

Teoria da Resposta ao Item e avaliações educacionais

Caio L. N. Azevedo, IMECC/Unicamp

Agradecimentos à Raquel da Cunha Valle, Fundação Carlos Chagas e ao Prof. Dalton Andrade, INE-UFSC, por cederem parte do material utilizado

IX Mostra Estadual de Ciência, Tecnologia e Inovação/14ª Semana Nacional de Ciência e Tecnologia, Cuiabá, Outubro de 2017

Graduação em Estatística na Unicamp

- <http://www.ime.unicamp.br/graduacao/estatistica>.

Disciplinas de fundamentos, de cunho metodológico e análise de dados. Algumas das áreas de atuação do estatístico são: controle de qualidade, modelagem de crédito e risco, pesquisa e desenvolvimento em indústrias farmacêuticas, ensaios clínicos, pesquisas de opinião, pesquisas de mercado e marketing, pesquisa acadêmica, avaliação psicométrica.

Pós-Graduação em Estatística na Unicamp

- <http://www.ime.unicamp.br/pos-graduacao/estatistica>. O Programa recebeu a nota 5 na última avaliação do CAPES. Todos os professores credenciados no programa são pesquisadores ativos: Análise de Sobrevivência, Biologia Computacional, Bioinformática clínica, Modelos Lineares Generalizados, Modelos Mistos, Métodos Não-paramétricos, Estatística Genética, Epidemiologia, Inferência Estatística Frequentista, Inferência Estatística Bayesiana, Métodos não Paramétricos, Séries Temporais e Finanças, Teoria Assintótica, Robustez, Modelos Lineares Generalizados, Modelos Mistos, Teoria de Resposta ao Item, Métodos Estatísticos Computacionais, Processos Markovianos, Inferência para Processos Estocásticos, Sistemas de Partículas Interagentes, Meios Aleatórios, Processos Pontuais, Teoria de Filas e Probabilidade Aplicada em Linguística.

- Traço latente: característica não-observável.
- Exemplo: conhecimento, grau de maturidade, nível de depressão, predisposição de adquirir um produto, grau de aceitação à pena de morte (entre outros).
- Conhecimento: “domínio” sobre algum assunto.
- Como medir conhecimento (traços latentes de forma geral)?
 - Quantificar
 - Interpretar

■ Teoria Clássica de Medidas

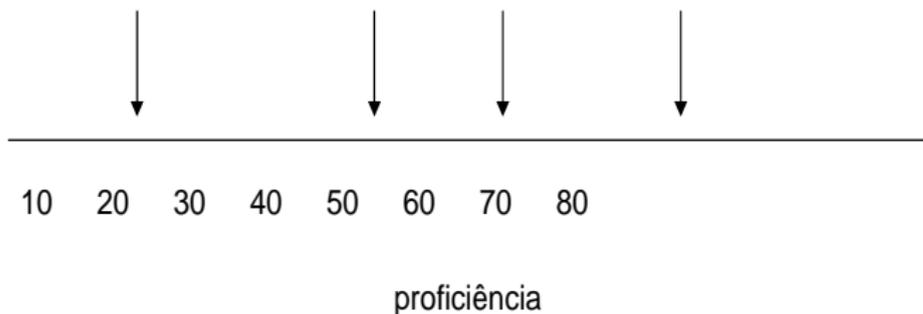
- Escore total (respondente): no. de acertos varia de 0 a I (no. de itens).
- Índice de dificuldade (item): proporção ou porcentagem de acerto varia de 0 a 1.
- Índice de discriminação (item): diferença entre a porcentagem de acerto do grupo superior e a porcentagem de acerto do grupo inferior varia de -1 a 1.
 - Grupo superior: 27% dos respondentes com os escores mais altos.
 - Grupo inferior: 27% dos respondentes com os escores mais baixos.

- O que significa um indivíduo obter uma nota maior que outro ?
- Os indivíduos tem de ser submetidos ao mesmo teste (ou testes equivalentes).
- Como comparar indivíduos submetidos à testes diferentes?
- Como analisar as características de um teste ?
- Como avaliar o que, de fato, um indivíduo sabe ?
- Como medir o ganho de conhecimento de indivíduos, ao longo do tempo ?

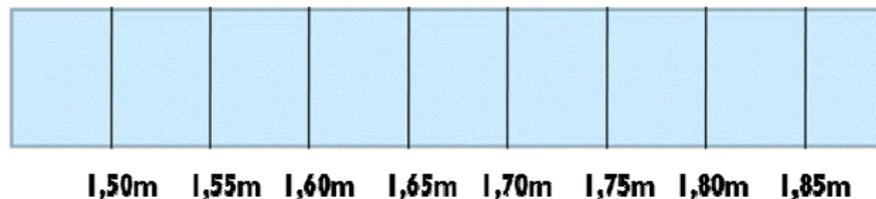
- Proficiências: traços latentes (construtos)
 - Não é possível observá-las diretamente.
 - Diferentemente da altura ou renda, é necessário se estabelecer uma métrica (régua).
 - Conhecimento, predisposição para manifestar alguma doença, grau de maturidade de uma empresa na implementação da gqt, nível de depressão.
 - Escala de proficiência: régua na qual se podem colocar os “traços latentes dos indivíduos” .
 - Cumulativo (cognição, nível de depressão) ou não cumulativo (opnião, grau de aceitação).

