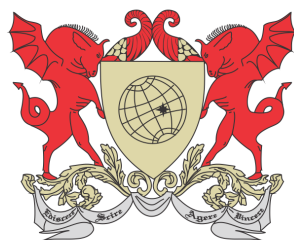


UNIVERSIDADE FEDERAL DE VIÇOSA  
DISSERTAÇÃO DE MESTRADO



VANESSA DE FREITAS SANTOS

UMA INTRODUÇÃO À TEORIA DE RESPOSTA  
AO ITEM E SUA UTILIZAÇÃO NO ENEM

FLORESTAL  
MINAS GERAIS – BRASIL  
2018

VANESSA DE FREITAS SANTOS

**UMA INTRODUÇÃO À TEORIA DE RESPOSTA AO  
ITEM E SUA UTILIZAÇÃO NO ENEM**

Dissertação apresentada à Universidade Federal de Viçosa,  
como parte das exigências do Programa de Pós-Graduação  
Mestrado Profissional em Matemática em Rede Nacional,  
para obter o título *Magister Scientiae*.

FLORESTAL  
MINAS GERAIS – BRASIL  
2018

**Ficha catalográfica preparada pela Biblioteca da Universidade Federal de Viçosa - Câmpus Florestal**

T

Santos, Vanessa de Freitas, 1987-  
S237i Uma introdução à teoria de resposta ao item e sua  
2018 aplicação no Enem / Vanessa de Freitas Santos. – Florestal, MG,  
2018.  
vi, 29f. : il. (algumas color.) ; 29 cm.

Inclui anexos.

Orientador: Danielle Franco Nicolau Lara.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Teoria de Resposta ao Item. 2. TRI. 3. ENEM.  
4. Teoria Clássica dos Testes. 5. TCT. I. Universidade Federal de Viçosa. Instituto de Ciências Exatas e Tecnológicas. Mestrado em Matemática - Profissional. II. Título.

153.94

VANESSA DE FREITAS SANTOS

**UMA INTRODUÇÃO À TEORIA DE RESPOSTA AO  
ITEM E SUA UTILIZAÇÃO NO ENEM**

Dissertação apresentada à Universidade Federal de Viçosa,  
como parte das exigências do Programa de Pós-Graduação  
Mestrado Profissional em Matemática em Rede Nacional,  
para obter o título *Magister Scientiae*.

APROVADA: 28 de fevereiro de 2018.

---

Profa. Camila Ferreira de Souza

---

Prof. Mehran Sabetti

---

Profa. Danielle Franco Nicolau Lara  
(Orientadora)

# Agradecimentos

---

Agradeço à Deus, pelas inúmeras e incontáveis oportunidades de ser cada dia melhor. Aos meus pais, por me ensinarem que com bondade, compaixão e perseverança tudo se ajeita. À Ritinha, por fazer de mim quem sou. À Márcia e Ivany, pelo carinho e exemplo. Aos meus avós, tios e primos, por serem meu porto-seguro. Às minhas amigas, por me fazerem tão bem. Ao Dalton, Adriano e Cácio, pela paciência, aprendizado e estima. Aos queridos do Bernoulli, pela oportunidade e confiança. À Daniele e Mehran por acreditarem e não desistirem de mim.

# Resumo

---

SANTOS, Vanessa de Freitas, M.Sc., Universidade Federal de Viçosa, fevereiro de 2018. **Uma introdução à Teoria de Resposta ao Item e sua utilização no ENEM.** Orientadora: Danielle Franco Nicolau Lara. Coorientador: Mehran Sabetti.

O presente trabalho visa introduzir e refletir acerca dos conceitos fundamentais da Teoria de Resposta ao Item (TRI), sendo esses embasados pela explicitação das concepções e das definições pertencentes à Psicometria e à Teoria Clássica dos Testes (TCT). Apresentados seus principais fundamentos, este estudo objetiva difundir informações relativas ao Exame Nacional do Ensino Médio (Enem) e a utilização da TRI como método de correção de sua avaliação objetiva.

# Abstract

---

SANTOS, Vanessa de Freitas, M.Sc., Universidade Federal de Viçosa, February, 2018.  
**An Introduction to Item Response Theory and its Use in Enem.** Adviser:  
Danielle Franco Nicolau Lara. Co-adviser: Mehran Sabetti.

The present work aims to introduce and reflect on the fundamental concepts of Item Response Theory (IRT), which are based on the explication of the conceptions and definitions belonging to Psychometry and Classical Theory of Tests (CTT). The main objective of this study was to disseminate information about the “Exame Nacional do Ensino Médio” (Enem) and the use of IRT as a method to correct its objective evaluation.

# Lista de Figuras

---

3.1 Esquema ilustrativo do modelo da Teoria Clássica dos Testes . . . . .	5
4.1 Esquema ilustrativo do modelo da Teoria Clássica dos Testes . . . . .	7
4.2 Exemplo da Curva Características de um Teste com 05 itens distintos . . .	8
4.3 Exemplo de uma CCI com parâmetros identificados . . . . .	9
4.4 Exemplo de Curvas características de itens $i$ 's com diferentes acertos casuais, no qual $c_1 > c_2 > c_3$ . . . . .	9
4.5 Exemplo de Curvas características de itens $i$ 's com diferentes dificuldades, no qual $b_3 > b_2 > b_1$ . . . . .	10
4.6 Exemplo de Curvas características de itens $i$ 's com diferentes discriminações, no qual $a_1 > a_2 > a_3$ . . . . .	10
5.1 Fonte: Thiago Ferreira Miranda, Projeto de Iniciação científica: “Estimação de parâmetros: técnicas de calibração e equalização dos itens do ENEM”, UFPA, Agosto/2016 . . . . .	17
5.2 Fonte: BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Guia do Participante – Entenda a sua nota no Enem. Brasília, DF, 2012. Pag.08 (adaptada) . . . . .	18
5.3 Exemplo de uma situação em que a nota TRI é calculada com o mesmo número de acertos em uma prova. . . . .	20



# Sumário

---

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Psicometria</b>	<b>2</b>
2.1	Breve História . . . . .	2
2.2	Medidas Psicométricas . . . . .	3
<b>3</b>	<b>Teoria Clássica dos Testes</b>	<b>4</b>
<b>4</b>	<b>Teoria de Resposta ao Item</b>	<b>6</b>
4.1	Breve História . . . . .	6
4.2	Fundamentos . . . . .	6
4.2.1	Curva Característica de um item . . . . .	7
4.2.2	Curva Característica de um Teste . . . . .	7
4.2.3	Parâmetros de um item . . . . .	8
4.2.4	Estimação e Invariância dos parâmetros . . . . .	10
4.3	Modelos TRI . . . . .	13
4.3.1	Modelo Logístico de um Parâmetro . . . . .	13
4.3.2	Modelo Logístico de dois Parâmetros . . . . .	13
4.3.3	Modelo Logístico de três Parâmetros . . . . .	14
<b>5</b>	<b>Exame Nacional do Ensino Médio</b>	<b>15</b>
5.1	Introdução . . . . .	15
5.2	Teoria de Resposta ao Item No ENEM . . . . .	16
5.3	Escalas ENEM . . . . .	18
5.4	Coerência das Notas . . . . .	19
<b>6</b>	<b>Atividades de Aplicação da TRI no ambiente escolar</b>	<b>21</b>
<b>7</b>	<b>Conclusão</b>	<b>24</b>
<b>A</b>	<b>Anexo Um: Cartilha</b>	<b>25</b>
<b>B</b>	<b>Anexo Dois: Simulados na escala do ENEM</b>	<b>28</b>
	<b>Bibliografia</b>	<b>29</b>

# Introdução

---

A importância do medir é incontestável. Uma afirmação sem dados é apenas uma consolidação empírica e, muitas vezes errônea. As medidas se fazem presentes em todas as ciências. Um médico possui informações sobre a saúde e desenvolvimento de um bebê por meio de suas medidas, antes mesmo de nascer. Um engenheiro civil solicita a análise laboratorial do solo em que fará sua obra para que calcule a profundidade da fundação que servirá de alicerce de sustentação para toda a estrutura que irá construir. Se souber exatamente quantas pessoas irão à sua festa, poderá planejar melhor a quantidade de comida a ser servida. Se não souber, ela poderá não ser suficiente para todos.

É possível medir facilmente a altura ou peso de alguém, basta fazer uso de algum instrumento que já carregue a métrica desejada para determinar tais medidas, como uma fita métrica graduada ou uma balança em quilogramas. Medições desse tipo são realizadas de forma *direta*. Isso significa que, apenas observando um determinado evento, é possível medi-lo (fenômeno observável). Porém, nos estudos psicométricos, muitas vezes se fez necessária uma medida de avaliação *indireta*, ou seja, baseada não no evento propriamente dito, mas sim em sua manifestação (fenômeno não observável). Como exemplo podemos citar o potencial empreendedor de um sujeito, seu nível de ansiedade ou, até mesmo, seu nível de sabedoria em determinada área do conhecimento. Aos fenômenos não observáveis damos o nome de *construtos*, *traços latentes* ou *habilidade*. Mensurar habilidades é um dos principais focos de estudo da Psicometria. Com o tempo, medir os traços latentes se tornou essencial não só na Psicometria, mas em muitas outras ciências, análises ou estudos. Surge, então, a necessidade de criar Teorias capazes de fornecer e calcular tais medidas. Nesse trabalho, citaremos duas delas: uma amplamente usada, porém com limitações, mas não menos importante, chamada *Teoria Clássica de Testes*; e outra que será o foco central do vigente trabalho, chamada *Teoria de Resposta ao Item*.

