

MANUAL SOBRE A CONSTRUÇÃO DA NOTA POR TEORIA DA RESPOSTA AO ITEM (TRI)



**GOVERNO DO ESTADO
DE SÃO PAULO**

SECRETARIA DE GESTÃO PÚBLICA
UNIDADE CENTRAL DE RECURSOS HUMANOS

CONCURSO DE PROMOÇÃO



**GOVERNO DO ESTADO
DE SÃO PAULO**
SECRETARIA DE GESTÃO PÚBLICA
UNIDADE CENTRAL DE RECURSOS HUMANOS

MANUAL SOBRE A CONSTRUÇÃO DA NOTA POR TEORIA DA RESPOSTA AO ITEM (TRI)

Fundação para o Vestibular da Universidade Estadual Paulista “Júlio de Mesquita Filho”

Diretora Presidente

Sheila Zambello de Pinho

Superintendente Acadêmica

Tânia Cristina Arantes Macedo de Azevedo

Superintendente de Planejamento

Carlos Augusto Araújo Valadão

Superintendente Administrativo

Ricardo Samih Georges Abi Rached

Elaboração: Fundação Vunesp

Diretoria Acadêmica

Núcleo de Avaliação Educacional

Dalton Francisco de Andrade

Heliton Ribeiro Tavares

Diagramação

Marcelo Alt

Capa

Cintia Tinti

SÃO PAULO 2014



Sumário

1. – Introdução	01
2. – Calculando as notas	02
3. – Relação entre a nota da TRI e o número de acertos.....	08
4. – Avaliações que utilizam a TRI	10
5. – Entendendo tecnicamente a nota da TRI.....	10
Referências bibliográficas	12



1. – Introdução

Este texto foi elaborado com o objetivo de descrever e esclarecer suas dúvidas sobre a obtenção da nota do Concurso, baseada na metodologia denominada *Teoria da Resposta ao Item (TRI)*. Essa metodologia tem sido adotada em diversos exames, e seu detalhamento pode ser encontrado em diversos textos complementares, tais como o livro *Teoria da Resposta ao Item: Conceitos e Aplicações* (Andrade, Tavares & Valle, 2000) e *Entenda sua nota no Enem – Guia do Participante* (Inep, 2012). Este último serviu de base para a elaboração deste manual.

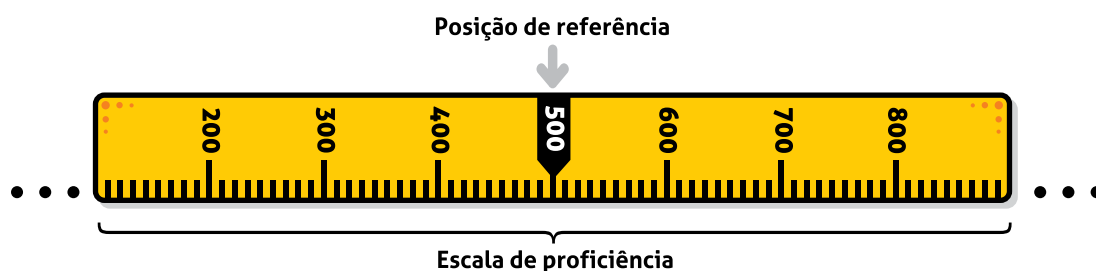
Nas avaliações convencionais, normalmente, as nossas notas são calculadas de “zero a 10” ou de “zero a 100” ou ainda de “zero” ao número de questões (também chamadas de *itens*). A forma de calcular a nota geralmente é simples, bastando somar as questões corretas na prova, podendo estas notas ser “zeros” (erros) e “uns” (acerto), ou pesos que representem o nível de dificuldade atribuído na composição da prova. Imaginemos que um professor elabora uma prova com 10 questões para avaliar o conhecimento dos seus participantes em Matemática, em que cada questão vale um ponto. Ao final da prova, o *participante A* e o *participante B* acertaram seis questões, porém não acertaram exatamente as mesmas questões. Será que os participantes deveriam receber a mesma nota? Será que os dois participantes possuem o mesmo conhecimento em Matemática?

