

# MANUAL

FGV PROJETOS

SOBRE A ATRIBUIÇÃO DE NOTAS  
EM PROVAS DE MÚLTIPLA ESCOLHA  
UTILIZANDO A

# TRI

TEORIA DA RESPOSTA

AO ITEM (TRI)



# APRESENTAÇÃO



O objetivo deste manual é esclarecer possíveis dúvidas do candidato em relação à forma como é calculada a sua nota, baseada na metodologia denominada Teoria da Resposta ao Item (TRI). Essa é uma metodologia cada vez mais empregada em certificações e avaliações em grande escala, como ENEM, PISA e TOEFL.

Diferentemente das avaliações convencionais, na TRI, a nota do candidato não é calculada levando-se em conta somente o número de questões corretas, mas também o conjunto de acertos nas questões que formam a prova. As notas são atribuídas em uma escala que tem como objetivo medir o grau de conhecimento (ou proficiência) do candidato em determinada área. A TRI ainda permite comparar o desempenho de participantes em provas e anos diferentes; assim, podemos acompanhar o progresso da proficiência dos candidatos em determinada área do conhecimento. Essas e outras características importantes dessa metodologia serão discutidas a seguir.

Este documento traz uma breve introdução sobre a TRI, destacando suas principais particularidades quando comparada às avaliações tradicionais. São discutidas ainda, a forma de cálculo da nota pela TRI, bem como o processo de construção da escala de proficiência. Por fim, são apresentadas algumas perguntas e respostas que tratam das dúvidas mais frequentes sobre a TRI.