# Psicometria

---

Epistemologicamente, a *Psicometria* representa a teoria e a técnica de medida dos processos mentais, é moldada por um conjunto de técnicas para medir, quantificar ou mensurar fenômenos psíquicos. Ela assume o modelo quantitativista da *Psicologia*<sup>1</sup> especialmente aplicado às áreas da própria Psicologia e da Educação.

De acordo com *Pasquali* (2003) a *Psicometria fundamenta-se na teoria da medida em ciências para explicar o sentido que têm as respostas dadas pelos sujeitos a uma série de tarefas e propor técnicas de medida dos processos mentais*. Embora muitos definam a *Psicometria* como um ramo da estatística - talvez por seus precursores serem, em sua maioria, estatísticos de formação - ela deve ser concebida como um ramo da Psicologia que interage com a Estatística.

## 2.1 Breve História

Historicamente a *Psicometria* tem suas origens na *Psicofísica*<sup>2</sup> de Ernst Heinrich Weber (1795-1878) e de Gustav Fechner (1801-1887), psicólogos alemães interessados na experiência subjetiva. Ainda no século XIX se iniciou o desenvolvimento de instrumentos de avaliação psicológica, uma vez que a ciência positivista avançava e obter medidas objetivas para o desenvolvimento de estudos e análises clínicas se tornavam uma necessidade. Nesse sentido, era preciso criar métodos que avaliassem as propriedades psicométricas das avaliações. O inglês *Francis Galton* (1822-1911) foi o primeiro a criar testes para medir processos mentais e, por isso é considerado por muitos o criador da *Psicometria*. Ele acreditava que aplicar um exame padronizado em um grande número de indivíduos poderia revelar as diferenças mentais existentes entre eles. O estudo psicométrico dos testes passou por diferentes etapas, tendo início em torno da década de 1880 com Galton, e divagando pelas décadas de 1890 com *James McKeen Cattell* (1860-1944), que desenvolveu medidas de diferenças individuais e inaugurou a terminologia *mental test* (teste mental); 1900 com *Alfred Binet* (1857-1911) - com interesse na avaliação das aptidões humanas visando principalmente a predição na área acadêmica e área da saúde - e *Charles Spearman* (1863-1945), que

---

<sup>1</sup>Ciência relacionada aos estados e processos mentais do ser humano e de suas interações com um ambiente físico e social.

<sup>2</sup>Medida de processos diretamente observáveis, do estímulo e resposta do organismo.

lançou os fundamentos da teoria de Psicometria clássica em suas obras; e 1910 na era dos testes de inteligência. No entanto, foi na década de 1930 que *Leon Louis Thurstone* (1887-1955) deu real vida à Psicometria, diferenciando-a da psicofísica e definindo-a como a medida do comportamento do organismo por meio de processos mentais. Thurstone, além de desenvolver a análise fatorial múltipla, fundou em 1936 a Sociedade Psicométrica Americana, juntamente com a revista *Psychometrika*, ambas dedicadas ao estudo e progresso da Psicometria.

A Psicometria atual possui duas vertentes: a chamada clássica, com Teoria Clássica dos Testes (TCT) e a chamada moderna com a Teoria de Resposta ao Item (TRI). A TRI possui como essência o estudo individual dos itens componentes de um teste ao contrário da sua antecedente, a TCT, na qual o teste por completo é o mais importante.

## 2.2 Medidas Psicométricas

Qual a definição de medida? Apesar de não ser consenso entre os pesquisadores, uma das definições mais aceitas foi dada por *Stanley Smith Stevens* (1906-1973) em 1946. Segundo ele, medir consiste em *assinalar números a objetos e eventos de acordo com alguma regra*.

A medida em Psicologia, se insere dentro da *Teoria de Medida*<sup>3</sup> em geral. Essa última, por sua vez, desenvolve uma discussão epistemológica em torno da utilização do “número” no estudo científico dos fenômenos naturais, como na Psicometria. Portanto, cria-se uma conexão entre sistemas teóricos de diferentes compreensões, tendo a teoria da medida a função de justificar e explicar o sentido que essa conexão possui.

---

<sup>3</sup>Mais informações podem ser encontradas em PASQUALI, Luis. *Psicometria: Teoria dos testes na Psicologia e na Educação*. Vozes, 2011.

## Teoria Clássica dos Testes

---

Na Teoria Clássica de testes (TCT) as habilidades são, em geral, medidas através de uma soma de respostas dadas a uma série de itens. A cada questão respondida, o sujeito recebe uma pontuação e seu escore final total ( $T$ ) é dado pela soma desses pontos, ou seja, o  $T$  em um teste com 10 itens é a soma dos itens corretamente acertados, por exemplo. Se for dado 1 ponto para cada item respondido de maneira correta e 0 pontos para o incorreto e o sujeito acertou 6 itens e errou 4, seu escore  $T$  seria 6.

Como nessa teoria o escore gerado é bruto (escore total menos os erros), ele representa uma medida. A habilidade medida é obtida em uma escala específica, exclusiva de sua medição. Portanto, a habilidade está sempre associada à escala utilizada.

De acordo com Pasquali (2003), uma grande preocupação da Psicometria era o erro contido nessa soma já que se trata de uma análise empírica. Nessa Teoria, a dificuldade de um item depende da porcentagem de acertos nesse item. Quanto mais próximo de 100%, mais fácil ele é.

O modelo da TCT foi elaborado por Spearman e detalhado por *Harold Gulliksen* (1903-1996):

$$T = V + E$$

No qual:

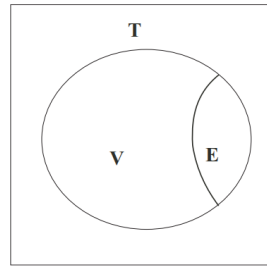
$T$  = escore bruto ou empírico do sujeito (soma dos pontos obtidos no teste)

$V$  = escore verdadeiro (amplitude real daquilo que se quer medir no sujeito)

$E$  = erro cometido na medida

Observe que o escore verdadeiro seria o próprio escore bruto se não houvesse erro na medida. Além disso, o escore bruto é a soma do escore verdadeiro com o erro e, conseqüentemente,  $E = T - V$ , bem como  $V = T - E$ . Dessa forma, o escore bruto de um sujeito difere do seu escore verdadeiro e essa diferença é devida ao erro, ou melhor, essa diferença é o próprio conceito de erro, como ilustrado na Figura 3.1.

Fica, agora, óbvio entender que o grande desafio da TCT consiste em elaborar



**Figura 3.1:** Esquema ilustrativo do modelo da Teoria Clássica dos Testes

estratégias (estatísticas) para controlar ou avaliar a magnitude do erro. Os erros são devidos a inúmeros fatores inusitados, tais como defeitos do próprio teste, estereótipos e vieses do sujeito, fatores históricos e ambientais aleatórios

A prática da Teoria Clássica, apesar de muito importante em alguns aspectos, cria problemas que não podem ser ignorados. Um deles está ligado ao fato do teste depender dos itens que o compõem. Assim, a dificuldade da prova depende dos respondentes, e não da qualidade particular dos itens. Além disso, como o traço latente medido está relacionado à escala utilizada, os resultados de testes diferentes não são comparáveis. Surge, então a necessidade de criar uma Teoria que pudesse responder pelas limitações da TCT.

# Teoria de Resposta ao Item

---

A Teoria de resposta ao item (TRI), também chamada de Psicometria Moderna, surge para preencher as limitações da Teoria Clássica, melhorando a qualidade da avaliação, principalmente por consequência de uma mudança no foco de estudo. Antes estudava-se o teste como um todo, o que na TRI recai sobre cada um de seus itens. Agora o objetivo é identificar as características intrínsecas a um item. Dessa forma, as conclusões não dependem apenas do teste ou questionário, mas de cada item que o compõe.

## 4.1 Breve História

A TRI tem suas origens no trabalho de Richardson (1936), Lawley (1943) e Tucker (1946). Porém Lord (1952,1953) foi o pioneiro no desenvolvimento de um modelo capaz de medir um traço latente modelado por uma distribuição normal acumulada. Seu estudo teve uma rápida expansão nos anos 60 com o desenvolvimento de novas tecnologias e com o trabalho de Rasch (1960).

No Brasil, a teoria foi usada pela primeira vez em 1995 na análise de dados dos SAEB (Sistema Nacional de Ensino Básico), que permitiu a comparação e criação de uma escala única para os desempenhos dos alunos do 5º e 9º anos do Ensino Fundamental e 3ª série do Ensino Médio. A partir dos resultados obtidos no SAEB, outras avaliações de larga escala como o SARESP (Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo) e mais atualmente, em 2009, o ENEM (Exame Nacional do Ensino Médio) passaram a fazer uso dessa teoria.