- O que significa “medir” traços latentes?
 - É o processo por meio do qual números são atribuídos aos objetos (itens e respondentes), tais que a relação entre objetos é representada pela relação entre números.
- Uma das maiores contribuições da TRI foi propor modelos que permitem que itens e indivíduos possam ser posicionados na mesma régua.

Escala de proficiência : régua



Régua para altura



Medida: “brincando” com a altura

- 1 Na cama, eu frequentemente sinto frio nos pés.
- 2 Eu frequentemente desço as escadas de dois em dois degraus.
- 3 Eu acho que me daria bem em um time de basquete.
- 4 Como policial, eu impressionaria muito.
- 5 Na maioria dos carros eu me sinto desconfortável.
- 6 Eu literalmente olho para meus colegas de cima para baixo.
- 7 Você é capaz de pegar um objeto no alto de um armário, sem usar escada?
- 8 Você abaixa quando vai passar por uma porta?
- 9 Você consegue guardar a bagagem no porta-malas do avião?
- 10 Você regulava o banco do carro para trás?
- 11 Normalmente quando você está andando de carona lhe oferecem o banco da frente?
- 12 Quando você e várias pessoas vão tirar fotos, formando-se três fileiras, onde ninguém ficará agachado, você costuma ficar atrás?
- 13 Você tem dificuldade para se acomodar no ônibus?
- 14 Em uma fila, por ordem de tamanho, você é sempre colocado atrás?

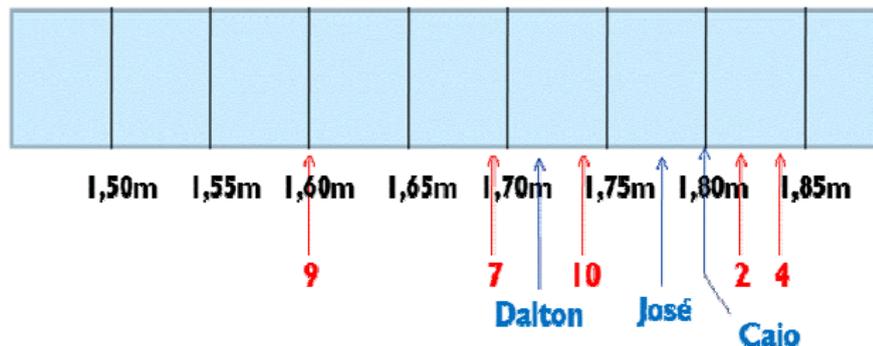
Medida: “brincando” com a altura

- 1 Na cama, eu frequentemente sinto frio nos pés.
- 2 **Eu frequentemente desço as escadas de dois em dois degraus.**
- 3 Eu acho que me daria bem em um time de basquete.
- 4 **Como policial, eu impressionaria muito.**
- 5 Na maioria dos carros eu me sinto desconfortável.
- 6 Eu literalmente olho para meus colegas de cima para baixo.
- 7 **Você é capaz de pegar um objeto no alto de um armário, sem usar escada?**
- 8 Você abaixa quando vai passar por uma porta?
- 9 **Você consegue guardar a bagagem no porta-malas do avião?**
- 10 **Você regulava o banco do carro para trás?**
- 11 Normalmente quando você está andando de carona lhe oferecem o banco da frente?
- 12 Quando você e várias pessoas vão tirar fotos, formando-se três fileiras, onde ninguém ficará agachado, você costuma ficar atrás?
- 13 Você tem dificuldade para se acomodar no ônibus?
- 14 Em uma fila, por ordem de tamanho, você é sempre colocado atrás?