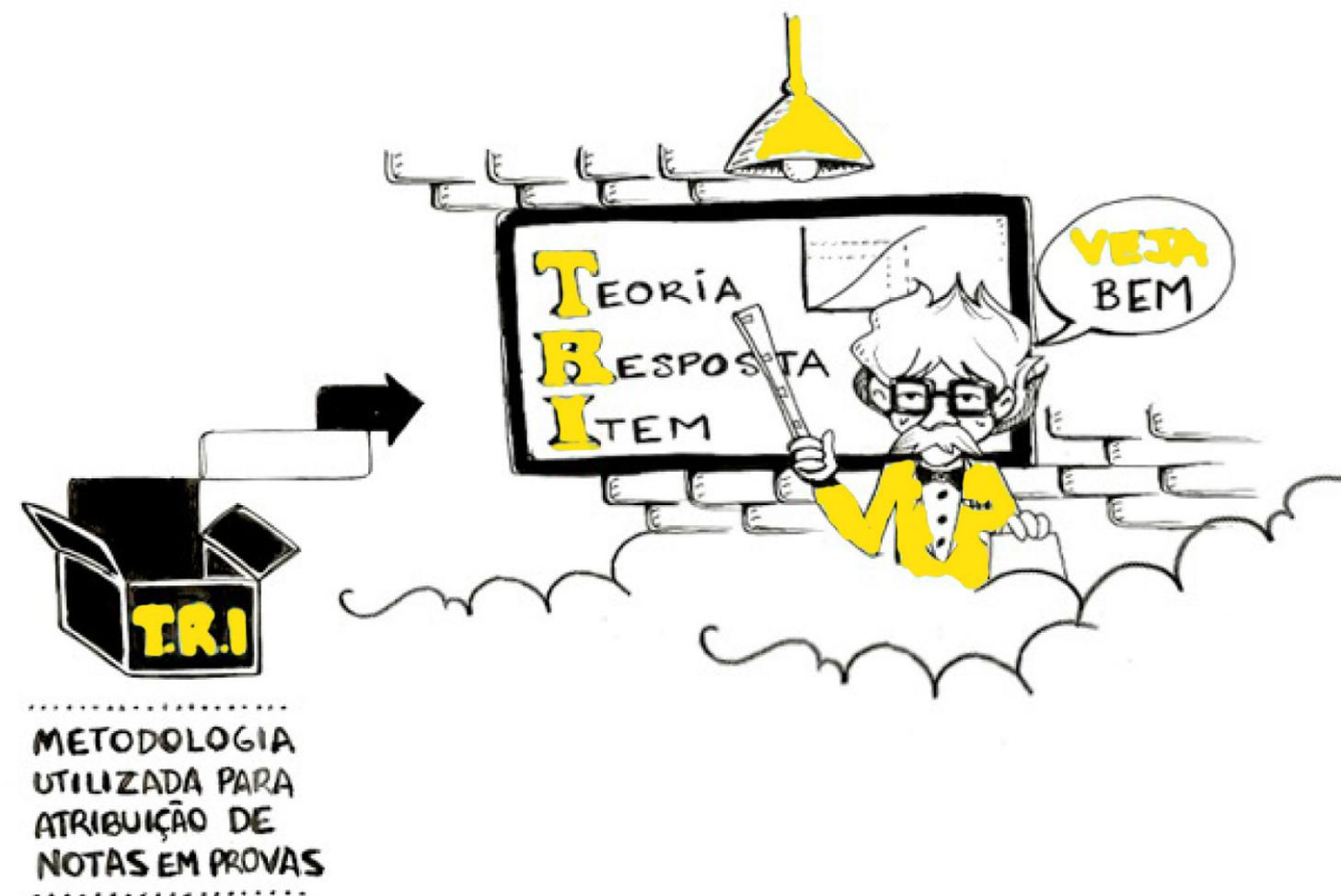
# INTRODUÇÃO



O objetivo de avaliações como a da certificação é medir o grau de conhecimento do candidato em determinadas áreas (como, por exemplo, Matemática, Língua Portuguesa ou Geografia). Não existe uma forma direta de aferir este conhecimento, como podemos, por exemplo, medir a temperatura do ambiente através de um termômetro. Diferentemente disso, o grau de conhecimento (proficiência) do candidato é estimado por meio das respostas dadas por ele em uma prova. Essa estimativa é a sua nota.

Nas avaliações tradicionais, as notas dos candidatos são calculadas baseando-se unicamente no número de questões respondidas corretamente. Dois candidatos que tenham acertado exatamente o mesmo número de questões terão a mesma nota, independentemente de quais questões cada um deles tenha acertado. Nesse método, não é levada em conta a dificuldade de cada questão nem a coerência dos acertos (e erros).

A Teoria da Resposta ao Item (TRI) emprega um método diferente para estimar a proficiência em uma determinada área do conhecimento. Essa metodologia vai além da análise de cada resposta separadamente, levando em conta todas as respostas de uma só vez. Assim, a nota obtida por um candidato não depende apenas do número de acertos, mas também da coerência desses acertos.



# EXEMPLO

## QUESTIONÁRIO DA TEMPERATURA



Para ilustrar as ideias principais que estão por trás da TRI, vamos fazer um paralelo com a seguinte situação: desejamos estimar a temperatura em uma cena externa de certo filme; obviamente, não podemos usar um termômetro. Assim como no caso de uma prova, há uma quantidade desconhecida que se deseja medir, mas não podemos fazer essa medição diretamente. Na falta do termômetro, as respostas ao questionário ao lado dariam uma boa ideia da temperatura.

Nesse exemplo, queremos usar as respostas do questionário para estimar a temperatura. Nas provas de conhecimento, queremos usar as respostas do candidato para estimar seu grau de conhecimento (proficiência).

Algumas perguntas do questionário da temperatura são claramente informativas de temperaturas mais baixas, como as de números 1, 2, 6 e 7. Outras são informativas de temperaturas mais altas, como as perguntas 4, 5, 8, 9 e 10. Já a pergunta 3 é pouco informativa: é possível haver temperaturas altas e baixas com uma resposta positiva (ou negativa), pois existem dias ensolarados frios e quentes.

1. Há pessoas usando casacos?
2. Há pessoas usando luvas?
3. O sol está brilhando, sem nuvens?
4. Há pessoas usando camisas de manga curta?
5. Há homens andando sem camisa?
6. Há vestígios de geada no chão ou no teto dos carros?
7. Quando as pessoas falam, aparece uma nuvenzinha? O hálito condensa?
8. Há pessoas tomando sorvete?
9. Há carros circulando com a janela aberta?
10. Há pessoas se abanando?

