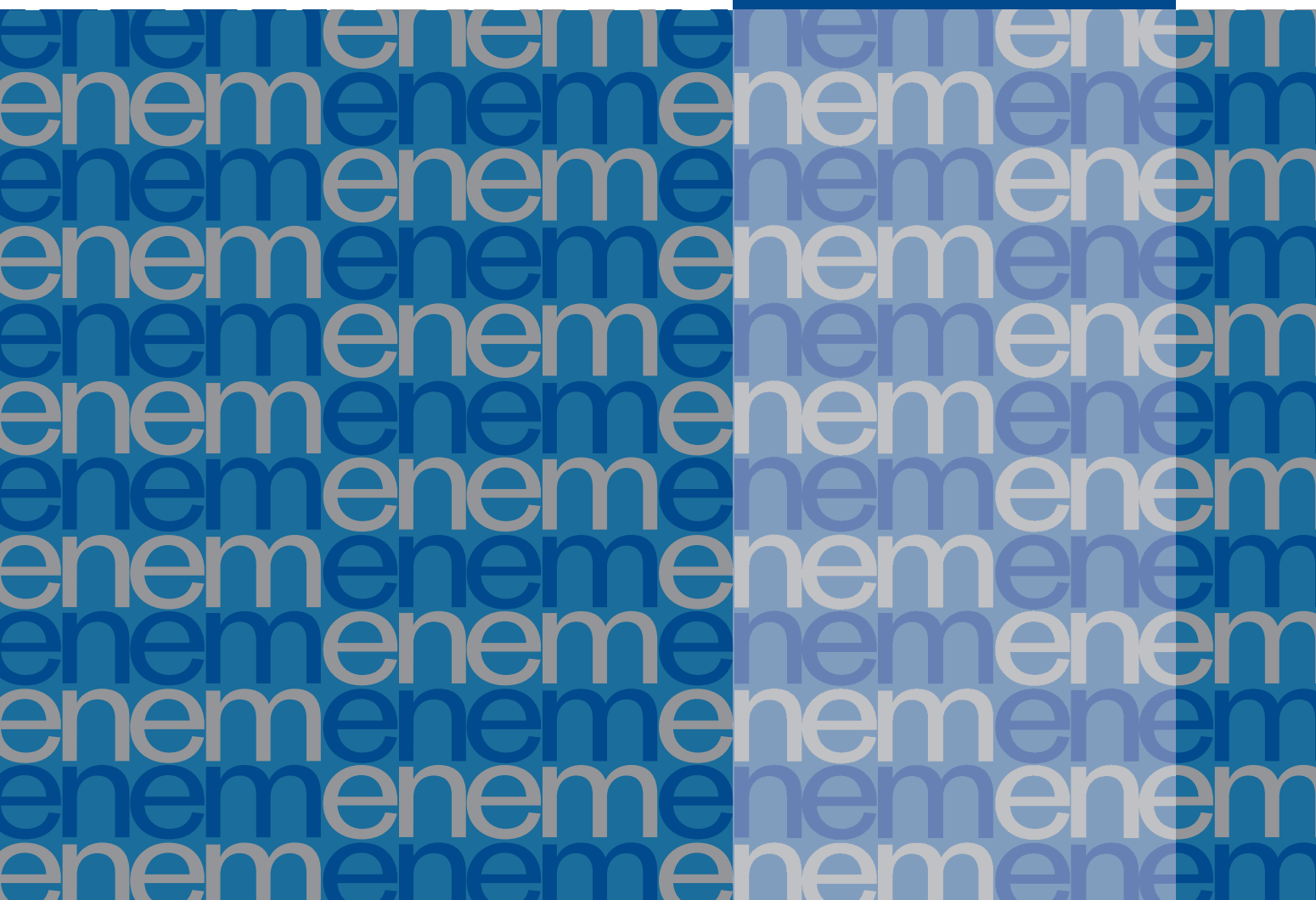


**EXAME NACIONAL
DO ENSINO MÉDIO – ENEM
PROCEDIMENTOS
DE ANÁLISE**

DIRETORIA DE AVALIAÇÃO
DA EDUCAÇÃO BÁSICA
DAEB



REPÚBLICA FEDERATIVA DO BRASIL

MINISTÉRIO DA EDUCAÇÃO | **MEC**

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS
EDUCACIONAIS ANÍSIO TEIXEIRA | **INEP**

DIRETORIA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA | **DAEB**

enem2021

**EXAME NACIONAL
DO ENSINO MÉDIO – ENEM**
PROCEDIMENTOS
DE ANÁLISE

Brasília-DF
Inep/MEC
2021



Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep)
É permitida a reprodução total ou parcial desta publicação, desde que citada a fonte.

