numéricos.

Fonte: Brasil (2009).

<sup>29</sup> A matriz de referência do Enem completa encontra-se em: [https://download.inep.gov.br/download/enem/matriz\\_referencia.pdf](https://download.inep.gov.br/download/enem/matriz_referencia.pdf). Acesso em: 21 abr. 2021.

Quando se quer construir um item para compor um teste em uma avaliação, normalmente se tem como primeira etapa a escolha da habilidade que se pretende avaliar - conforme a matriz de referência das habilidades de cada teste - e apenas em seguida constrói-se uma situação-problema para o item. Nesse sentido, por exemplo, quando um aluno responde corretamente a um item, cuja habilidade a ser associada a ele é a H2 - Identificar padrões numéricos ou princípios de contagem -, descrita acima no Quadro 9, é esperado que ele tenha adquirido a referida habilidade. Esses itens – construídos a partir das habilidades da matriz de referência- e as proficiências/habilidades dos examinandos poderão ser comparados e é dessa forma que se pode avaliar quais habilidades o examinando adquiriu ao longo do tempo até realização do teste, como foi apresentado na Seção 4.1.

Nas etapas subsequentes de validação de conteúdo, revisão técnico-pedagógica e da língua portuguesa, alguns itens são reformulados ou ainda descartados, pois segundo Júdice e Karino (2015, p. 9) mesmo que o processo tenha sido aprimorado ao longo dos anos “ainda é alto o descarte de itens inadequados. Muitas vezes eles não servem porque não distinguem bem quem sabe de quem não sabe; noutras porque não tratam do que o aluno deve dominar (competência)”. Desse modo, realizar todos esses procedimentos para a construção de um item, como apresentado no fluxograma acima, são importantes para garantir a qualidade do item, seu poder avaliativo e conseqüentemente reflete em testes mais fidedignos.

Nesse sentido, mesmo que o item seja aprovado nas etapas de construção, validação e revisão, é necessária uma avaliação da qualidade psicométrica dos itens, e isso é concebido a partir dos resultados da etapa de *pré-teste* (ou pré-testagem), que consiste na aplicação de um conjunto de itens a uma amostra de sujeitos com características semelhantes às da população para a qual a prova se destina. Conforme Brasil (2012a, p. 21) essa etapa “é a forma empírica de se avaliar a qualidade técnico-pedagógica e psicométrica dos itens [...] e ela tem como objetivo captar subsídios importantes para aumentar a precisão da prova” e, quando utilizado a técnica da TRI, poderá selecionar os itens, que foram previamente testados e aprovados no pré-teste, para compor distintas provas, inclusive em períodos diferentes, para que os resultados possam ser comparados de forma estatisticamente mais adequada.

Desse modo, quando se quer arquivar itens de qualidade para compor os testes em uma avaliação é fundamental a realização desse pré-teste, mas para que esses itens sejam arquivados é necessário que cada item passe por avaliação qualitativa e quantitativa rigorosa por especialistas (REEVE et al., 2007 apud COHEN; SWERDLIK; STURMAN, 2014). Esses arquivos ou armazém de itens é denominado de *banco de itens*, que para Campos (2013, p.35),

tem por objetivo “armazenar itens de qualidade técnica que permitam a montagem de provas capazes de estimar com maior precisão a proficiência dos estudantes” com referência às suas competências, habilidades e características.

Segundo Muñiz (2018, p .269, tradução nossa) “durante as fases iniciais da construção do banco de itens, recomenda-se que o número de itens inicialmente seja pelo menos o dobro do que finalmente considera-se que poderia fazer parte da versão final do instrumento de medição”, pois alguns itens poderão não ser utilizados por ser incompatíveis com as características desse instrumento. Além disso, se a função de um banco de itens consiste em fornecer informações necessárias ao desenvolvedor de uma avaliação para construir os testes com características específicas, então esse arquivo deve conter itens que cubram todos os níveis da escala de habilidade, e isso normalmente significa que um banco de itens deve ter uma boa quantidade de itens fáceis, medianos e difíceis (PASQUALI, 2018).

Além disso, o banco de itens pode ser composto de itens de diversos formatos, tais como itens de múltiplas escolhas, de certo ou errado, do tipo *Likert*; mas o importante é que as suas estatísticas psicométricas descritas pela TCT e TRI – e outras informações gerais – estejam cadastradas para cada item. Para tanto, como um banco de itens contém uma quantidade grande de itens cadastrados, a procura de itens com as características desejada para compor um teste, torna-se facilitada quando é criada uma nomenclatura, como por exemplo: C1T2D3, que se refere a um item de C1: competência 1, T2: tema 2, D3: descritor 3 (BARRETO; MENDES; OLIVEIRA, 2009).

Mesmo que as informações cadastradas para o item possam variar, conforme planejado pelo construtor do banco de itens, ela deve/pode conter a identificação do item (nome/código) e as estatísticas da TRI e TCT, como: Curva Característica do Item (CCI) e a Curva de Informação do Item (CII); parâmetros dos itens da TRI: índice de dificuldade, discriminação e acerto ao acaso; percentual de acerto em cada alternativa do item e correlação item total da TCT; índice de dificuldade (DIF) e discriminação (DISC) da TCT; Análise gráfica do item (AGI) da TCT . Além das informações supracitadas, Pasquali (2018) diz que ainda deve/pode conter a função diferencial do item em termos de grupos, carga fatorial, e nela ainda pode figurar comentários sobre o item<sup>30</sup>.

No Saeb, por exemplo, avaliam-se os itens com informações como os parâmetros dos itens pela TRI e também elementos da TCT, entre eles o índice de dificuldade (DIF), o índice

---

<sup>30</sup> Para mais detalhes sobre os elementos da TCT apresentados nesse parágrafo, ver em PASQUALI (2011).

de discriminação (DISC) do item, a proporção de resposta correta em cada item, o coeficiente bisserial do item e também de cada alternativa de resposta (inclusive para respostas em branco e também para as inválidas), além da análise gráfica do item - AGI. De forma semelhante, os itens do Saresp também são avaliados segundo alguns elementos da TCT e da TRI, como apresentado na Figura 28 a seguir:

Figura 28 – Estatísticas de um item da prova do 5º ano no SARESP 2011, e ancorada no nível Básico (175 a < 225).

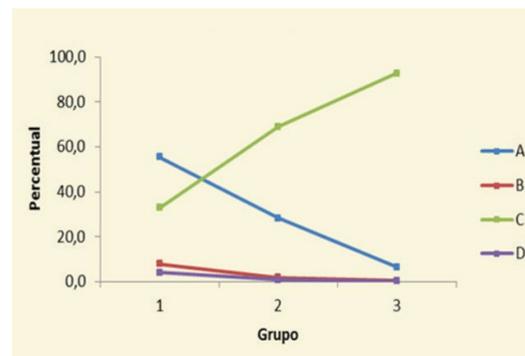
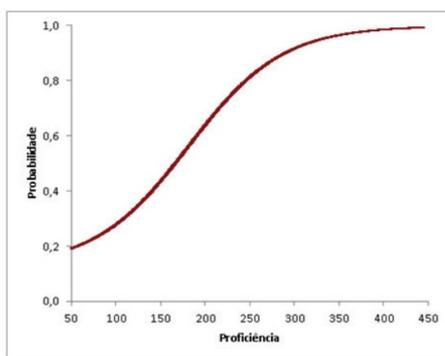
1. O desenho a seguir representa a parte do muro da escola que foi pintado



A fração que pode representar a parte pintada é

(A)  $2/3$   
 (B)  $2/4$   
 (C)  $2/5$   
 (D)  $2/6$

ÍNDICES			PERCENTUAIS DE ACERTO				PARÂMETROS DA TRI <sup>31</sup>		
GAB	DIF	DISC	A	B	C	D	<i>a</i>	<i>b</i>	<i>c</i>
C	0,33	0,60	28,5	3,2	66,8	1,6	0,02132	188,25	0,128



Fonte: adaptado de BORTOLUCCI et al. (2013).

É a partir dos dados, como apresentados na Figura 28, estimados por *softwares* computacionais especializados, que a qualidade psicométrica dos itens é avaliada com o

<sup>31</sup> No documento oficial esses parâmetros estão apresentados na escala (0,1), ou seja,  $a=1,066$ ,  $b=-1,235$  e  $c=0,128$  (BORTOLUCCI et al., 2013, p. 76).

objetivo de que eles possam ser arquivados em um banco de itens para posteriormente compor um teste. Nota-se ainda que, quando se tratam dessas análises aos itens, a TCT e a TRI são teorias se completam e desse modo pode-se destacar a importância de cada uma delas para uma avaliação.

Como foi dito, a construção de um banco de itens constitui uma ferramenta importante para os desenvolvedores de testes, pois é a partir desse arquivo que se pode selecionar os itens com as características pretendidas para compor um determinado instrumento de medição. Nesse sentido, o Inep criou o Banco Nacional de Itens (BNI), que é como uma coleção de itens de natureza específica disponíveis para a construção de instrumentos de avaliação, seja prova ou questionário. Esse banco de itens fornece insumos, assegurando acesso a uma grande quantidade de itens com qualidade técnico-pedagógica e psicométrica que irão compor testes em avaliações de larga escala e exames (BRASIL, 2018a).

O BNI é essencial, pois é desse arquivo que são extraídos os itens para a elaboração das provas das avaliações de larga escala como o Exame Nacional de Desempenho de Estudantes (Enade), do Exame Nacional de Certificação de Competências de Jovens e Adultos (Encceja), do Exame Nacional do Ensino Médio (Enem), do Programa Internacional de Avaliação de Estudantes (Pisa), do Sistema de Avaliação da Educação Básica (Saeb), entre outros.

Como já foi dito, os itens são criados e avaliados por um grupo de especialistas para que possam compor um banco de itens, mas com o tempo é necessária uma manutenção e atualização nesse banco - inserindo ou retirando itens-, pois com o tempo eles podem perder suas características de bons itens. Nesse sentido, Pasquali (2018, p.204) afirma que essa manutenção feita com a utilização da Teoria de Reposta ao Item (TRI) constitui

uma tarefa facilitada, porque ela permite estabelecer os parâmetros dos itens [dificuldade, discriminação e acerto ao acaso] [...] independentemente da amostra de sujeitos utilizadas levando em conta as características semelhantes dos sujeitos [a quem irá se destinar o teste]; daí é possível sempre incluir novos itens diretamente comparáveis com os já incluídos no banco de itens [...]. A técnica [...] consiste em aplicar os novos itens juntamente com uma amostra de itens já incluídos no banco de itens a uma amostra razoável de sujeitos e estimar os parâmetros dos novos itens.

Essa é uma das grandes vantagens da utilização da TRI, já que o banco de itens necessita de atualizações, acréscimos ou supressões de tempos em tempos, em relação aos itens que o

compõe, com intuito de que eles possam ter as características pretendidas e qualidade para uma avaliação específica.

Portanto, como visto acima, criar um banco de itens, que é uma aplicação da Teoria de Resposta ao Item, constitui uma etapa importante para a eficiência e qualidade do instrumento de medida, principalmente em avaliações de larga escala.

## 5 O CONHECIMENTO DO PROFESSOR DE MATEMÁTICA ACERCA DA TEORIA DE RESPOSTA AO ITEM

O estudo apresentado neste capítulo relaciona-se a uma pesquisa que tem por objetivo colher as impressões do(a)s professore(a)s do ensino básico, que lecionam matemática nas redes de ensino no estado de Alagoas, acerca do seu conhecimento sobre a Teoria de Resposta ao Item (TRI), que é amplamente utilizada diversas avaliações, principalmente as externas em larga escala.

### 5.1 Metodologia

A pesquisa foi realizada com uma amostra aleatória de 67 professores que lecionam matemática nas redes de ensino no estado de Alagoas. Ela tem uma abordagem quantitativa e caracteriza-se como sendo exploratória com levantamento de dados sobre o que os professores conhecem acerca da TRI.

As informações e dados foram obtidos a partir de um questionário (ver **APÊNDICE E**) e continha ao todo 8 questões de informações gerais<sup>32</sup> e 17 questões específicas ao tema, aplicado de forma on-line por meio de um formulário eletrônico (Google Forms)<sup>33</sup> durante o período de 07/07/2021 a 12/07/2021.

Cada questão ligada ao tema da pesquisa apresentou alternativas de respostas objetivas, exigindo obrigatoriamente a seleção de uma opção. Em algumas delas continha um campo aberto e opcional para que o professor justificasse a sua escolha, permitindo expressar sua opinião ou conhecimento de maneira livre. Ademais, apenas na questão 4 (Qual a rede de ensino que você leciona?) o professor tinha a possibilidade de selecionar mais de uma opção.

---

<sup>32</sup> Dentre essas oito questões existiu uma questão de controle, a questão 4 (Você leciona matemática na educação básica?), ao qual foi verificado que dos 71 professores que responderam ao questionário quatro deles marcaram a opção “não” e foram excluídos da pesquisa, já que a população em estudo são os professores de matemática do ensino básico que atuam na rede de ensino em Alagoas.

<sup>33</sup> Devido a pandemia de Covid-19 a pesquisa não pode ser realizada complementarmente de forma presencial, limitando ao pesquisador a realizá-la de forma on-line. Ademais, no período da pesquisa, a maioria das escolas do estado de Alagoas não estava tendo aulas presenciais, o que obstaculizou o contato com os professores de matemática.

A partir das informações sobre o conhecimento dos professores que ensinam matemática em Alagoas, esta pesquisa tem o objetivo adicional de servir como um norte para que se possa propor ações ou políticas públicas que envolva a qualificação profissional desse professor, principalmente relacionadas às formações continuadas propostas pela(s) sua(s) respectiva(s) rede(s) de ensino – incluindo parcerias com as universidades-, já que entender sobre a Teoria de Resposta ao Item (TRI) constitui um tema essencial na formação do professor do ensino básico.

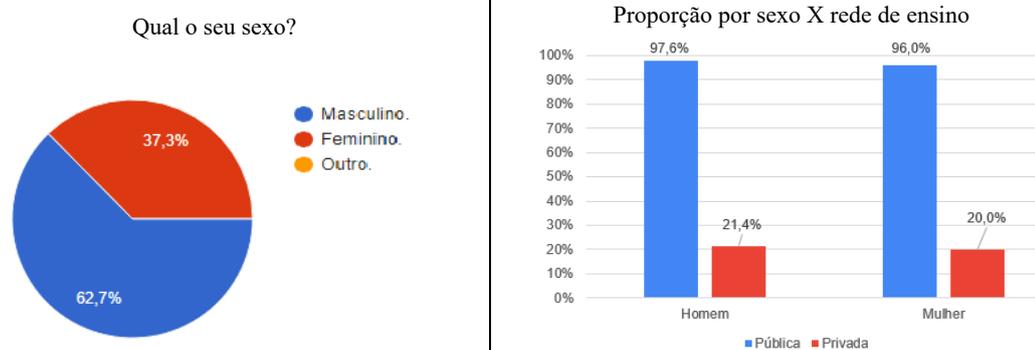
## 5.2 Resultados e discussões

Segundo Marconi e Lakatos (2002, p. 35) “a importância dos dados está não neles mesmos, mas no fato de proporcionarem respostas às investigações”. Nesse sentido, serão apresentados a seguir os resultados e discussões sobre o perfil dos professores que participaram da pesquisa e também sobre o seu conhecimento acerca da Teoria de Resposta ao Item (TRI).

### 5.2.1 Análise das informações gerais sobre o professor de matemática

As informações contidas nesta subseção versam sobre o perfil do professor de matemática, coletada a partir de um questionário composto por perguntas como a sua formação, tempo de ensino na educação básica, dentre outras. Os resultados estão apresentados nos gráficos das Figura 29 a Figura 32 e descrevem algumas características dos professores que participaram da pesquisa.