# EXEMPLO

## QUESTIONÁRIO DA TEMPERATURA



Na TRI, as questões de uma prova são submetidas a uma análise estatística para identificar se elas são mais informativas de proficiências altas ou baixas. Isso está associado ao nível de dificuldade da questão, como será visto mais adiante.

Voltando ao questionário da temperatura, para algumas perguntas, esperaríamos certa coerência nas respostas. Por exemplo, se a resposta à pergunta 2 é “sim”, esperaríamos também “sim” à pergunta 1. De modo análogo, se a resposta à pergunta 5 é “sim”, esperaríamos resposta “não” às perguntas 1 e 2.

A TRI leva em consideração a coerência dos acertos (e erros) dos candidatos. Espera-se que um candidato que acerta as questões mais difíceis também acerte questões mais fáceis.

No caso da temperatura, existem algumas escalas alternativas para mensurá-la (Fahrenheit, Celsius, Kelvin, etc.). Ainda que a escolha de uma determinada escala possa ter um componente cultural arbitrário (países latinos têm a tendência de usar Celsius, cientistas têm a tendência de usar Kelvin, etc.), essas escalas são todas equivalentes e existem fórmulas para converter uma escala em outra.

No caso de provas de conhecimento, também é possível construir uma escala-padrão para que seja possível comparar o desempenho de candidatos em anos diferentes, que fizeram provas diferentes – embora a metodologia tradicional de avaliação não faça isso.

Na avaliação tradicional, a escala de notas é própria de cada prova, geralmente estabelecendo nota zero para quem erra todas as questões e dez para quem acerta todas. Se dois candidatos obtiveram a mesma nota em Matemática na avaliação tradicional, mas fizeram provas diferentes, não é possível afirmar que eles têm o mesmo nível de conhecimento em Matemática, pois uma prova pode ter sido mais fácil do que a outra.

Já na TRI, é construída uma escala de medida da proficiência, que pode ser usada para medir (e comparar) as proficiências de candidatos que fizeram provas diferentes, desde que sejam garantidas algumas exigências técnicas para esse fim.

# NOTA

COMO A NOTA É CALCULADA PELA TRI?



O ponto de partida para calcular a nota dos candidatos na TRI é a análise de cada questão. Para cada questão são atribuídos, usando métodos estatísticos, três números, chamados de parâmetros *a*, *b* e *c*.

Esses parâmetros permitem calcular a probabilidade de um candidato com uma determinada proficiência (que representaremos por *N*) acertar a questão. Os três parâmetros (*a*, *b*, *c*) determinados para cada questão são descritos a seguir:

# NOTA

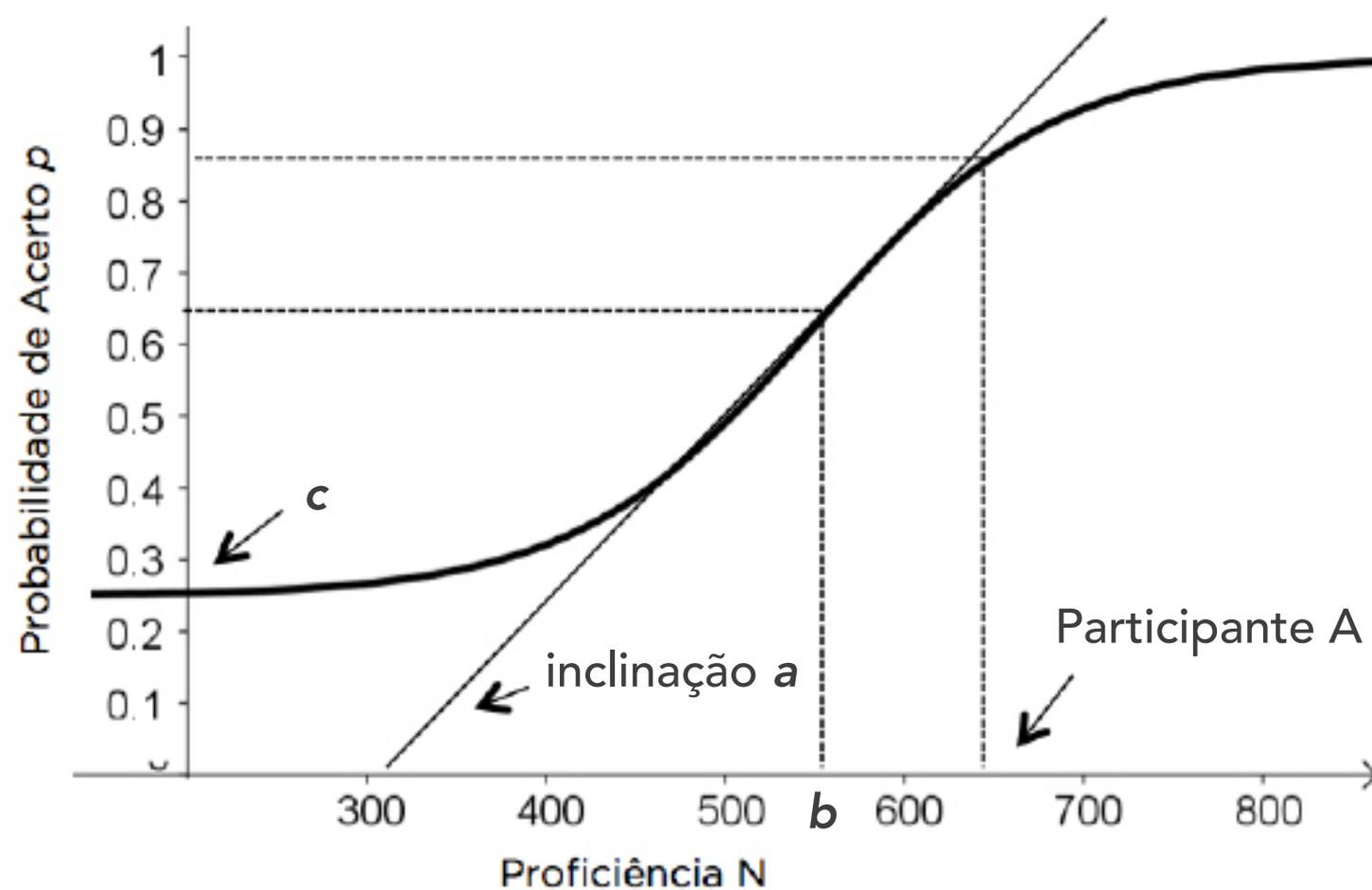
## COMO A NOTA É CALCULADA PELA TRI?

- Discriminação ( $a$ ) – determina a capacidade da questão de precisar um nível de proficiência tal que candidatos com proficiência acima desse nível têm alta probabilidade de acertar a questão, e candidatos com proficiência abaixo desse nível têm baixa probabilidade de acertá-la. Valores mais altos de  $a$  caracterizam uma melhor discriminação e, portanto, uma melhor qualidade da questão. Como visto no exemplo da temperatura, a pergunta 3 “O sol está brilhando, sem nuvens?” tem baixa discriminação, pois a probabilidade de o sol estar brilhando aumenta muito lentamente com o aumento da temperatura (ou seja, há uma probabilidade razoável de o sol estar brilhando em um dia frio, bem como de ele não estar brilhando em um dia quente).
- Dificuldade ( $b$ ) – mede o grau de dificuldade da questão; é expresso na mesma escala das proficiências atribuídas aos candidatos. Isso significa que candidatos com proficiência  $N$  maior do que  $b$  tendem mais a acertar do que errar a questão, e candidatos com proficiência  $N$  menor do que  $b$  tendem mais a errar do que a acertar essa questão. Portanto, quanto maior o valor de  $b$ , maior o nível de conhecimento exigido do candidato para responder corretamente. Uma prova com qualidade deve conter questões com variados graus de dificuldade, para que seja possível avaliar candidatos com proficiências baixas, médias e altas.
- Acerto casual ( $c$ ) – representa a probabilidade de um candidato com proficiência muito baixa responder corretamente a essa questão. Em provas de múltipla escolha, pode ocorrer de um candidato acertar alguma questão sem saber como resolvê-la, por pura sorte, e isso é considerado no modelo estatístico. Como representa uma probabilidade, seus valores ficam entre 0 e 1.



# NOTA

COMO A NOTA É CALCULADA PELA TRI?