Posicionamento dos respondentes e dos itens na mesma escala



Posicionamento dos respondentes e dos itens na mesma escala

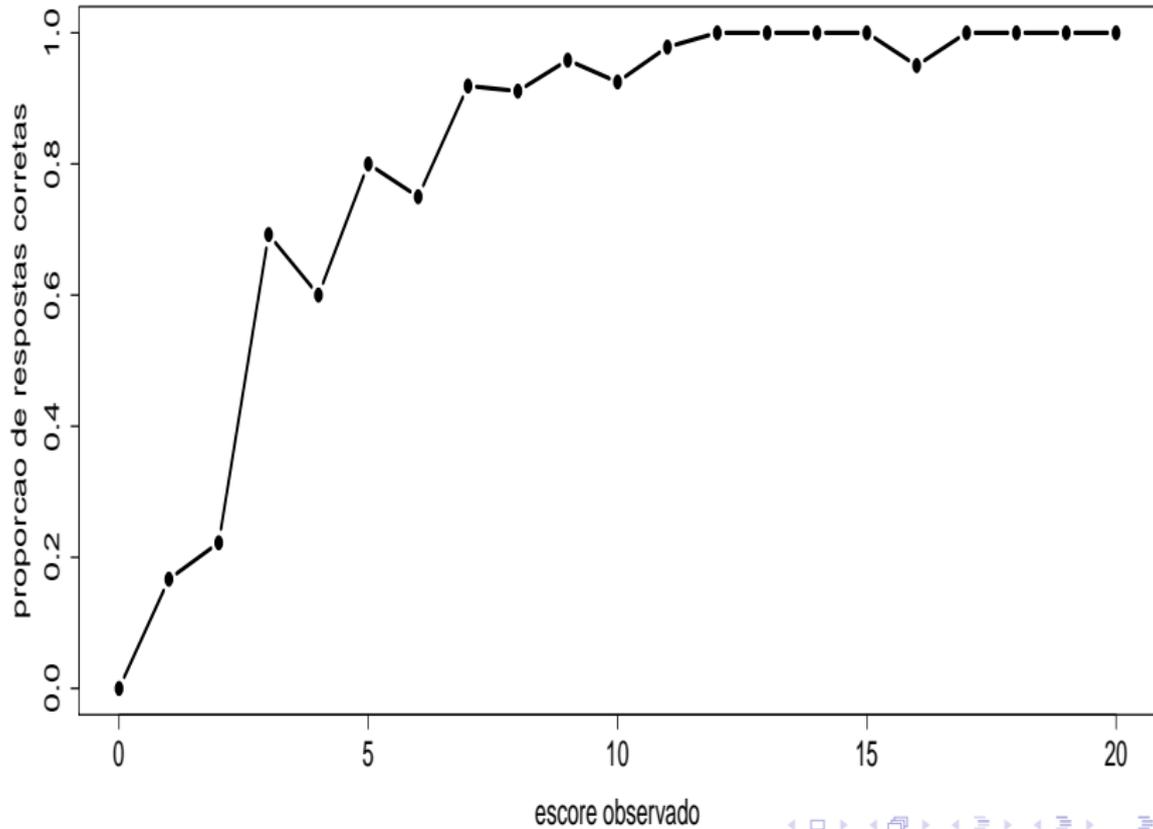


1

- Teoria psicométrica desenvolvida para suprir necessidades na área educacional. É composta por conjunto de modelos que consideram variáveis latentes.
- Modelos de Resposta ao Item (MRI) : representam o relacionamento entre traços latentes de indivíduos e itens de um instrumento de medida (prova, questionário). Tal modelagem consiste na probabilidade de obter um certo escore em cada item.
- Existe um grande número de classes de MRI : dicotômicos e politômicos, um e múltiplos grupos, multidimensionais, longitudinais multiníveis, entre outros.
- Aplicada em diversas áreas: educação, marketing, psiquiatria, genética etc.

- Surgiu, formalmente, a partir dos trabalhos de Lord (1952) e Rasch (1960).
- ENEM (desde 2009).
- No Brasil vem sendo usada extensamente em avaliação educacional
SAEB, ENADE, ENEM, ...
- No mundo: TOEFL, GRE, PISA, ...

- A probabilidade do indivíduo obter uma certa resposta a um item é modelada como função da proficiência do indivíduo e dos parâmetros que representam algumas propriedades dos itens.
- Modelos para traços latentes cumulativos: quanto maior a proficiência do indivíduo, maior a probabilidade de uma resposta correta.
- Modelos para traços latentes não-cumulativos (desdobramento): quanto mais próxima a proficiência do indivíduo estiver do atributo de interesse, maior a probabilidade do indivíduo apresentar tal atributo.
- Propriedade da invariância: os parâmetros dos itens e as proficiências são invariantes, exceto pela escolha da escala (métrica).



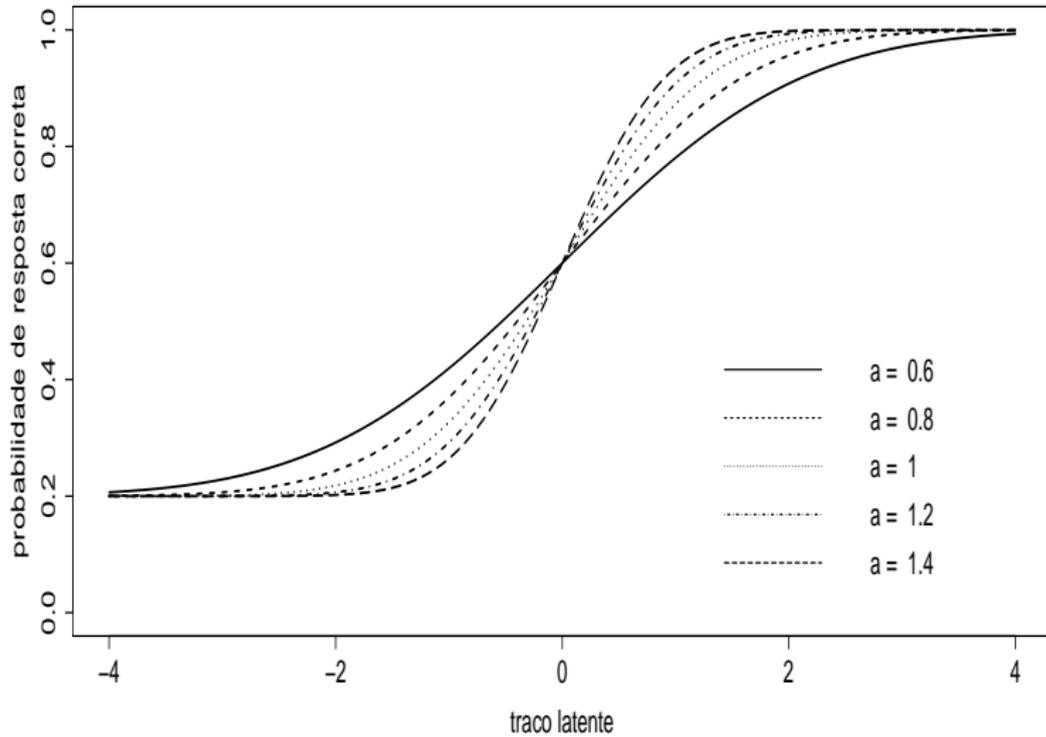
- **Modelo de resposta ao item** : Seja Y_{ij} a resposta do indivíduo j ao item i .

