



MINISTÉRIO DA EDUCAÇÃO
INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA
DIRETORIA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA – DAEB

NOTA TÉCNICA

Assunto: Procedimento de cálculo das notas do Enem.

O Exame Nacional do Ensino Médio (Enem) é uma avaliação cujos resultados podem ser utilizados para: (1) compor a avaliação de medição da qualidade do Ensino Médio no país; (2) a implementação de políticas públicas; (3) a criação de referência nacional para o aperfeiçoamento dos currículos do Ensino Médio e (4) o desenvolvimento de estudos e indicadores sobre a educação brasileira. Neste exame, busca-se aferir as competências e habilidades desenvolvidas pelas estudantes ao fim da escolaridade básica. Esta aferição é realizada por meio de uma redação e de provas objetivas que avaliam quatro áreas do conhecimento: Linguagens, Códigos e suas Tecnologias, Ciências Humanas e suas Tecnologias, Ciências da Natureza e suas Tecnologias e Matemática e suas Tecnologias. Neste documento, pretende-se explicar como é o procedimento de cálculo das notas dos participantes do Enem.

A correção das notas de redação

Em 2011, o Inep publicou em seu site o documento intitulado “A Redação do Enem”. O referido documento objetiva esclarecer aos participantes do Enem o que é avaliado durante a correção das redações, como são atribuídas as pontuações e como é o procedimento de correção.

O que é avaliado durante a correção da redação tem como base a matriz de referência do exame que foi instituída em 2009 e também foi publicada no edital de 2011. As matrizes desenvolvidas pelo Inep são estruturadas a partir de competências e habilidades que se espera que os participantes do exame tenham desenvolvido ao longo de sua formação. No caso da redação, cinco competências são avaliadas:

- I- Demonstrar domínio da norma padrão da língua escrita.
- II- 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.

III- Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista.

IV- Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação.

V- Elaborar proposta de solução para o problema abordado, respeitando os valores humanos e considerando a diversidade sociocultural.

A avaliação da redação é realizada de forma holística, por dois corretores independentes, tendo como base essas cinco competências. Para cada competência, o corretor precisa atribuir uma pontuação que varia de 0 a 1000. Depois, somam-se as notas atribuídas nas cinco competências e divide-se por cinco, obtendo-se a média. O resultado é a nota do primeiro corretor. O mesmo procedimento é realizado para obter a nota do segundo corretor. A nota final é a média aritmética das duas notas obtidas. No caso de discrepância igual ou maior a 300 pontos entre as notas atribuídas pelos dois corretores, haverá outra correção por um professor supervisor. Essa terceira nota é a que prevalecerá. A terceira correção configura-se como um recurso de ofício.

No documento mencionado, são descritos os níveis esperados em cada competência e a respectiva nota atribuída. Para uma melhor compreensão do processo de correção, recomenda-se a leitura do material.

A correção das provas objetivas

Para cada participante do Enem são calculadas quatro proficiências. Cada uma delas baseada nas respostas dadas aos 45 itens de cada prova. O procedimento de correção das provas objetivas inicia-se com a leitura dos cartões-resposta para a produção de uma base de dados. Vários procedimentos de controle são realizados pelo consórcio contratado com vistas a garantir que a leitura dos cartões seja fidedigna. Após estes procedimentos, a base de dados é consolidada e encaminhada para a equipe de análise de dados.

Na primeira etapa da análise é realizada uma avaliação do comportamento de todos os itens da prova, os parâmetros já conhecidos dos itens são reavaliados e são estimados os parâmetros daqueles itens que ainda não tinham parâmetros conhecidos. Apenas depois da consolidação destes parâmetros é que se inicia a fase de estimação das proficiências dos participantes.