DIRETORIA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA (DAEB)

COORDENAÇÃO-GERAL DE INSTRUMENTOS E MEDIDAS (CGIM)

DIRETORIA DE ESTUDOS EDUCACIONAIS (DIRED)

COORDENAÇÃO DE EDITORAÇÃO E PUBLICAÇÕES (COEP)

REVISÃO

Jair Santana Moraes

NORMALIZAÇÃO:

Clarice Rodrigues da Costa

REVISÃO GRÁFICA

Lilian dos Santos Lopes

Váleria Borges

PROJETO GRÁFICO E CAPA

Marcos Hartwich

DIAGRAMAÇÃO E ARTE-FINAL

Daniel Caixeta

Dados Internacionais de Catalogação na Publicação (CIP)

Brasil. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.
Exame Nacional do Ensino Médio – Enem : procedimentos de análise
[recurso eletrônico]. – Brasília, DF : Inep, 2021.

22 p. : il.

ISBN 978-65-5801-010-4

1. Exame Nacional do Ensino Médio. 2. Proficiência. 3. Teoria da Reposta
ao Item. I. Título.

CDU 371.27



SUMÁRIO

INTRODUÇÃO.....	5
PROCEDIMENTOS INICIAIS DE ANÁLISE.....	9
ANÁLISE POR MEIO DA TCT.....	11
ANÁLISE DE DIMENSIONALIDADE	13
ANÁLISE POR MEIO DA TRI.....	15
Calibração.....	17
Equalização	18
Análise de DIF e ajuste.....	18
Cálculo das proficiências.....	19
Escala	20
CONSIDERAÇÕES FINAIS.....	23
REFERÊNCIAS	25



INTRODUÇÃO

Criado em 1998, o Exame Nacional do Ensino Médio (Enem) apresenta, desde sua origem, como finalidade avaliar o desempenho individual dos concluintes do ensino médio brasileiro. A partir de 2009, o Enem ampliou seu escopo de objetivos quando se tornou um dos principais mecanismos de seleção para ingresso nas universidades federais do País. No exame, busca-se aferir as competências e habilidades desenvolvidas pelos estudantes ao fim da escolaridade básica. Essa 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.

Em avaliação educacional, usualmente, mede-se a proficiência ou o conhecimento de pessoas em certa área. Isso pode ser realizado por meio de um teste (uma prova), composto por vários itens (questões) que englobam as diversas competências e habilidades associadas à área em estudo. Nas avaliações de sala de aula, normalmente, as notas são calculadas de “0 a 10” ou de “0 a 100”, e a forma de calcular a nota é simples, bastando somar as questões corretas na prova. Assim, estudantes que acertam o mesmo número de questões obtêm a mesma nota.

Esse modelo de atribuição de notas, baseado na Teoria Clássica dos Testes (TCT), embora facilmente entendido e bastante utilizado, não usa a totalidade de informações que podem ser extraídas do conjunto de respostas do examinando; não leva em conta, por exemplo, se o examinando mantém um padrão de resposta coerente com sua proficiência, isto é, se acerta mais itens de nível de dificuldade inferior ou igual à sua proficiência e menos itens de nível de dificuldade superior à sua proficiência.

Desde 2009, os procedimentos de análise dos itens e de cálculo das proficiências no Enem passaram a ter 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 1970 e 1980 com o desenvolvimento de *softwares* que permitiram a implementação dos complexos modelos matemáticos relacionados à TRI. Na década de 1990, houve uma expansão do uso da TRI em testes de avaliação educacional e, atualmente, a maioria dos programas de avaliação em larga escala no mundo tem como base essa teoria (Yen; Fitzpatrick, 2006).