$$Y_{ij} | (\theta_j, \zeta_i) \sim \text{Bernoulli}(p_{ij}),$$

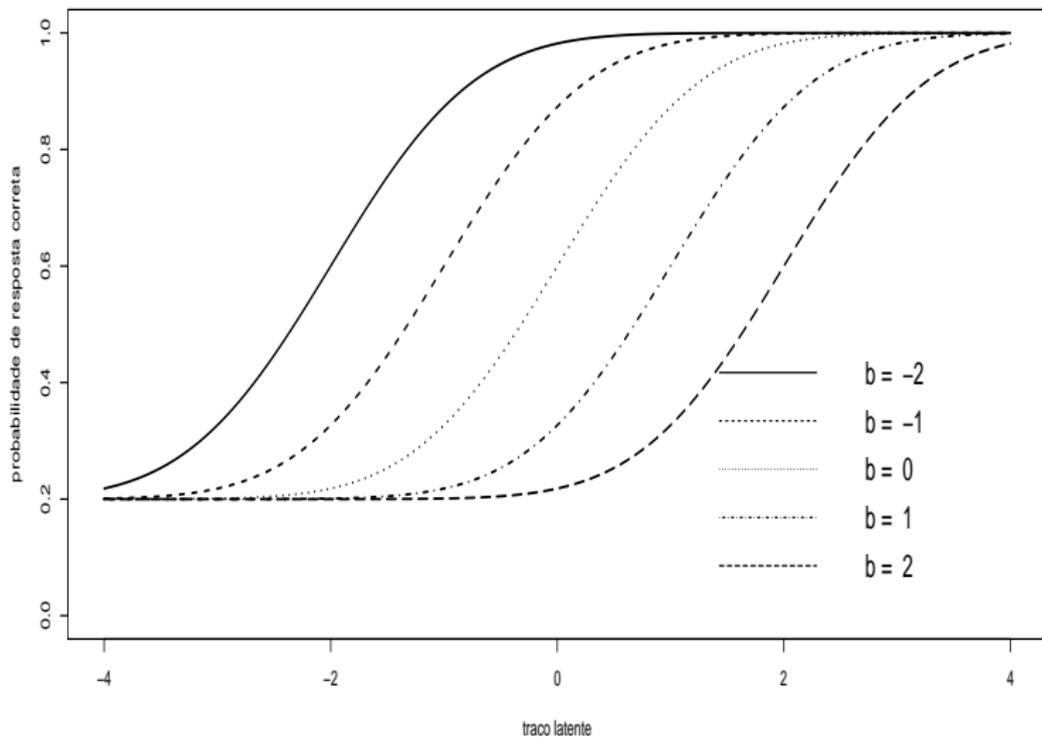
$$p_{ij} = P(Y_{ij} = 1 | \theta_j, \zeta_i) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}$$

- θ_j : traço latente do indivíduo j . Suposição usual $\theta_j \sim N(0, 1)$.
- $\zeta_i = (a_i, b_i, c_i)'$.
- a_i : parâmetro de discriminação (escala) do item i .
- b_i : parâmetro de dificuldade (posição) do item i .
- c_i : probabilidade aproximada (assintótica) de indivíduos com traço latente baixo do item i .

Curvas do modelo L3P



Curvas do modelo L3P



ENEM

- Aplicado pela primeira vez em 1998.
- Até 2009 o ENEM tinha estrutura multidimensional:
 - É (era) um exame interdisciplinar, onde os vários conhecimentos associados aos conteúdos do ensino fundamental e médio são avaliados de uma só vez por um único teste.
 - O grande diferencial do exame pode ser atribuído aos itens que o compõem.
 - Cada um deles era elaborado de modo a avaliar até 5 competências, mesclando os conhecimentos de diferentes disciplinas. Nojosa (2001)
 - As notas eram obtidas a partir da Teoria Clássica de medidas (escores padronizados).