## 4.2 Fundamentos

A TRI é um conjunto de modelos matemáticos que buscam representar a probabilidade para um sujeito dar uma resposta certa a um item específico  $P(\theta)$  em função dos parâmetros do item e do nível da habilidade  $\theta$  do respondente. Nela, quanto maior a habilidade, maior a probabilidade de acerto no item. A partir de um conjunto de respostas de um grupo de respondentes a um conjunto de itens, a TRI possibilita a estimação dos parâmetros dos itens e dos sujeitos em uma escala de medida. Os itens e os sujeitos estão na mesma escala, assim o nível de uma característica exigido pelo item pode ser comparado ao nível da característica que o

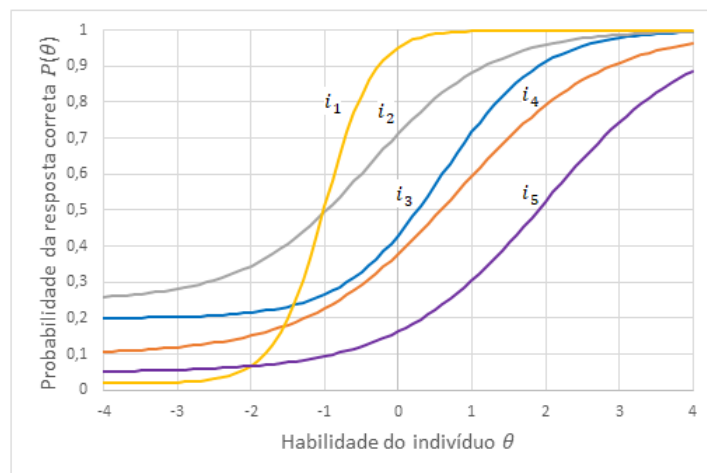
indivíduo possui, o que facilita a interpretação da escala gerada.

### 4.2.1 Curva Característica de um item

Como já mencionado anteriormente, na TRI a probabilidade de acerto de um indivíduo em um determinado item  $i$ , denominada  $P_i(\theta)$ , depende exclusivamente da habilidade  $\theta$  desse indivíduo e das especificidades do item. Sabemos, também, que quanto maior o  $\theta$ , maior deverá ser  $p(\theta)$ , este último variando de 0 a 1. Porém, por maior que seja o  $\theta$ , não se pode garantir que  $P(\theta)$  será 1. Seguindo o mesmo raciocínio, também não se pode garantir que por menor que seja o  $\theta$ ,  $P(\theta)$  não será 0.

Todas essas características fizeram com o que empiricamente modelassem a função  $P(\theta)$  como monotônica crescente, não-linear e assintótica nas extremidades (já que não poderia alcançar 0 ou 1). Por isso, geralmente são utilizadas as funções logarítmicas e a de distribuição normal acumulada.

Cada item possui particularidades inerentes a ele, por isso, cada um possui uma curva específica, conhecida como curva característica do item (CCI). Portanto, curvas itens diferentes são diferentes (Figura 4.1).



**Figura 4.1:** Esquema ilustrativo do modelo da Teoria Clássica dos Testes

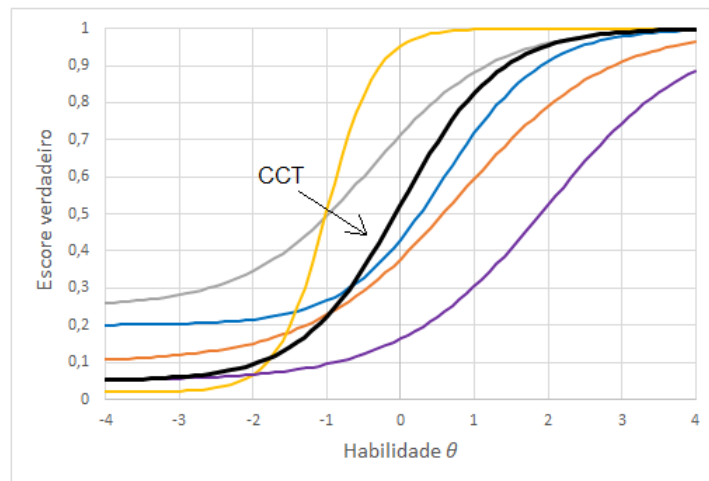
A CCI vincula as diferentes probabilidades de acerto de sujeitos distintos às diferentes habilidades  $\theta$  que eles manifestam. O valor de  $\theta$  pode variar de  $-\infty$  a  $+\infty$ . É importante ressaltar sua escala é arbitrária, na qual o que importa é apenas a relação de ordem resistente entre seus pontos, e não necessariamente sua magnitude (Andrade, Tavares e Valle; 2000).

### 4.2.2 Curva Característica de um Teste

Quando somadas todas as CCI's dos itens que compõem um teste, obtém-se a chamada *Curva Característica de um Teste* (CCT), para cada valor de  $\theta$  somam-se os valores de  $P(\theta)$  de cada item nesse nível de habilidade.

A CCT apresenta a probabilidade de desempenho de um sujeito (score verdadeiro) dado um valor de  $\theta$ . Ela pode ser encontrada através da equação:





**Figura 4.2:** Exemplo da Curva Características de um Teste com 05 itens distintos

$$CCT = \sum_{i=1}^n P_i(\theta) \quad (4.1)$$

No qual  $n$  representa o número de itens que constituem o teste.

A importância do escore verdadeiro estimado por meio da TRI para a interpretação de resultados é que ele está na mesma escala que as pontuações empíricas, as mesmas advindas da TCT. Como já sabemos que os valores de  $\theta$  são medidos em uma escala diferente, o que as CCT's realmente mostram é uma relação entre o escore verdadeiro e a escala de habilidade.

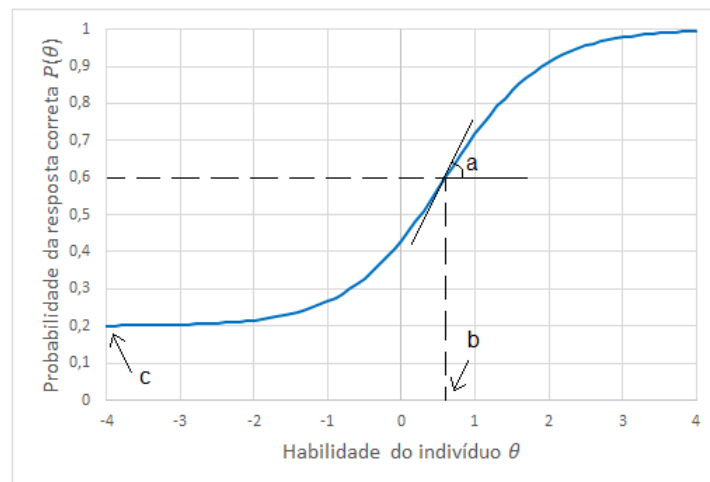
A justificativa do maior uso da habilidade  $\theta$  do que o valor do escore verdadeiro reside justamente em uma das limitações que a TCT nos impõem: o fato de que  $\theta$  nos permite uma independência do teste utilizado. Em contrapartida, o escore verdadeiro sempre dependerá do teste aplicado.

### 4.2.3 Parâmetros de um item

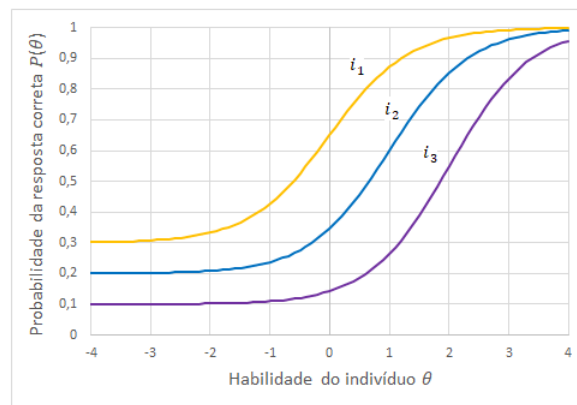
Às particularidades psicométricas de um item damos o nome de *parâmetros*. São eles a capacidade de *discriminação* de um item, sua *dificuldade* e a probabilidade de *acerto casual*. Os parâmetros independem do traço latente que está sendo mensurado, já que são intrínsecos aos itens e podem ser identificados na CCI (Figura 4.3).

O acerto casual (representado por  $c$ ) de um item indica a probabilidade de um indivíduo de baixa ou baixíssima habilidade responder corretamente a um item, ou seja, ele indica através de uma escolha arbitrária das alternativas (“chute”), a probabilidade de acerto naquele item. Seu valor pode ser encontrado ao se considerar  $P(\theta) \neq 0$  e  $\theta = -\infty$ .

Já a dificuldade  $b$  de um item, também conhecida como *posição* do item, representa a habilidade necessária para que um indivíduo obtenha a probabilidade de acerto do item igual a  $\frac{1+c}{2}$ . Ela é medida na mesma escala da habilidade do indivíduo, peculiaridade essa que gera uma vantagem da TRI sobre a Psicometria clássica, já que



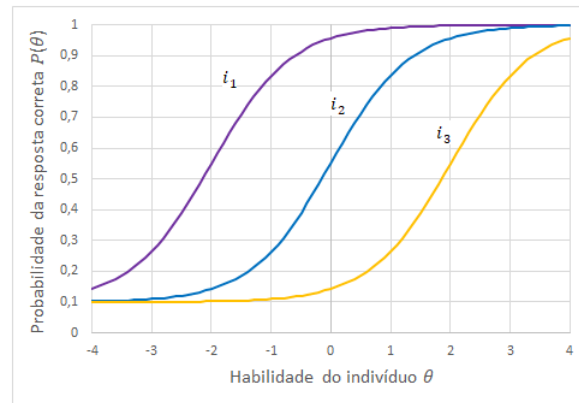
**Figura 4.3:** Exemplo de uma CCI com parâmetros identificados



**Figura 4.4:** Exemplo de Curvas características de itens  $i$ 's com diferentes acertos casuais, no qual  $c_1 > c_2 > c_3$

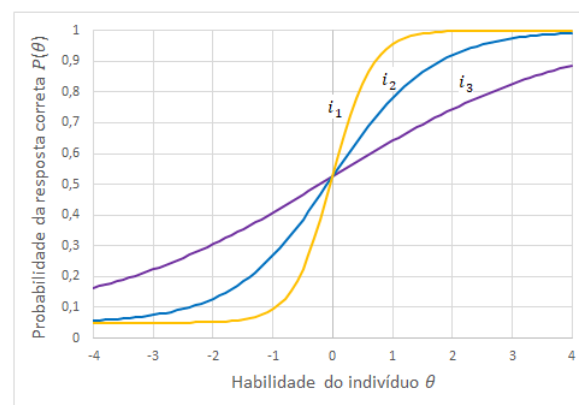
permite que os itens sejam posicionados na mesma escala em que estão posicionados seus respondentes e, portanto, pode-se criar uma escala de *habilidades interpretada* de acordo com as variações das habilidades dos sujeitos. Na CCI, representa o valor de  $\theta$  no ponto de inflexão da curva. Nos modelos em que o parâmetro  $c$  não é considerado, representa o valor de  $\theta$  quando a probabilidade do indivíduo responder corretamente o item é igual a 0,5.