No Brasil, o uso da TRI em avaliações educacionais teve início com o Sistema Nacional de Avaliação da Educação Básica (Saeb) em 1995 e, posteriormente, foi implementado também em outros testes desenvolvidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), como o Exame Nacional para Certificação de Competências de Jovens e Adultos (Encceja). No âmbito internacional, a TRI é largamente utilizada em testes desenvolvidos em diversos países: Estados Unidos, França, Holanda, Coreia do Sul e China, por exemplo. Além disso, é usada no Programa Internacional de Avaliação de Estudantes (Pisa) e no Test of English as a Foreign Language (Toefl), teste internacional de proficiência em língua inglesa.

A TRI não é uma teoria que busca substituir a TCT. Pelo contrário, é importante que se busque utilizar os avanços oferecidos em cada uma delas. No Enem, o cálculo da proficiência com o uso da TRI permite acrescentar outros aspectos além do quantitativo de acertos (utilizado na 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 formas distintas a depender de quais itens estão certos e podem, assim, ter habilidades diferentes. Portanto, a estimativa da proficiência está relacionada ao número de acertos, aos parâmetros dos itens e ao padrão de respostas. A TRI permite a comparabilidade dos resultados dos examinandos entre as diferentes aplicações dos testes. A comparação dos resultados entre avaliações é possível na medida em que uma escala é estabelecida. Esses aspectos serão explicitados com mais detalhes ao longo deste documento.

O objetivo deste documento é discorrer a respeito dos procedimentos realizados para a análise psicométrica das provas objetivas do Enem, buscando esclarecer aspectos teóricos e práticos do processo, bem como apontando literatura e documentos pertinentes para aprofundamento. Serão descritos os procedimentos iniciais de análise, com abordagem de aspectos relativos aos pré-testes e à consolidação das bases de dados; a análise por meio da TCT; uma breve explanação de pressupostos da TRI (como a análise de dimensionalidade); a análise por meio da TRI, incluindo o cálculo dos parâmetros dos itens e das proficiências.

O documento utilizou como base para sua construção notas técnicas construídas no âmbito do Inep e relatórios internos. A título de referência e visualização do histórico do

processo de trabalho no que diz respeito às análises psicométricas do Enem, segue abaixo lista de notas técnicas e documentos relacionados ao tema, elaborados pelo Inep:

- 2011: Nota técnica: Procedimento de cálculo das notas do Enem.
- 2012: Nota técnica: Teoria da Resposta ao Item.
- 2012: *Guia do participante: entenda sua nota no Enem.*
- Relatórios de análise do Enem (realizados para cada edição e disponíveis por meio do Serviço de Acesso a Dados Protegidos – Sedap).



PROCEDIMENTOS INICIAIS DE ANÁLISE

Antes de iniciar a explanação dos passos de análise, faz-se importante discorrer brevemente acerca dos pré-testes. A elaboração de um bom teste educacional exige um processo rigoroso de elaboração e de validação teórica, tendo por base a matriz de referência, os critérios técnico-pedagógicos e o conhecimento prévio dos parâmetros dos itens. Este é conseguido por meio da realização de pré-testes. O pré-teste consiste na aplicação de um conjunto de itens a uma amostra de alunos com características semelhantes às da população para a qual a prova se destina. É a forma empírica de se avaliar a qualidade técnico-pedagógica e psicométrica dos itens. Essa etapa tem como objetivo captar subsídios importantes para aumentar a precisão da prova que será aplicada aos milhões de participantes do Enem. A partir de diversas aplicações de pré-testes, é construído um banco nacional de itens, que abriga os itens disponíveis para composição da prova do Enem e que são selecionados segundo critérios técnicos.

No que concerne à aplicação da prova do Enem, todas as etapas de análise dos dados de respostas dos participantes e cálculo de suas notas são realizadas de forma independente entre a equipe de psicometria do Inep e as equipes do consórcio aplicador, um grupo de empresas com comprovada experiência na aplicação e análise de provas em larga escala e selecionadas mediante processo licitatório. Durante todas as etapas de análise, são realizadas conferências 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 depois de constatado que os resultados obtidos pelos grupos (Inep e empresas contratadas) são os mesmos.