- A partir de 2009 o ENEM passou a ter uma estrutura unidimensional (uma competência dominante por item) e as respectivas notas obtidas (parcialmente) através da TRI (provas objetivas) + teoria clássica de medidas (redação).

- Site oficial do ENEM:

https://enem.inep.gov.br/#/antes?_k=5pwsnt.

- Manual explicativo do ENEM:

http://download.inep.gov.br/educacao_basica/enem/manuais/2012/guia-do-participante-Notas.pdf.

- Manual da redação do ENEM 2017:

http://download.inep.gov.br/educacao_basica/enem/guia_participante/2017/manual_de_redacao_do_enem_2017.pdf.

- Provas objetivas, sendo 45 questões cada
 - Ciências Humanas e suas Tecnologias
 - Ciências da Natureza e suas Tecnologias
 - Linguagens, Códigos e suas Tecnologias
 - Matemática e suas Tecnologias

- Redação
 - Texto dissertativo-argumentativo a partir de uma situação-problema (política, social ou cultural) 30 linhas no máximo.

- Como a redação será avaliada?

Os dois professores avaliarão seu desempenho de acordo com os seguintes critérios:

Competência 1:	Demonstrar domínio da modalidade escrita formal da língua portuguesa.
Competência 2:	Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa.
Competência 3:	Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista.
Competência 4:	Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação.
Competência 5:	Elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos.

Régua do ENEM (2009)

Considera-se como “ponto de referência” a média dos alunos que fizeram o ENEM em 2009 (500) e o desvio-padrão (unidade de medida) de 100. Assim, temos a escala (500,100).



Escalas...

O que é
uma escala?

No contexto educacional, a escala de proficiência é uma forma de estabelecer uma unidade de medida padrão do conhecimento. Sem uma unidade de medida preestabelecida, é difícil fazer comparações e julgamento de valor.

Para o Enem foram desenvolvidas quatro escalas, uma para cada área do conhecimento.

1,53 m



Como medir
a proficiência em
Matemática?



Para medir o conhecimento, aplicam-se provas com itens pertencentes a uma escala de proficiência.

CCI (ENEM)

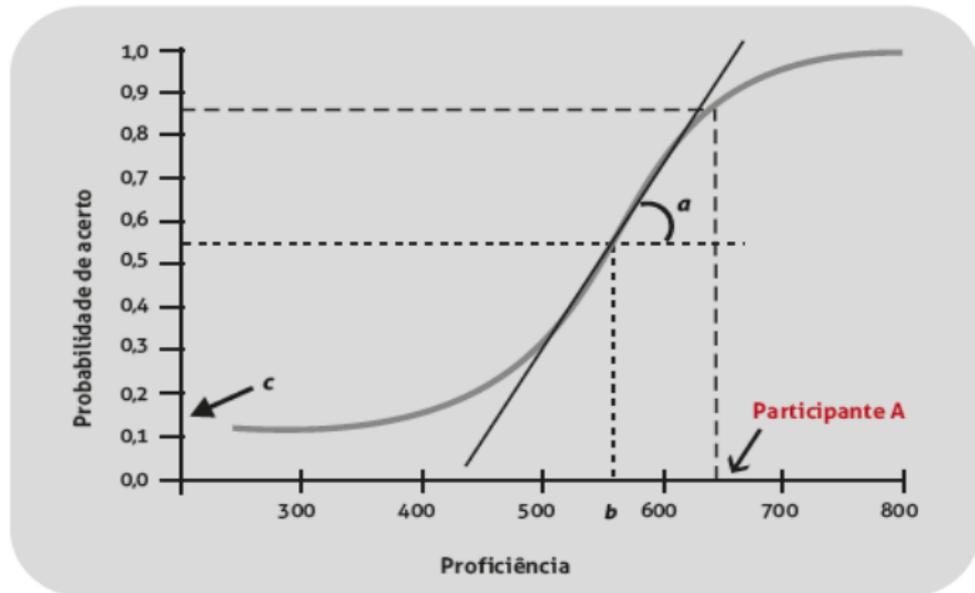
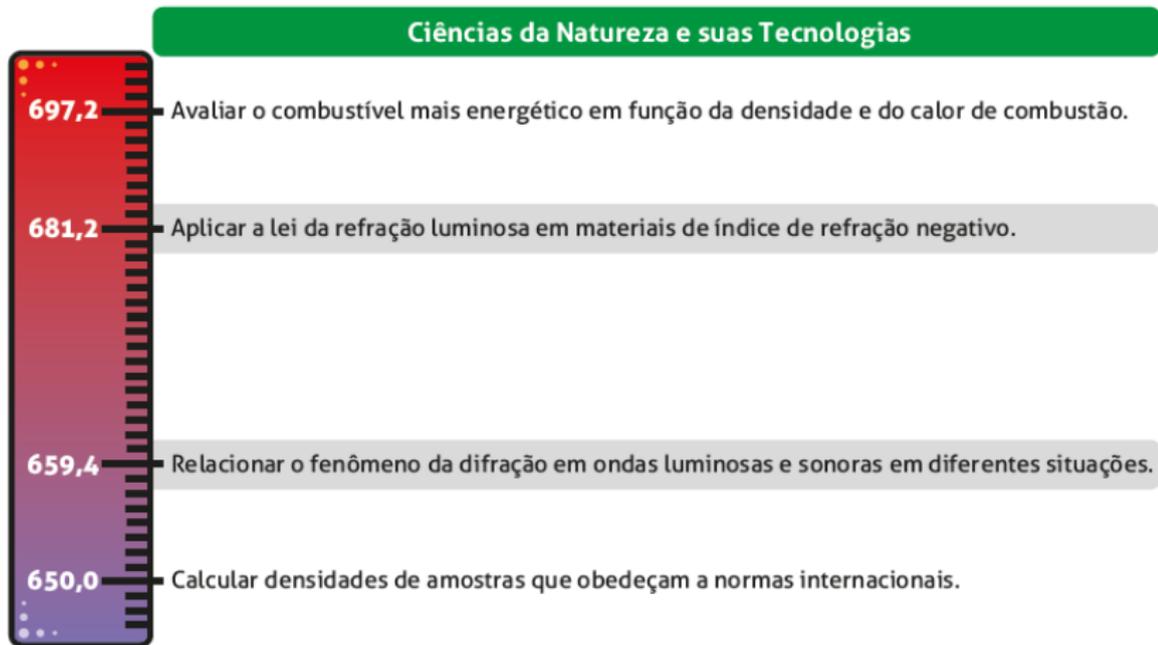