Os procedimentos de análise dos itens e de cálculo das proficiências no Enem têm como base a Teoria de Resposta ao Item (TRI). Os conceitos básicos da teoria psicométrica baseada no item tiveram início com o trabalho de Lawley (1943) e foram posteriormente complementados com os trabalhos desenvolvidos por Lord (1952). Todavia, o crescimento do uso da teoria somente veio a ocorrer nas décadas de 70 e 80 com o desenvolvimento de *softwares* que permitiram a implementação dos modelos matemáticos relacionados à TRI. Na década de 90, houve uma expansão no uso da TRI em testes de avaliação educacional e, atualmente, a maioria dos programas de avaliação em larga escala no mundo tem como base esta teoria (Yen & Fitzpatrick, 2006).

A TRI não é uma teoria que busca substituir a Teoria Clássica dos Testes (TCT). Pelo contrário, é importante que se busque utilizar os avanços oferecidos em cada uma delas. A TRI é considerada a forma mais avançada de se mensurar um traço-latente (no caso, conhecimento). No Enem, o cálculo da proficiência a partir do uso da TRI permite acrescentar outros aspectos além do quantitativo de acertos, tais como os parâmetros dos itens e o padrão de resposta do participante. Assim, duas pessoas com a mesma quantidade de acertos na prova são avaliadas de forma distintas a depender de quais itens estão certos e errados e podem, assim, ter habilidades diferentes.

No Enem, foi estabelecida uma escala em 2009 para possibilitar o acompanhamento e comparação do desempenho dos participantes ao longo dos anos. Da mesma forma que se convencionou a escala “metro” para mensuração do comprimento, estabeleceu-se para esta comparação, a escala “Enem”. Nessa escala, foram considerados os concluintes regulares de 2009 como grupo de referência, definindo-se a média desse grupo como 500 e seu desvio-padrão 100.

Por exigir um domínio de matemática e estatística avançado, não é simples explicar como são realizados os cálculos dos parâmetros e das proficiências. Todavia, para aqueles que querem se aprofundar no assunto, há uma vasta literatura da área: Andrade, Tavares e Valle (2000); Baker e Kim (2004); Hambleton, Swaminathan e Rogers (1991); Klein (2003) e Pasquali (1997).

A TRI modela a probabilidade de um indivíduo responder corretamente a um item como função dos parâmetros do item e da proficiência (habilidade) do respondente. Essa relação é expressa por meio de uma função monotônica crescente que indica que quanto maior a proficiência do avaliado, maior será a sua probabilidade de acertar o item (ver, por exemplo, Andrade & cols, 2000; Baker & Kim, 2004; Hambleton & cols, 1991; Klein, 2003; Pasquali, 1997). **Estão erradas as explicações que tratam os parâmetros dos itens como pesos** e supõem que a divulgação dos “pesos” possibilitaria a reprodução dos cálculos a partir apenas de ponderações.

No Enem, a função monotônica é uma função logística de 3 parâmetros desenvolvida por Birbaum (1968). De acordo com esse modelo, três características do item são consideradas para cálculo da proficiência do aluno: poder de discriminação (parâmetro a), dificuldade (parâmetro b) e a probabilidade de acerto ao acaso (parâmetro c). Assim, **a estimação da proficiência está relacionada ao número de acertos, aos parâmetros dos itens e ao padrão de respostas**. Apesar de não ser simples e exigir estimativas dos parâmetros realizada por métodos estatísticos avançados, o cálculo da proficiência é **objetivo**, e participantes com exatamente o mesmo padrão de respostas apresentam exatamente as mesmas proficiências. O método utilizado para os cálculos das proficiências é denominado *Expected a Posteriori* (EAP).

Buscaremos esclarecer as três principais confusões relacionadas à interpretação dos resultados obtidos pela TRI e dos obtidos pela Teoria Clássica cuja nota está relacionada apenas ao percentual de acertos:

- **Mínimo e Máximo:** as proficiências na TRI são estimadas em uma escala métrica que não possui mínimo e máximo pré-estabelecidos. Esses valores variam de acordo com as características dos itens que compõem a prova de cada edição do Exame. Dessa forma, ao acertar todos os itens da prova não significa ter uma proficiência igual a 1000, como é comum imaginar. Para ilustrar este conceito,

suponha que um teste de resistência com várias etapas seja elaborado para um grupo de pessoas e que a etapa mais fácil seja uma corrida de 100 metros e a mais difícil uma corrida de dois quilômetros. A avaliação para uma pessoa que não conseguiu cumprir com a tarefa mais fácil é de que ela não é capaz de correr 100 metros, mas não é possível inferir que ela seja incapaz de correr (nota zero), pois no teste não havia etapas que avaliassem se ela era capaz de correr menos de 100 metros. Da mesma forma, a única afirmação que se pode fazer a respeito de uma pessoa que conseguiu cumprir todas as etapas é que ela é capaz de correr dois quilômetros. No entanto, se existisse nesta mesma avaliação uma corrida de três quilômetros, esta mesma pessoa que correu os dois quilômetros poderia correr ou não, os três quilômetros. Percebe-se, portanto, que a avaliação sofre influência das etapas programadas para o teste. O mesmo ocorre com a avaliação do conhecimento, por isso os valores de mínimo e máximo são diferentes a cada avaliação. Os valores de mínimo e máximo representam o mínimo e o máximo que o teste pode avaliar. Assim, uma pessoa que erra todas as questões recebe o valor mínimo do teste, e não uma nota zero, pois não se pode afirmar, a partir do teste, que ela possui zero conhecimento.

- **Comparação do número de acertos com a proficiência.** Há uma correspondência entre o número de acertos e a proficiência, mas a proficiência não é o percentual de itens acertados. Outros fatores são considerados nas análises, tais como os parâmetros dos itens acertados e o padrão de resposta. Inclusive, de acordo com a teoria, pessoas com o mesmo percentual de acertos podem obter proficiências diferentes, a depender do padrão de resposta. Voltando à ilustração do teste de resistência anterior, não é coerente uma pessoa que consegue correr 600 metros não concluir uma tarefa de 300 metros; por isso, a performance avaliada será distinta a depender das tarefas concluídas.
- **Comparação das proficiências obtidas em cada escala de proficiência.** A partir das explicações anteriores é fácil compreender porque cada área do conhecimento possui uma escala própria e não é possível comparar as proficiências entre as escalas. Também não é correto inferir que mais acertos em uma área do conhecimento significam ter uma proficiência mais alta nesta área. Comparar as proficiências obtidas em escalas distintas não é correto, pois seria equivalente a dizer que uma pessoa que consegue correr 100 metros é igual a uma pessoa que consegue nadar 100 metros.

Considerando que o cálculo das proficiências de acordo com a TRI exige um conhecimento avançado de estatística e a utilização de *software* próprio, o Inep, com o objetivo de ter a máxima confiança nos resultados, exige que os cálculos sejam realizados de forma independente por três grupos distintos (especialistas do Cespe/UnB, especialistas da Cesgranrio e os especialistas do Inep). Este procedimento de tripla conferência garante a qualidade dos resultados do Enem. Todos os profissionais com larga experiência na área e com formação em estatística, matemática ou psicometria. Somente com 100% de concordância entre os resultados obtidos pelos três grupos para cada participante é que o resultado é divulgado.

Este documento teve o objetivo de explicitar que a TRI tem bases científicas e que ela garante uma avaliação do conhecimento do participante de forma mais justa do que a Teoria Clássica. Não é simples explicitar os detalhes dos cálculos devido à exigência de conhecimento mais avançado em matemática e estatística. Todavia, o Inep tem adotado mecanismos para garantir um alto grau de confiabilidade nos resultados divulgados.

Brasília, dezembro de 2011.

Camila Akemi Karino
Coordenadora-geral de Instrumentos e Medidas - Inep

Maria Tereza Serrano Barbosa
Diretora de Avaliação da Educação Básica - Inep

REFERÊNCIAS

- Andrade, D. F. de, Tavares, H. R. & Valle, R. da C. (2000). *Teoria de 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.
- Birbaum, A. (1968). *Some latent trait models and their models ant 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.
- Pasquali, L. (1997). *Psicometria: teoria e aplicações*. Brasília: Editora Universidade de Brasília.
- 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.