Após entendermos a construção da nota no Concurso pela TRI, nós voltaremos a essas e outras perguntas. A nota pela TRI não é calculada levando-se em conta somente o número de questões corretas, mas também o conjunto das questões que formam a prova. Em outras palavras, também leva em conta as questões que o participante erra. Essa nota é atribuída em uma métrica, escala, criada especialmente para o Concurso, com o objetivo de medir o conhecimento (proficiência) do participante.

A Secretaria de Gestão criou uma escala que depende de dois valores:

- de posição ou de referência, para a qual foi atribuído o valor 500, que representa o desempenho médio dos participantes do Concurso;
- de dispersão, a qual foi atribuído o valor 100, que representa uma medida de variabilidade média das notas desses participantes em relação ao desempenho médio 500. Este valor é conhecido como desvio padrão, conceito da área de Estatística amplamente difundido. Vale a pena ressaltar que estes valores não têm relação com os valores de posição e de dispersão utilizados no Enem ou em qualquer outra avaliação.

A partir desses dois valores, podemos dizer que um participante com nota 600 apresenta proficiência uma unidade de desvio padrão acima da proficiência média dos participantes. Essa escala poderia ser representada graficamente por uma régua, como a apresentada a seguir, na qual as proficiências e os itens podem ser todos posicionados, e seus pontos, interpretados pedagogicamente, como veremos mais à frente.



Fonte: Entenda a sua Nota no ENEM – Guia do Participante. MEC-Inep, 2012.

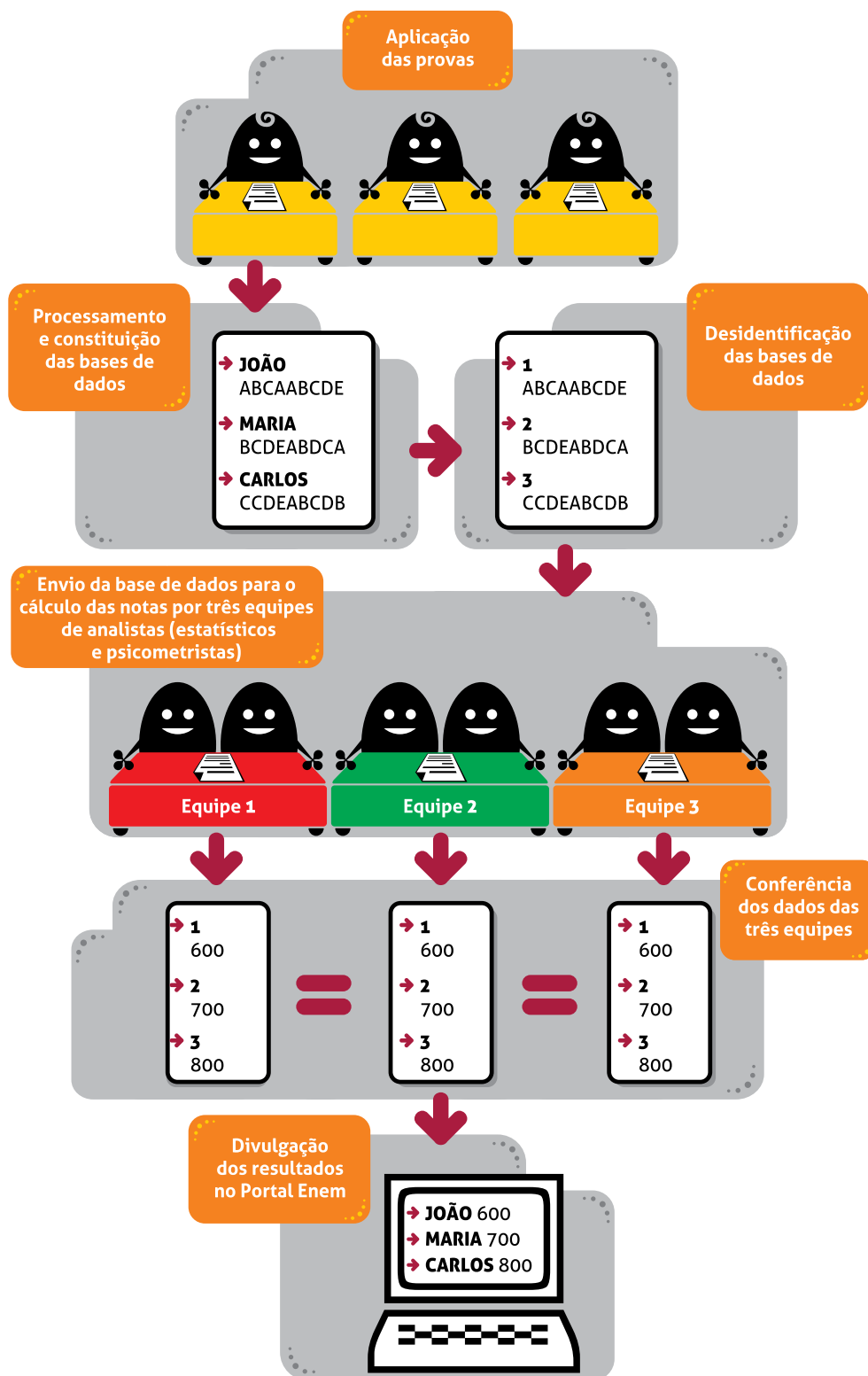
A sua nota na escala (régua) pode assumir qualquer valor no conjunto dos números reais, como por exemplo, 632,6; 421,8 etc. Outro aspecto importante no cálculo da nota é que as notas mínima (nenhum acerto) e máxima (acerto em todos os itens) dependem do grau de dificuldade da prova; então está claro que a nota mínima não terá que ser “zero”, e a nota máxima não terá que ser 1.000.