Os valores dos parâmetros  $a$ ,  $b$  e  $c$  determinam a curva característica da questão. A figura à esquerda mostra um exemplo típico.<sup>1</sup> Os valores dos parâmetros  $a$ ,  $b$  e  $c$  podem ser observados, bem como a influência deles na qualidade da questão. Observe que, no mesmo eixo horizontal, estão posicionados os valores do parâmetro  $b$  (dificuldade da questão) e da proficiência (próxima de 650) de um hipotético candidato (Participante A), cuja probabilidade de acerto para essa questão pode ser lida na figura como sendo aproximadamente 0,85 (85%). Repare que a probabilidade do Participante A acertar a questão é alta, pois a proficiência dele é maior do que a dificuldade ( $b$ ) da questão.

Você pode fazer a leitura no gráfico da probabilidade de acerto dessa questão para outros valores da proficiência. Por exemplo, um candidato com proficiência próxima de 400 tem probabilidade de acerto de aproximadamente 0,3 (30%).

<sup>1</sup> A expressão que define a curva característica de uma questão é  $p = c + \frac{1-c}{1+e^{-a(N-b)}}$ .

# NOTA

## COMO A NOTA É CALCULADA PELA TRI?

A ideia geral da TRI é que, uma vez calibradas as questões (isto é, uma vez conhecidos os seus parâmetros), determina-se, para cada candidato, a partir do conjunto de suas respostas, o valor da proficiência  $N$  que melhor explica a ocorrência dos acertos e dos erros que ele obteve, isto é, estima-se a proficiência mais provável do candidato, tendo em vista as questões que ele acertou e as questões que ele errou. Esse procedimento, assim como a calibração das questões, exige o uso de métodos matemáticos e computacionais sofisticados, que não serão abordados neste manual. O leitor interessado poderá encontrar descrições detalhadas desses métodos nas referências citadas no final do manual.

O processo de estimação das proficiências citado leva em conta a coerência dos acertos (e erros) do candidato no conjunto de questões que compõem a prova. Por exemplo:

- É mais provável que indivíduos que acertaram apenas as questões mais fáceis e erraram as mais difíceis tenham uma proficiência baixa.
- É mais provável que indivíduos que acertaram tanto as questões mais fáceis quanto questões difíceis tenham uma proficiência mais alta.
- É mais provável que indivíduos que acertaram questões difíceis e erraram questões fáceis tenham uma proficiência baixa (devido aos acertos casuais).



# NOTA

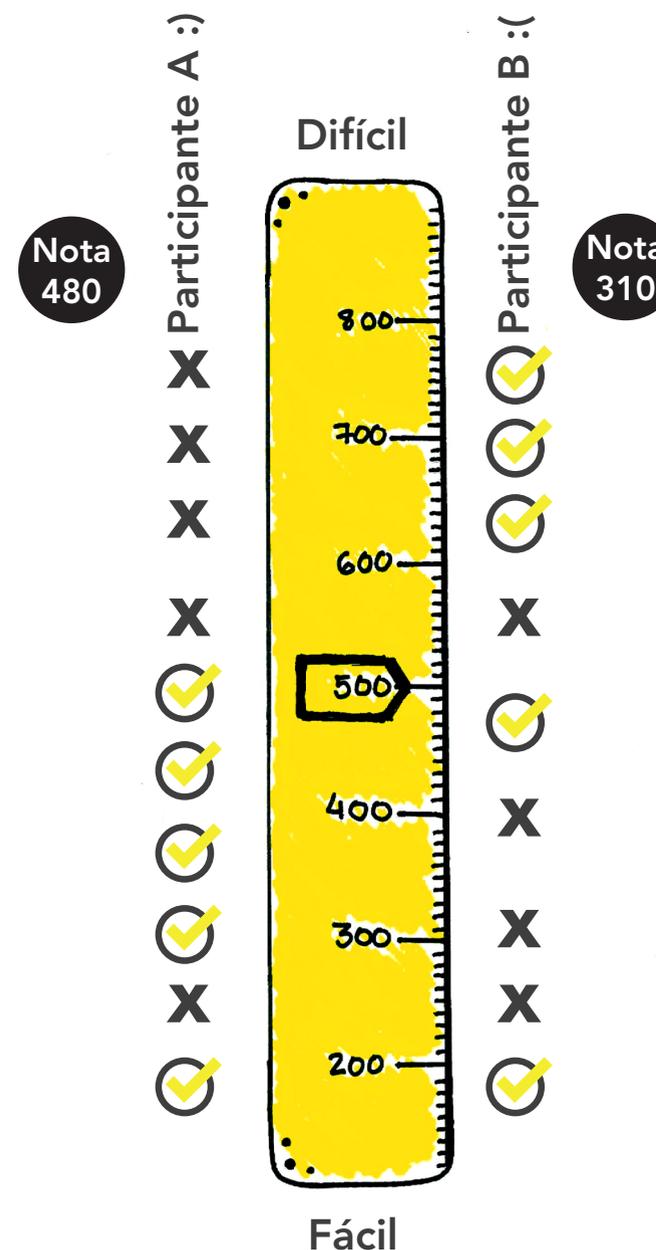
## COMO A NOTA É CALCULADA PELA TRI?