Gráfico 1 - Curva Característica do Item

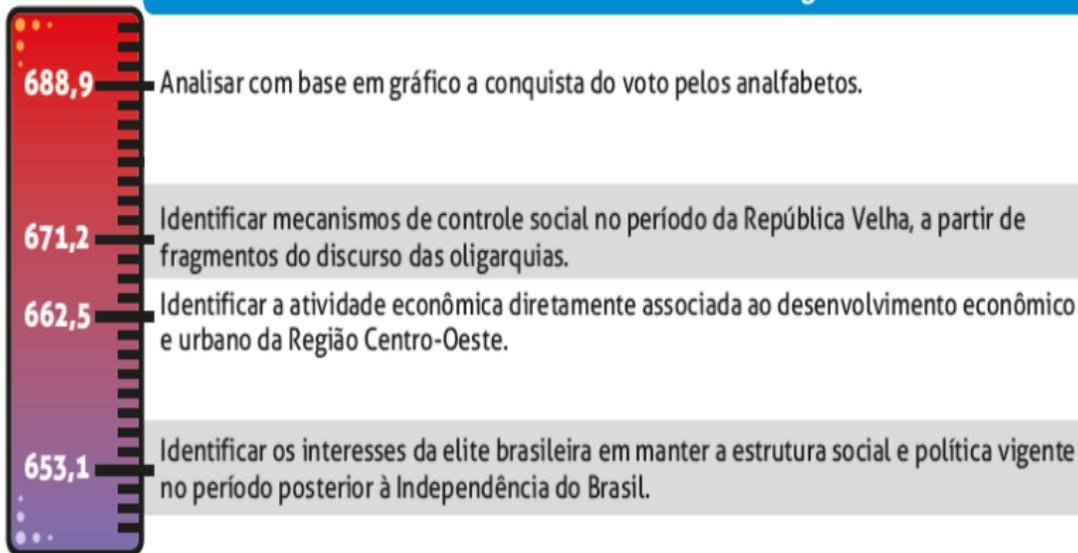
Fonte: Elaborado pela Daeb/Inep para este Guia.

Regua (Ciências da Natureza)



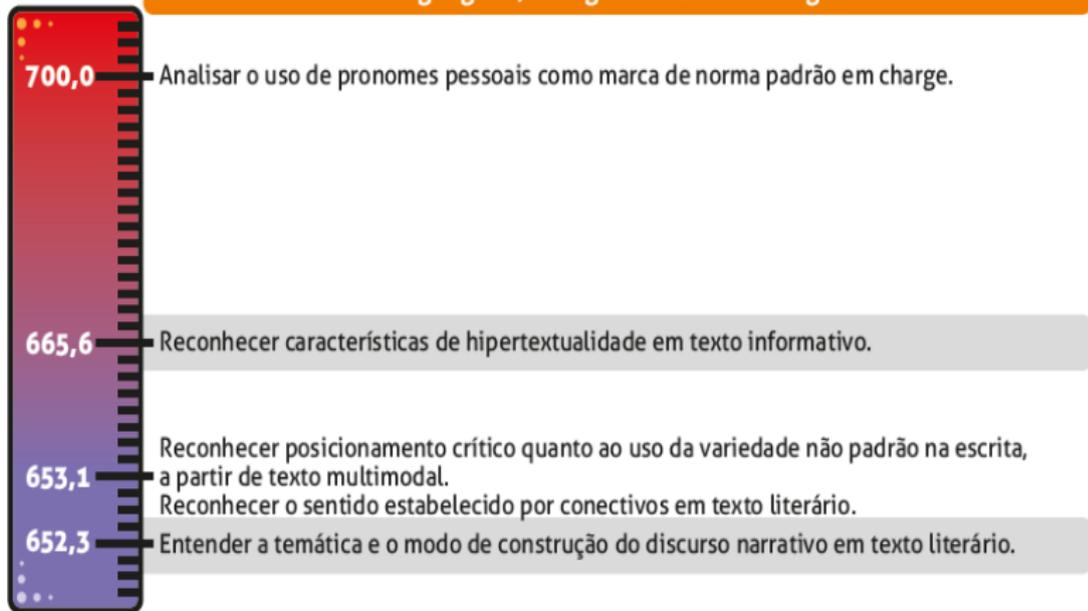
Regua (Ciências Humanas)