Em uma situação similar, podemos pensar em uma prova física que envolve um conjunto de obstáculos, em que a nota do participante será dada pela soma das alturas dos obstáculos. Mesmo sem conseguir ultrapassar nenhum obstáculo, a nota não seria nula. Ainda, para um conjunto de participantes de alto desempenho, o ideal seria apresentar obstáculos maiores, possibilitando a obtenção de notas mais precisas, e até superiores às aquelas obtidas com o primeiro conjunto de obstáculos.

2. – Calculando as notas

Já está claro que a nota não será calculada usando somente o número de acertos, a nota mínima não terá que ser “zero”, e a nota máxima não terá que ser 1.000, mas como é calculada a nota?

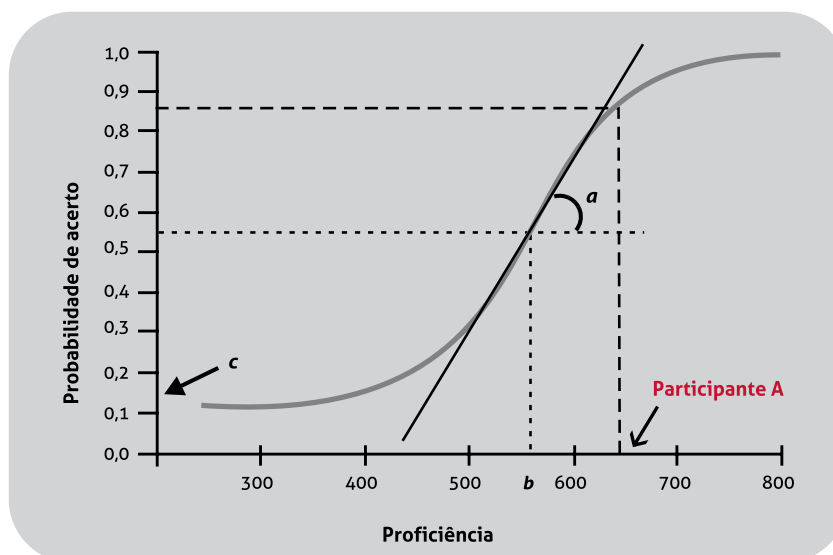
A nota do Concurso deverá responder à seguinte pergunta: na escala considerada, qual a nota que, com grande (maior) probabilidade, gerou o conjunto de respostas do participante? Ela é calculada por computador, usando o método estatístico denominado Teoria da Resposta ao Item (TRI), em que cada questão é um item, uma parte específica de um todo. Essa teoria considera para o cálculo da nota a consistência da resposta segundo o grau de dificuldade de cada questão. O cálculo da sua nota exige 6 etapas, apresentadas no Infográfico a seguir:



Fonte: Entenda a sua Nota no ENEM – Guia do Participante. MEC-Inep, 2012.