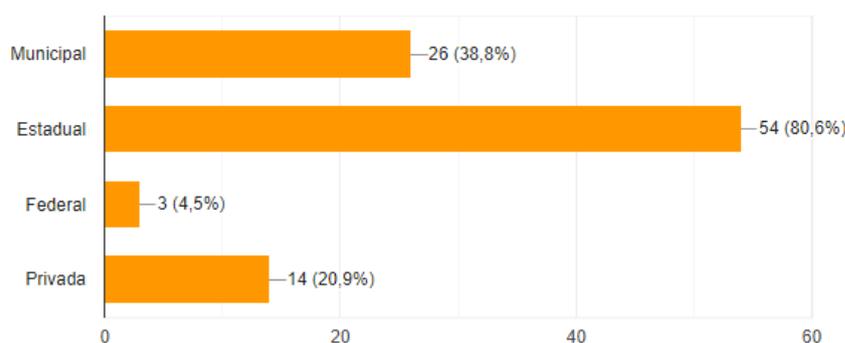
Figura 29 - Distribuição de professores por sexo (esquerda) e por sexo X rede de ensino (direita).



Fonte: elaboração do autor

Inicialmente observa-se na Figura 29 que dos 62,7% dos professores do sexo masculino, 97,6% lecionam na rede pública e 21,4% atuam na rede privada. Já dos 37,3% dos professores do sexo feminino, tem-se que 96% lecionam na rede pública e 20% na rede privada. Nenhum participante escolheu a opção “outro” em relação ao seu sexo. Ademais, a proporção geral dos professores de matemática por rede de ensino pode ser observada na Figura 30, a seguir:

Figura 30 - Quantidade e percentual de professores de matemática por rede de ensino.

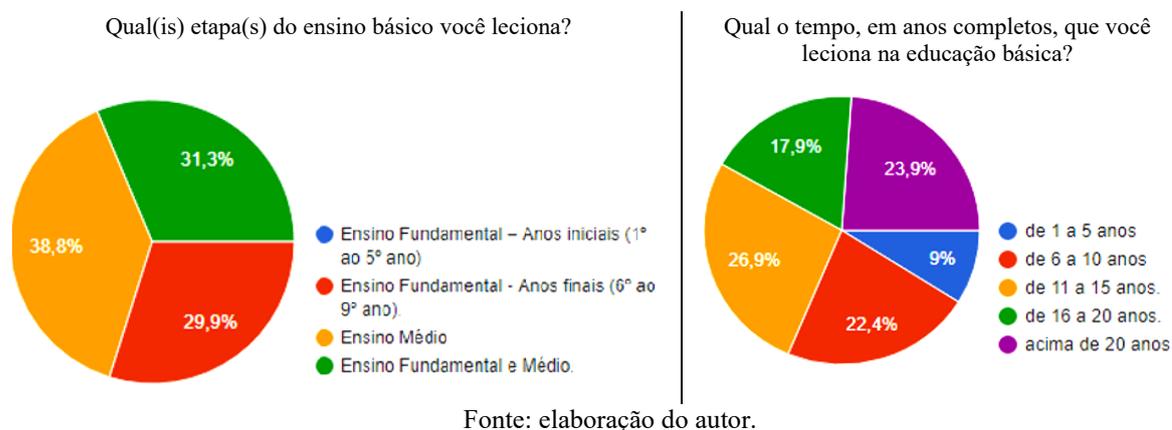


Fonte: elaboração do autor.

Nota-se da Figura 30 que a maioria dos professores que responderam ao questionário atuam na rede estadual de ensino (80,6%), e desses, 48,1% atuam em outras redes de forma concomitante. Dos professores das redes municipais de ensino (38,8%), tem-se que 80,8% também atuam em outras redes de ensino. Além disso, dos professores que lecionam na rede privada (20,9%) tem-se que 85,7% também atuam em outras redes. Por fim, todos os professores que responderam lecionar no ensino federal (4,5%) atuam exclusivamente nessa rede de ensino.

Nos gráficos da Figura 31, a seguir, são apresentados o percentual relativo às etapas da educação básica que os professores lecionam (à esquerda) e também o tempo que esses professores atuam na educação básica (à direita). Foi observado nessa análise que apenas 29,9% atuam exclusivamente no ensino fundamental e que os demais professores (70,1%) atuam no ensino médio (apenas no ensino médio ou no ensino médio e ensino fundamental).

Figura 31 - Etapas do ensino básico (à esquerda) e tempo de ensino (à direita) dos professores de matemática.

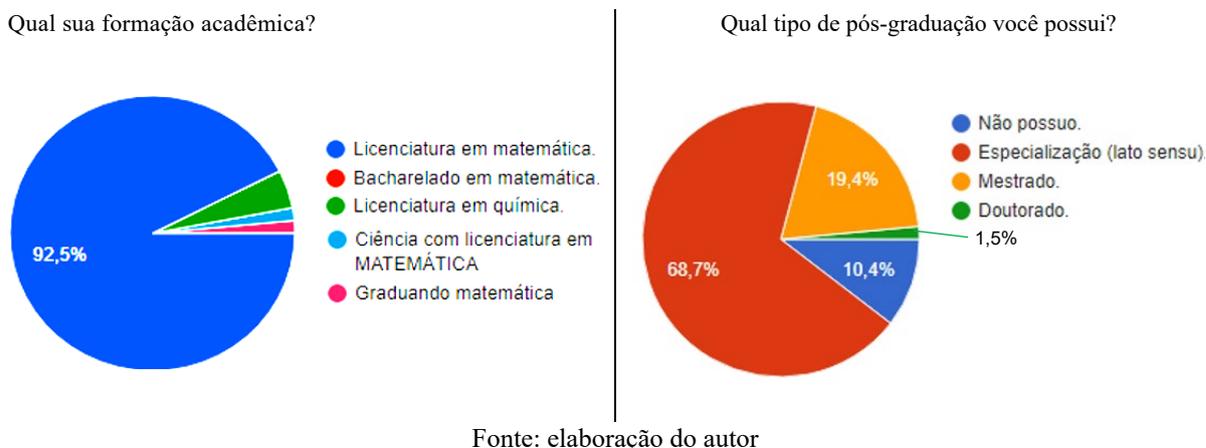


Além dos dados supracitados, tem-se, do gráfico mais à direita, que 68,7% dos professores atuam há mais de 10 anos na educação básica, indicando que eles tiveram contato por diversas edições de avaliações que utilizam a TRI, como o Saeb (a cada dois anos para o 5º e 9º ano do ensino fundamental, e também para o 3º ano do ensino médio) e o Enem (a cada ano para o ensino médio). Daí, espera-se que esses professores conheçam acerca do funcionamento da TRI nas avaliações externas em larga escala.

Mesmo que o dado observado acima faça referência aos professores com atuação de mais de 10 anos de atuação no ensino básico, pode-se afirmar também que os demais professores (31,3%) tiveram - ou deveriam ter tido- contato com algumas dessas avaliações externas que utilizam a TRI como metodologia para ao menos realizar estimativas das proficiências dos estudantes.

Dos gráficos a seguir (Figura 32), tem-se que dos professores que responderam ao questionário, 92,5% são formados em Licenciatura em Matemática. Dos demais professores, dois deles possui formação em Química (licenciatura), um professor é formado em Ciências (com habilitação em Matemática) e um graduando em Matemática. Além disso, desses professores que responderam ao questionário, 89,6% possui algum tipo de pós-graduação.

Figura 32 - Formação dos professores que ensinam matemática.



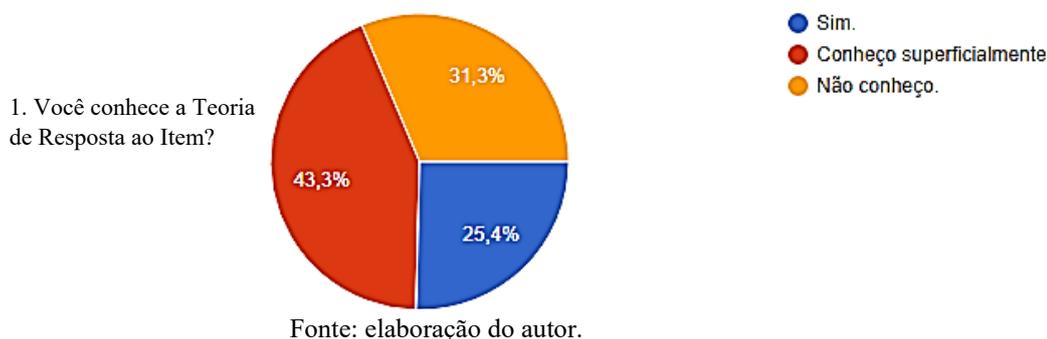
Portanto, a partir das informações apresentadas acima pode-se ter uma caracterização do perfil do professor que participou da pesquisa. Uma análise sobre o conhecimento desse professor acerca da TRI será apresentada na subseção a seguir.

### 5.2.2 Análise das respostas dos professores sobre o seu conhecimento acerca da Teoria de Resposta ao Item (TRI)

Entender sobre a TRI deve/deveria ser fundamental para um professor que leciona no ensino básico, já que diversas avaliações em larga escala utilizam essa metodologia para avaliar o conhecimento dos alunos em testes, mas na realidade muitas vezes essa questão não é confirmada, talvez seja pela precária formação acadêmica inicial e/ou continuada do professor -relativo a esse tema ou afins- ou ainda pela falta de tempo para seus estudos complementares de forma autônoma devido ao acúmulo excessivo de carga de trabalho.

Nesse sentido, foram realizadas 17 perguntas ao professor, relacionadas ao seu conhecimento acerca da TRI e dos temas afins. A questão 1 versa sobre esse conhecimento e os resultados estão apresentados no gráfico da Figura 33 a seguir.

Figura 33 - Proporção sobre o conhecimento dos professores de matemática em relação a TRI.



Analisando as informações contidas no gráfico acima, pode-se constatar que 74,6% tem um conhecimento superficial ou não conhece sobre o tema. É impactante perceber que quase 1/3 dos professores (31,3%) sequer conhece sobre a TRI. Além das estatísticas apresentadas acima, alguns professores, que afirmaram ter um conhecimento superficial, justificaram o seu conhecimento em relação a TRI como:

- a) “Tenho a ideia do que seja, mas nunca tive nenhuma formação sobre.”
- b) “Acho que o ENEM é elaborado assim.”
- c) “Li sobre o assunto em alguns textos, mas nada profundo.”
- d) “Esse método estatístico consegue diminuir as chances de que o candidato tenha uma boa nota a partir de chutes, uma vez que identifica incoerências e, assim, atribui notas mais justas.”
- e) “Apenas sei que se um aluno responde corretamente um item considerado difícil e erra um considerado fácil, a nota gerada será menor do que a esperada.”
- f) “Conheço muito pouco, sei que nas provas do SAEB utilizam-se desse conhecimento para avaliar mais qualitativamente o desempenho do estudante.”
- g) “Conheço porque é o método utilizado para correção das provas do ENEM.”

Dos professores que responderam ter conhecimento superficial apenas 51,7% justificaram suas respostas. Desses, poucos explicaram algo sobre a TRI. Alguns afirmaram que já leram sobre o tema, mas sem dar mais explicações, como visto em “a” e “c”. Um único professor, o “f”, afirmou saber que essa é a teoria adotada pelo Saeb e que ela é utilizada para

avaliar os alunos de forma qualitativa. Outros professores, como “b” e “g”, apenas reconhecem ser o método utilizado no Enem. Por fim, os professores “d” e “e” explicaram que a coerência no padrão de respostas dos alunos reflete em melhores notas.

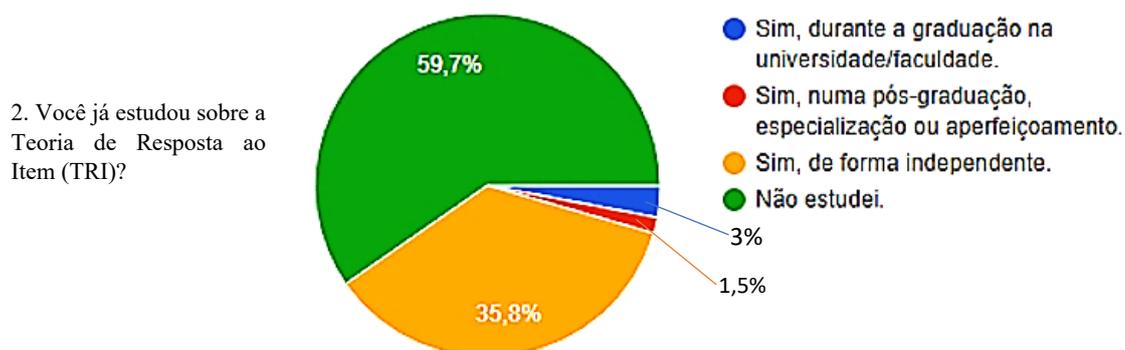
Para os professores que responderam conhecer a TRI, segue abaixo algumas dessas respostas:

- a) “Tipo de avaliação que o ENEM usa.”
- b) “Observa o nível de dificuldade do aluno tendo questões fáceis, médio e difícil, nisso o professor observa se ele chutou ou não, se errar muitas questões fáceis e acertar as difíceis deduz que ele chutou.”
- c) “Já vi algumas avaliações que utilizam este método.”
- d) “Valida a resposta de acordo com o nível de conhecimento do aluno.”
- e) “Avaliação das questões pelo nível de dificuldade. Tentando evitar o famoso chute.”
- f) “A teoria de resposta ao item TRI é por exemplo usada no ENEM. Uma prova bem elaborada traz questões fáceis, médias e difíceis. A TRI atribui pontos a essas questões. Ela penaliza o aluno que acerta questões difíceis e erra questões fáceis. Ao errar as fáceis ele mostra que o acerto das difíceis foi chute ou fraude.”
- g) “Conheço, principalmente, por conta da metodologia de elaboração de questões para o Exame Nacional do Ensino Médio (ENEM). Em vários momentos, tive que estudar essa teoria para compreender a estrutura do ENEM e poder preparar os estudantes da melhor maneira possível para o exame.”

Das respostas dadas no questionário, apenas 64,7% dos professores, que afirmaram conhecer a TRI, deram alguma justificativa para essa questão. Alguns professores como o “a”, “f” e “g”, justificaram suas respostas mencionando que a TRI é uma teoria utilizada no Enem, mas apenas um professor, o “f”, deu alguma explicação simplificada de como essa teoria funciona nesse exame. Outros professores (“b” e “e”) afirmaram que a teoria reconhece, a partir do seu padrão de resposta, se o aluno “chutou” (escolheu uma resposta ao acaso). Já os professores “c” e “d” responderam de forma mais genérica a questão.

O gráfico da Figura 34, a seguir, apresenta o percentual de professores que estudaram ou não a TRI em algum momento de sua formação.

Figura 34 - Professores x Estudo da TRI durante a formação.



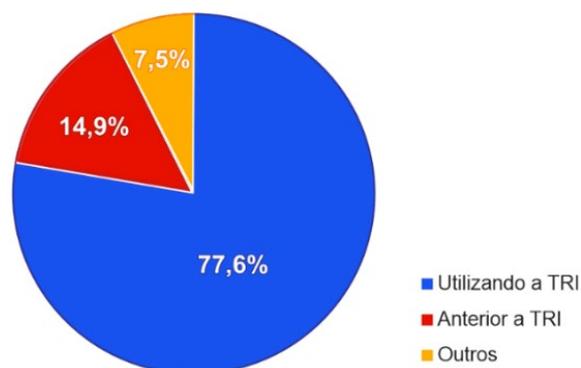
Fonte: elaboração do autor

Nota-se que aproximadamente 60% nunca estudaram sobre o tema em questão e esse dado é importante, pois a grande maioria desses professores (72,5%) leciona na educação básica há mais de 10 anos. Apenas 4,5% estudaram na graduação ou pós-graduação, indicando uma possível lacuna na construção dos programas dos cursos de graduação ou pós-graduação, relativo a esse tema. Ademais, essa questão pode ser observada no gráfico acima, onde 35,8% dos professores estudaram de forma independente sobre a TRI e essa escolha possivelmente se deu pelo fato desses professores não terem estudado esse tema durante a sua formação acadêmica.