O parâmetro discriminação de um item é representado por  $a$ . Ela indica o quanto um item consegue distinguir um indivíduo de outro, tomando um nível como referência (definido pela dificuldade do item). Logo, sua função é especificar qual indivíduo está abaixo ou acima desse nível. Seu valor é representado pela inclinação da CCI em relação ao eixo das abcissas e é proporcional ao valor da derivada da tangente da curva no ponto de inflexão. Como espera-se que a probabilidade de acertar um item aumente com o aumento da habilidade, não são esperados valores negativos para  $a$  nessa Teoria. Um baixo valor de  $a$  indica que alunos de habilidades muito distintas possuem aproximadamente a mesma probabilidade de acertar o item. Por outro lado, um alto valor de  $a$  indica que discriminam os alunos basicamente em



**Figura 4.5:** Exemplo de Curvas características de itens  $i$ 's com diferentes dificuldades, no qual  $b_3 > b_2 > b_1$ .

dois grupos: os que possuem habilidades abaixo do valor da dificuldade do item e os que possuem habilidade acima (Andrade, Tavares e Valle; 2000).



**Figura 4.6:** Exemplo de Curvas características de itens  $i$ 's com diferentes discriminações, no qual  $a_1 > a_2 > a_3$ .

#### 4.2.4 Estimação e Invariância dos parâmetros

Para que um teste seja bem representado, deve-se adequar os modelos disponíveis aos dados que possui. O primeiro passo é definir precisamente a habilidade ou *dimensão* que se deseja medir. Em seguida, criar itens que condizem adequada e fielmente com essa habilidade. Só aí passamos para uma das mais importantes etapas da TRI: a *estimação dos parâmetros* ou *calibração* dos itens.

Para calibrar os itens e verificar sua unidimensionalidade, aplica-se os itens do teste para uma amostra representativa da população para qual o teste foi concebido. Só depois os parâmetros serão conhecidos e os itens fixados em uma mesma métrica.

Mas como são estimados? Como já mencionado, o modelo TRI se baseia na probabilidade de acerto de um item, probabilidade essa que depende apenas da habilidade do indivíduo e dos parâmetros que qualificam o item. Porém, geralmente ambos são desconhecidos. Dessa forma, para que os parâmetros sejam estimados, se faz necessário manipular algum método. O mais conhecido e utilizado é deno-

minado *Máxima Verossimilhança*, que maximiza a probabilidade de ocorrência dos dados, gerando valores iniciais hipotéticos quase sempre advindos dos índices da psicometria clássica, aferindo valores temporários para as habilidades dos indivíduos e re-estimando os parâmetros de maneira mais precisa, e assim sucessivamente até não conseguir mais melhorar as curvas traçadas aos dados empíricos. Dessa forma, o método da Máxima Verossimilhança consiste em estimar os parâmetros de um modelo utilizando as estimativas que tornam máximo o valor da função de verossimilhança. Isso é equivalente a encontrar o valor para o parâmetro que torna mínima a função de log-verossimilhança negativa (João Luis Batista, 2009), como veremos a seguir.

Considere uma população e uma variável aleatória discreta  $X$  - relacionada a essa população por uma função  $p(x, \theta)$  de probabilidade, em que  $\theta$  é o parâmetro desconhecido - ou uma variável aleatória contínua  $X$  - relacionada a essa população por uma função  $f(x, \theta)$  de densidade de probabilidade. Se retirarmos uma amostra aleatória simples de  $X$  de tamanho  $n$ ,  $(X_1, X_2, \dots, X_n)$  e considerarmos como  $x_1, x_2, \dots, x_n$  os valores verdadeiramente observados, a função de verossimilhança  $L$  será definida por:

$$L(\theta; x_1, \dots, x_n) = f(x_1; \theta) \times \dots \times f(x_n; \theta) \prod_{i=1}^n f(x_i; \theta). \quad (4.2)$$

Se  $X$  é uma variável aleatória discreta

$$L(\theta; x_1, \dots, x_n) = p(x_1; \theta) \times \dots \times p(x_n; \theta) \prod_{i=1}^n p(x_i; \theta).$$

(4.3)

O estimador de máxima verossimilhança de  $\theta$  é, portanto, o valor que irá maximizar  $L(\theta; x_1, \dots, x_n)$ .

Vamos, agora, resumidamente, descrever o processo de estimação de parâmetros dos itens\*.<sup>4</sup>

Para isso precisaremos, primeiro, adotar e enunciar duas suposições que usaremos em todo o nosso estudo:

1. As respostas de indivíduos diferentes são independentes
2. Quando fixada sua habilidade, os itens são respondidos de forma independente por cada indivíduo (Independência local).

Devemos, também, indicar algumas notações necessárias para o desenvolvimento dos modelos. Sejam  $\theta_j$  a habilidade do indivíduo  $j$  e  $U_{ji}$  a variável aleatória que representa a resposta binária do indivíduo  $j$  ao item  $i$ , padronizando 0 como resposta errada e 1 como resposta certa. Ainda, sejam  $U_j = (U_{j1}, U_{j2}, \dots, U_{jl})$  o vetor

<sup>4</sup>Para maiores minúcias consultar [1], capítulo 3.

aleatório de respostas do indivíduo  $j$ ;  $U_{..} = (U_1, \dots, U_n)$  o conjunto integral de respostas;  $u_{ji}, u_j, eu_{..}$  as observações;  $\theta = (\theta_1, \dots, \theta_n)$  o vetor de habilidades dos  $n$  indivíduos e  $\zeta = (\zeta_1, \dots, \zeta_l)$  o conjunto de parâmetros dos itens.

Por (1) e (2), podemos escrever a verossimilhança  $L(\zeta) = P(U_{..} = u_{..} | \theta, \zeta)$ , como

$$L(\zeta) = \prod_{j=1}^n P(U_j = u_j | \theta_j, \zeta) = \prod_{j=1}^n \prod_{i=1}^l P(U_{ji} = u_{ji} | \theta_j, \zeta_i). \quad (4.4)$$

Utilizando as notações  $P_{ji} = P(U_{ji}=1 | \theta_j, \zeta_i)$  e  $Q_{ji} = 1 - P_{ji}$ , teremos:

$$P(U_{ji} = u_{ji} | \theta_j, \zeta_i) = P(U_{ji} = 1 | \theta_j, \zeta_i)^{u_{ji}} P(U_{ji} = 0 | \theta_j, \zeta_i)^{1-u_{ji}} = P_{ji}^{u_{ji}} Q_{ji}^{1-u_{ji}}$$

$$L(\zeta) = \prod_{j=1}^n \prod_{i=1}^l P_{ji}^{u_{ji}} Q_{ji}^{1-u_{ji}} \quad (4.5)$$

A log-verossimilhança pode ser escrita como

$$\log L(\zeta) = \sum_{j=1}^n \sum_{i=1}^l \{u_{ji} \log P_{ji} + (1 - u_{ji}) \log Q_{ji}\} \quad (4.6)$$

Os EMV (Estimadores de Máxima verossimilhança) de  $\zeta_i, i = 1, \dots, l$ , são os valores que maximizam a verossimilhança ou, também, a solução da equação:

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = 0, \quad i = 1, \dots, l \quad (4.7)$$

Daí teremos,

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^n \left\{ \frac{u_{ji} - P_{ji}}{P_{ji} Q_{ji}} \right\} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right) \quad (4.8)$$

Sejam  $P_{ji}^* = \{1 + e^{-Da_i(\theta_j - b_i)}\}^{-1}$  e  $Q_{ji} = 1 - P_{ji}^*$ .

Por comodidade, adotaremos  $W_{ji} = \frac{P_{ji}^* Q_{ji}^*}{P_{ji} Q_{ji}}$ .