O modelo da TRI considera três informações (denominadas de *parâmetros*) essenciais para avaliar a qualidade do item e, conseqüentemente, a qualidade da medida:

- *parâmetro de discriminação (a)*: é a capacidade que cada questão possui para diferenciar participantes que sabem dos participantes que não sabem o conteúdo pedagógico da questão (item);
- *parâmetro de dificuldade (b)*: associado à dificuldade do conteúdo pedagógico da questão, quanto maior seu valor, mais difícil (exigente) é a questão. Ele é expresso na mesma escala da proficiência. Em uma prova de qualidade, devemos ter questões de diferentes níveis de dificuldade para avaliar adequadamente os participantes em todos os níveis de conhecimento;
- *parâmetro de acerto casual (c)*: em provas de múltipla escolha, um participante que não domina o conteúdo de uma determinada questão da prova pode responder corretamente a esse item por acerto casual. Com isso, esse parâmetro representa a probabilidade de um participante acertar a questão não dominando seu conteúdo pedagógico.



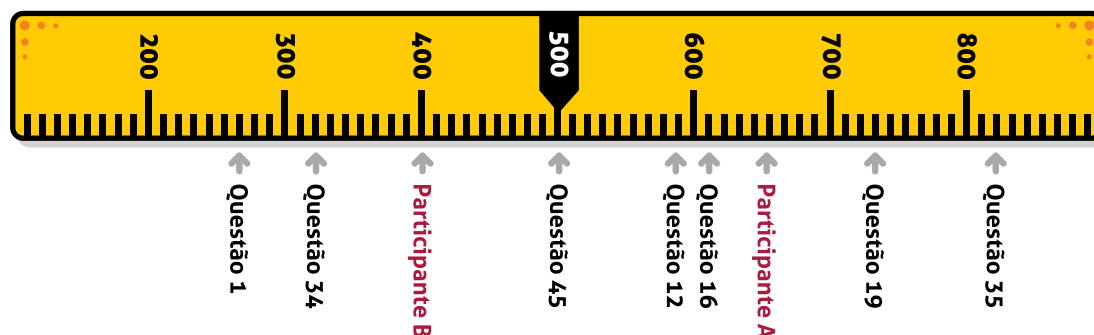
Fonte: Entenda a sua Nota no ENEM – Guia do Participante. MEC-Inep, 2012.

O gráfico mostra o perfil de uma questão em relação aos seus parâmetros. É natural pensar que participantes com desempenho muito abaixo da média tenham baixa probabilidade de acertar o item, enquanto participantes muito acima da média tenham alta probabilidade de acertar o item. Essa probabilidade é medida no intervalo de “zero” a “um”. A curva obtida pela TRI mostra a relação entre a probabilidade de acerto e a proficiência do respondente, para um particular conjunto de parâmetros dos itens. Podemos então notar que somente respondentes com proficiência acima do valor do parâmetro b , de dificuldade, é que terão alta probabilidade de responder corretamente à questão. No nosso exemplo, o participante A com proficiência em torno de 650 tem probabilidade de 0,85, aproximadamente, de responder corretamente à questão representada. Em outras palavras, espera-se que 85% dos participantes com proficiência 650 acertem essa questão, ou ainda, que o conteúdo dessa questão seja de domínio da grande maioria das pessoas que têm proficiência 650 ou mais.

A partir das respostas dos candidatos, serão obtidos (estimados) os valores dos parâmetros a , b e c dos itens (questões). A partir do conhecimento dos valores desses parâmetros, as questões são posicionadas na escala (régua), permitindo, assim, sua interpretação pedagógica. O posicionamento de cada questão nos níveis da escala (régua) se dá a partir de critérios probabilísticos que garantem que somente participantes com proficiência igual ou maior do que aquele nível possuem alta probabilidade de responder corretamente às questões que estão naquele nível e em níveis inferiores.

Questões com conteúdos pedagógicos menos exigentes serão posicionadas na parte inferior da régua e aquelas com conteúdos pedagógicos mais exigentes serão posicionadas na parte superior da régua.