Na figura a seguir, destaca-se a opinião do professor sobre a forma de avaliar os estudantes nas avaliações externas em larga escala, cujo resultado segue abaixo:

Figura 35 - Opinião sobre a melhor forma de mensurar o desempenho dos estudantes em avaliações externas em larga escala.

3. Antes as avaliações externas em larga escala no Brasil mensuravam o desempenho do aluno a partir do número de questões acertadas (essa perspectiva está relacionada a Teoria Clássica dos Testes-TCT). Atualmente esse desempenho é medido a partir das técnicas da TRI (Teoria de Resposta ao Item). Na sua opinião qual é a melhor forma para obter esse desempenho?



Fonte: elaboração do autor

A introdução da TRI nas avaliações externas em larga escala constituiu um marco, pois a utilização dessa teoria possibilitou, por exemplo, fazer comparações entre diferentes provas ao longo do tempo com intuito de avaliar a evolução do ensino e aprendizagem na educação. Para tanto foi perguntado ao professor sobre a sua opinião em relação à melhor forma de calcular o desempenho dos estudantes nessas avaliações. Entre as respostas, 7,5% dos professores citaram que existem outras formas melhores para calcular esse desempenho, quais sejam:

- a) “Questões abertas.”
- b) “Um método que oportunize a todos.”
- c) “Não tenho aprofundamento suficiente pra opinar.”
- d) “Não conheço.”

Observa-se nessas respostas que dois professores afirmaram não conhecer nenhum método para calcular o desempenho de um estudante em uma avaliação de larga escala. O professor “a” acredita que questões abertas avaliariam melhor um estudante, mas esse tipo de questão tanto pode ser considerado na TCT ou na TRI, então aparentemente ele não conhece acerca dessas teorias. Por fim, o professor “b” deu uma resposta vaga, não podendo assim ser interpretada objetivamente.

Abaixo, seguem-se as justificativas de alguns desses professores que responderam a alternativa “Anterior a TRI”:

- a) “O famoso chute.”

- b) “Como não conheço o TRI, não sei como funciona esse tipo de avaliação.”
- c) “Pois tem como diagnosticar de forma precisa o conhecimento a partir da análise das respostas. Acredito que estudar o erro é tão importante quanto estudar os acertos.”
- d) “Muitas vezes os alunos desenvolvem raciocínios lógicos que ia auxiliam a responder questões que são consideradas difíceis, nesse caso não considero relevante o aluno perder sua pontuação só porque ele não conseguiu resolver uma questão considerada fácil.”

O professor “a” prefere algum modelo que aparentemente valoriza e pontua o “chute” do estudante, já que em alguns modelos da TRI, como o 3PL, o acerto ao acaso é considerado na análise quando se quer calcular a proficiência do estudante. Já o professor “b” escolheu a opção “anterior a TRI” por não conhecer essa teoria, mas afirmou nos outros itens do questionário possuir interesse em estudar e/ou compreender sobre os conceitos fundamentais sobre o tema. O professor “c”, por sua vez, parece não conhecer nenhuma das teorias, pois sua resposta se enquadra mais com as características da TRI, que estima os resultados a partir do padrão de respostas (acertos e erros) do examinando, já que na TCT a preocupação principal situa-se no escore total (soma das pontuações das questões acertadas) e não na análise do conjunto das respostas dadas em cada item.

Por fim, o professor “d” menciona discordar de uma das características do modelo 3PL da TRI, em que o acerto casual é considerado na análise de desempenho do estudante. Esse professor acredita que a coerência no padrão de respostas do respondente a um teste não deve ser levando em conta e discorda da TRI (modelo 3PL) por achar errado o “aluno perder sua pontuação só porque ele não conseguiu resolver uma questão considerada fácil”. Na verdade, o aluno não perde a pontuação plenamente, mas o que ocorre é que em alguns casos ele não terá uma pontuação máxima como se espera, quando observada o padrão nas suas respostas.

Por outro lado, seguem abaixo algumas justificativas dos professores que responderam que a utilização da TRI avalia melhor o desempenho dos estudantes.

- a) “Avalia melhor.”
- b) “A TRI avalia se o aluno ‘chutou’ e isso é muito importante em uma real avaliação de desempenho cognitivo.”

- c) “Com isso, é possível saber pontualmente em quais habilidades o aluno tem domínio ou dificuldade.”
- d) “Leva em consideração o conhecimento que o estudante possui.”
- e) “Acredito que utilizando a TRI leve em conta as habilidades demonstradas pelo aluno.”
- f) “Sim, a Teoria de Resposta ao item pode ser um modelo de avaliação muito benéfica.”
- g) “Vê em que nível está o aluno e se ele apreendeu os conteúdos.”
- h) “Se dá o valor particularizado de cada questão, de acordo com o nível de proficiência exigido. Além da dedução de questões que podem ter sido respondidas aleatoriamente.”
- i) “Penso que a TRI seja melhor num sentido de abarcar mais variáveis (por exemplo, o nível de complexidade da questão) e não somente a quantidade de acertos e erros. A TRI traz uma abrangência maior por considerar aspectos qualitativos em detrimento a aspectos unicamente quantitativos (isso facilita até a possibilidade de comparação de resultados - as quais, dentro da TCT não é eficaz por desconsiderar a complexidade das questões).”

Analisando as respostas dos professores, percebe-se que nenhum deles mencionou em suas respostas de que a escolha da TRI para avaliações de larga escala se justifica pelo fato desse tipo de teoria permitir comparar a evolução da aprendizagem dos estudantes ao longo do tempo, além de permitir que resultados em provas distintas possam ser comparados e ainda que pequenas amostras em um teste têm expressiva representatividade.

Assim, das respostas apresentadas acima, observa-se que apenas o professor “i” apresentou uma resposta mais completa, evidenciando algumas características da TRI e da TCT. Tem-se ainda que, dos 76,6% dos professores que afirmaram ser melhor avaliar um estudante pela TRI, mais da metade deles (53,8%) não justificaram a sua escolha. Nesse sentido, dos professores que justificaram a sua escolha na TRI, constatou-se que 41,67% afirmou que essa teoria leva em conta a análise do chute (acerto causal) do estudante no teste, como em “b” e “h”. Já 12,5% dos que justificaram a escolha na TRI afirmaram que essa escolha se deu pelo fato de que com essa teoria pode-se saber em qual nível de conhecimento o estudante está a partir da sua proficiência, como em “c” e “g”, mas destaca-se que nenhum deles menciona como esse nível é obtido. Ademais, outras respostas (45,83%) se mostraram vazias ou genéricas, como em “a”, “d” e “f”.

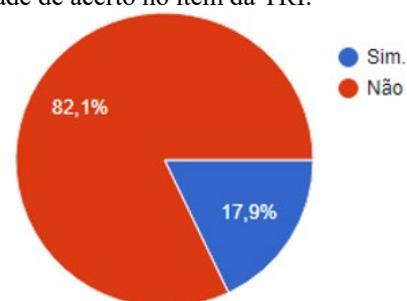
Conhecer sobre os modelos utilizados na TRI constitui uma importante ferramenta para compreender o funcionamento dessa teoria. Para tanto, quando se quer saber a probabilidade de um sujeito acertar um item, em função da sua habilidade e dos parâmetros dos itens ( $a$ ,  $b$  e  $c$ ), é necessário conhecer os modelos propostos por essa teoria. Na questão a seguir foi apresentado ao professor o modelo logístico de três parâmetros (3PL), pois esse modelo é o mais utilizado nas avaliações externa em larga escala no Brasil. Então, infere-se que esse professor deveria conhecer esse modelo.

Entretanto, observando o gráfico da Figura 36, a seguir, nota-se que a maioria dos professores (82,1%) não conhece a função apresentada na questão, e esse dado reforça o baixo conhecimento desses professores sobre o tema, já que a função, em geral, é apresentada na parte inicial nas publicações sobre a TRI. Por outro lado, esse resultado pode ser justificado, pois a base de estudo sobre a TRI por muitos professores é o manual do Enem denominado “Entenda sua nota no Enem”, que contém 33 páginas e traz essa função apenas em uma de suas últimas páginas sem nenhum detalhamento adicional (ver BRASIL, 2012a).

Figura 36 - Conhecimento do professor sobre a função probabilidade de acerto no item da TRI.

4. A expressão abaixo representa a probabilidade de um respondente acertar um item em função da sua habilidade/proficiência e dos parâmetros do item em um teste. Essa função é muito utilizada em avaliações que envolvem a TRI. Você conhece essa função?

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

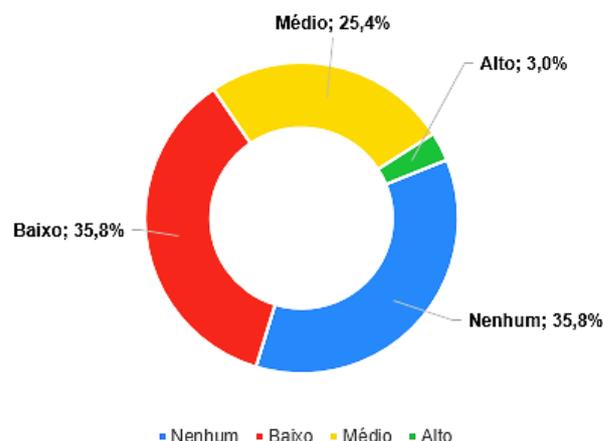


Fonte: elaboração do autor

Essa questão tem relevância para este estudo, pois, como foi mencionado, compreender como é calculada a probabilidade de um respondente, com uma certa habilidade, acertar um item constitui uma etapa importante para a consolidação do entendimento sobre essa teoria. Além disso, da função probabilidade apresentada acima (Figura 36), observa-se que ela é escrita em função de  $\theta$ ,  $a$ ,  $b$  e  $c$ , então para interpretar de forma expressiva os resultados dessa probabilidade de acerto em um item é importante compreender o que cada um desses termos significa. Para tanto, foi perguntado ao professor acerca do seu conhecimento sobre essa questão e o resultado está apresentado na figura a seguir:

Figura 37 - Conhecimento do professor sobre a variável ( $\theta$ ) e parâmetros dos itens ( $a$ ,  $b$  e  $c$ ).

5. A função apresentada na questão anterior possui parâmetros e uma variável. Qual o seu conhecimento acerca dos significados desses parâmetros e dessa variável?



Fonte: elaboração do autor

Dos professores que responderam ter um alto conhecimento sobre essa questão (3%), destaca-se que nenhum deles apresentou justificativa. Para os professores que disseram ter conhecimento médio (25,4%), apenas três deram alguma justificativa, como pode ser visto a seguir:

- a) “Subsequentes.”
- b) “Na equação  $ax^2 + bx + c = 0$ ,  $x$  é a incógnita enquanto  $a$ ,  $b$ ,  $c$  são parâmetros.”
- c) “Conhecendo os parâmetros dá para entender o significado.”

Observa-se das respostas dadas que nenhum deles apresentou uma justificativa coerente com o que se pede nessa questão. Em “a” a resposta do professor apresenta-se vaga e incompleta, daí não é possível avaliá-la. Em “b” o professor não compreendeu o que se pedia e no campo da justificativa entendeu que se referia à equação do segundo grau, o que não foi perguntado na questão. Pode-se inferir que esse professor marcou ter conhecimento médio, não sobre a questão relativa a TRI, e sim sobre equação do segundo grau, que não era o tema dessa pesquisa. Em “c” o professor acredita que só dá para saber sobre o significado do parâmetro se for dado o valor numérico, o que não é verdade. Mesmo assim esses professores acenaram ter conhecimento médio sobre a variável e os parâmetros da expressão indicada na questão, o que parece não ser um nível adequado para eles.

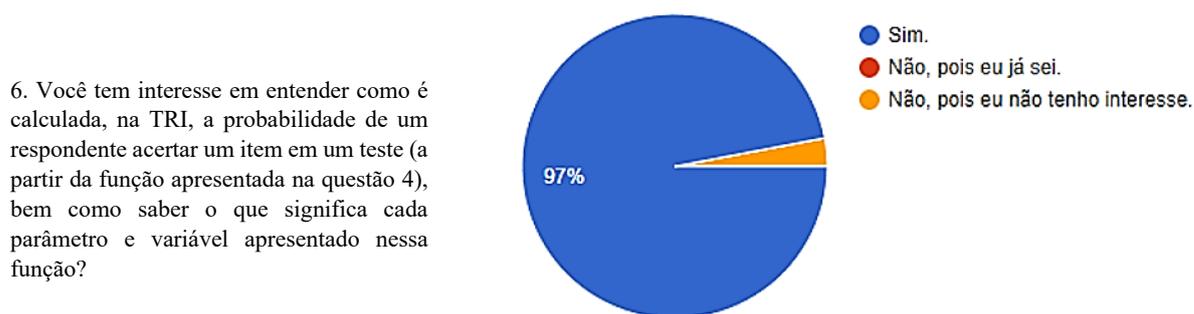
Para os que responderam ter baixo ou nenhum conhecimento (71,6%), segue algumas respostas:

- a) “Não conheço.”

- b) “O número de Euler é famoso, já os outros não sei o significado deles.”
- c) “Não tenho conhecimento da fórmula.”
- d) “vi a fórmula quando estou respondendo o esse questionário”
- e) “Não faço ideia desses parâmetros”
- f) “Eu não conheço essa fórmula, não tenho o conhecimento para desenvolver”
- g) “eu só conhecia o método não a expressão matemática”
- h) “Infelizmente, em meus estudos independentes sobre TRI nunca deparei com essa fórmula, sendo assim, não conheço o significado dos parâmetros e da variável apresentados na fórmula.”

Destaca-se que nenhum dos professores que justificaram sua resposta, nos níveis “baixo” ou “nenhum” (71,6%), mostraram saber o que significa tal variável e os parâmetros da fórmula, indicando um percentual alto sobre a falta de conhecimento acerca dessa questão. Por outro lado, como vemos na figura a seguir, a maior parte dos professores pesquisados (97%) indicou ter interesse em estudar/aprender sobre essa questão.

Figura 38 - Interesse do professor de matemática em entender como se calcula a probabilidade de acerto em um item, incluindo a compreensão do significado da variável e dos parâmetros.



