Os procedimentos de análise dos itens e de cálculo das proficiências no Enem e no Saeb 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 trabalhados 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 é considerada a forma mais avançada de se mensurar um traço-latente (no caso, conhecimento). No Enem e no Saeb, o cálculo da proficiência a partir do uso da TRI permite acrescentar outros aspectos além do quantitativo de acertos (utilizado na Teoria Clássica dos Testes, TCT), 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. Assim, a estimação da proficiência está relacionada ao número de acertos, aos parâmetros dos itens e ao padrão de respostas. Entretanto, os parâmetros dos itens não podem ser interpretados como pesos, e está errada a inferência de que a divulgação dos parâmetros possibilitaria a reprodução dos cálculos a partir de simples ponderações.

Todas as etapas de análise dos dados de respostas dos participantes e cálculo de suas notas são realizadas de forma independente entre as equipes do Inep e o consórcio aplicador, no caso de 2019, com o Enem, as empresas Cesgranrio e Fundação Getúlio Vargas, e no Saeb somente a Cesgranrio. Realiza-se conferências ao final de cada etapa de análise objetivando verificar se os resultados obtidos internamente pelas equipes do Inep e, em segunda instância, os resultados produzidos pelas equipes técnicas do consórcio estão idênticos. Considera-se uma etapa concluída apenas após serem constatados que os resultados obtidos pelos grupos (Inep e empresas contratadas) são os mesmos. Os resultados são divulgados apenas quando há 100% de concordância nos valores obtidos pelas equipes técnicas. Essa conferência com o Consórcio juntamente ao rigor científico dos métodos estatísticos utilizados garantem a confiabilidade dos resultados.

Para cada participante do Enem são calculadas quatro proficiências, uma para cada área de conhecimento. No Saeb, são apenas duas proficiências (língua portuguesa e matemática), para estudantes do quinto, nono e terceiro ano do ensino médio. 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.

A equipe de análise psicométrica tem acesso às informações das provas, tais como estrutura e itens, apenas após sua aplicação e divulgação dos gabaritos. De posse das estruturas dos testes, inicia-se a montagem dos arquivos que serão utilizados nas análises da TRI no programa Bilog-MG.

**Análise por meio da TRI**

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.

O modelo utilizado no Enem e no Saeb para itens de múltipla escolha é o logístico de 3 parâmetros definido por:

Em que:

xj é a resposta ao item j,

xj = 1 se a resposta é correta

xj = 0 se a resposta não é correta

aj, onde aj > 0, é o parâmetro de inclinação do item j, também chamado de parâmetro de discriminação do item;

bj é o parâmetro de posição do item, também chamado de parâmetro de dificuldade;

cj, onde 0 < cj < 1, é o parâmetro da assíntota inferior do item j, refletindo as chances de um examinando de proficiência muito baixa selecionar a opção correta; e

θi é o traço latente do indivíduo

Primeiramente, a estimação dos parâmetros dos itens, também chamada de calibração, é realizada. Os parâmetros utilizados são:

a) parâmetro de discriminação: é o poder de discriminação que cada questão possui para diferenciar os participantes que dominam dos participantes que não dominam a habilidade avaliada naquela questão (item);

b) parâmetro de dificuldade: associado à dificuldade da habilidade avaliada na questão, quanto maior seu valor, mais difícil 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;

c) parâmetro de acerto casual: em provas de múltipla escolha, um participante que não domina a habilidade avaliada em uma determinada questão da prova pode responder corretamente a esse item por acerto casual. Assim, esse parâmetro representa a probabilidade de um participante acertar a questão não dominando a habilidade exigida.

O BILOG-MG implementa uma extensão da TRI para múltiplos grupos de examinandos não equivalentes, o que permite estimar os parâmetros de itens de uma mesma área, tanto no Enem quanto no Saeb, conjuntamente e em uma mesma escala (por exemplo, vários pré-testes do Enem e a aplicação do Enem daquela edição, diferentes aplicações do SAEB e pré-testes do SAEB). Para a estimação dos parâmetros, utiliza-se o procedimento de estimação pela máxima verossimilhança marginal.

O primeiro passo para análise via TRI se dá pela construção dos arquivos que serão utilizados no software Bilog-MG, sendo eles: arquivos de bases de dados, arquivo de programação, arquivo de parâmetros, a também o arquivo de gabaritos. Todos os arquivos são construídos de modo padronizado, com normas previamente estabelecidas, separadamente pelas equipes de análise, conferidas e alinhadas, posteriormente, via teleconferência.

As bases de dados com as respostas dos examinandos sem identificação de respondentes são preparadas em formato apropriado para utilização do software e inclui, não só as respostas dos examinandos da aplicação atual, como também respostas de grupos de respondentes de pré-testes.