Ciências Humanas e suas Tecnologias



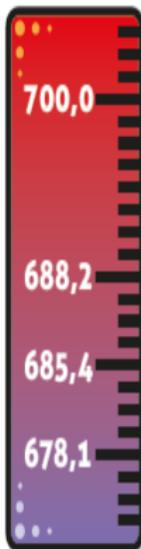
Regua (Linguagens e Códigos)

Linguagens, Códigos e suas Tecnologias



Regua (Matemática)

Matemática e suas Tecnologias



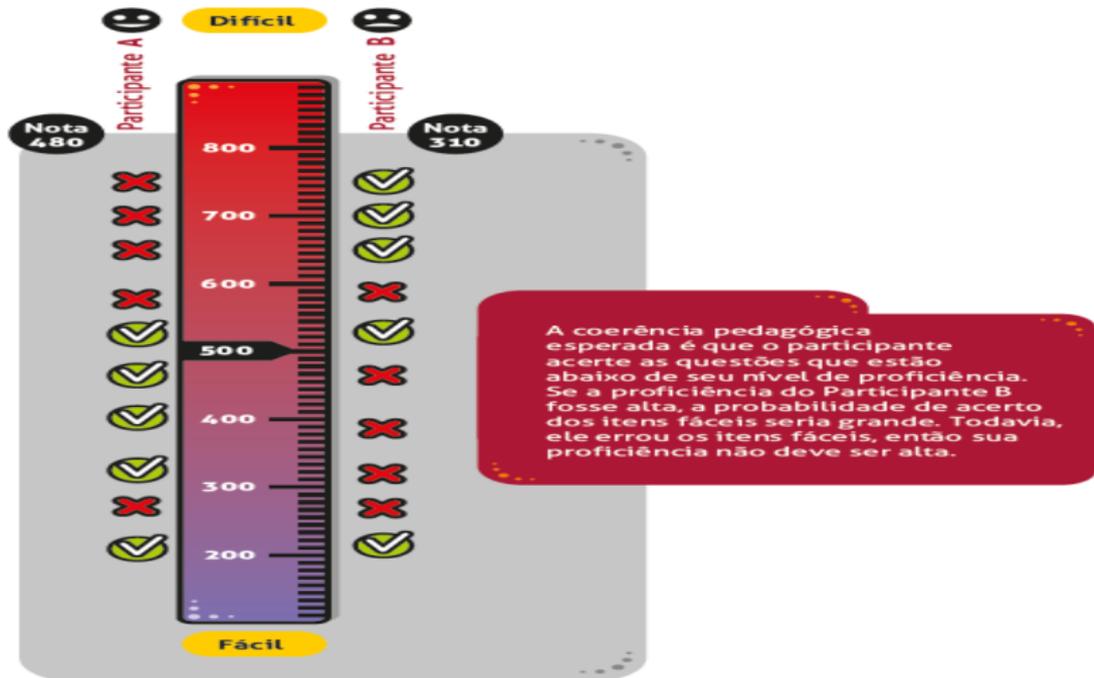
Determinar uma expressão algébrica baseada nos termos de uma sequência de figuras geométricas (quadrados), atendendo a um padrão de formação, a partir da relação entre o número de lados e o número de figuras.

688,2 Calcular o volume de material necessário para a fabricação de um objeto vazado (cubo).

685,4 Calcular a probabilidade de um evento utilizando dados contidos em tabela de dupla entrada.

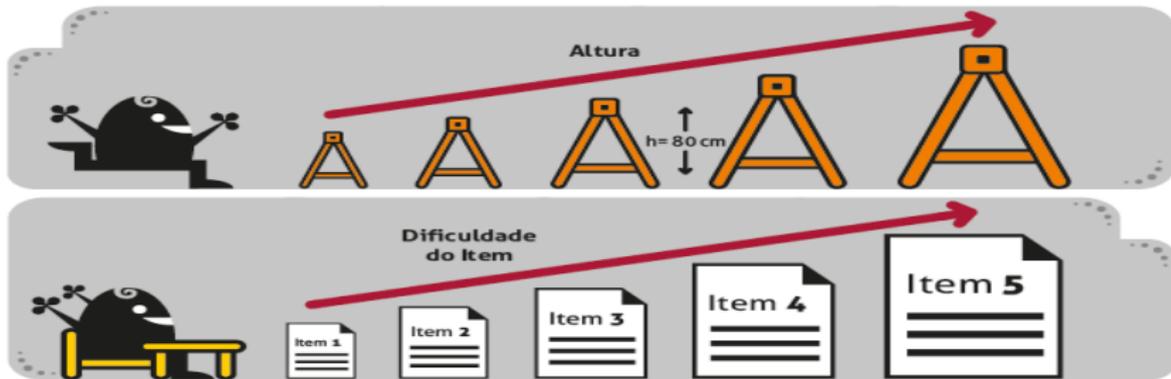
678,1 Determinar, em uma sequência de listas numéricas justapostas na forma de um empilhamento triangular, o padrão da soma de seus termos.