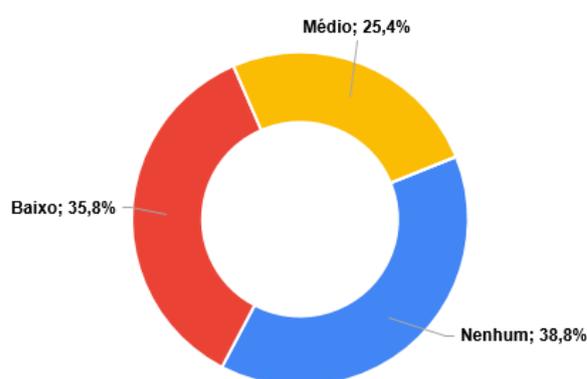
Fonte: elaboração do autor

A função, apresentada na questão 4 do questionário, produz um gráfico denominado Curva Característica do Item (CCI), que descreve a probabilidade de um sujeito, com uma certa habilidade, acertar um item. Nesse sentido, esse tema também foi abordado no questionário,

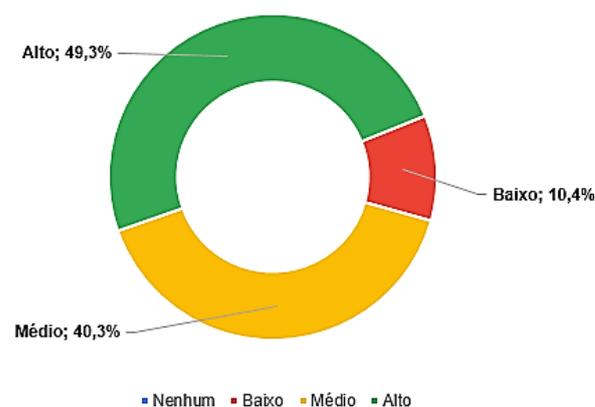
pois entender como se interpreta ou se constrói a CCI também constitui uma etapa importante para a compreensão dos fundamentos básicos da TRI.

Figura 39 - Conhecimento do professor de matemática sobre a Curva Característica do Item.

7. Nos testes que utilizam a TRI cada item (questão) possui um gráfico que descreve a probabilidade de acerto em função da habilidade do respondente, chamado de Curva Característica do Item. Sobre o seu conhecimento acerca desse gráfico, assinale:



8. Qual interesse você tem em entender como se interpreta ou constrói o gráfico da Curva Característica ao Item da TRI?



Fonte: elaboração do autor

O gráfico (à esquerda) da Figura 39 mostra que poucos professores conhecem como se interpreta ou constrói a CCI. Pode-se notar que 74,6% dos professores afirmaram ter “baixo” ou “nenhum” conhecimento sobre essa questão, o que pode configurar um problema, já que a CCI é considerada como uma unidade fundamental para os modelos da TRI e a sua compreensão é basilar quando se quer construir um teste utilizando essa teoria.

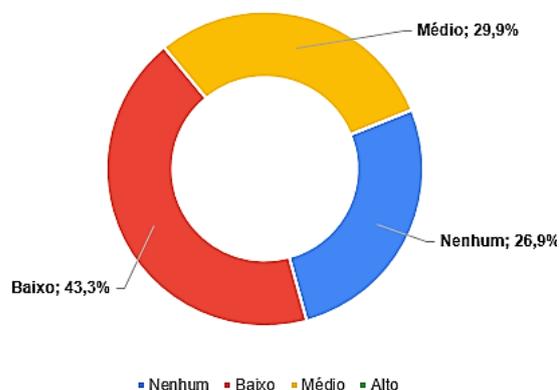
Por outro lado, quando observado o gráfico Figura 39 (à direita), nota-se que 89,3% tem um alto ou médio interesse em entender sobre essa questão. Portanto, mesmo que a maioria dos professores (74,5%) tenham afirmado baixo ou nenhum conhecimento sobre essa questão, eles responderam ter interesse em compreender mais sobre esse tema. Destaca-se ainda que quase a metade (49,35%) respondeu ter alto interesse. Se análise for ampliada para as respostas alto e médio, tem-se um total de 89,6% dos professores com interesse em compreender sobre os aspectos gerais da CCI.

Outra questão abordada no questionário trata do conhecimento acerca das estimativas das proficiências e dos parâmetros dos itens. Sabe-se que o objetivo principal de algumas avaliações é estimar as proficiências dos sujeitos, seja para classificar um estudante em um exame, como em um vestibular, ou ainda utilizar essa estimativa para avaliar de forma

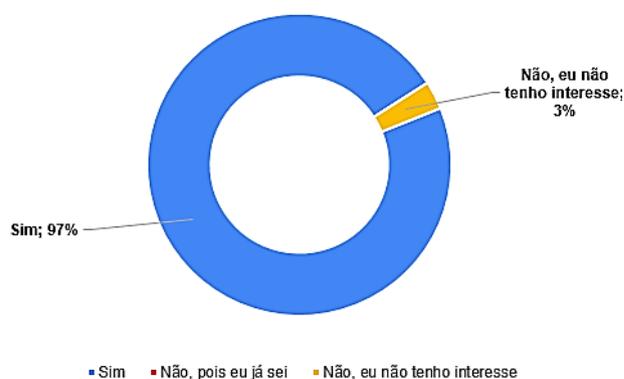
qualitativa o quanto esse sujeito aprendeu até o momento do teste, isto é, quais as habilidades esse sujeito possui. Ademais, estimar os parâmetros dos itens constitui também uma etapa importante, pois a partir desses parâmetros pode-se avaliar, por exemplo, a qualidade dos itens para compor um teste. Os resultados dessa questão estão apresentados na figura a seguir:

Figura 40 - Conhecimento do professor de matemática sobre as estimativas da habilidade e parâmetros dos itens na TRI.

9. Na TRI uma das etapas mais importantes é a estimação dos parâmetros de cada item (discriminação, dificuldade e acerto ao acaso) e também a estimação das proficiências (às vezes chamada de habilidade ou simplesmente nota) dos respondentes a um teste. Sobre o seu conhecimento acerca de como é realizada essas estimativas, assinale:



10. Você gostaria de compreender como são realizadas as estimativas destacadas na questão anterior?



■ Sim ■ Não, pois eu já sei ■ Não, eu não tenho interesse

Fonte: elaboração do autor

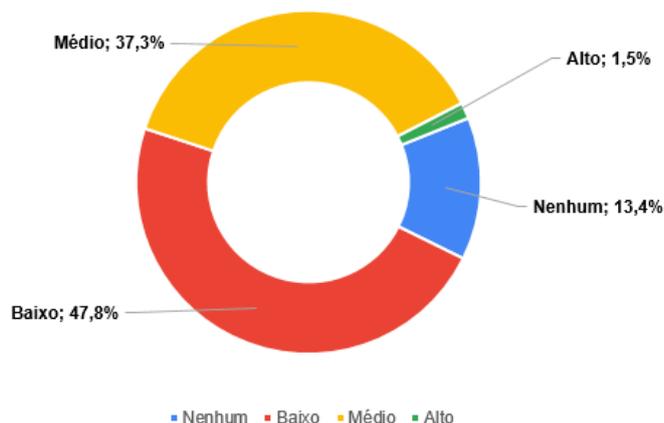
Do gráfico (à esquerda) da Figura 40, nota-se que nenhum professor afirmou ter alto conhecimento sobre essa questão. Daqueles que responderam possuir baixo ou nenhum conhecimento, figuram a maioria dos professores (70,2%). Por outro lado, quando observado o gráfico (à direita) dessa figura percebe-se que 97% gostaria de compreender como essas estimativas são realizadas.

Para uma parte dos professores saber como se estima a proficiência, os parâmetros ou ainda, entender sobre a curva característica dos itens, talvez não tem tanta relevância na sua prática pedagógica - já que essas estimativas e representações gráficas são realizadas por programas computacionais-, mas destaca-se a importância dos professores, ao menos, compreenderem como interpretar os resultados dos estudantes a partir das suas proficiências, quando submetidos as avaliações externas de larga escala, pois é a partir dessas interpretações que se pode fazer intervenções, tais como propor ações com intuito de melhorar os resultados

da sua escola, por exemplo. O conhecimento acerca da interpretação dos resultados das avaliações com o uso da TRI é denotado na figura abaixo:

Figura 41 - Conhecimento do professor de matemática sobre a interpretação dos resultados das avaliações com uso da TRI.

11. Nas as avaliações externas (em larga escala) as proficiências (ou as proficiências médias) são apresentadas para que os resultados sejam interpretados não apenas de forma quantitativa, mas também de forma qualitativa pelos professores, secretarias de educação, entre outros. Assinale sobre o seu conhecimento acerca de como esses resultados podem ser interpretados.



Fonte: elaboração do autor

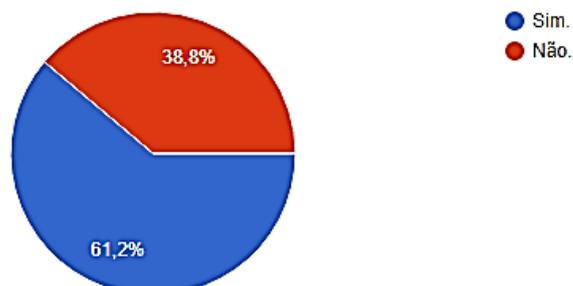
Nota-se que apenas 1,5% dos professores afirmou possuir alta (ou plena) capacidade de interpretar os resultados em uma avaliação externa de larga escala. Aqueles que responderam ter um conhecimento “médio” para essa questão figuram 37,3% dos professores, de modo que se pode inferir que o professor deste nível consegue interpretar de alguma maneira os resultados em uma avaliação de larga escala, mesmo que não plenamente. Já os que indicaram baixo ou nenhum conhecimento sobre a questão totalizam 61,2% dos professores. Isso mostra que mais da metade dos professores não conseguem ao menos interpretar os resultados das avaliações e isso constitui um problema grave na educação básica, já que essas interpretações poderiam servir de subsídio para que as escolas propusessem ações para a melhoria de sua realidade educacional.

Nesse questionário também foi avaliado o conhecimento do professor acerca da coerência nos padrões de respostas dadas a um participante de um teste, pois nas avaliações, quando se utiliza a TRI com modelos que contém o parâmetro de acerto casual, como no caso do Enem<sup>34</sup>, nem sempre quem acerta mais itens tem maior habilidade/proficiência. O resultado para essa questão é apresentado no gráfico da figura a seguir:

<sup>34</sup> Ver APÊNDICE A - Coerência pedagógica das respostas no Enem.

Figura 42 - Coerência pedagógica nas respostas dos participantes em um teste.

12. Suponha que em um teste utilizando a TRI (com o mesmo modelo do Enem) um respondente A, que acertou 15 questões, teve uma maior classificação que um sujeito B, que acertou 16 questões. Você saberia explicar essa situação, segundo essa teoria?



Fonte: elaboração do autor

Percebe-se que 61,2% dos professores afirmaram compreender bem a situação trazida na questão. Entretanto, alguns deles apresentaram uma justificativa que não é coerente com os conceitos da TRI, o que pode indicar uma contradição com a resposta dada. Desse modo, o percentual daqueles que não compreendem essa situação pode ser superior a essa marca. Seguem algumas dessas justificativas:

- a) “Porque seria analisado a idade, distorção idade série e o nível da questão (grau de dificuldade).”
- b) “Os erros foram mais grosseiros do que acertou mais.”
- c) “O aluno acertou uma média de questões de acordo ao nível de dificuldade.”
- d) “O que prevaleceu foi o peso de cada questão e não somente o quantitativo de acertos.”
- e) “O sujeito que acertou 15 questões acertou mais questões que poucos acertaram”
- f) “Dentro das minhas limitações, creio que o que acertou menos questões numericamente tenha acertado questões mais complexas ou, então, tenha errado menos questões de nível de complexidade mais baixo. A questão do chute também é levada em consideração, mas não consigo explicar como isso é calculado.”

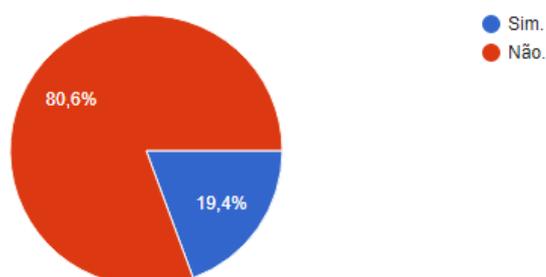
Por outro lado, dos professores que responderam “sim” e deram algum tipo de justificativa, tem-se que 44,4% parece ter algum entendimento ou conhece plenamente acerca dessa questão. Seguem algumas dessas justificativas:

- a) “O respondente A acertou questões de acordo com seu nível de aprendizado. Ele conseguiu acertar aquelas questões mais fáceis e errando as mais difíceis. No entanto o aprendente B fez exatamente o contrário.”
- b) “O (A) acertou os itens em sequência, fácil, médio e difícil. O (B) deve ter errado itens fáceis e acertado itens difíceis, o que indica que ele chutou o item difícil.”
- c) “O aluno que acertou 15 questões teve nota superior pois essas questões se enquadravam em níveis próximos [ou abaixo] de [sua] proficiência.”
- d) “As questões da prova foram graduadas em fáceis médias e difíceis, o candidato A foi mais coerente que o candidato B nesta graduação. Por exemplo o candidato B acertar uma questão difícil e errou uma fácil e por isso recebeu uma pontuação menor para o seu acerto.”
- e) “O que acertou menos foi mais coerente nos acertos e erros”

Outro tema abordado no questionário foi o da mudança de escala de habilidade, pois quando se utiliza alguns *softwares* computacionais, como o EIRT ou o *software* R a escala utilizada é a tradicional, com média 0 e desvio padrão 1, cuja notação é  $(0,1)$ . Porém, quando se quer apresentar os resultados em uma outra escala, é importante compreender como são transformadas as proficiências e os parâmetros dos itens para uma outra qualquer  $(m,d)$ , com intuito de apresentar esses resultados conforme planejado pelo construtor do teste. Sobre esse conhecimento seguem os resultados na Figura 43, a seguir:

Figura 43 - Percentual de professores em relação a sua compreensão da mudança de escala na TRI.

13. Os *softwares* computacionais utilizados na TRI normalmente realizam as estimativas na escala  $(0,1)$ . Assim, quando se quer apresentar os resultados em uma outra escala normalmente é necessário fazer uma transformação para a escala pretendida. Você saberia transformar a proficiência de um respondente a um teste da escala  $(0,1)$  para a escala do Enem  $(500,100)$ ?



Fonte: elaboração do autor

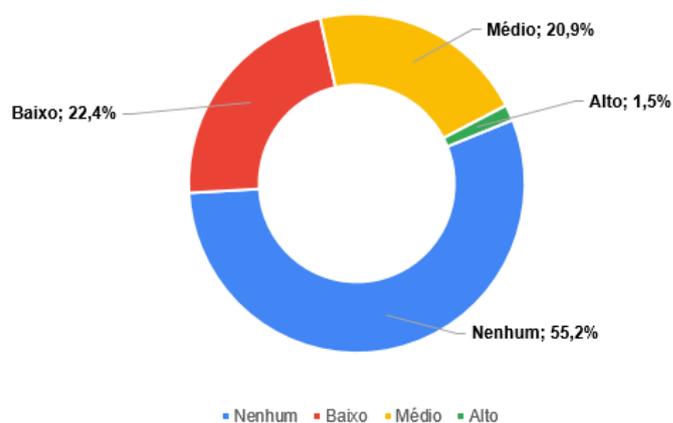
Observa-se que a maioria dos professores (80,6%) não dominam sobre essa questão, e isso impossibilitaria a interpretação de alguns textos relacionados a TRI. Ademais, a ausência desse entendimento dificultaria a interpretação – ou a construção- da Curva Característica do

Item e também a compreensão do valor numérico da proficiência para os diferentes testes que utilizam a TRI, já que a escolha da escala em cada avaliação é arbitrária e normalmente apresentam valores diferentes.

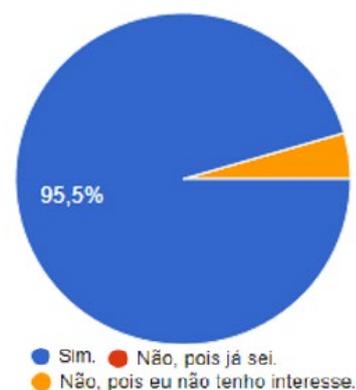
O questionário apresenta uma questão que versa sobre o conhecimento do professor em relação aos *softwares* computacionais utilizados para dos cálculos da TRI. Os dados obtidos podem ser vistos na Figura 44, a seguir:

Figura 44 - Conhecimento sobre os *softwares* utilizados na TRI

14. Os cálculos que são realizados com TRI, em geral, utilizam-se de *softwares* computacionais. Quanto ao seu conhecimento acerca desses softwares, assinale:



15. Você teria interesse em aprender como usar algum *software* utilizado para os cálculos da TRI, como na estimação das proficiências/habilidades e parâmetros dos itens em um teste?