A figura ao lado ilustra a situação de dois candidatos (Participante A e Participante B) que, numa prova de 10 questões, tiveram a mesma quantidade de acertos. Entretanto, é possível observar que os acertos (e erros) do Participante A são mais coerentes do que os acertos (e erros) do Participante B.

O Participante A acertou a maioria das questões mais fáceis e errou as questões mais difíceis. Já o Participante B acertou a maioria das questões difíceis e errou a maioria das questões fáceis (acertos casuais).

A coerência esperada é que o participante acerte as questões que estão abaixo de seu nível de proficiência.

Se a proficiência do participante B fosse alta, a probabilidade de acerto dos itens fáceis seria grande. Todavia, ele errou os itens fáceis, então sua proficiência não deve ser alta.



Fonte: Entenda sua nota no ENEM – Guia do Participante MEC – Inep 2012 (adaptado).

# PROFICIÊNCIAS

## A ESCALA DE PROFICIÊNCIAS



A escala de proficiências é uma forma de estabelecer uma unidade-padrão de medida do conhecimento. Os itens da prova são posicionados nessa escala, assim como os participantes. A coerência esperada é que um participante com proficiência  $N$  domine os conteúdos das questões com dificuldade  $b$  menor do que  $N$  (acertando a maioria delas) e não domine os conteúdos das questões com dificuldade  $b$  maior do que  $N$  (errando a maioria delas).

É importante ressaltar, entretanto, que toda questão respondida corretamente contribui para aumentar a nota do candidato. O que acontece é que, quando ele acerta uma questão difícil tendo errado questões mais fáceis, a contribuição desse acerto é menor. Portanto, não deixe questões sem responder, pois questões deixadas em branco serão consideradas como erradas. O acerto, mesmo que casual, sempre contribui para aumentar a nota do candidato.

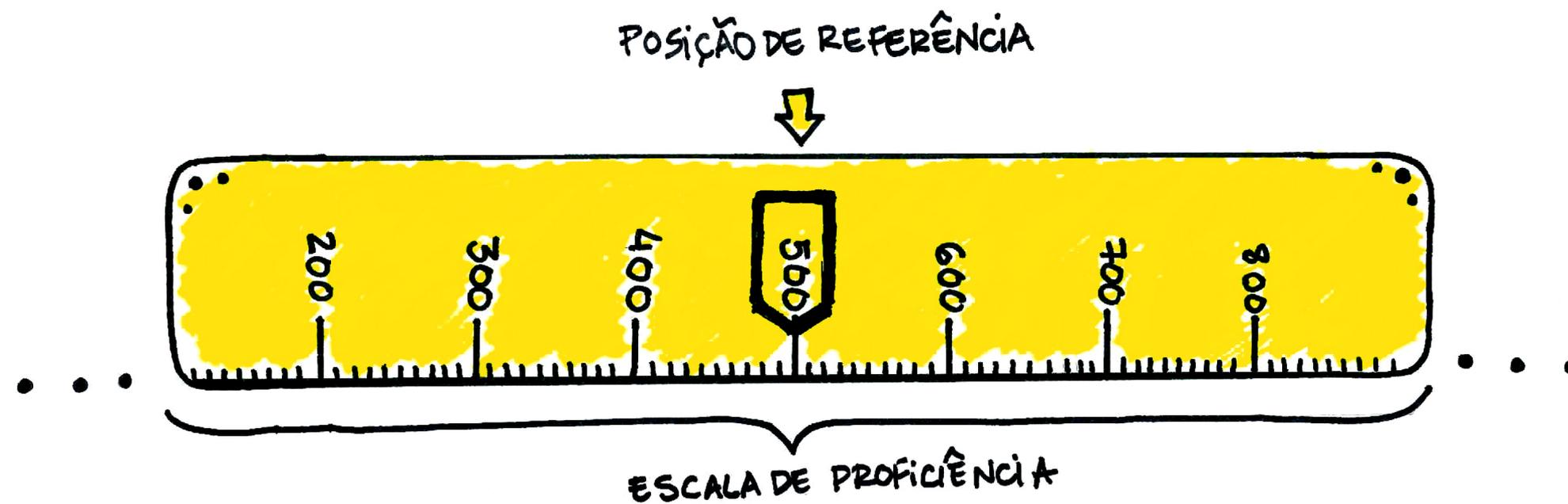
A escala que a TRI utiliza para a medição das proficiências dos candidatos é construída a partir de dois valores:

- Valor de referência – representa o desempenho médio dos candidatos que participam do processo de construção da escala, isto é, no nosso caso, os candidatos que participarem do processo de certificação para Gerente de Organização Escolar (GOE) 2014. Esse valor de referência pode ser escolhido arbitrariamente. Para a Certificação GOE, será utilizado o valor 500.
- Valor de dispersão – representa uma medida da variabilidade (o desvio padrão) do desempenho dos candidatos em relação ao valor de referência. Esse valor de dispersão também pode ser escolhido arbitrariamente. Para a Certificação GOE, será utilizado o valor 100.

Esse par de valores (500,100) que será utilizado na construção das nossas escalas de medida é o mesmo utilizado, por exemplo, pelo ENEM. Assim, um candidato que recebeu nota 400 em uma prova teve um desempenho um desvio padrão abaixo do valor de referência (desempenho médio dos candidatos que participaram do processo de construção da escala); já um candidato com nota 650 teve um desempenho um desvio padrão e meio acima do valor de referência.

# PROFICIÊNCIAS

A ESCALA DE PROFICIÊNCIAS



Fonte: Entenda sua nota no ENEM – Guia do Participante MEC – Inep 2012 (adaptado).

# DÚVIDAS FREQUENTES



## QUAIS SÃO AS NOTAS MÍNIMA E MÁXIMA EM UMA PROVA AVALIADA COM A TRI?

A escala de notas na TRI não é construída com base no percentual de acertos, com a nota mínima (normalmente 0) correspondendo a 0 acertos e a máxima (normalmente 10 ou 100) correspondendo ao acerto em todas as questões. A escala da TRI é construída de modo que a média e o desvio padrão das notas tenham valores fixados (normalmente em 500 e 100). As notas mínima e máxima dependem do modo como o desempenho dos candidatos é distribuído. Os valores para as notas mínima e máxima ficam normalmente em torno de 200 e de 800, respectivamente.

## COMO A NOTA PELA TRI SE RELACIONA COM O NÚMERO DE ACERTOS NA PROVA?

A nota pela TRI não depende apenas do número de acertos, mas de quão consistentes foram esses acertos. Candidatos com o mesmo número de acertos têm, em geral, notas diferentes pela TRI (a menos que tenham acertado exatamente as mesmas questões). Também é possível que um candidato com mais acertos que outro tenha nota menor pela TRI. No entanto, se um candidato acertar todas as questões que outro acertar e mais algumas, sua nota pela TRI será necessariamente maior do que a do outro candidato.

## PROVAS CORRIGIDAS COM A TRI SÃO MAIS FÁCEIS OU MAIS DIFÍCEIS DO QUE AS CONVENCIONAIS?

Não há nenhuma relação entre o nível de dificuldade da prova e o uso da TRI. A metodologia afeta somente a forma pela qual a nota do candidato é calculada. Como em qualquer prova, é desejável que a prova contenha questões de todos os níveis de dificuldade.