Padrões de resposta



Padrões de resposta (altura)

A ilustração a seguir também exemplifica a coerência esperada.



Fazendo uma analogia, imagine uma competição de corrida com barreiras cuja altura aumenta ao longo do percurso. Imagine, ainda, um atleta que durante os treinos tenha conseguido saltar barreiras de até 80 centímetros. Espera-se que esse atleta, durante a competição, consiga saltar facilmente as barreiras com altura menor que 80 centímetros e tenha dificuldade para saltar barreiras com altura superior. Da mesma forma, espera-se que um participante apresente, durante a prova, coerência pedagógica em suas respostas.

Traço latente x escore

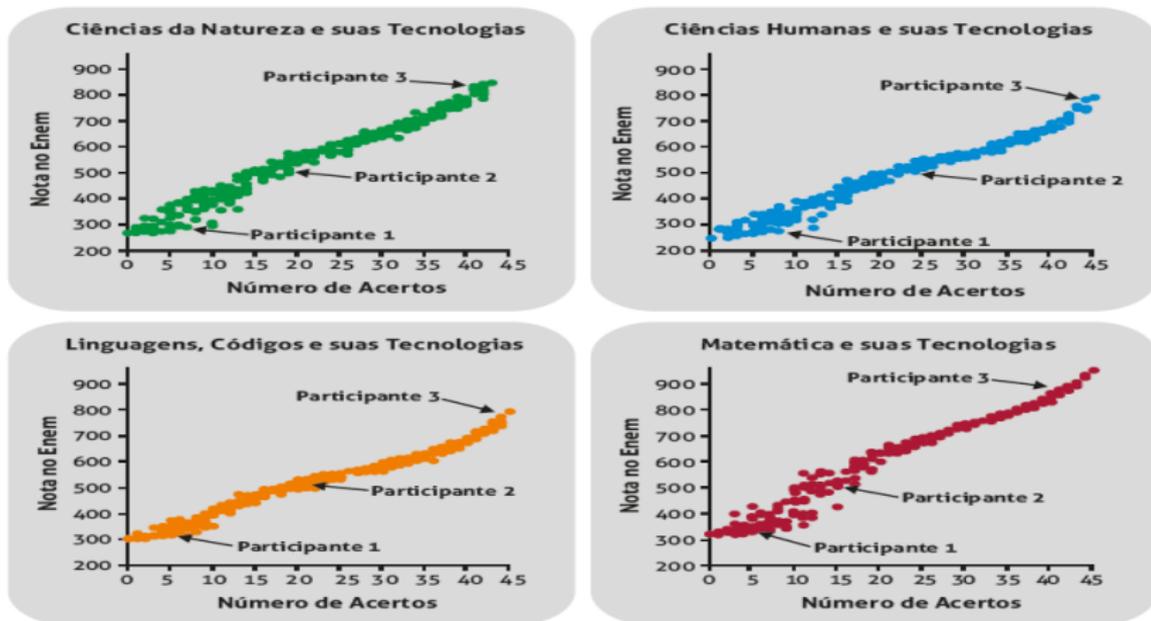


Gráfico 2 – Relação entre o número de acertos e a nota calculada pela TRI

Fonte: Elaborado pela Daeb/Inep para este Guia.

Composição da nota final



6 Composição das notas

Ao final, o participante receberá cinco notas, uma para cada área da prova objetiva e a nota da redação. Cada faculdade/universidade tem regras próprias para utilizar as notas de ingresso, colocando peso nas notas das áreas que julgue ser mais importantes para o curso que está sendo pleiteado ou fazendo uma média aritmética simples.

Referências

- Andrade, D. F., Tavares, H. R. & Valle, R. C. (2000). Teoria da Resposta ao Item: conceitos e aplicações. Associação Brasileira de Estatística
(<http://www.ime.unicamp.br/~cnaber/LivroTRI.pdf>).
- Azevedo, C. L. N. (2003). Métodos de estimação na Teoria da Resposta ao Item. Dissertação de Mestrado, IME-USP
(http://www.teses.usp.br/teses/disponiveis/45/45133/tde-05102004-163906/publico/Disser_METRI.pdf).

Referências

- Entendendo a sua nota no ENEM, INEP-MEC,
http://download.inep.gov.br/educacao_basica/enem/manuais/2012/guia-do-participante-Notas.pdf.
- Nojosa, R. T. (2001). Modelos multidimensionas na Teoria da resposta ao Item. Dissertação de Mestrado, DE-UFPE (artigo da dissertação: <http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/2193>).