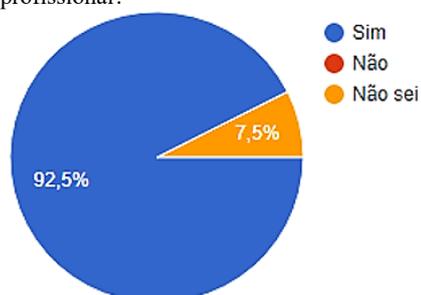
Fonte: elaboração do autor

Do gráfico acima (à esquerda) observa-se que mesmo que 77,6% dos professores tenham acenando ter baixo ou nenhum conhecimento sobre esses *softwares*. Por outro lado, destaca-se que a grande maioria (95,5%) tem interesse em aprender como utilizá-los para calcular as proficiências e parâmetros dos itens, como observado na figura acima, no gráfico da direita. Isso indica que se esses professores tiverem acesso a esse conhecimento, poderiam desenvolver habilidades necessárias para produzir os seus próprios testes utilizando a TRI.

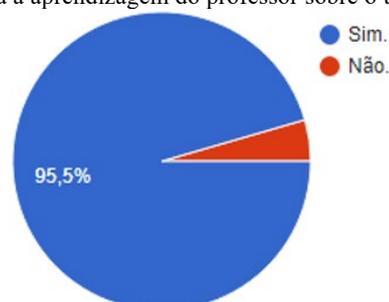
Por fim, nas questões 16 e 17 focaram em dois aspectos: se o entendimento do professor acerca dos conceitos fundamentais da TRI poderia contribuir para a melhoria da sua prática pedagógica e também se um livro, cartilha ou guia gratuito poderia ser útil para sua aprendizagem sobre esse tema. Os resultados estão apresentados a seguir.

Figura 45 - Percentual dos professores que responderam que entender sobre a TRI contribui para melhoria da sua prática profissional na TRI

16. Em sua opinião entender sobre os conceitos fundamentais da TRI irá contribuir para a melhoria de sua prática profissional?



17. Você acha que um livro gratuito, uma cartilha ou um guia, explicando de forma objetiva e didática sobre a TRI, facilitaria a aprendizagem do professor sobre o tema?



Fonte: elaboração do autor

Nota-se do gráfico da Figura 45 (à esquerda), que 92,5% dos professores acreditam que entender os conceitos fundamentais da TRI contribuirá para a melhoria da sua prática profissional. Em vista disso, 95,5% dos professores responderam que um material de apoio pedagógico, como um guia, cartilha ou livro gratuito, versando sobre a TRI, facilitaria a aprendizagem do professor e conseqüentemente poderia impactar positivamente para uma melhoria da realidade do seu contexto educacional.

Como visto acima, a partir das respostas dadas pelos professores ao questionário da pesquisa, pode-se inferir que mesmo que a grande parte dos professores não compreendam os temas fundamentais relacionados a Teoria de Resposta ao Item, eles têm um alto interesse em aprender sobre esses temas. Portanto, o resultado desse estudo poderá ser utilizado como base para discussões acerca desse tema, e ainda servir de referência para que se possa propor ações nas unidades de ensino, ou ainda políticas públicas realizadas pelas secretarias de educação no estado de Alagoas, objetivando a melhoria da educação.

## 6. CONCLUSÃO E TRABALHOS FUTUROS

A Teoria da Resposta ao Item (TRI) constitui uma nova abordagem para as análises dos testes na área educacional, pois permite resolver certos problemas de medição inatacáveis pela Teoria Clássica de Testes (TCT). Segundo Muñiz (2018) a TRI não contradiz os pressupostos fundamentais da Teoria Clássica dos Testes, mas traz novos elementos e suposições que implicou em uma grande reviravolta para a medição dos construtos em testes psicológicos ou educacionais.

Pasquali (1996, p. 194) diz que essa teoria psicométrica é

predominante no dito Primeiro Mundo de hoje. Embora ela seja teoricamente complexa e praticamente exigente em seus procedimentos analíticos, parece imprescindível que todos os que trabalham com testes psicológicos [ou educacionais] tenham conhecimento da mesma e dela façam uso na elaboração de seus instrumentos.

Porém, atualmente, com a disponibilidade de *softwares* computacionais para a TRI, sejam pagos ou gratuitos, os cálculos se tornaram facilitados e com isso poupando tempo para que as pessoas envolvidas nos processos de análise em uma avaliação se concentrem nas interpretações qualitativas dos resultados de um teste. Nesse sentido, foi mostrado que o *software* EIRT pode ser considerado como uma excelente ferramenta para ajudar os professores na compreensão dos resultados de um teste quando se utiliza a TRI, bem como pode auxiliá-lo na construção de seus próprios testes.

Para Couto e Primi (2011), a construção de testes com a utilização de teorias psicométricas - como a TRI - constitui uma tarefa necessária para se ter instrumentos de medidas de qualidade, e é desejável que os construtores e usuários tenham algum conhecimento sobre o funcionamento da TRI em testes, seja para utilizar na construção dos seus próprios testes ou pesquisas, ou ao menos compreender os conceitos quando são apresentados em trabalhos alheios.

Contudo, ainda há um grande desconhecimento acerca da TRI no Brasil e isso vem a complicar no que tange a construção e interpretação de instrumentos psicológicos, principalmente na área educacional (PASQUALI, 1996). Assim, mesmo que nos últimos anos as revistas especializadas estejam repletas de pesquisas sobre diversos aspectos da TRI muitos

professores sequer conhecem os temas afins a essa teoria, o que constitui um problema. Segundo Marconi e Lakatos (2002, p. 26) um “problema é uma dificuldade, teórica ou prática, no conhecimento de alguma coisa de real importância, para a qual se deve encontrar a solução”.

Nesse sentido, o estudo apresentado nesse trabalho revelou que os professores participantes, em sua maioria, possuem um baixo (ou nenhum) conhecimento sobre a Teoria de Resposta ao Item (74,6%), evidenciando um problema que precisa ser solucionado – seja de forma autônoma pelo professor ou ainda por meio formações continuadas promovidas por secretarias de educação, por exemplo, pois a compreensão acerca dessa teoria deveria ser um dos temas centrais em propostas de qualificação docente. Ademais, a compreensão pelos professores em relação as avaliações que utilizam a TRI, como o Enem e o Saeb, constitui um tema fundamental na educação básica, pois é a partir das interpretações dos resultados nessas avaliações que se pode propor ações para a melhoria da educação.

Nesse estudo, percebeu-se ainda que a grande maioria dos professores sequer estudaram esse tema na universidade/faculdade (95,5%), o que também constitui um aparente problema, pois essa realidade encontrada poderia ter se apresentado de maneira oposta caso a TRI fosse ofertada durante a graduação (ou pós-graduação) desses professores. Uma possível solução para esse problema seria a realização de formação continuada promovida pelas secretarias de educação como previsto na LDB, Art. 62, § 1º, que diz “A União, o Distrito Federal, os Estados e os Municípios, em regime de colaboração, deverão promover a formação inicial, a continuada e a capacitação dos profissionais de magistério” ou ainda em colaboração com as universidades, como observado no Art. 63 dessa lei, que diz que “os institutos superiores de educação manterão: III - programas de educação continuada para os profissionais de educação dos diversos níveis” (BRASIL, 1996).

De toda sorte, essa pesquisa pôde trazer à luz a informação de que a maioria dos professores pesquisados, afirmou ter interesse em estudar/entender sobre os conceitos fundamentais que permeiam essa teoria e que essa aprendizagem contribuirá na melhoria de sua prática profissional, mostrando-se ser um tema relevante para esses professores.

Espera-se que este trabalho contribua para auxiliar o professor da educação básica na compreensão acerca da TRI, seja para a interpretação dos resultados em uma avaliação ou ainda na utilização para construção dos seus próprios testes. Ademais, o resultado da pesquisa também poderá ser utilizado como base de dados para se propor ações relativas às formações continuadas sobre o tema para os professores de matemática das redes de ensino em Alagoas.

Como trabalhos futuros, propõe-se:

- Ampliar a população e amostra da pesquisa para os professores dos demais componentes curriculares, que compõem o ensino básico, sobre o seu conhecimento acerca da TRI.
- Ampliar o estudo sobre equalização utilizado em diversas avaliações.
- Estudar a utilização de outros *softwares* para realizar as estatísticas relativas à TRI.
- Construir um *e-book* sobre a TRI, a partir do referencial teórico deste trabalho, para servir como um guia de estudos para os professores que atuam na educação básica.

## REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, Dalton F.; DA PONTE, Tadeu. **Você nunca vai tirar zero na prova do Enem, saiba por quê.** [entrevista cedida a] Bruna Souza Cruz. **Uol**, São Paulo, 31 de out. 2014. Disponível em: <<https://educacao.uol.com.br/noticias/2014/10/31/voce-nunca-vai-tirar-zero-na-prova-do-enem-saiba-por-que.htm>>. Acesso em: 22 abr. 2021.

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

ANDRADE, D. F.; KARINO, C. A. **Nota Técnica: Teoria de Resposta ao Item.** Diretoria de Avaliação da Educação Básica (DAEB), Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), Ministério da Educação. Brasil, 2012. Disponível em: <[https://download.inep.gov.br/educacao\\_basica/enem/nota\\_tecnica/2011/nota\\_tecnica\\_tri\\_ene\\_m\\_18012012.pdf](https://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_tri_ene_m_18012012.pdf)>. Acesso em: 10 jun. 2021.

ARIAS, María R. M.; LLOREDA, María V. H.; LLOREDA, María J. H. **Psicometría.** Madrid: Alianza Editorial, S. A. Madrid, 2014. ISBN: 978-84-206-8859-6

BAKER, Frank B. **The Basics of Item Response Theory.** 2ed. ed. University of Wisconsin: ERIC Clearinghouse on Assessment and Evaluation, 2001.

BAKER, Frank B.; KIM, Seock-Ho. **Item Response Theory: Parameter Estimation Techniques.** 2. ed. rev. e expandida. New York: Marcel Dekker, 2004.

\_\_\_\_\_. **The Basics of Item Response Theory Using R.** New York, NY: Springer, 2017.

BARRETO, H. P. D.; MENDES, L. V. A.; OLIVEIRA, Z. M. F. **Banco de Itens: um recurso para a prática avaliativa do professor universitário.** In: RODRIGUES JÚNIOR, José F. (Org.). Avaliação do Estudante Universitário. 1. ed. Brasília: SENAC, 2009.

BEKMAN, Roberto M. **Aplicação dos blocos incompletos balanceados na teoria de resposta ao item.** Estudos em Avaliação Educacional, nº 24, jul-dez/2001. Disponível em: <<https://doi.org/10.18222/ea02420012202>>. Acesso em: 10 jun. 2021.

BIRNBAUM, A. **Some latent trait models and their use in inferring an examinee's ability.** In F. M. Lord & M. R. Novick, Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

BLASIS, Eloisa de; FALSARELLA, Ana M.; ALAVARSE, Ocimar M. **Avaliação e Aprendizagem: Avaliações externas: perspectivas para a ação pedagógica e a gestão do ensino**. São Paulo: CENPEC: Fundação Itaú Social, 2013.

BORTOLUCCI, R. S.; FINI, M. E.; TREVISAN, L. M. V.; AZEVEDO, T. C. A. M. **Relatório Pedagógico SARESP 2012: Matemática. Material de Apoio Pedagógico**. São Paulo, 2013.

BRASIL. Ministério da Educação. **Lei nº 9.394, de 20 de dezembro de 1996**. Estabelece as diretrizes e bases da educação nacional. Brasília, 1996. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/19394.htm](http://www.planalto.gov.br/ccivil_03/leis/19394.htm)>. Acesso em: 10 set. 2021.

\_\_\_\_\_. Ministério da Educação. PDE: **Plano de Desenvolvimento da Educação: Prova Brasil: ensino fundamental: matrizes de referência, tópicos e descritores**. Brasília, DF: Inep, 2008. Disponível em: <[http://portal.mec.gov.br/dmdocuments/prova%20brasil\\_matriz2.pdf](http://portal.mec.gov.br/dmdocuments/prova%20brasil_matriz2.pdf)>. Acesso em: 10 mai. 2021.

\_\_\_\_\_. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Matriz de Referência ENEM**. Brasília: INEP/MEC, 2009.

\_\_\_\_\_. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Guia de elaboração e revisão de itens**. v.1. Brasília: Ministério da Educação, 2010.

\_\_\_\_\_. **Entenda a sua nota no Enem – Guia do Participante**. Brasília: Ministério da Educação, 2012a. Disponível em: <[http://download.inep.gov.br/educacao\\_basica/enem/manuais/2012/guia-do-participante-Notas.pdf](http://download.inep.gov.br/educacao_basica/enem/manuais/2012/guia-do-participante-Notas.pdf)>. Acesso em: 30 mar. 2021.

\_\_\_\_\_. **Metodologia usada para correção define notas máxima e mínima**. Brasília: Ministério da Educação, 2012b. Disponível em: <<http://portal.mec.gov.br/ultimas-noticias/212-educacao-superior-1690610854/17389-metodologia-usada-para-correcao-define-notas-maxima-e-minima>>. Acesso em: 24 jun. 2021.

\_\_\_\_\_. **ENEM: Interpretação pedagógica das escalas de proficiência**. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2014.

\_\_\_\_\_. **Avaliação Nacional da Alfabetização: relatório 2013-2014: volume 1: da concepção à realização**. – Brasília, DF: Inep, 2015.

\_\_\_\_\_. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Inep 80 anos: 1937-2017**. Brasília: Viva Editora, 2018a.

\_\_\_\_\_. **Saeb 2017 (Press Kit)**. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2018b. Disponível em: <[https://download.inep.gov.br/educacao\\_basica/saeb/2018/documentos/presskit\\_saeb2017.pdf](https://download.inep.gov.br/educacao_basica/saeb/2018/documentos/presskit_saeb2017.pdf)>. Acesso em: 10 mai. 2021.

\_\_\_\_\_. **Relatório Saeb 2017 [recurso eletrônico]**. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2019.

CAEd/UFJF. **Guia de Elaboração de Itens – Matemática**. Universidade Federal de Juiz de Fora. Juiz de Fora: 2008.

CAMPOS, Fernanda C. dos S. **Elaboração da prova do Enade no modelo do banco nacional de itens**. Dissertação (Mestrado Profissional em Gestão e Avaliação da Educação Pública) – Universidade Federal de Juiz de Fora. Juiz de Fora, 2013.

CARLOS, Pablo R. de O. **Uma análise do desempenho dos estudantes no exame nacional do ensino médio e as contribuições para o ensino aprendizagem de física**. Dissertação (Mestrado Nacional Profissional em Ensino de Física Profissional) – Universidade Federal de Juiz de Fora. Juiz de Fora, 2016.

COHEN, Ronald Jay; SWERDLIK, Mark E.; STURMAN, Edward D. **Testagem e avaliação psicológicas: introdução a testes e medidas**. Tradução: Maria Cristina G. Monteiro. – 8. ed. – Porto Alegre: AMGH, 2014.

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

DOS ANJOS, Adilson; ANDRADE, D. F. **Teoria da Resposta ao Item com uso do R**. João Pessoa, 2012.