Dessa forma, podemos reescrever a equação  $\frac{\partial \log L(\zeta)}{\partial \zeta_i}$  como:

$$\frac{\partial \log L(\zeta)}{\partial \zeta_i} = \sum_{j=1}^n \left\{ (u_{ji} - P_{ji}) \frac{W_{ji}}{P_{ji}^* Q_{ji}^*} \right\} \left( \frac{\partial P_{ji}}{\partial \zeta_i} \right).$$

Assim, as equações de estimação dos parâmetros serão dadas por

$$a_i = (1 - c_i) \sum_{j=1}^n (u_{ji} - P_{ji}) (\theta_j - b_i) W_{ji} = 0$$

$$b_i = -a_i (1 - c_i) \sum_{j=1}^n (u_{ji} - P_{ji}) W_{ji} = 0$$

$$c_i = \sum_{j=1}^n (u_{ji} - P_{ji}) \frac{W_{ji}}{P_{ji}^*} = 0$$

A *invariância dos parâmetros* é uma das grandes diferenças dos modelos da TRI em relação aos da TCT. Quando utilizadas amostras adequadas da população, os resultados encontrados independem da amostra utilizada ou novamente retirada. “Uma das grandes vantagens da TRI sobre a Teoria Clássica é que ela permite a comparação entre populações, desde que submetidas a provas que tenham alguns itens comuns, ou ainda a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente diferentes” (*Dalton Francisco Andrade*, 2000).

### 4.3 Modelos TRI

Existe uma grande diversidade de modelos utilizados na TRI, alguns criados para possibilitar a análise de itens politômicos, outros dicotômicos outros para mensuração de uma (*unidimensionais*) ou de várias habilidades (*multidimensionais*). Neste trabalho, trataremos dos conceitos dos modelos logísticos que medem um único traço latente mais utilizados, dando ênfase à sua aplicação em itens de múltipla-escola ou abertos, ambos *dicotomizados*<sup>5</sup>, que se diferem principalmente nos parâmetros considerados para avaliação.

#### 4.3.1 Modelo Logístico de um Parâmetro

É também conhecido como *modelo de Rasch*, em homenagem ao matemático dinamarquês *George Rasch* (1901-1980), pioneiro nessa proposição. É um modelo no qual é considerada apenas a dificuldade  $b$  do item como parâmetro, utilizado para diferenciar um item de outro. A dificuldade do item condiz com o nível de habilidade mínimo necessário para que o indivíduo acerte aquele item que, nesse modelo, é formalmente definida como o valor de  $\theta$  para o qual  $P(\theta)$  é de 50%. Sua formulação é a seguinte:

$$P(\theta, b) = \frac{e^{(\theta-b)}}{1 + e^{(\theta-b)}}$$

É aplicado principalmente em modelos de saúde mental, no desenvolvimento de instrumentos de qualidade de vida, entre outros.

#### 4.3.2 Modelo Logístico de dois Parâmetros

O Modelo logístico de dois parâmetros, além de considerar o parâmetro dificuldade  $b$  do item, também considera a sua discriminação  $a$ . Sua formulação é mostrada a seguir:

$$P(\theta, b) = \frac{e^{a(\theta-b)}}{1 + e^{a(\theta-b)}}$$

Sua aplicação é reservada a testes em que a discriminação varia em função da habilidade do sujeito.

---

<sup>5</sup>Itens corrigidos como certos ou errados.

### 4.3.3 Modelo Logístico de três Parâmetros

O Modelo logístico de três parâmetros difere dos demais por também considerar a possibilidade de “chute”, ou seja, considera o parâmetro acerto casual  $c$  em sua concepção.

$$P(\theta, a, b, c) = c + (1 - c) \frac{e^{a(\theta - b)}}{1 + e^{a(\theta - b)}}$$

Por penalizar respostas não consistentes ao item, é utilizado principalmente em testes de medição de proficiências educacionais e de mensuração do domínio de determinada competência.

# Exame Nacional do Ensino Médio

---

## 5.1 Introdução

O termo “medidas educacionais” refere-se aos sistemas de mensuração do rendimento acadêmico, tomando por base dados estatísticos. A avaliação da aprendizagem ganhou nova conotação nas últimas duas décadas, afastando-se do conceito mais quantitativo das medidas, buscando aproximar-se de critérios mais qualitativos. As medidas educacionais ocupam-se principalmente com:

- Cálculo de medidas estatísticas (média, desvio padrão, etc);
- Coleta e análise de dados;
- Estudo e comparação de populações.

Nesse sentido, em 1998, como medida educacional dos anos finais escolares, é criado o Exame Nacional do Ensino Médio (Enem) que nasce como um exame voluntário com o objetivo de ser um indicador do perfil de saída dos alunos do Ensino Médio, motivado por discussões internacionais acerca da educação e impulsionado pelo compromisso firmado pelo Brasil e mais de 150 países na Declaração Mundial de Educação para todos, de Jomtie, no qual foram propostos compromissos como inclusão de novas e diferentes pessoas na Educação, ampliação do conceito de aprendizagem e a erradicação do analfabetismo.

A partir de 2009, a reformulação do exame garantiu a sua utilização como forma de seleção unificada para os processos das principais universidades do país (Sisu – Sistema de Seleção Unificada). Esse exame tem também o objetivo de aferir se os participantes demonstram, ao final do Ensino Médio, individualmente, domínio dos princípios científicos e tecnológicos que presidem a produção moderna e se detêm conhecimento das formas contemporâneas de linguagem.

Os resultados do Exame:

- criam parâmetros para a autoavaliação do participante, com vista à continuidade de sua formação e à sua inserção no mercado de trabalho;
- criam referência nacional para o aperfeiçoamento dos currículos do Ensino Médio;



- constituem mecanismo único, alternativo ou complementar para acesso à Educação Superior, especialmente a ofertada pelas instituições federais de Educação Superior;
- possibilitam acesso a programas governamentais de financiamento ou apoio ao estudante da Educação Superior;
- constituem instrumento de seleção para ingresso nos diferentes setores do mundo do trabalho;
- baseiam o desenvolvimento de estudos e indicadores sobre a educação brasileira.

O modelo de exame adotado pelo Enem foi desenvolvido com ênfase na aferição das habilidades e competências que compõem as estruturas mentais, com as quais construímos continuamente o conhecimento, e não somente no uso da memória, que, mesmo sendo de fundamental importância, não pode ser considerada o único elemento responsável pela compreensão do mundo por parte dos indivíduos.

Diferentemente dos modelos e dos processos avaliativos tradicionais, a prova do Enem é contextualizada, colocando o estudante diante de situações-problema que demandam do candidato, além do conhecimento de conceitos, a capacidade de aplicá-los. O Enem não supervaloriza a capacidade de assimilar e acumular informações, mas incentiva o aprender a pensar, a refletir e a “saber como fazer”; valoriza, portanto, a autonomia do jovem na hora de fazer escolhas e de tomar decisões.

Seus itens são elaborados atendendo à Matriz de Referência do Enem<sup>6</sup> tendo como ponto de partida os cinco eixos cognitivos:

- I- Dominar linguagens (DL);
- II - Compreender fenômenos (CF);
- III - Enfrentar situações-problema (SP);
- IV - Construir argumentação (CA);
- V - Elaborar propostas (EP).

Esses eixos estão relacionados a cada uma das quatro áreas do conhecimento: Linguagens, Códigos e suas Tecnologias; Ciências Humanas e suas Tecnologias; Ciências da Natureza e suas Tecnologias; Matemática e suas Tecnologias. Cada área é estruturada com 30 habilidades associadas aos eixos cognitivos.

## 5.2 Teoria de Resposta ao Item No ENEM

Com sua reformulação em 2009, o Enem passa a adotar a Teoria de Resposta ao Item como modelo para cálculo da nota de sua prova objetiva. Sua escolha se deve à promessa de que, por considerar o vetor de respostas do aluno (coerência) e os parâmetros dos itens que constituem a prova (encontrados previamente), a nota

---

<sup>6</sup>Para download da Matriz de Referência do Enem acesse [http://download.inep.gov.br/educacao\\_basica/enem/downloads/2012/matriz\\_referencia\\_enem.pdf](http://download.inep.gov.br/educacao_basica/enem/downloads/2012/matriz_referencia_enem.pdf)

calculada seria mais justa. Além disso, a TRI permite a comparação de populações diferentes ao longo dos anos ou testes. Dessa forma se fez possível um estudo longitudinal e comparativo do desempenho dos participantes.

O modelo utilizado conta com a função logística de 3 parâmetros, já citada anteriormente. Com ele, se faz possível encontrar a probabilidade de resposta correta de um indivíduo para cada item que compõem a prova. Encontradas as probabilidades, calcula-se sua nota de fato.

Existem, na literatura, alguns métodos para se obter a nota de um indivíduo utilizando a TRI. O Enem, assim como o Saeb, faz uso do denominado *Expected a Posteriori* (EAP), que tem como objetivo aplicar a função de probabilidade a priori para calcular a nota, e que possui a informação do desvio padrão e da média dos concluintes do ensino médio da rede pública participantes do Enem de 2009.

A EAP considera os parâmetros dos itens como fixos e, pela suposição da independência entre as habilidades de diferentes indivíduos, podemos fazer as estimações separadas para cada indivíduo. Para encontrarmos a posteriori para a habilidade do indivíduo, vamos atribuir que a distribuição a priori para  $\theta_j, j = 1, \dots, n$  é Normal com vetor de parâmetros  $\eta = (\mu, \sigma^2)$  já conhecidos.

Teremos:

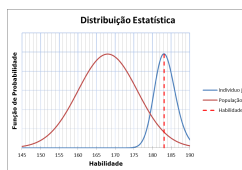
$$g_j^*(\theta_j) \equiv g(\theta_j|u_j, \zeta, \eta) \propto P(u_j|\theta_j, \zeta)g(\theta_j|\eta) \tag{5.1}$$

Para entender as variáveis ver subtópico 4.2.4.