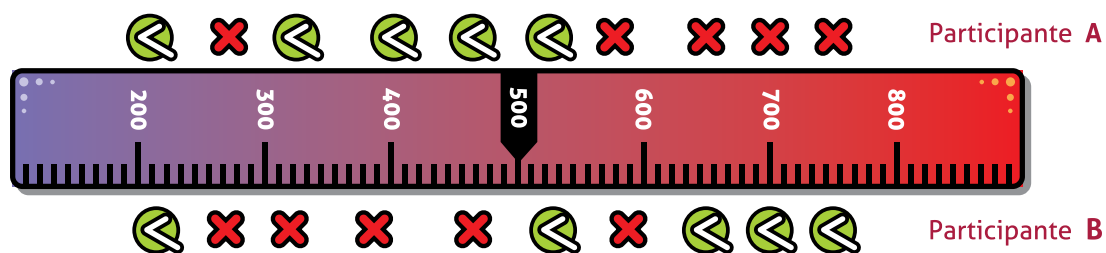
A partir da Figura a seguir, podemos dizer que o *participante A* deve dominar os conteúdos contemplados nas questões 1, 34, 45, 12 e 16, e não dominar aqueles contemplados nas questões 19 e 35. Já o *participante B*, espera-se que domine somente os conteúdos das questões 1 e 34. Com isso, podemos dizer que o *participante A* deve dominar, além dos conteúdos contemplados nas questões 1 e 34, que são de domínio do *participante B*, também aqueles das questões 12, 16 e 45; portanto, sua nota deverá ser maior.



NOTA: A escala tem uma interpretação pedagógica e as questões são posicionadas na régua (escala) de acordo com uma coerência pedagógica. O participante também será posicionado na régua, ou seja, terá sua nota conforme seu conhecimento pedagógico. Isto quer dizer que sua nota não depende das notas dos demais participantes, depende apenas do posicionamento das questões na régua.

Fonte: Entenda a sua Nota no ENEM – Guia do Participante. MEC-Inep, 2012.

No cálculo da nota, o modelo da TRI considera a coerência das respostas corretas do participante. Participantes que acertaram questões difíceis também devem acertar as questões fáceis, já que para dominar conteúdos mais difíceis, é necessário dominar os conteúdos mais fáceis. Por exemplo, na régua de Matemática, caso haja, um participante que possui conhecimento de trigonometria também possui conhecimento em equações do segundo grau – na régua, as questões de trigonometria serão posicionadas acima das questões de segundo grau. Se o participante acerta questões de trigonometria e erra questões de equações do segundo grau, não existe coerência nas respostas e, dessa forma, os acertos acrescentarão menos na nota do participante. Na régua a seguir temos as respostas a 10 questões já posicionadas, dada por dois participantes, ambos acertaram a mesma quantidade de questões, mas o *participante A* acertou as questões mais fáceis dentro de uma coerência pedagógica, logo sua nota será maior do que do *participante B*, pois como ele acertou questões mais difíceis e errou as questões mais fáceis, o modelo entende que foram acertadas no chute. Este controle é devido ao parâmetro de acerto casual.



A nota do participante depende somente do seu conhecimento no momento da prova, e dos itens respondidos, com a escala predefinida. Ela não depende do desempenho dos outros participantes e nem do desempenho médio dos participantes.



NOTA: quando dizemos que o participante acertou uma questão no “chute”, não significa que sua nota irá diminuir, mas ela não tem tanto valor como se o participante tivesse acertado os itens com coerência do conteúdo pedagógico. Então, sempre é melhor responder à questão do que deixar em branco, pois uma questão certa sempre aumenta a nota e uma questão deixada em branco será corrigida como errada.

Fonte: Entenda a sua Nota no ENEM – Guia do Participante. MEC-Inep, 2012.

Após essa explicação, nós podemos voltar a situação da prova com 10 questões na qual os *participantes A* e *B* acertaram seis questões cada. Se o *participante A* acertou as questões mais fáceis – posicionadas na parte inferior da régua – e o *participante B* acertou duas questões fáceis, duas intermediárias e duas difíceis, quem terá a maior nota? Se considerarmos o mesmo modelo matemático da TRI que é usado na prova objetiva do ENEM, o *participante A* terá maior nota por ter respondido às questões com mais coerência. Podemos afirmar que o *participante A* domina mais o conteúdo abordado na prova feita pelo elaborador do que o *participante B*.