FLETCHER, Philip Ralph. **Da teoria clássica dos testes para os modelos de resposta ao item**. Rio de Janeiro: ENCE-IBGE, 2010.

HAMBLETON, Ronald K.; SWAMINATHAN, H.; ROGERS, D. Jane. **Fundamentals of Item Response Theory**. 1.ed. Newbury Park, CA: SAGE, 1991.

GERMAIN, Stephane; VALOIS, Pierre; ABDOUS, Belkacem. **Eirt - Item Response Theory Assistant for Excel**. 2007. Disponível em: <<https://libirt.psychometricon.net/eirt/eirt-en/index.html>>. Acesso em: 20 jun. 2021.

JÚDICE, Renato; KARINO, Camila. **História da TRI: entenda a régua de avaliação no Enem**. Educação & Avaliação. *E-book*. São Paulo: Geekie, 2015.

KARINO, C. A.; BARBOSA, M. T. S. **Nota Técnica: Procedimento de cálculo das notas do Enem**. Diretoria de Avaliação da Educação Básica (DAEB), Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), Ministério da Educação. Brasil, 2011. Disponível em: <[http://download.inep.gov.br/educacao\\_basica/enem/nota\\_tecnica/2011/nota\\_tecnica\\_procedimento\\_de\\_calculo\\_das\\_notas\\_enem\\_2.pdf](http://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_procedimento_de_calculo_das_notas_enem_2.pdf)>. Acesso em: 10 jun. 2021.

KLEIN, Ruben; FONTANIVE, Nilma S. **Avaliação em larga escala: uma proposta inovadora**. Em Aberto, n. 66, v. 15, p.28-34, Brasília: Inep/Mec, abr./jun. 1995.

KLEIN, Ruben. **Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB)**. Revista ENSAIO, n. 40, v.11, p.283-296, jul./set. 2003.

LORD, F. **A Theory of Test Scores**. Psychometric Monograph. n. 7. 1952.

LORD, F. M.; NOVICK, M. R. **Statistical theories of mental test scores**. Reading, MA: Addison-Wesley, 1968.

LUCKESI, Cipriano C. **Avaliação da aprendizagem escola: estudos e proposições**. 19. ed. São Paulo: Cortez, 2008.

LUECHT, Richard M; HAMBLETON, Ronald K. **ITEM RESPONSE THEORY: A Historical Perspective and Brief Introduction to Applications**. In: CLAUSER, Brian E; BUNC, Michael B. *The History of Educational Measurement: Key Advancements in Theory, Policy, and Practice*. 1. ed. New York: Routledge, 2021.

MARCONI, Marina de A. M; LAKATOS, Eva M. **Técnicas de pesquisa: planejamento e execução de pesquisas, elaboração, análises e interpretação de dados**. 5. ed. São Paulo: Atlas, 2002.

MOREIRA JUNIOR, F. J. **Aplicações da teoria da resposta ao item (TRI) no Brasil**. Revista Brasileira de Biometria, Jaboticabal, v.24, n.4, p.137-170, 2011.

MUÑIZ, José. **Introducción a la Psicometría - Teoría clásica y TRI**. 1. ed (edição eletrônica). Madrid: Ediciones Pirámide, 2018.

NAKANO, Tatiana de C.; PRIMI, Ricardo; NUNES, Carlos H. S. da S. **Análise de itens e teoria de resposta ao item (TRI)**. In: HUTZ, C. S.; BANDEIRA, D. R.; TRENTINI, C. (Org.). *Psicometria*. Porto Alegre: Artmed, 2015.

OECD. **Aprendendo para o Mundo de Amanhã: Primeiros resultados do PISA 2003**. São Paulo: Editora Moderna Ltda, 2005.

PASQUALI, Luiz. **A teoria de resposta ao item - TRI: Uma introdução**. In: PASQUALI, L. (Org.). *Teoria e métodos de medida em ciências do comportamento*. 1 ed. Brasília: MEC/SEDIAE-INEP, 1996, v. 1, p. 173-195.

\_\_\_\_\_. **Psicometria: teoria dos testes na psicologia e na educação**. 4. ed. Petrópolis, RJ: Vozes, 2011.

\_\_\_\_\_. **TRI - Teoria de Resposta ao Item: teoria, procedimentos e aplicações.** 1 ed. - Curitiba: Appris, 2018.

RABELO, Mauro. **Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro.** Coleção Profmat. Rio de Janeiro: SBM, 2013.

URBINA, S. **Fundamentos da testagem psicológica.** 1ª ed. Porto Alegre: Artmed, 2007.

VALENTINI, Felipe; LAROS, Jacob Arie. **Teoria de resposta ao item na avaliação psicológica.** In: AMBIEL, Rodolfo et al. **Avaliação Psicológica: guia de consulta para estudantes e profissionais de psicologia.** São Paulo: Casa do Psicólogo, 2011.

VALLE, R. da C. **Teoria de resposta ao item.** Estudos em Avaliação Educacional, São Paulo, n. 21, p. 7–92, 2000.

\_\_\_\_\_. **A construção e interpretação de escalas de conhecimento: considerações gerais e uma visão do que vem sendo feito no Saesp.** Estudos em Avaliação Educacional, São Paulo, n. 23, p. 71–92, 2001.

VALOIS, P. et al. **An open source tool to verify the psychometric properties of an evaluation instrument.** Procedia - Social and Behavioral Sciences 15, p. 552-556. Elsevier Ltd, 2011.

VIANNA, Heraldo M. **Avaliações Nacionais em Larga Escala: análise e propostas.** Estudos em Avaliação Educacional, n. 27, p. 41-76, jan.-jun./2003.

## APÊNDICE A - Coerência pedagógica das respostas no Enem

Em avaliações que utilizam a TRI, onde o acerto casual é considerado, a medida de proficiência/habilidade não é calculada observando somente o número de itens acertados, mas também é analisado o padrão de respostas do examinando após a realização do teste. Isso significa que “espera-se que participantes que acertaram as questões difíceis devam também acertar as questões fáceis, pois, entende-se que a aquisição do conhecimento ocorre de forma cumulativa, de modo que habilidades mais complexas requerem o domínio de habilidades mais simples” (BRASIL, 2012a, p. 17).

No entanto, quando se quer fazer essas análises entre o conhecimento que um examinando possui, a partir das suas respostas, é necessário posicionar cada item, que é associado a alguma habilidade, na escala de proficiência (ver Figura 46), conforme um processo de ancoragem - descrito na Seção 4.1-, cujo posicionamento do item nessa escala pode ser encontrado a partir dos seus parâmetros. Para tanto, o Enem estima conjuntamente esses parâmetros utilizando o método de Máxima Verossimilhança Marginal (MVM) a partir das respostas dos examinandos em um pré-teste. Daí, em um outro momento, os sujeitos são posicionados nessa mesma escala, de acordo com as suas respectivas proficiências, para que se possa realizar essas análises comparativas de forma quantitativa e qualitativa, já que os itens e os sujeitos estão na mesma métrica, conforme a sua posição/ancoragem e a proficiência, respectivamente.

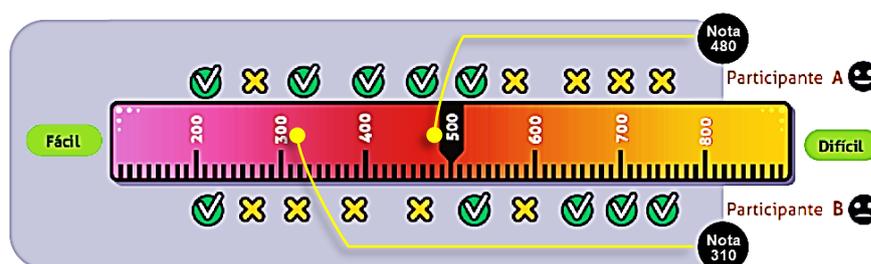
Figura 46 – Escala de habilidade de Matemática e suas tecnologias de níveis entre 742 e 767 para uma certa prova do Enem.

Questão 30	766,7	Determinar a variação percentual de perímetro de figura plana cujos vértices são centros de círculos tangentes quando se varia o raio de alguns desses círculos.
Questão 11	763,2	Resolver problema envolvendo cálculo de áreas de figuras planas, para determinação de custo de material.
Questão 44	757,1	Calcular operações com números racionais para enquadramento de Índice de Massa Corporal (IMC).
Questão 25	756,2	Determinar, usando operações com números naturais, a frequência anual de um evento excluindo um período do ano.
Questão 8	747,7	Avaliar as probabilidades de ocorrência de eventos realizados em duas etapas em um jogo.
Questão 17	745,8	Calcular a diferença entre as médias aritméticas de dois conjuntos de valores apresentados por meio de uma tabela.
Questão 5	742,9	Calcular medidas de tendência central a partir de dados contidos em um quadro simples.

Fonte: adaptado de BRASIL (2014).

Nesse sentido, a ilustração abaixo mostra a situação de dois participantes em uma prova contendo 10 questões que foram ancoradas em uma escala de proficiência. Nota-se que eles acertaram o mesmo número de questões, mas o participante A teve uma nota/proficiência bem maior que o participante B, devido a coerência no padrão das suas respostas. A saber: o participante A acertou as questões mais fáceis e errou as mais difíceis; já o participante B acertou as mais difíceis e errou as mais fáceis.

Figura 47 – Comparação do padrão de respostas e notas/proficiência de dois examinandos na escala do Enem - (500,100).

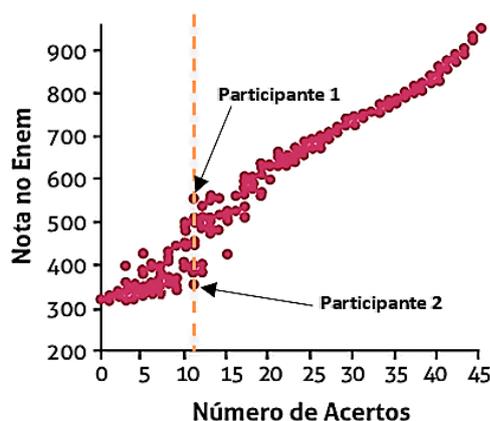


Fonte: adaptado de BRASIL (2012a).

Desse modo, a partir do seu padrão de respostas é esperado que o participante acerte as questões que estão abaixo de seu nível de proficiência, como acontece com o participante A. Então, entende-se que as questões mais difíceis provavelmente foram acertadas “no chute” pelo participante B. Entretanto, isso não significa que a nota/proficiência deste participante irá diminuir, mas ela não terá tanto valor como esperado. A saber, questões sem marcação (em “branco”) são consideradas como erradas nesse exame, então é preferível marcar todas as questões (BRASIL, 2012a).

Uma análise congênere em relação ao padrão de respostas dos examinandos na prova de Matemática e suas tecnologias do Enem 2011 pode ser vista na Figura 48 a seguir, cujo gráfico relaciona a nota e o número de acertos do participante.

Figura 48 - Relação entre o número de acertos e a nota no Enem 2011 em Matemática e suas tecnologias.



Fonte: adaptado de BRASIL, 2012a.

Nesse gráfico acima pode-se notar, por exemplo, que cada participante que acertou 11 questões teve a sua nota/proficiência variando entre 350 e 550 (aproximadamente), e isso é justificado pela coerência no padrão das suas respostas. Nesse sentido, quando se estima as proficiências dos examinados, utilizando a TRI, dificilmente terá como resultado uma proficiência igual a outra, e essa é uma característica destacada para testes que buscam classificar os examinandos, pois evita um grande número de empates no resultado final.

No Enem a prova é composta por 45 questões para cada área do conhecimento – contendo questões fáceis, intermediárias e difíceis e que são selecionadas a partir do Banco Nacional de Itens (BNI)<sup>35</sup>-, mas as notas em áreas distintas não são comparáveis diretamente, pois mesmo a escala sendo a mesma, cuja média é 500 e o desvio-padrão é 100, as questões foram calibradas separadamente para cada uma das quatro áreas de conhecimento e dessa forma impossibilitando a comparação entre as diferentes áreas. Então, não é correto inferir que acertar mais questões em uma área do conhecimento implicará em uma proficiência mais alta que em uma outra, pois, segundo Karino e Barbosa (2011), seria equivalente a comparar uma pessoa que consegue correr 100 metros a outra que consegue nadar 100 metros.

Por outro lado, uma vantagem da TRI aplicada a esse exame é que possibilita realizar comparações entre as notas/proficiência entre as diferentes provas na mesma área. Essas comparações em uma mesma área podem ser realizadas, pois as diferentes provas foram

<sup>35</sup> Mais detalhes ver a Seção 4.2 - Banco de itens para avaliações educacionais.

construídas a partir de questões de um banco de itens comum, que foram calibrados na mesma escala utilizando a TRI e uma única matriz de competências/habilidades.

Além disso, nesse exame, as proficiências dos respondentes a um teste, que são estimadas na TRI pelo método *Expected a Posteriori* (EAP), pode teoricamente assumir qualquer valor real, mas, conforme Nakano, Primi e Nunes (2015, p. 102) “comumente oscila entre -4 e +4 (ainda que valores mais altos ou mais baixos possam ser encontrados)”. Então, se for levado essa informação para a escala utilizada no Enem – (500,100), na prática, a grande maioria das proficiências (notas) dos estudantes figura, aproximadamente, entre 100 e 900 (que corresponde a 4 desvios abaixo e acima da média 500 na escala de proficiência).

Destaca-se também que “as proficiências na TRI são estimadas em uma escala métrica que não possui mínimo e máximo pré-estabelecidos. Esses valores variam de acordo com as características dos itens que compõem a prova de cada edição do Exame” (KARINO; BARBOSA, 2011, n.p); desse modo, como as questões das provas em cada edição não são as mesmas e possuem graus de dificuldades diferentes, normalmente são observados valores distintos para essas notas mínimas e máximas, como pode ser observado na tabela abaixo. Assim, não faz sentido dizer que a nota mínima em cada área de conhecimento é 0 e a máxima é 1000, como alguns podem pensar<sup>36</sup>. A saber: a maior nota em matemática no Enem foi 1008,3 (em 2015) e a menor foi 277,2 (em 2012).

Seguem abaixo, na Tabela 12, as notas máximas e mínimas nas edições do Enem de 2009 a 2020 em Matemática e suas Tecnologias.

Tabela 12 - Notas mínimas e máximas em Matemática e suas Tecnologias no Enem de 2009 a 2020.

	Nota por ano											
	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Mínimo	296,0	313,4	321,6	277,2	322,4	318,5	280,2	309,7	310,4	360,0	359,0	327,1
Máximo	985,1	973,2	953,0	955,2	971,5	973,6	1008,3	991,5	993,9	996,1	985,5	975,0

Fonte: Inep/Mec.

<sup>36</sup> No Enem apenas a nota de redação varia de 0 a 1000, mas que não é calculada pela TRI. Para mais detalhes ver: <[https://download.inep.gov.br/publicacoes/institucionais/avaliacoes\\_e\\_exames\\_da\\_educacao\\_basica/a\\_redacao\\_do\\_enem\\_2020\\_-\\_cartilha\\_do\\_participante.pdf](https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_exames_da_educacao_basica/a_redacao_do_enem_2020_-_cartilha_do_participante.pdf)>. Acesso em: 25 jun. 2021.

A nota do Enem é estimada pela TRI em uma faixa/intervalo de proficiência de  $[0, +\infty)$ , conforme os documentos oficiais que subsidiam esse exame<sup>37</sup>, mas na prática “não há um zero absoluto, ainda que os candidatos tenham entregue a prova completamente em branco” (BRASIL, 2012b), pois segundo Karino e Barbosa (2011, n.p) “uma pessoa que erra todas as questões recebe o valor mínimo do teste, e não uma nota zero, pois não se pode afirmar, a partir do teste, que ela possui zero conhecimento”.