A função que apresentaremos a seguir é de grande importância uma vez que fornece a estimativa da habilidade dado o vetor de respostas do indivíduo  $j$ .

$$g_j^*(\theta) = \frac{P(u_j|\theta, \zeta)g(\theta|\eta)}{\int_{\mathbb{R}} P(u_j|\theta|\eta)d\theta} \tag{5.2}$$

Cada indivíduo possui sua própria função densidade, que também será utilizada na estimação da proficiência.



**Figura 5.1:** Exemplo da função de distribuição do indivíduo  $j$  e distribuição populacional dado por uma Distribuição Normal (0,1).

Para estimar  $\theta_j$ , podemos adotar alguma característica de  $g_j^*(\theta_j)$ . No caso da EAP estimamos pela média da posteriori que consiste em obter a esperança da posteriori.

$$g(\theta_j|u_j, \zeta, \eta) = \frac{P(u_j|\theta, \zeta)}{P(u_j|\zeta, \eta)} \tag{5.3}$$

A estimação de  $\theta_j$  consiste em obter a esperança da distribuição ilustrada na

figura 5.1, que pode ser escrita como

$$\theta_j \equiv E[\theta|u_j, \zeta, \eta] = \frac{\int_{\mathbb{R}} \theta g(\theta|\eta) P(u_j|\theta, \zeta) d\theta}{\int_{\mathbb{R}} g(\theta|\eta) P(u_j|\theta, \zeta) d\theta} \quad (5.4)$$

Para encontrar a nota do indivíduo é preciso resolver a equação e, por envolver o cálculo de integrais de funções bastante complexas, o Enem faz uso do método de quadratura Gaussiana, que foi criado por *Stroud e Secrest* (1966), os quais suscitam que sejam criados intervalos para essas funções, as quais seriam resolvidas individualmente.

### 5.3 Escalas ENEM

A proficiência do aluno no Enem é medida em cada uma de suas áreas do conhecimento, seguindo a matriz de referência do exame. Para calcular a nota dos participantes, o Inep<sup>7</sup> criou uma escala para cada uma dessas áreas, conhecidas como *escalas de proficiência*. Nelas, dois valores são essenciais: o valor de referência e o de dispersão das notas.

Para a referência foi atribuído o valor 500. Essa medida representa a média do desempenho dos alunos do ensino médio da rede pública, que realizaram e concluíram o exame no ano de 2009. Já para a dispersão ou desvio padrão das notas foi atribuído o valor 100, que representa a variabilidade média das notas desses alunos em relação ao valor de referência. As escalas são comumente representadas por uma régua.

Cada escala possui uma interpretação pedagógica. Nelas, as questões são posicionadas de acordo com sua coerência pedagógica e parâmetros previamente conhecidos, já calculados na pré-testagem. O participante, por sua vez, também é posicionado na régua, portanto terá sua nota conforme seu conhecimento pedagógico. Isso significa que sua nota e a interpretação de seu desempenho não depende das notas dos demais participantes, mas do seu posicionamento das questões na régua (5.2).

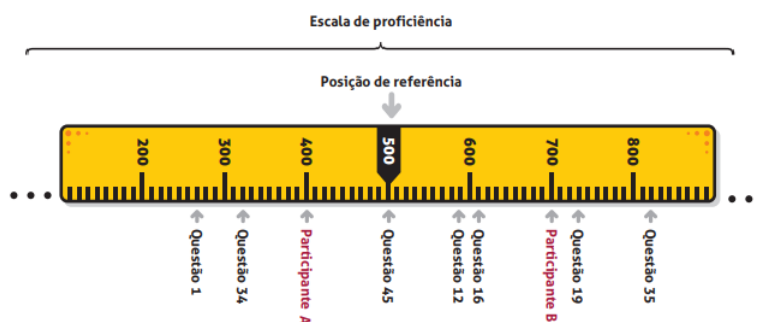


Figura 5.2: Escala de proficiência

Na régua ilustrada o participante A obteve a nota 400, proficiência com uma unidade de desvio padrão (100 pontos) abaixo da proficiência média dos alunos de 2009 nessa área do conhecimento. De acordo com seu posicionamento na régua, ele

<sup>7</sup>O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) é uma autarquia federal vinculada ao Ministério da Educação (MEC), cuja missão é subsidiar a formulação de políticas educacionais.

deve dominar as habilidades avaliadas nas questões 1 e 34, uma vez que elas estão abaixo da posição em que ele se encontra e não dominar as contempladas nas questões 45, 12, 16, 19 e 35, já que estão fixadas acima da posição em que se encontra. Por outro lado, o participante B obteve a nota 700, proficiência com duas unidades de desvio padrão (200 pontos) acima da proficiência média dos alunos de 2009 nessa área do conhecimento. Ele deve compreender as habilidades avaliadas nas questões que são de domínio do Participante A e também as das questões 45, 12 e 16 e não dominar apenas as contempladas nas questões 19 e 35.

É importante ressaltar que, como a escala é formada apenas pelos itens que a compõem, ela não possui início ou fim. Portanto, os valores de máximo e mínimo de uma prova são variáveis e não dependentes de quem a responde.

A escala é alimentada com novos itens a cada ano e os resultados gerados por ela são sempre comparáveis. Porém, em termos pedagógicos, não é possível compararmos notas de escalas (áreas) diferentes. Melhor dizendo: uma nota 900 em matemática não pode ser avaliada como maior ou melhor do que uma nota 800 em Ciências da Natureza, já que esses valores são advindos de diferentes métricas com representações pedagógicas distintas.

## 5.4 Coerência das Notas

Como já explicitado neste trabalho, na metodologia TRI a nota não é calculada considerando-se apenas o número de acertos ou erro de uma avaliação, ela leva em conta a coerência das respostas do participante naquela prova, de forma que não há pontuação predeterminada por questão, ou seja, na TRI os cálculos são feitos levando em consideração a coerência pedagógica das respostas do participante.

A Teoria leva em conta que o conhecimento é cumulativo, portanto não é coerente que alguém erre as questões mais fáceis e acerte várias difíceis. Se o conjunto das respostas não mostrar certa consistência, a nota do respondente não será tão alta quanto a de outro aluno com a mesma quantidade de acertos, mas coerentes.

A Figura 5.3 ilustra uma situação em que as questões de uma mesma prova foram organizadas por nível de dificuldade (do mais fácil para o mais difícil). Em todos os casos apresentados, os candidatos acertaram exatamente 20 questões, porém, receberam notas diferentes.

Analisando os casos de Luiza, candidata com as respostas mais coerentes (acertou todas as questões fáceis e algumas médias, na sequência), e de Rafael, candidato com as respostas menos coerentes (acertou somente as questões mais difíceis), é possível perceber como a coerência tem alta influência no cálculo de uma nota utilizando a TRI. Isso significa que errar uma questão que você tinha alta probabilidade de acertar – uma questão errada entre vários acertos – tem grande impacto em sua nota. Por outro lado, acertar uma questão que você tem baixa probabilidade de acertar – uma questão correta entre vários erros – também soma pontos em sua média, porém, menos do que uma questão coerente.





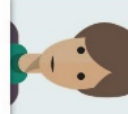










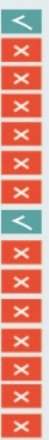
ALUNOS	DIFICULDADE BAIXA	DIFICULDADE MÉDIA	DIFICULDADE ALTA	NOTA
 Maria	Maria acertou os 20 itens mais fáceis, ou seja, obedeceu um comportamento coerente com a régua do Enem, pois é esperado que o aluno acerte os itens mais fáceis e não consiga superar os itens a partir de um determinado nível de dificuldade. 			<b>615,8</b>
 Pedro	Pedro acertou 20 itens de dificuldade média. Como supõe-se que para alcançar o nível médio ele deveria resolver também os itens mais fáceis, sem comportamento não foi tão coerente com a régua do Enem e, por isso, sua nota será mais baixa. 			<b>376,3</b>
 Vitor	Vitor acertou os 20 itens mais difíceis, um comportamento muito incoerente, e, por isso, sua nota foi bastante baixa. De acordo com a TRI, como ele não tem conhecimento para resolver os fáceis, os acertos dos difíceis são interpretados como "chutes". 			<b>301,5</b>
 Carla	Um comportamento próximo do real foi o de Carla, que dos 20 itens que acertou, a maioria tinha uma menor dificuldade. Esse comportamento é razoavelmente coerente e, por isso, sua nota, de acordo com a TRI, ficou maior que as de Pedro e Vitor. 			<b>587,1</b>

Figura 5.3: Exemplo de uma situação em que a nota TRI é calculada com o mesmo número de acertos em uma prova.