O procedimento de correção das provas objetivas inicia-se com a leitura dos cartões de resposta das provas 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 esses procedimentos, a base de dados é consolidada e encaminhada para a equipe de análise de dados. A base de dados a que as equipes de psicometria (profissionais que analisam a prova e realizam os cálculos das proficiências) têm acesso não possui informações que permitam identificar os examinandos, mas apenas um código sequencial fictício (máscara).

A equipe de análise psicométrica tem acesso às informações das provas, tais como estrutura e itens que as compõem, somente depois de sua aplicação e divulgação dos gabaritos, sendo esse um exemplo das medidas tomadas para garantia de segurança. De posse das estruturas dos testes, começa a montagem dos arquivos que serão utilizados nos programas que operacionalizam as análises por meio da TCT, utilizando o *software* R e as análises pela TRI, utilizando o *software* Bilog-MG. As análises e os cálculos de proficiência são realizados de forma independente para cada área de conhecimento.



ANÁLISE POR MEIO DA TCT

Numa primeira etapa de análise psicométrica, a TCT é utilizada, principalmente, para subsidiar a equipe pedagógica na verificação de inconsistências de gabarito, observar de forma simples a reação da população aos itens, auxiliar em futuras montagens de provas, entre outros. Durante a análise clássica, calculam-se: os percentuais de acerto do item; os percentuais de escolha de cada alternativa; o percentual de acerto dos grupos superior e inferior, definidos pelos 27% de candidatos com maiores e menores desempenhos (respectivamente); o índice de discriminação clássico, definido pela diferença entre os percentuais de acerto dos grupos superior e inferior; e os coeficientes de correlação bisserial para cada alternativa.

Os cálculos dos percentuais de escolha e dos coeficientes bisseriais de cada alternativa são muito importantes para a análise pedagógica do item, pois permitem saber quais foram as alternativas erradas mais escolhidas pelos alunos e, também, quais grupos de desempenho predominantemente escolheram cada uma delas.

O coeficiente correlação bisserial é uma medida de associação entre o desempenho no item e o desempenho no teste. Esse coeficiente estima a correlação entre a variável de desempenho no teste e uma variável latente (não observável) com distribuição normal que, por hipótese, representa a “habilidade” que determina o acerto ou erro no item. O índice de correlação bisserial de cada alternativa é calculado por meio da seguinte fórmula:

$$r_{bis} = \frac{M_p - M}{S} \times \frac{p}{h(p)}$$

Na qual:

- M_p é a média da medida de desempenho no teste para os alunos que acertaram o item;
- M é a média geral da medida de desempenho no teste para todos os alunos;
- S é o desvio-padrão da medida de desempenho no teste para todos os alunos;
- p é o percentual de acerto no item;
- $h(p)$ é o valor da densidade da distribuição normal com média 0 (zero) e variância 1 (um) no ponto em que a área da curva à esquerda desse ponto é igual a p .

Espera-se que o coeficiente bisserial do item que representa a alternativa correta seja positivo, ou seja, que pela fórmula a média dos alunos que acertaram o item (M_p) seja maior que a média geral (M). Já para as alternativas erradas é esperado que o coeficiente bisserial seja negativo, isto é, que a média dos alunos que escolhem a alternativa errada (M_p) seja menor que a média geral (M). Dessa forma, alternativas erradas que apresentam coeficientes bisseriais positivos devem ser examinadas de forma cuidadosa, pois podem indicar algum equívoco como erro de gabarito, mais de uma solução correta para a questão ou nenhuma solução correta. Após as verificações, segue-se para as próximas etapas de análise.



ANÁLISE DE DIMENSIONALIDADE

O modelo de TRI utilizado no Enem possui dois pressupostos que devem ser destacados: a independência local e a unidimensionalidade. Na independência local, tem-se que a probabilidade de resposta correta para os itens é estatisticamente independente. No que se refere à unidimensionalidade, há o postulado que há apenas uma habilidade (ou traço latente) responsável pela realização da tarefa (responder aos itens). Contudo, é ponto pacífico a consideração de que a ação humana é multideterminada, logo, para satisfazer o postulado da unidimensionalidade, é suficiente aceitar que haja um fator dominante responsável pelo conjunto de itens (Pasquali, 2004).

Para as provas do Enem, realiza-se a análise de dimensionalidade, e a técnica estatística utilizada é a Análise Fatorial de Informação Plena (*full information factor analysis*), descrita em Bock, Gibbons e Muraki (1988). A dimensionalidade é estudada também por meio dos autovalores da matriz de correlação tetracórica dos itens (Fundação Cesgranrio, 2020).