Nesse sentido, o professor de estatística da UFSC, Dalton Francisco de Andrade, que fez parte da equipe técnica na construção do “Guia do Participante - Entenda a sua nota no Enem do Enem”<sup>38</sup>, em entrevista para o site Uol, em 31/10/2014, explica que se em uma prova “a questão mais fácil [...] [tiver], por exemplo, o valor de 300 e a mais difícil 800. Quem fizer o exame e errar tudo vai ganhar 300 pontos, pois este era o valor mínimo do exame. Tudo depende dos itens inseridos dentro dessa escala” (ANDRADE; DA PONTE, 2014). Da mesma forma pode-se inferir que nessa prova o candidato que acertar todas as questões teria uma nota 800.

Essa característica do Enem, descrita acima, pode ser observada comparando as informações da Tabela 12 e da Tabela 13 a seguir, onde pode-se notar que no Enem 2015 um total de 5.773 participantes tiraram as notas no intervalo  $[0, 100)$  (ver Tabela 13), mas, mesmo assim, esses participantes obtiveram a nota mínima de 208,2 nesse exame, como apresentado na Tabela 12. Essa característica é exclusiva desse exame e não pode ser empregada em todas as avaliações que utilizam a TRI.

Tabela 13 – Quantidade de participantes, no Enem 2015, em Matemática e suas Tecnologias, por faixa de proficiência.

Faixa de Proficiência	Participantes do ENEM
$[0, 100)$	5.773
$[100, 200)$	-
$[200, 300)$	40
$[300, 400)$	1.661.196
$[400, 500)$	2.237.043
$[500, 600)$	1.029.088
$[600, 700)$	464.520

<sup>37</sup> Microdados do Enem disponível em: <<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>>. Acesso em: 25 jun. 2021.

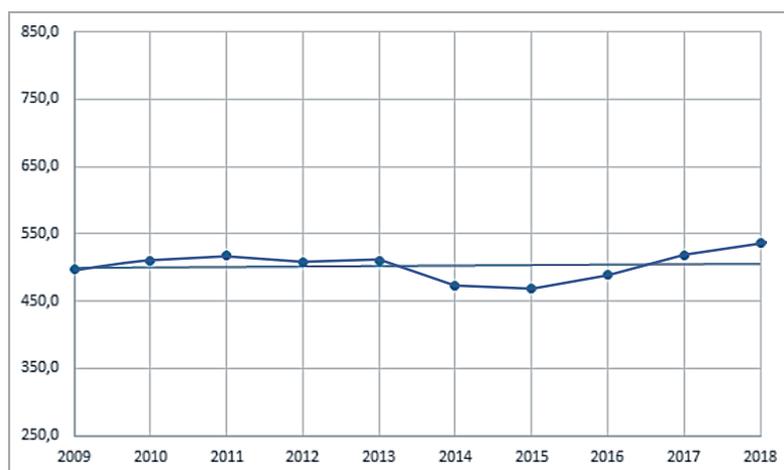
<sup>38</sup> BRASIL, 2012a.

Faixa de Proficiência	Participantes do ENEM
[700, 800)	180.620
[800, 900)	42.129
[900, 1.000)	5.040
[1.000, +∞)	91

Fonte: microdados do Enem disponível em:  
<[https://download.inep.gov.br/microdados/microdados\\_enem2015.zip](https://download.inep.gov.br/microdados/microdados_enem2015.zip)>. Acesso em: 25 jun. 2021.

Outra análise característica da TRI, que é observada nesse exame, é a comparação das proficiências médias dos estudantes em uma série temporal, que serve de parâmetro para avaliar como o resultado médio dos estudantes está evoluindo ao longo dos anos, e pode ser visto a seguir:

Figura 49 – Proficiências médias dos estudantes em Matemática e suas tecnologias (2009-2018).



Fonte: adaptado de Inep/Mec<sup>39</sup>.

Portanto, como observado nesta seção, entender como a TRI é utilizada no Enem, seja na construção das escalas, na estimação dos parâmetros dos itens em pré-testes para compor o BNI, ou ainda como é interpretada a relação entre o padrão de respostas e a proficiência dos estudantes, constitui uma importante etapa para a consolidação dos conceitos fundamentais dessa teoria aplicada nesse exame.

<sup>39</sup> Disponível em:

<[https://download.inep.gov.br/educacao\\_basica/enem/downloads/2018/presskit/presskit\\_enem-resultados2018.pdf](https://download.inep.gov.br/educacao_basica/enem/downloads/2018/presskit/presskit_enem-resultados2018.pdf)>. Acesso em: 20 jun. 2021.

## APÊNDICE B – Instalação do software EIRT

O programa EIRT, que é um suplemento (*add-in*) do Excel, está disponível para *download* em <https://libirt.psychometricon.net/releases/>, cujo arquivo de instalação é denominado EIRT-2.0.0.exe, em que 2.0.0 é o número da versão<sup>40</sup>.

O programa de instalação inicialmente pede para selecionar o idioma. Para o resto da instalação, você pode manter o valor padrão. Perto do final da instalação, uma macro do Excel é executada para instalar o programa no menu do Excel, e daí é necessário aceitar a ativação da macro para que a instalação seja concluída.

As informações completas sobre o *software* EIRT pode ser acessada em: <https://libirt.psychometricon.net/eirt/eirt-en/index.html>.

---

<sup>40</sup> As versões mais antigas do *software* EIRT estão disponíveis em: [<https://libirt.psychometricon.net/releases/old/>](https://libirt.psychometricon.net/releases/old/).

**APÊNDICE C – Matriz das Respostas dos 100 sujeitos aplicadas ao teste hipotético.**

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
<b>GABARITO</b>	<b>A</b>	<b>B</b>	<b>D</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>D</b>	<b>B</b>	<b>D</b>	<b>D</b>	<b>A</b>	<b>B</b>	<b>D</b>	<b>D</b>
SUJEITO 1	A	D	D	C	A	A	D	B	D	B	A	B	D	D	A	B	A	B	C	B
SUJEITO 2	D	B	A	C	A	A	C	D	A	B	A	C	D	B	B	A	A	B	D	D
SUJEITO 3	A	B	D	C	A	A	B	C	D	D	A	A	D	B	D	D	A	B	D	A
SUJEITO 4	D	C	D	C	A	C	A	B	A	D	C	C	C	C	A	D	C	B	C	A
SUJEITO 5	D	B	D	C	A	B	C	D	B	A	A	B	D	B	D	C	C	D	B	D
SUJEITO 6	D	D	B	C	A	A	C	D	B	A	A	A	C	B	B	B	C	B	D	D
SUJEITO 7	D	B	A	D	B	A	C	D	B	A	A	C	C	B	B	C	A	B	D	D
SUJEITO 8	D	B	D	D	A	D	D	C	D	C	D	C	D	D	D	A	A	B	B	D
SUJEITO 9	A	D	C	C	A	B	C	D	A	B	C	A	D	B	A	B	A	B	D	A
SUJEITO 10	D	D	D	B	A	C	B	C	C	A	A	A	D	B	B	D	A	B	D	D
SUJEITO 11	A	D	D	D	A	A	B	D	B	C	A	A	C	B	A	B	C	B	B	D
SUJEITO 12	B	B	A	A	A	A	B	C	A	D	A	A	D	B	D	A	C	D	B	D
SUJEITO 13	D	A	B	C	A	A	D	D	A	C	A	C	D	B	B	A	C	B	D	D
SUJEITO 14	A	D	A	C	A	A	D	C	B	A	C	A	D	C	C	A	A	B	B	D
SUJEITO 15	A	D	C	D	A	B	B	D	A	D	C	A	C	B	A	A	C	B	C	A
SUJEITO 16	D	B	C	B	A	D	D	B	A	C	D	B	B	C	A	A	A	B	A	D
SUJEITO 17	A	B	B	C	A	A	B	D	D	C	A	C	D	B	D	D	A	B	D	D
SUJEITO 18	D	B	D	C	A	A	A	C	B	A	D	A	D	B	C	D	A	B	D	D
SUJEITO 19	D	B	D	C	A	A	B	A	D	C	A	A	D	B	D	A	A	B	D	D
SUJEITO 20	A	B	D	C	A	A	C	D	B	D	D	B	D	D	D	D	A	B	D	D
SUJEITO 21	D	D	A	C	A	D	D	B	A	C	C	A	D	B	A	C	A	B	B	D
SUJEITO 22	D	B	D	C	A	A	B	A	A	A	D	A	D	D	A	C	A	B	D	D
SUJEITO 23	D	B	D	C	A	C	B	D	D	B	D	A	D	D	A	A	C	B	A	D
SUJEITO 24	D	B	D	C	A	A	D	C	A	D	C	A	C	C	A	A	A	B	D	D
SUJEITO 25	D	D	A	C	A	A	B	D	C	D	C	C	D	C	A	A	A	D	B	B
SUJEITO 26	C	B	D	C	A	B	B	C	D	A	A	C	D	B	C	A	A	B	B	D
SUJEITO 27	A	D	A	D	A	A	B	C	D	A	A	C	D	B	A	A	A	B	A	D
SUJEITO 28	A	B	D	C	A	D	D	D	D	C	A	A	D	B	D	D	A	B	D	D
SUJEITO 29	D	D	B	C	A	D	D	D	A	B	C	A	C	D	A	A	A	C	A	D
SUJEITO 30	A	B	A	B	A	A	B	D	C	C	A	A	D	B	D	D	A	B	D	D
SUJEITO 31	D	B	A	C	A	A	B	C	A	D	A	C	D	B	B	A	A	B	B	D
SUJEITO 32	D	B	D	A	A	B	B	A	D	C	A	A	D	B	D	A	A	B	D	D
SUJEITO 33	D	B	A	A	A	C	A	C	B	A	D	A	C	A	D	A	C	B	B	A
SUJEITO 34	D	D	A	C	A	B	B	D	B	C	C	D	C	C	A	A	A	D	A	D
SUJEITO 35	A	B	A	C	A	A	D	D	D	C	D	C	D	B	D	D	A	B	D	D
SUJEITO 36	D	D	D	C	A	D	D	C	B	C	D	A	D	B	D	B	A	B	D	C
SUJEITO 37	A	B	B	C	A	A	B	C	A	C	D	A	C	D	B	B	A	B	B	D

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
SUJEITO 38	A	B	D	A	A	C	B	A	A	C	A	C	D	B	D	C	A	A	A	D
SUJEITO 39	A	B	D	C	A	A	C	C	D	C	A	A	D	B	A	B	A	B	D	D
SUJEITO 40	D	B	A	C	A	A	D	C	D	B	A	B	C	B	A	A	C	C	B	D
SUJEITO 41	D	B	B	C	A	A	B	C	D	C	A	A	D	B	A	D	A	B	D	D
SUJEITO 42	A	B	C	D	A	B	C	D	C	D	D	B	D	B	C	A	C	B	B	D
SUJEITO 43	D	B	D	C	A	A	B	C	D	A	D	A	B	B	B	B	A	B	D	D
SUJEITO 44	A	B	D	C	A	A	D	A	A	B	A	A	D	B	B	A	A	B	B	D
SUJEITO 45	A	B	A	C	A	C	C	C	B	C	A	C	D	B	D	C	A	B	D	D
SUJEITO 46	A	B	B	C	A	C	B	D	D	C	D	B	D	B	A	A	C	B	D	D
SUJEITO 47	A	C	D	C	A	B	C	A	D	C	C	C	D	B	B	B	A	D	C	B
SUJEITO 48	D	D	B	C	A	D	B	C	B	C	A	B	C	B	B	D	A	B	A	D
SUJEITO 49	D	D	D	C	A	A	B	D	D	D	A	C	D	B	D	D	A	B	D	D
SUJEITO 50	A	B	A	B	A	A	C	C	D	C	D	A	D	A	A	A	A	B	D	D
SUJEITO 51	D	B	D	C	C	A	B	C	D	C	A	A	D	B	A	D	A	B	A	D
SUJEITO 52	D	D	D	C	A	D	C	B	C	D	B	A	C	A	A	A	A	D	A	D
SUJEITO 53	A	B	D	C	A	A	D	D	A	C	A	B	D	B	A	A	C	B	D	D
SUJEITO 54	D	D	C	B	A	D	A	C	C	A	A	B	D	D	A	D	A	A	B	A
SUJEITO 55	A	D	A	C	A	B	B	C	D	D	C	D	B	A	D	A	A	B	D	D
SUJEITO 56	D	B	C	B	A	A	B	C	A	C	A	A	D	B	B	A	A	B	D	D
SUJEITO 57	D	D	A	D	C	B	B	D	D	D	A	B	D	C	A	D	C	A	B	D
SUJEITO 58	A	B	D	B	A	B	B	D	D	C	D	C	C	C	D	A	C	D	C	B
SUJEITO 59	A	D	A	C	A	A	B	C	A	C	D	C	D	D	A	A	A	D	B	D
SUJEITO 60	A	B	A	D	A	A	C	C	D	A	A	A	D	B	C	A	A	B	X	A
SUJEITO 61	A	B	D	C	A	A	C	C	D	A	A	A	D	B	A	A	A	B	D	D
SUJEITO 62	D	D	A	C	A	A	B	C	B	A	D	A	D	D	D	D	A	B	D	D
SUJEITO 63	D	D	A	D	A	B	C	D	B	D	B	A	B	D	B	C	A	B	B	D
SUJEITO 64	D	D	D	C	A	A	B	C	B	D	D	B	C	B	D	B	A	B	D	A
SUJEITO 65	A	B	D	C	A	A	B	C	A	C	A	B	D	B	D	C	C	B	B	D
SUJEITO 66	B	A	C	A	A	B	C	C	D	C	A	C	B	D	A	A	C	D	D	D
SUJEITO 67	A	B	A	C	A	B	D	C	C	A	D	B	D	D	B	D	C	B	D	D
SUJEITO 68	A	D	D	A	A	D	D	C	C	X	B	D	D	D	D	B	B	B	D	A
SUJEITO 69	A	D	A	B	A	C	B	C	C	C	D	A	D	B	D	A	A	B	D	D
SUJEITO 70	A	D	A	C	A	B	D	D	A	B	B	A	B	D	C	A	C	B	D	D
SUJEITO 71	A	B	A	C	A	A	D	A	D	A	D	A	D	B	D	C	A	B	D	D
SUJEITO 72	A	D	A	C	A	A	B	D	D	D	A	A	D	B	D	A	C	B	D	D
SUJEITO 73	A	B	B	C	A	A	D	A	C	A	A	A	D	B	D	D	A	B	D	A
SUJEITO 74	A	B	A	C	A	A	B	C	D	C	A	A	D	B	A	C	A	B	D	D
SUJEITO 75	A	B	B	B	A	A	D	B	B	C	A	B	X	D	A	B	C	B	D	B
SUJEITO 76	A	B	D	A	A	A	B	D	D	C	A	A	D	D	D	D	A	B	D	D
SUJEITO 77	D	C	D	C	A	A	C	A	D	C	C	A	B	D	C	A	C	B	C	D

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
SUJEITO 78	A	B	B	B	A	B	A	A	C	C	C	A	D	B	B	A	A	B	B	D
SUJEITO 79	A	B	D	C	A	A	B	C	A	C	A	A	D	D	A	A	A	B	B	D
SUJEITO 80	A	B	D	C	A	A	B	A	D	C	A	A	D	B	D	A	A	B	D	D
SUJEITO 81	A	B	D	C	A	D	D	A	B	D	C	A	C	D	A	C	A	B	A	D
SUJEITO 82	B	B	D	C	A	A	C	C	A	C	B	B	D	B	C	D	A	B	D	A
SUJEITO 83	D	D	A	B	A	A	D	D	A	D	D	A	D	B	A	A	D	B	D	A
SUJEITO 84	A	X	D	B	A	C	B	C	D	A	A	A	D	A	A	A	A	B	A	D
SUJEITO 85	D	C	D	C	A	A	B	D	D	C	A	B	C	B	D	C	A	B	D	D
SUJEITO 86	A	D	B	D	A	B	C	D	A	D	C	B	C	B	B	D	C	C	A	D
SUJEITO 87	B	A	A	D	B	B	B	D	A	B	C	A	B	B	B	B	A	B	B	C
SUJEITO 88	D	D	A	D	C	A	D	D	A	D	B	B	B	B	D	A	A	B	A	D
SUJEITO 89	D	D	B	D	B	C	B	C	D	A	A	A	C	B	D	A	C	B	B	D
SUJEITO 90	A	D	A	D	A	A	B	A	A	D	A	A	D	B	B	D	C	B	B	D
SUJEITO 91	A	D	B	B	A	B	D	A	B	A	A	B	C	B	A	C	C	B	D	C
SUJEITO 92	D	D	A	B	D	B	B	D	B	C	C	A	D	B	C	D	A	B	D	D
SUJEITO 93	C	D	A	C	A	A	B	C	B	D	A	C	D	B	D	C	A	B	D	D
SUJEITO 94	D	D	B	C	A	D	D	D	B	C	D	A	D	D	C	B	C	B	B	D
SUJEITO 95	A	B	C	B	D	A	B	C	D	C	C	C	C	D	B	D	A	B	B	A
SUJEITO 96	A	D	D	A	A	D	D	B	A	C	A	B	C	C	C	A	A	B	C	B
SUJEITO 97	D	D	A	B	C	A	B	A	A	C	C	D	C	C	A	D	A	C	D	D
SUJEITO 98	A	B	D	C	A	A	B	C	D	A	A	A	D	B	A	D	A	B	B	D
SUJEITO 99	A	B	D	C	A	A	B	C	D	C	A	A	D	B	D	D	A	B	D	D
SUJEITO 100	D	C	B	D	B	B	C	D	A	B	C	B	C	A	B	B	C	D	A	A

Nota - Sinal convencional utilizado:

X Questão sem marcação (em branco).