## OUVI FALAR DE CANDIDATOS QUE, NA AVALIAÇÃO PELA TRI, ACERTARAM MAIS QUESTÕES EM UMA DISCIPLINA DO QUE NOUTRA, MAS QUE AS NOTAS PELA TRI SE COMPORTARAM DE MANEIRA INVERSA. ISSO PODE MESMO OCORRER?

Sim. Como a nota pela TRI é ajustada de modo que a média e o desvio padrão sejam fixados, se a prova de uma disciplina é fácil para os candidatos, então será necessária uma quantidade grande de acertos para se atingir um determinado ponto da escala. Assim, é possível que um número menor de acertos em uma prova difícil leve a uma nota pela TRI maior do que um número maior de acertos em uma prova fácil.

# DÚVIDAS FREQUENTES



**É VERDADE QUE, NA TRI, AS QUESTÕES MAIS DIFÍCEIS VALEM MAIS?**

Não. Na verdade, não há um valor (ou peso) específico para cada questão. O valor maior ou menor que a questão tem para cada candidato depende do restante de seus acertos. Para um candidato que acertou muitas questões, acertar uma questão difícil vale mais do que para um candidato que errou muitas questões (o modelo matemático tem uma tendência a interpretar, para este último candidato, o acerto como casual, valorizando-o menos, assim). Inversamente, acertar uma questão fácil vale mais para um candidato que acertou poucas questões.

**QUE ESTRATÉGIA DEVO USAR PARA FAZER UMA PROVA QUE VAI SER CORRIGIDA COM A TRI?**

Para o candidato, não há nenhuma estratégia especial a ser usada pelo fato de a correção ser feita pela TRI. Como em qualquer prova, é recomendável começar pelas questões mais fáceis para o candidato, deixando aquelas em que ele tem mais dificuldade para uma segunda passagem pela prova.

**A CORREÇÃO PELA TRI PUNE O “CHUTE”? NO CASO DE DÚVIDA EM UMA QUESTÃO, É MELHOR DEIXÁ-LA EM BRANCO?**

Não, não há nenhuma vantagem em se deixar em branco uma questão. O modelo da TRI usualmente empregado em avaliações trata igualmente questões em branco e questões erradas. Desse modo, convém responder a todas as questões, mesmo àquelas que o candidato não sabe responder ou não tem certeza da resposta.

**É VERDADE QUE, COM A TRI, É POSSÍVEL COMPARAR O DESEMPENHO DE DOIS CANDIDATOS SUBMETIDOS A PROVAS DE NÍVEIS DE DIFICULDADE DIFERENTES DE UMA MESMA DISCIPLINA?**

Sim, essa é uma das vantagens mais importantes de utilizar a TRI em uma avaliação. Notas convencionais, baseadas em percentual de acertos, não podem ser comparadas quando os candidatos são submetidos a provas diferentes. Mas o modelo matemático usado pela TRI torna isso possível, desde que as provas tenham alguma parte em comum, ou contenham algumas questões aplicadas conjuntamente a um outro grupo de candidatos.

# REFERÊNCIAS BIBLIOGRÁFICAS



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

AZEVEDO, Caio L. N. e GAMERMAN, Dani. *Introdução à Teoria de Resposta ao Item*, 1º Congresso Brasileiro de Teoria de Resposta ao Item, Florianópolis, 2009.

BAKER, F. B.; KIM, S. *Item Response Theory: parameter estimation techniques*. 2. ed. revised and expanded. New York: Marcel Dekker, 2004.

ENEM – Guia do Participante – MEC – Inep 2012.

MACHADO, Ledo V. *Avaliação de Larga Escala e Proficiência Matemática*. Universidade Federal do Rio de Janeiro, Dissertação de Mestrado – 2010.

VAN DER LINDEN, W. J.; HAMBLETON, R. K. *Handbook of Modern Item Response Theory*. New York: Springer-Verlag, 1997.

# MAIS INFORMAÇÕES



**FGVPROJETOS.FGV.BR**

<http://fgvprojetos.fgv.br/certificacao/goe>

# BOA PROVA!

 **FGV PROJETOS**



**CERTIFICAÇÃO  
PARA GERENTE DE  
ORGANIZAÇÃO  
ESCOLAR**