IMPORTANTE: *para os participantes com o mesmo número de acertos na prova, terá a maior nota quem acertar as questões de forma mais coerente pedagogicamente.*

3. – Relação entre a nota da TRI e o número de acertos

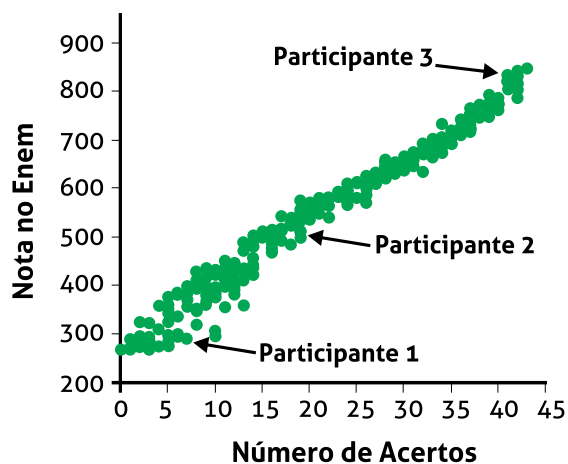
Apesar de a nota do Concurso não ser calculada diretamente pelo número de acertos, existe uma clara relação entre o número de acertos e a nota calculada pela TRI. Isto quer dizer que se um participante teve número de acertos alto, também terá nota alta no Concurso, e o participante que teve pouco acerto terá nota baixa no Concurso, notas essas relacionadas com os valores mínimo e máximo de cada prova. Para exemplificar, os gráficos a seguir mostram a relação da nota pela TRI e o número de acertos de 300 participantes que fizeram a prova do ENEM 2011. Observe que se a nota da TRI fosse exatamente o número de acertos, os pontos nos gráficos deveriam estar alinhados, representados em uma reta. As variações desses pontos em relação à reta mostram que participantes com o mesmo número de acertos podem ter notas diferentes no ENEM.

- Poderíamos nos perguntar: se existe alta relação entre o número de acertos e a nota da TRI, para que usar a nota da TRI? Não seria mais simples usar o número de acertos? Uma das vantagens de se utilizar a TRI é que a sua nota nos possibilita fazer comparações entre notas de diferentes provas de uma mesma Área/Disciplina, já o número de acertos de provas diferentes não é diretamente comparável. Com ela também podemos construir e interpretar pedagogicamente a escala de proficiência.

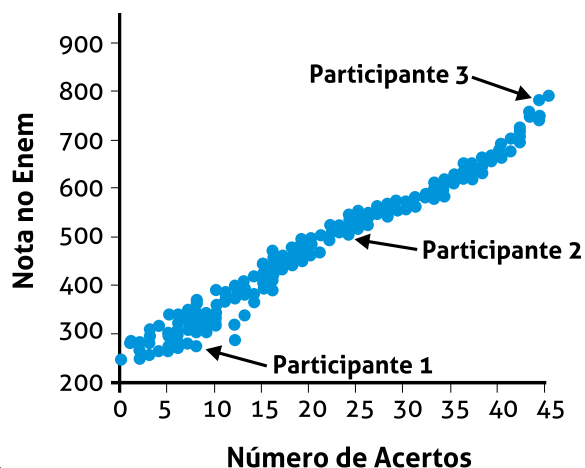
A seguir, alguns exemplos da relação entre a nota da TRI e o número de acertos nas provas de 2011.

- **Ciências da Natureza e suas tecnologias:** o **Participante 1** teve 6 acertos, e sua nota no ENEM foi 300,8; o **Participante 2** teve 19 acertos, e sua nota foi 500,5 (lembre que a nota 500 representa a nota média dos concluintes regulares do ensino médio da rede pública que realizaram o ENEM 2009), ou seja, ele acertou menos da metade da prova e seu desempenho foi equivalente ao desempenho médio dos concluintes regulares do ensino médio que realizaram o ENEM 2009; e o **Participante 3** teve 41 acertos com nota 825,0 (a nota máxima, ou seja, para quem acertou 45 questões foi de 867,2);
- **Ciências Humanas e suas tecnologias:** o **Participante 1** teve 8 acertos, e sua nota no ENEM foi 279,4; o **Participante 2** teve 24 acertos, mais da metade das questões, e sua nota foi 507,0; e o **Participante 3** teve 44 acertos, e sua nota foi 784,4 (a nota máxima, ou seja, para quem acertou 45 questões foi de 793,1);
- **Linguagens, códigos e suas tecnologias:** o **Participante 1** teve 5 acertos, e sua nota no ENEM foi 316,8; o **Participante 2** teve 21 acertos, e sua nota foi 498,9; e o **Participante 3** teve 44 acertos, e sua nota foi 756,1 (a nota máxima, ou seja, para quem acertou 45 questões foi de 795,5);
- **Matemática e suas tecnologias:** o **Participante 1** teve 5 acertos, e sua nota no ENEM foi 330,6; o **Participante 2** teve 15 acertos, e sua nota foi 504,0, como já relatado anteriormente, a área de Matemática exige menor conhecimento dos **Participantes** para que a nota alcance 500, isso está sendo evidenciado pelo **Participante 2**, que acertou somente 15 questões e sua nota foi maior que 500; e o **Participante 3** teve 41 acertos, e sua nota foi 878,7 (a nota máxima, ou seja, para quem acertou 45 questões foi de 953,0);