**APÊNDICE D – Matriz dicotômicas contendo as respostas dos 100 sujeitos ao teste.**

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
SUJEITO 1	1	0	1	1	1	1	0	0	1	0	1	0	1	0	0	0	1	1	0	0
SUJEITO 2	0	1	0	1	1	1	0	0	0	0	1	0	1	1	0	0	1	1	1	1
SUJEITO 3	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0
SUJEITO 4	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
SUJEITO 5	0	1	1	1	1	0	0	0	0	0	1	0	1	1	1	0	0	0	0	1
SUJEITO 6	0	0	0	1	1	1	0	0	0	0	1	1	0	1	0	0	0	1	1	1
SUJEITO 7	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1
SUJEITO 8	0	1	1	0	1	0	0	1	1	1	0	0	1	0	1	0	1	1	0	1
SUJEITO 9	1	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0
SUJEITO 10	0	0	1	0	1	0	1	1	0	0	1	1	1	1	0	1	1	1	1	1
SUJEITO 11	1	0	1	0	1	1	1	0	0	1	1	1	0	1	0	0	0	1	0	1
SUJEITO 12	0	1	0	0	1	1	1	1	0	0	1	1	1	1	1	0	0	0	0	1
SUJEITO 13	0	0	0	1	1	1	0	0	0	1	1	0	1	1	0	0	0	1	1	1
SUJEITO 14	1	0	0	1	1	1	0	1	0	0	0	1	1	0	0	0	1	1	0	1
SUJEITO 15	1	0	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0
SUJEITO 16	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1
SUJEITO 17	1	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1
SUJEITO 18	0	1	1	1	1	1	0	1	0	0	0	1	1	1	0	1	1	1	1	1
SUJEITO 19	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1
SUJEITO 20	1	1	1	1	1	1	0	0	0	0	0	0	1	0	1	1	1	1	1	1
SUJEITO 21	0	0	0	1	1	0	0	0	0	1	0	1	1	1	0	0	1	1	0	1
SUJEITO 22	0	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	1	1	1	1
SUJEITO 23	0	1	1	1	1	0	1	0	1	0	0	1	1	0	0	0	0	1	0	1
SUJEITO 24	0	1	1	1	1	1	0	1	0	0	0	1	0	0	0	0	1	1	1	1
SUJEITO 25	0	0	0	1	1	1	1	0	0	0	0	0	1	0	0	0	1	0	0	0
SUJEITO 26	0	1	1	1	1	0	1	1	1	0	1	0	1	1	0	0	1	1	0	1
SUJEITO 27	1	0	0	0	1	1	1	1	1	0	1	0	1	1	0	0	1	1	0	1
SUJEITO 28	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
SUJEITO 29	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1
SUJEITO 30	1	1	0	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
SUJEITO 31	0	1	0	1	1	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1
SUJEITO 32	0	1	1	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1
SUJEITO 33	0	1	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0
SUJEITO 34	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1
SUJEITO 35	1	1	0	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1
SUJEITO 36	0	0	1	1	1	0	0	1	0	1	0	1	1	1	1	0	1	1	1	0
SUJEITO 37	1	1	0	1	1	1	1	1	0	1	0	1	0	0	0	0	1	1	0	1
SUJEITO 38	1	1	1	0	1	0	1	0	0	1	1	0	1	1	1	0	1	0	0	1
SUJEITO 39	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
SUJEITO 40	0	1	0	1	1	1	0	1	1	0	1	0	0	1	0	0	0	0	0	1
SUJEITO 41	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
SUJEITO 42	1	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1
SUJEITO 43	0	1	1	1	1	1	1	1	1	0	0	1	0	1	0	0	1	1	1	1
SUJEITO 44	1	1	1	1	1	1	0	0	0	0	1	1	1	1	0	0	1	1	0	1
SUJEITO 45	1	1	0	1	1	0	0	1	0	1	1	0	1	1	1	0	1	1	1	1
SUJEITO 46	1	1	0	1	1	0	1	0	1	1	0	0	1	1	0	0	0	1	1	1
SUJEITO 47	1	0	1	1	1	0	0	0	1	1	0	0	1	1	0	0	1	0	0	0
SUJEITO 48	0	0	0	1	1	0	1	1	0	1	1	0	0	1	0	1	1	1	0	1
SUJEITO 49	0	0	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1
SUJEITO 50	1	1	0	0	1	1	0	1	1	1	0	1	1	0	0	0	1	1	1	1
SUJEITO 51	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1
SUJEITO 52	0	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1
SUJEITO 53	1	1	1	1	1	1	0	0	0	1	1	0	1	1	0	0	0	1	1	1
SUJEITO 54	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	1	1	0	0	0
SUJEITO 55	1	0	0	1	1	0	1	1	1	0	0	0	0	0	1	0	1	1	1	1
SUJEITO 56	0	1	0	0	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1
SUJEITO 57	0	0	0	0	0	0	1	0	1	0	1	0	1	0	0	1	0	0	0	1
SUJEITO 58	1	1	1	0	1	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0
SUJEITO 59	1	0	0	1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	0	1
SUJEITO 60	1	1	0	0	1	1	0	1	1	0	1	1	1	1	0	0	1	1	0	0
SUJEITO 61	1	1	1	1	1	1	0	1	1	0	1	1	1	1	0	0	1	1	1	1
SUJEITO 62	0	0	0	1	1	1	1	1	0	0	0	1	1	0	1	1	1	1	1	1
SUJEITO 63	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1
SUJEITO 64	0	0	1	1	1	1	1	1	0	0	0	0	0	1	1	0	1	1	1	0
SUJEITO 65	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1	0	1
SUJEITO 66	0	0	0	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	1	1
SUJEITO 67	1	1	0	1	1	0	0	1	0	0	0	0	1	0	0	1	0	1	1	1
SUJEITO 68	1	0	1	0	1	0	0	1	0	0	0	0	1	0	1	0	0	1	1	0
SUJEITO 69	1	0	0	0	1	0	1	1	0	1	0	1	1	1	1	0	1	1	1	1
SUJEITO 70	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1
SUJEITO 71	1	1	0	1	1	1	0	0	1	0	0	1	1	1	1	0	1	1	1	1
SUJEITO 72	1	0	0	1	1	1	1	0	1	0	1	1	1	1	1	0	0	1	1	1
SUJEITO 73	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	0
SUJEITO 74	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1
SUJEITO 75	1	1	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	1	1	0
SUJEITO 76	1	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1
SUJEITO 77	0	0	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0	1	0	1
SUJEITO 78	1	1	0	0	1	0	0	0	0	1	0	1	1	1	0	0	1	1	0	1
SUJEITO 79	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	1	0	1
SUJEITO 80	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1
SUJEITO 81	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1



## **APÊNDICE E – Questionário: O conhecimento do professor de matemática acerca da Teoria de Resposta ao Item**

Este questionário faz parte de um estudo que tem por objetivo colher as impressões do(a)s professore(a)s do ensino básico, que lecionam matemática em Alagoas, acerca de seu conhecimento sobre a Teoria de Resposta ao Item (TRI), que é amplamente utilizada diversos tipos de avaliações. Em algumas questões haverá espaço para que você apresente uma justificativa, caso deseje.

As informações fornecidas terão sua privacidade garantida pelos pesquisadores responsáveis. Os participantes da pesquisa não serão identificados em nenhum momento, mesmo quando os resultados desta pesquisa forem divulgados em qualquer forma.

### **1ª parte: Informações gerais**

#### **1. Qual o seu e-mail?**

#### **2. Qual o seu sexo?**

- a) Masculino.
- b) Feminino.
- c) Outro.

#### **3. Você leciona matemática na educação básica?**

- a) Sim.
- b) Não.

#### **4. Qual rede de ensino você leciona?**

(você poderá marcar mais de uma opção)

- Municipal.
- Estadual.
- Federal.
- Privada.

**5. Qual(is) etapa(s) do ensino básico você leciona?**

- a) Ensino Fundamental – Anos iniciais (1º ao 5º ano).
- b) Ensino Fundamental - Anos finais (6º ao 9º ano).
- c) Ensino Médio.
- d) Ensino Fundamental/Médio.

**6. Qual o tempo, em anos completos, que você leciona na educação básica?**

- a) de 1 a 5.
- b) de 6 a 10.
- c) de 11 a 15.
- d) de 16 a 20.
- e) acima de 20 anos.

**7. Qual sua formação acadêmica?**

- a) Magistério.
- b) Licenciatura em matemática.
- c) Bacharelado em matemática.
- d) Licenciatura em física.
- e) Pedagogia

Outro:
--------

**8. Qual tipo de pós-graduação você possui? Assinale o nível mais alto.**

- a) Não possuo.
- a) Especialização (lato sensu).
- b) Mestrado.
- c) Doutorado.

d) Outro:
-----------

**2ª parte: Conhecimentos sobre a Teoria de Resposta ao Item (TRI)****1. Você conhece a Teoria de Resposta ao Item?**

- a) Sim.
- b) Conheço superficialmente.
- c) Não conheço.

Justificativa:

**2. Você já estudou sobre a Teoria de Resposta ao Item (TRI)?**

- a) Sim, durante a graduação na universidade/faculdade.
- b) Sim, numa pós-graduação, especialização ou aperfeiçoamento.
- c) Sim, de forma independente e autônoma.
- d) Não estudei.

**3. Antes as avaliações externas em larga escala no Brasil mensuravam o desempenho do aluno a partir do número de questões acertadas (essa perspectiva está relacionada a Teoria Clássica dos Testes-TCT). Atualmente esse desempenho é medido a partir das técnicas da TRI (Teoria de Resposta ao Item). Na sua opinião qual é a melhor forma para obter esse desempenho?**

- a) Anterior a TRI.
- b) Utilizando a TRI.
- c) outra (cite-a no campo de justificativa)

Justificativa:

**4. A expressão abaixo representa a probabilidade de um respondente acertar um item em função da sua habilidade/proficiência e dos parâmetros do item em um teste. Essa função é muito utilizada em avaliações que envolvem a TRI. Você conhece essa função?**

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

- a) Sim.  
b) Não.

**5. A função apresentada na questão anterior possui parâmetros e uma variável. Qual o seu conhecimento acerca dos significados desses parâmetros e dessa variável?**

Nenhum	Baixo	Médio	Alto
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Justificativa:

**6. Você tem interesse em entender como é calculada, na TRI, a probabilidade de um respondente acertar um item em um teste (a partir da função apresentada na questão 4), bem como saber o que significa cada parâmetro e variável apresentado na função?**

- a) Sim  
b) Não, pois eu já sei.  
c) Não, pois eu não tenho interesse.

**7. Nos testes que utilizam a TRI cada item (questão) possui um gráfico que descreve a probabilidade de acerto em função da habilidade do respondente, chamado de Curva Característica do Item. Sobre o seu conhecimento acerca desse gráfico, assinale:**

Nenhum	Baixo	Médio	Alto
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**8. Qual interesse você tem em entender como se interpreta ou constrói o gráfico da Curva Característica ao Item da TRI?**

Nenhum	Baixo	Médio	Alto
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**9. Na TRI uma das etapas mais importantes é a estimação dos parâmetros de cada item (discriminação, dificuldade e acerto ao acaso) e também a estimação das proficiências (às vezes chamada de habilidade ou simplesmente nota) dos respondentes a um teste. Sobre o seu conhecimento acerca de como é realizada essas estimativas, assinale:**

Nenhum	Baixo	Médio	Alto
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**10. Você gostaria de compreender como são realizadas as estimativas destacadas na questão anterior?**

- a) Sim.
- b) Não, pois eu já sei.
- c) Não, eu não tenho interesse.

Justificativa:

**11. Nas as avaliações externas (em larga escala) as proficiências (ou as proficiências médias) são apresentadas para que os resultados sejam interpretados, não apenas de forma quantitativa, mas também de forma qualitativa pelos professores, secretarias de educação, entre outros. Assinale sobre o seu conhecimento acerca de como esses resultados podem ser interpretados.**

Nenhum	Baixo	Médio	Alto
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**12. Suponha que em um teste utilizando a TRI (com o mesmo modelo do Enem) um respondente A, que acertou 15 questões, teve uma maior classificação que um sujeito B, que acertou 16 questões. Você saberia explicar essa situação, segundo essa teoria?**

- a) Sim.
- b) Não.

Justificativa:

**13. Os softwares computacionais utilizados na TRI normalmente realizam as estimativas na escala (0,1). Assim, quando se quer apresentar os resultados em uma outra escala normalmente é necessário fazer uma transformação para a escala pretendida. Você saberia transformar a proficiência de um respondente a um teste da escala (0,1) para a escala do Enem (500,100)?**

- a) Sim.
- b) Não.

**14. Os cálculos que são realizados com TRI, em geral, utilizam-se de softwares computacionais. Quanto ao seu conhecimento acerca desses softwares, assinale:**

Nenhum                      Baixo                      Médio                      Alto

**15. Você teria interesse em aprender como usar algum software utilizado para os cálculos da TRI, como na estimação das proficiências/habilidades e parâmetros dos itens em um teste?**

- a) Sim
- b) Não, pois eu já sei.
- c) Não, pois eu não tenho interesse.

**16. Em sua opinião entender sobre os conceitos fundamentais da TRI irá contribuir para a melhoria de sua prática profissional?**

- a) Sim.
- b) Não.
- c) Não sei.

**17. Você acha que um livro gratuito, uma cartilha ou um guia, explicando de forma objetiva e didática sobre a TRI, facilitaria a aprendizagem do professor sobre o tema?**

- a) Sim.
- b) Não.