## Atividades de Aplicação da TRI no ambiente escolar

---

Um dos objetivos deste trabalho é orientar estudantes e professores do Ensino Médio a respeito da TRI, em particular, para o entendimento acerca do funcionamento do ENEM. Para atingir os objetivos propostos aceitei, em 2016, o convite do Sistema de Ensino para o qual trabalho a me especializar nessa Teoria. Ainda nesse ano, tive a honra em conhecer e participar de um treinamento introdutório sobre TRI com os exímios Dalton Andrade e Adriano Borgatto, ambos excepcionais conhecedores do método. Ao final do treinamento, fui convidada pelos mesmos a participar de duas disciplinas com esse tema que ministravam no *Programa de Pós-graduação em Métodos e Gestão em Avaliação* (PPGMGA) da Universidade Federal de Santa Catarina (UFSC). Comecei então a aprofundar meus estudos e a frequentar eventos que possuíam como tema a TRI e a Avaliação Educacional em geral. Listo a seguir algumas participações que acredito merecerem destaque:

- Agosto de 2016: Palestra ministrada na UFMG pelo ex presidente do INEP Prof José Francisco Soares com o tema “Enem: os estudantes as provas e as escolas”.
- Setembro de 2016: Participação do congresso Educativa Sudeste com palestra e discussões com grandes nomes da Educação. Destaque para a palestra sobre a TRI do Renato Júdice de Andrade, também grande conhecedor do assunto.
- Novembro de 2016: Participação com artigo e apresentação de pôster no Congresso Brasileiro de Teoria de Resposta ao Item (CONBRATRI) em Campinas. Os trabalhos apresentados foram intitulados como “Posicionamento de Itens na Escala Enem” e encontram-se disponível em anexo.
- Novembro de 2017: Presença e participação no VII Congresso Internacional em Avaliação Educacional, sediado em Fortaleza. Nele, tive a oportunidade em escutar, discutir e participar de várias interlocuções acerca da TRI com precursores e notáveis entendedores e admiradores do método. Destaque para

Ruben Klein e Joaquim José Soares Neto, esse último atual presidente da Associação Brasileira de Avaliação Educacional (ABAVE).

No mês de maio deste ano completarão 05 anos em que trabalho no Sistema de Ensino, instituição cuja matriz se encontra em Belo Horizonte e que possui hoje cerca de 400 escolas parceiras em todo o Brasil, com segmentos do infantil ao pré-vestibular. Essa parceria é selada com a comercialização de publicações de produção própria (livros, revisionais, simulados, etc) e capacitação de professores e equipe educacional em geral. Hoje assumo a coordenação de uma equipe que é responsável pela correção das provas objetivas de nossos Simulados Enem pela TRI e ministro com certa frequência palestras, mini-cursos e treinamentos internos e externos para alunos, professores e equipe pedagógica de todo o Brasil sobre a TRI e sua utilização no Enem, além de todas suas possibilidades de estudo e feedback do desempenho escolar.

A atividade de Orientação do profissional da educação contou com dois momentos:

Primeiramente, a elaboração de uma cartilha que se encontra em anexo intitulada como “Rumo ao Enem: Entenda a sua nota calculada pela TRI”. Seu objetivo é apresentar conceitos, expor o método utilizado e desmistificar algumas crenças errôneas acerca da Teoria, muitas vezes divulgadas inadequadamente pela mídia.

No segundo momento, ministrei uma palestra sobre a Teoria de Resposta a Item para professores e estudantes de escolas públicas, alunos de licenciatura e professores da UFV – Campus Florestal no dia 27 de novembro de 2017. Nela, abordei a diferença entre TCT e TRI e suas principais aplicações e fundamentos. Esta palestra foi organizada da seguinte maneira:

- 10 minutos para explicitar conceitos que embasam a Teoria: Psicometria, definição de medida, fenômenos observáveis e não-observáveis.
- 20 minutos para introduzir os fundamentos da TRI e da TCT, incluindo um quadro comparativo das duas Teorias;
- 10 minutos para apresentar uma brincadeira, que visa tratar a altura como um traço latente ou uma característica não-observável. Nela, para se encontrar uma altura de uma pessoa é necessário responder a um questionário, no qual cada item possui uma “altura” e as respostas deveriam se limitavam a sim ou não;
- 10 minutos para instruir sobre o posicionamento de itens em uma mesma métrica, anunciar que um indivíduo pode ser colocado nessa mesma escala e apresentar possibilidades de interpretação dessa escala;
- 20 minutos para evidenciar conceitos essenciais da TRI, como a formulação do modelo, a curva característica de um item e o comportamento e importância de cada parâmetro.
- 20 minutos para divulgar a utilização da TRI no Enem e discutir acerca dos mitos e verdades dessa utilização e sua interpretação.

Para este encontro estiveram presentes coordenadores do PIBID, que se interessaram pelo assunto proposto. Ao final da palestra, foi distribuído aos alunos do Ensino Médio a cartilha citada anteriormente.

O conhecimento da TRI é importante para todos aqueles que se encontram envolvidos na Educação Básica e a difusão dos principais conceitos dessa teoria deveria ser exigido em todas as escolas de Ensino Médio, uma vez que o principal meio de ingresso ao vestibular é através de uma prova corrigida através desse método. Espero que, com este trabalho, professores do Ensino Básico tenham acesso ao tema e possam compartilhar o conhecimento com seus alunos.



## Conclusão

---

Medir traços latentes e mensurar habilidades são os principais objetivos e focos de estudo da Psicometria. Para isso, ela fundamenta-se em teorias de medidas que explicam qual o entendimento deve-se ter às respostas obtidas através de estímulos calculados.

Atualmente, a Psicometria se divide em duas vertentes: a clássica, com a Teoria Clássica dos Testes (TCT) e a moderna, com a Teoria de Resposta ao Item (TRI). A segunda é um grande avanço em relação à primeira no que diz respeito à invariância das características dos itens e às estimativas de proficiências comparáveis, uma vez que permite que as habilidades de diferentes alunos sejam colocadas na mesma métrica que os parâmetros dos itens que compuseram ou compõem os testes realizados. Porém, a TRI não exclui a importância da TCT, que também possui suas peculiaridades.

No Enem, o uso da TRI foi implantado em 2009, visando uma avaliação mais justa e um maior controle do desempenho dos alunos do decorrer dos anos. Com o uso desse método é possível criar uma maior coerência entre as respostas dadas por um participante, de forma que a precisão da medida de sua habilidade seja mais adequada.

## Anexo Um: Cartilha

---





Esta cartilha foi produzida pela Vanessa Santos, aluna do PROFMAT da UFV/ Campus UFV-Florestal, com o intuito de orientar alunos do Ensino Médio acerca da teoria TRI. Com este documento o estudante aprende a entender sua nota no ENEM além de desmistificar algumas crenças errôneas acerca da Teoria, muitas vezes divulgadas inadequadamente pela mídia.

## Você sabia que, pela metodologia TRI a pontuação, em uma mesma área, de dois estudantes que acertaram o mesmo número de questões pode ser diferente?

Tradicionalmente, as notas geradas a partir de uma avaliação levam em conta somente o número de acertos e erros. Na metodologia TRI, a nota não é calculada considerando-se apenas esse critério, mas também a coerência das respostas do participante naquela prova, de forma que não há pontuação predeterminada por questão.

**Na TRI os cálculos são feitos levando em consideração a coerência pedagógica das suas respostas.**

Não é coerente que alguém erre as questões mais fáceis e acerte várias difíceis, concorda? Não faz sentido errar um cálculo simples de área do quadrado e, ao mesmo tempo, acertar uma questão complexa, envolvendo o cálculo de volume de um cone, por exemplo. Se o conjunto das respostas não mostrar certa consistência, a nota do respondente não será tão alta quanto a de outro aluno com a mesma quantidade de acertos, mas coerentes.

ALUNOS	DIFICULDADE BAIXA	DIFICULDADE MÉDIA	DIFICULDADE ALTA	NOTA
 Maria	✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓	✓✓✓✓✓✓ x x x x x x x x x x x x	x x x x x x x x x x x x x x x x x x	615,8
 Pedro	x x x x x x x x x x x x x x x x x x	✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓	✓✓✓✓ x x x x x x x x x x x x x x x x	376,3
 Vitor	x x x x x x x x x x x x x x x x x x	x x x x x x x x x x x x x x x x x x	✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓✓	301,5
 Carla	✓✓✓ x ✓✓✓ x ✓✓✓✓✓✓✓✓✓✓✓✓✓✓	x x x x x x x x x x x x x x x x x x	✓ x x x x x x x x x x x x x x x x x x	587,1

A imagem ilustra algumas situações possíveis. Nela, as questões de uma prova estão organizadas por nível de dificuldade (do mais fácil para o mais difícil). Em todos os casos apresentados, os candidatos acertaram exatamente 20 questões, mas receberam notas diferentes. Analisando os casos de Luiza, candidata com as respostas mais coerentes (acertou todas as questões fáceis e algumas médias, na sequência), e de Rafael, candidato com as respostas menos coerentes (acertou somente as questões mais difíceis), é possível perceber como a coerência impacta em sua nota. Ou seja, errar uma questão que você tinha alta probabilidade de acertar – uma questão errada entre vários acertos – tem grande impacto

em sua nota. Acertar uma questão que você tem baixa probabilidade de acertar – uma questão correta entre vários erros – também soma pontos em sua média, porém, menos do que uma questão coerente.