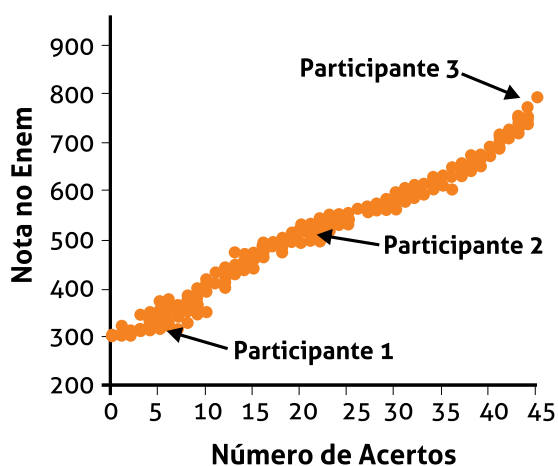
Ciências da Natureza e suas Tecnologias



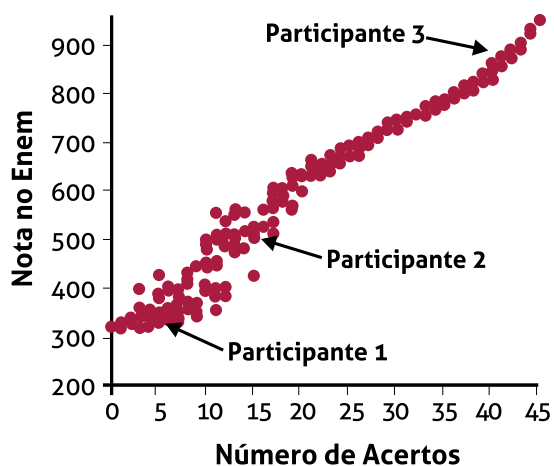
Ciências Humanas e suas Tecnologias



Linguagens, Códigos e suas Tecnologias



Matemática e suas Tecnologias



Fonte: Entenda a sua Nota no ENEM – Guia do Participante. MEC-Inep, 2012.

4. – Avaliações que utilizam a TRI

O uso da TRI em Avaliações Educacionais teve início no Brasil com o Sistema de Avaliação da Educação Básica (SAEB), em 1995, e, posteriormente, foi implementado também no Exame Nacional para Certificação de Competências de Jovens e Adultos (ENCCEJA) e, por último, no ENEM. No âmbito internacional, a TRI vem sendo utilizada largamente por diversos países: Estados Unidos, França, Holanda, Coreia do Sul, China, sem falar nos países participantes do PISA (Programa Internacional de Avaliação de Estudantes).

Um dos grandes exemplos de avaliação utilizando a TRI é o exame de proficiência em língua inglesa (TOEFL). Outro exame bastante importante e mais semelhante ao ENEM é o SAT (*Scholastic Aptitude Test* ou *Scholastic Assessment Test*). Este é um Exame Educacional padronizado dos Estados Unidos, aplicado a estudantes do Ensino Médio, que serve de critério para admissão nas universidades norte-americanas.

5. – Entendendo tecnicamente a nota da TRI

Apesar de alguns comentários já terem sido feitos anteriormente, destacaremos novamente as principais características do modelo de probabilidade da TRI para o cálculo da nota do Concurso.