Assim como ocorre com relação aos arquivos de bases de dados, o arquivo de programação, extensão .blm, é preparado contemplando informações do grupo de itens do exame/avaliação atual, bem como de pré-testes, de forma a garantir que os resultados calculados estejam em uma mesma escala, característica que permite sua comparabilidade entre testes diferentes (Prova 1, Prova de reaplicação/ Pessoas Privadas de Liberdade (PPL ),no caso do Enem; entre séries, no caso do SAEB) e entre anos de aplicações.

Os arquivos com parâmetros dos itens (extensão .prm) são preparados, em primeira instância, apenas com aqueles que já foram previamente estimados, e em segunda etapa, com os parâmetros de todos os itens calibrados.

A primeira etapa é a de calibração dos itens que ainda não apresentam parâmetros pré-estimados. Calculam-se os parâmetros a, b e c, todos na mesma escala dos demais itens com características pré-estimadas. Verifica-se também o ajuste dos dados ao modelo e o funcionamento diferencial dos itens (DIF) dos itens comuns aos pré-testes.

Um item apresenta DIF se respondentes de mesma proficiência em dois grupos diferentes apresentam um comportamento diferente de resposta para aquele item. Em termos de análise, considera-se que há DIF quando o máximo das diferenças absolutas entre as proporções esperadas de resposta (obtidas a partir do arquivo *.EXP* do software Bilog-MG) para os dois grupos excede 0,15 dentro do intervalo compreendido entre os percentis 5 e 95. Tanto no Enem quanto no Saeb, verifica-se se um item estimado no pré-teste apresenta o mesmo comportamento de parâmetros que esse mesmo item aplicado no exame/avaliação. Itens com comportamentos semelhantes são mantidos como itens comuns, aspecto fundamental para uma boa equalização dos resultados às escalas do Enem e do Saeb. Itens com comportamentos diferenciais para grupos de respondentes com mesma proficiência são considerados itens novos e assumem-se os parâmetros estimados independentemente.

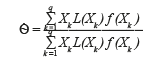
Os itens que apresentaram problemas de ajuste ou DIF são desafixados e considerados como novos e seus parâmetros reestimados em uma versão 2 de calibração. Os resultados dessa rodada de calibração são “input” para a estimação das proficiências dos participantes do Enem/SAEB.

Após a obtenção dos parâmetros calibrados na escala do Enem e do Saeb, calcula-se a proficiência de todos os examinandos, tendo como base os parâmetros dos itens e as respostas do participante. O método para cálculo da proficiência é o *Expected a Posteriori (EAP)*, através da seguinte fórmula:



Nota-se que f(Ѳ) é a função de probabilidade *a priori*; L(**u** | Ѳ) é uma função matemática associada ao padrão de respostas dos participantes e aos parâmetros dos itens, valores conhecidos; e Ѳ 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, pois depende da complexidade da função que está sendo usada — no caso dos exames e avaliações, essa função é bastante complexa. Para contornar esse problema, Stroud e Secrest (1966) sugerem a criação de intervalos para funções complexas e a resolução numérica de cada intervalo dessa função — esse método é conhecido como quadratura Gaussiana —, sendo que a expressão pode ser resolvida pela soma desses intervalos com o uso da seguinte fórmula:



Essa é a fórmula pela qual é obtida a nota dos participantes do Enem e do Saeb, representada nela por Ѳ. Os resultados de proficiências estimadas pelas equipes, separadamente, são comparados. Para confirmação que os resultados estão corretos, o protocolo de conferência exige que sejam idênticos até a quinta casa decimal, entre todas as equipes envolvidas.

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 do conjunto dos itens que compõem cada prova, a cada edição de aplicação.

Concluindo, apesar de não ser simples e exigir estimativas dos parâmetros realizada por métodos estatísticos avançados, o cálculo da proficiência é objetivo. Ao longo do texto, foi apresentado o modo de cálculo das proficiências dos candidatos, não sendo realizada distinção entre os participantes.

Com relação à disponibilização dos parâmetros dos itens do Enem e do Saeb, o Inep coloca à disposição dos pesquisadores por meio do Serviço de Acesso a Dados Protegidos (Sedap). Você pode encontrar informações sobre o Sedap em <http://portal.inep.gov.br/web/guest/dados/sedap>

**Programações do BILOG-MG**

O programa utilizado para o cálculo das proficiências do SAEB e do Enem é o BILOG-MG.

SAEB

No caso do SAEB utilizamos pelo menos dois grupos para fazer a equalização com a escala SAEB: o SAEB da edição anterior e o SAEB da edição atual. Mas cada edição é um caso. Tem edição que usa pré-testes também para fazer essa equalização, tem ano que usam duas edições anteriores do SAEB.

ENEM

No caso do Enem, em cada edição são utilizados vários grupos de Pré-teste para fazer a equalização com a escala Enem.

**Referências**

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.

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.

STROUD, A. H; SECREST, D. *Gaussian Quadrature Formulas.* Englewood Cliffs. NewJersey: Prentice-Hall, 1996