Como mencionado, um critério para a verificação da unidimensionalidade é a existência de um fator dominante. Assim, considera-se a hipótese de unidimensionalidade atendida quando o primeiro fator da análise fatorial apresentar uma explicação da variância muito maior do que a do segundo fator.



ANÁLISE POR MEIO DA TRI

A TRI modela a probabilidade de um participante responder corretamente a um item (questão), como função dos parâmetros do item e da proficiência desse participante. Essa relação é expressa por meio de uma função monotônica crescente que indica que quanto maior o conhecimento do participante, maior será sua probabilidade de acertar o item (Andrade; Tavares; Valle, 2000; Baker; Kim, 2004; Hambleton; Swaminathan; Rogers, 1991; Klein, 2003; Pasquali, 1997). No Enem, a função monotônica é uma função logística de três parâmetros, desenvolvida por Birnbaum (1968), expressa pela fórmula:

$$P(x_j = 1 | \theta, a_j, b_j, c_j) = c_j + \frac{(1 - c_j)}{1 + \exp[-a_j(\theta - b_j)]}$$

Na qual:

- $P(x_j = 1 | \theta, a_j, b_j, c_j)$ é a probabilidade do indivíduo com habilidade θ acertar o item j .
- a_j , em que $a_j > 0$ é o parâmetro de inclinação do item j , também chamado de parâmetro de discriminação do item;
- b_j é o parâmetro de posição do item, também chamado de parâmetro de dificuldade;
- c_j , em que $0 < c_j < 1$ é o parâmetro da assíntota inferior do item j , refletindo a probabilidade de um examinando de proficiência muito baixa selecionar a opção correta; e
- θ é o traço latente do indivíduo.

Conforme mencionado, o modelo da TRI utilizado no Enem é o logístico de três parâmetros. Estes se referem às informações expressas pelos itens, que no caso são relativos à sua dificuldade, sua discriminação e acerto ao acaso. Esses parâmetros podem ser definidos como:

- parâmetro “ a ” 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;
- parâmetro “ b ” de dificuldade: associado à dificuldade da questão, sendo que quanto maior seu valor, mais difícil é a questão. Ele é expresso na mesma escala da proficiência. Em uma prova, utilizam-se questões de diferentes níveis de dificuldade para avaliar os participantes em todos os níveis de proficiência;
- parâmetro “ c ” de probabilidade de acerto ao acaso: 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.

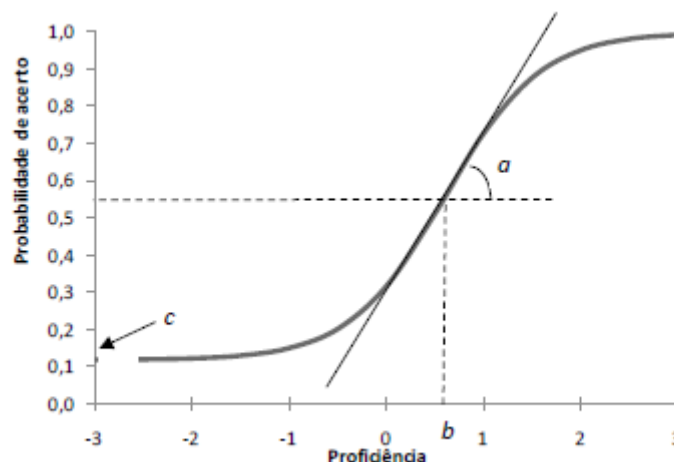


FIGURA 1

CURVA CARACTERÍSTICA DO ITEM (CCI) NO MODELO LOGÍSTICO DE TRÊS PARÂMETROS

Fonte: Fundação Cesgranrio (2020).

A Figura 1 ilustra a curva característica de um item fictício em uma escala com média “0” e desvio-padrão “1”. A CCI permite uma visualização gráfica das características do item, descrevendo a relação entre probabilidade de acerto e a proficiência. Verifica-se

que o parâmetro “ b ” se encontra na escala da proficiência. O parâmetro “ a ” apresenta valor proporcional à inclinação da curva no ponto b . O parâmetro “ c ” é representado pela assíntota horizontal inferior da curva. Cada item possui sua própria curva, que depende dos parâmetros que possui.