## MAS AFINAL, COMO POSSO INTERPRETAR A MINHA NOTA?

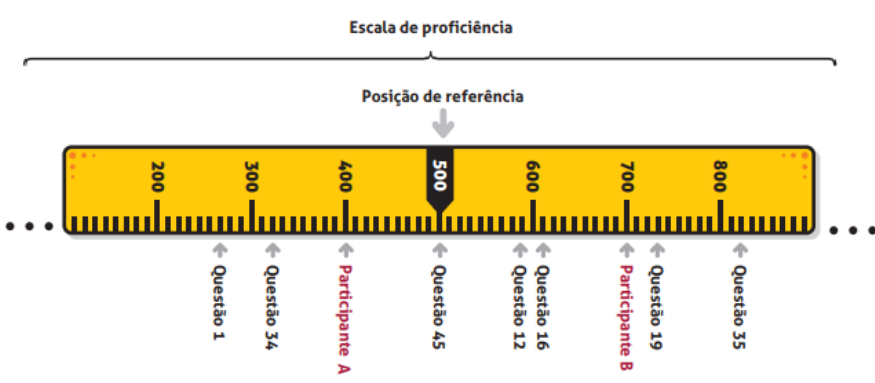
A nota do aluno é atribuída em uma escala criada especialmente pelo Enem, com o objetivo de medir o conhecimento (proficiência) dos participantes em quatro áreas distintas: Ciências da Natureza e suas Tecnologias, Matemática e suas Tecnologias, Ciências Humanas e suas Tecnologias e Linguagens, Códigos e suas Tecnologias. A escala

**Proficiência é a nota do participante**

**Cada área do conhecimento possui uma escala diferente. Quantitativamente, só são comparáveis as notas de uma mesma área**

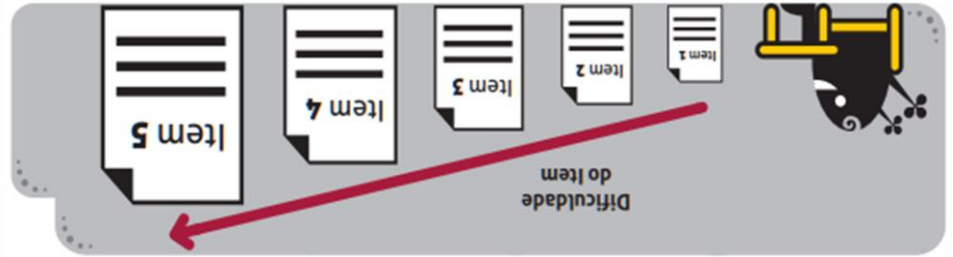
depende apenas de dois valores: **valor de referência**, pelo qual foi atribuído o valor 500 (desempenho médio dos concluintes do ensino médio da rede pública que realizaram o exame em 2009) e **valor de dispersão** (variabilidade média da nota ou desvio padrão). Além disso, a escala de cada área possui uma interpretação pedagógica. Nelas, as questões são posicionadas de acordo com uma coerência pedagógica e parâmetros previamente conhecidos (calculados na pré-testagem). O participante, por sua vez, também será posicionado na régua, portanto terá sua nota conforme seu conhecimento pedagógico. Dessa forma, sua nota não depende das notas dos demais participantes, mas apenas do posicionamento das questões na régua, assim como os valores de máximo e mínimo da prova.

ÁREA DO CONHECIMENTO: xxxxxx



**Participante A**  
Nota: 400  
Proficiência com uma unidade de desvio padrão (100 pontos) abaixo da proficiência média dos alunos de 2009 nessa área do conhecimento.  
Deve dominar as habilidades avaliadas nas questões 1 e 34 (abaixo de sua posição na régua) e não dominar as contempladas nas questões 45, 12, 16, 19 e 35 (acima de sua

**Participante B**  
Nota: 700  
Proficiência com duas unidades de desvio padrão (200 pontos) acima da proficiência média dos alunos de 2009 nessa área do conhecimento.  
Deve dominar as habilidades avaliadas nas questões que são de domínio do Participante A e também as das questões 45, 12 e 16, e não dominar apenas as contempladas nas questões 19 e 35. Portanto sua nota deve ser maior.



# ENTENDA A SUA NOTA CALCULADA PELA TRI

RUMO AO ENEM:

## MITOS E VERDADES: TRI NO ENEM

- ✓ **Chutar pode me prejudicar?** Não! Uma questão certa sempre aumenta sua nota e uma questão deixada em branco é corrigida como errada. Se a questão que acertou não possui a coerência pedagógica esperada, ela só não terá tanto valor como uma outra que seria esperada.
- ✓ **É possível tirar mais que 1000?** Sim! A prova é constituída por um conjunto de itens com parâmetros pré-estabelecidos pelo pré-teste. São esses parâmetros que definem os valores de mínimo e máximo da prova.
- ✓ **Existe nota zero?** O Enem só confere a nota zero àquele aluno que não assinalou nenhuma questão. Do contrário, a nota do aluno será sempre referente ao conhecimento mínimo que a prova consegue medir (nota mínima).
- ✓ **Posso comparar as minhas notas?** A nota de cada área do conhecimento é calculada em uma escala diferente. Portanto só é possível comparas as notas de áreas iguais. Por exemplo: Se tirei 700 em Matemática no ano de 2016 e 900 no ano de 2017, posso afirmar que melhorei meu desempenho, mas isso só é verdade ao se comparar notas da mesma área.
- ✓ **Alguma prova vale mais?** Não. Porém, alguns vestibulares consideram a nota final como a média simples das 4 áreas objetivas da prova. Portanto, como a régua de matemática já chegou de 1000 e a de Linguagens nunca alcançou 900, ao se calcular a média citada, a prova de matemática teria mais “valor”. Para que isso não ocorra, muitas faculdades recorrem à média ponderada, atribuindo pesos para as áreas mais importantes para o curso que está sendo pleiteado.
- ✓ **A redação é corrigida pela TRI?** Não. Existem estudos para futuramente se corrigir uma prova via TRI, mas a sua correção ainda é feita da maneira tradicional.
- ✓ **Consigo eu mesmo calcular a minha nota?** Não. Calcular uma nota via TRI requer reforço computacional, além de parâmetros pré-estabelecidos para as questões que são mensurados na pré-testagem das questões.

### PRINCIPAIS OBJETIVOS NO ENEM

- ❖ Medir conhecimento / proficiência em áreas do conhecimento
- ❖ Discriminar alunos bons dos ruins
- ❖ Considerar coerência pedagógica das respostas
- ❖ Tornar notas de uma mesma área comparáveis de ano a ano
- ❖ Criar uma escala de itens com interpretação pedagógica

$$P(U_{ij} = 1 | \theta_j) = c_j + (1 - c_j) \frac{1 + e^{-Da_j(\theta_j - b_j)}}{1}$$

O QUE É?  
 TRI é uma metodologia de correção de questões objetivas que considera para o cálculo da nota a consistência da resposta segundo o grau de dificuldade de cada questão.

## TEORIA DE RESPOSTAS AO ITEM (TRI)

# Anexo Dois: Simulados na escala do ENEM

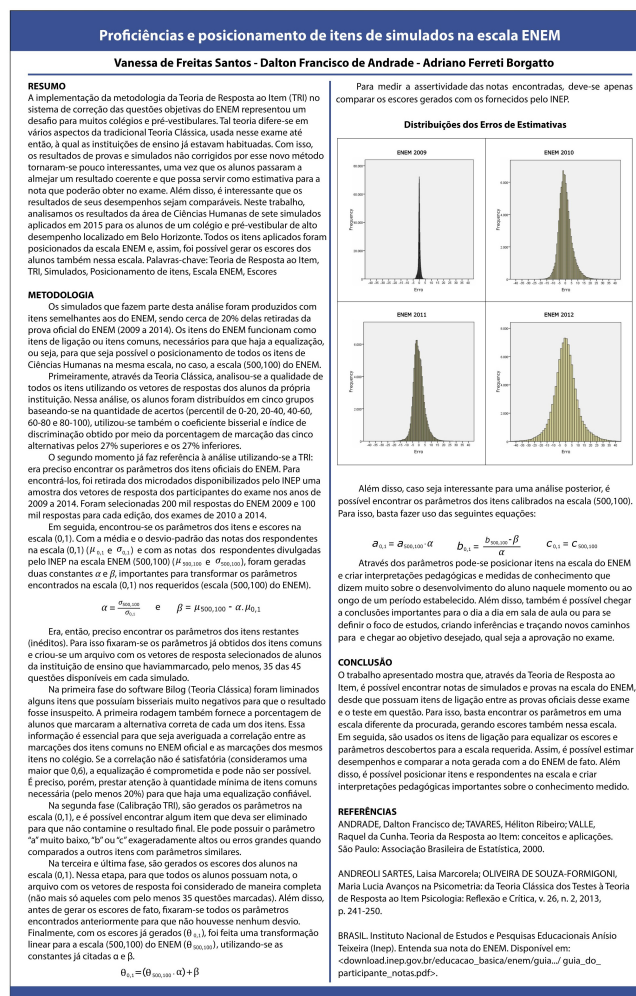


Figura B.1: Proficiência e posicionamento dos itens dos simulados na escala do ENEM.

# Bibliografia

---

- [1] Andrade, Dalton Francisco de, Heliton Ribeiro Tavares e Raquel da Cunha Valle: *Teoria de Resposta ao Item: conceitos e aplicações*. Associação Brasileira de Estatística (ABE), 2000.
- [2] Baker, Frank B e Seock Ho Kim: *Item response theory: Parameter estimation techniques*. CRC Press, 2004.
- [3] Brasil, Ministério da Educação: *Guia do Participante: Entenda sua nota no ENEM*. Instituto Nacional de Estudos e Pesquisas Educacionais, Brasília, DF, 2012.
- [4] Brasil, Ministério da Educação: *Matriz de Referência Enem: eixos cognitivos*. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Brasília, DF, 2012.
- [5] Brasil, Portaria n 468 de 3 de abril de 2017: *Dispõe sobre a realização do Exame Nacional do Ensino Médio – Enem, e dá outras providências*. Diário Oficial da União, 4 abr. 2017. Seção 1. p.40.
- [6] Lord, Frederic: *A theory of test scores*. Psychometric monographs, 1952.
- [7] Pasquali, Luiz: *Psicometria: teoria e aplicações: a teoria clássica dos testes psicológicos*. Ed. da UnB, 1997.
- [8] Pasquali, Luiz: *Psicometria: teoria dos testes na psicologia e na educação*. Editora Vozes Limitada, 2017.