A TRI modela a probabilidade de um participante responder corretamente a um item (questão) como função dos parâmetros do item e da proficiência do participante. Essa relação é expressa por meio de uma função monotônica crescente que indica que quanto maior o conhecimento do participante, maior será a sua probabilidade de acertar o item (ver, por exemplo, Andrade, Tavares & Valle, 2000; Baker & Kim, 2004; Hambleton, Swaminathan & Rogers, 1991; Klein, 2003; Pasquali, 1997). Normalmente, utiliza-se a função logística de 3 parâmetros desenvolvida por Birnbaum (1968), em que a probabilidade P de uma resposta correta do participante j ao item i ($u_{ji} = 1$), em função do parâmetro de proficiência θ_j e dos parâmetros do item a_i , b_i e c_i , é dada por

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

Para obter a nota do participante do Concurso basta estimar (encontrar) o valor do parâmetro θ que representa a nota do participante, levando em consideração o vetor de respostas às questões do participante e os parâmetros dos itens que compõem a prova. Como não é possível obter esse valor diretamente, através de uma fórmula simples, é preciso utilizar um método matemático mais sofisticado para calcular a nota. Existem, na literatura, alguns métodos para o cálculo da nota. Como exemplo, o ENEM e o SAEB utilizam o método denominado *Expected a Posteriori* (EAP). Este também será o método utilizado no Concurso.

O método EAP tem como princípio usar uma função de probabilidade *a priori*. Essa função tem a informação da nota média e do desvio padrão dos participantes. Com essa informação, as notas são calculadas a partir da fórmula apresentada a seguir:

$$E(\theta | \mathbf{u}) = \frac{\int_{\mathbf{R}} \theta L(\mathbf{u} | \theta) f(\theta) d\theta}{\int_{\mathbf{R}} L(\mathbf{u} | \theta) f(\theta) d\theta}.$$

Nota-se que $f(\theta)$ é a função de probabilidade *a priori*, $L(\mathbf{u} | \theta)$ é uma função matemática associada ao padrão de respostas dos participantes e aos parâmetros dos itens, valores conhecidos, e θ representa a proficiência do participante que será calculada. Para obter a nota do participante, é necessário resolver a expressão acima, que exige um processo de integração.

Na matemática, a resolução de integrais nem sempre é tão simples, depende da complexidade da função que está sendo usada. Para contornar esse problema, Stroud & Secrest (1966) sugerem criar intervalos para funções complexas e resolver numericamente cada intervalo dessa função – esse método é conhecido como quadratura Gaussiana – e a expressão pode ser resolvida por meio da soma desses intervalos usando a expressão a seguir:

$$\hat{\theta} = \frac{\sum_{k=1}^q X_k L(X_k) f(X_k)}{\sum_{k=1}^q L(X_k) f(X_k)}.$$

Essa é a fórmula de onde será obtida a nota pela TRI dos participantes do Concurso, representada nela por $\hat{\theta}$. Detalhes adicionais sobre as expressões apresentadas nesta seção podem ser consultadas em referências complementares.

Referências bibliográficas

- Andrade, D. F. de, Tavares, H. R. & Valle, R. da C. (2000). *Teoria da Resposta ao Item: conceitos e aplicações*. São Paulo: ABE – Associação Brasileira de Estatística.
- Baker, F. B. & Kim, S. (2004). *Item response theory: parameter estimation techniques*. Nova York: Marcel Dekker.
- Birnbaum, A. (1968). *Some latent trait models and their models and their use in inferring an examinee's ability*. In F. M. Lord and M.R. Novick (Eds.), *Statistical theories of mental test scores* (pp 397-479). Reading, MA: Addison-Wesley
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications.
- Klein, R. (2003). *Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB). Ensaio: Avaliação e Políticas Públicas em Educação, Rio de Janeiro, v.11, n.40, p. 283-296*.
- Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh, 61A, 273-287*.
- Lord, F. M. (1952). A theory of test scores. *Psychometric Monograph, No. 7*.
- Brasil. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Entenda a sua nota no Enem – Guia do Participante. Brasília: Ministério da Educação, 2012.
- Pasquali, L. (1997). *Psicometria: teoria e aplicações*. Brasília: Editora Universidade de Brasília.
- Stroud, A.H, Secrest, D. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs. NJ: Prentice-Hall.
- Yen, W. M. & Fitzpatrick, A. R. (2006). Item Response Theory. In R. L. Brennan, *Educational Measurement* (111-153). American Council on Education/Oryx Press Series on Higher Education.