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 com as respostas dos examinandos, arquivo de programação, arquivo de parâmetros, e 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 dos respondentes) são preparadas em formato apropriado para utilização do *software* e incluem não só as respostas dos examinandos na aplicação atual, como também respostas de grupos de respondentes de pré-testes de anos anteriores.

Assim como ocorre em relação aos arquivos de bases de dados, o arquivo de programação do Bilog-MG com extensão “.blm” é preparado contemplando informações do grupo de itens do exame 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 (tais como Prova 1, Prova de reaplicação/Pessoas Privadas de Liberdade) e entre anos de aplicações.

Os arquivos com parâmetros dos itens (extensão “.prm”, do Bilog-MG) são preparados, em primeira instância, apenas com aqueles que já foram previamente estimados e, em segunda etapa, para o cálculo das proficiências, com os parâmetros de todos os itens calibrados.

CALIBRAÇÃO

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. A calibração dos itens do Enem é realizada com base em uma amostra de respondentes concluintes da 3ª série do Ensino Médio Regular. Para cada área, é retirada uma amostra sistemática de 100 mil participantes, sendo os examinandos ordenados por escore total e CEP (a fim de garantir um melhor espalhamento do estudo entre as cinco regiões).

Para a calibração, utiliza-se o procedimento de estimação pela máxima verossimilhança marginal, introduzido em 1970 por Bock e Lieberman. Neste método estimam-se os parâmetros dos itens, assumindo-se uma certa distribuição para as habilidades e então, considerando os parâmetros dos itens conhecidos, estimam-se as proficiências (Andrade; Tavares; Valle, 2000).

EQUALIZAÇÃO

Um relevante processo que ocorre durante as análises é a equalização. De acordo com Andrade, Tavares e Valle (2000), este é um dos conceitos mais importantes da TRI e uma das grandes vantagens de seu uso em avaliação educacional. Equalizar, no caso da TRI, significa colocar parâmetros de itens vindos de provas distintas em uma escala comum, tornando-os comparáveis. No âmbito do Enem, a equalização se dá via itens comuns, que no caso são os itens advindos dos pré-testes. Esses itens funcionam como itens de ligação e garantem que os parâmetros dos itens procedentes de diferentes aplicações estejam na mesma escala e, portanto, sejam comparáveis.

O Bilog-MG implementa uma extensão da TRI para múltiplos grupos de examinandos, o que permite estimar os parâmetros de itens de uma mesma área do Enem conjuntamente e em uma mesma escala (isto é, diferentes aplicações do Enem e pré-testes do Enem). Para tanto, as distribuições *a priori* das habilidades podem apresentar, para cada grupo, diferentes valores de média e de desvio padrão. Os concluintes da rede pública de ensino de 2009 formam o grupo de referência e, portanto, este grupo assume a distribuição *a priori* normal padrão.

ANÁLISE DE DIF E AJUSTE

Em seguida à rodada inicial de calibração e equalização dos itens, verifica-se o ajuste do modelo aos dados e o funcionamento diferencial dos itens (Differential Item Functioning – DIF) em relação aos 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 esse item. Em termos de análise, considera-se que há DIF quando existem diferenças entre as proporções esperadas de resposta (obtidas com o arquivo “.exp” do *software* Bilog-MG) nos dois grupos dentro do intervalo compreendido entre os percentis 5 e 95. No Enem, verifica-se se um item estimado no pré-teste apresenta o mesmo comportamento de parâmetros quando aplicado no exame. Itens com comportamentos semelhantes são mantidos como itens comuns, aspecto fundamental para uma boa equalização dos resultados à escala do Enem. Itens com comportamentos diferenciais para grupos de respondentes com mesma proficiência são considerados itens novos e assumem-se os parâmetros estimados independentemente.

Considera-se que a função modelada para um item pode apresentar problemas no ajuste quando há diferenças entre as proporções esperadas de resposta e as proporções empíricas dentro do intervalo entre os percentis 5 e 95. Esses dados podem ser obtidos com base no arquivo “.exp” do Bilog-MG.

A Figura 2 apresenta gráfico que exemplifica um item com DIF.

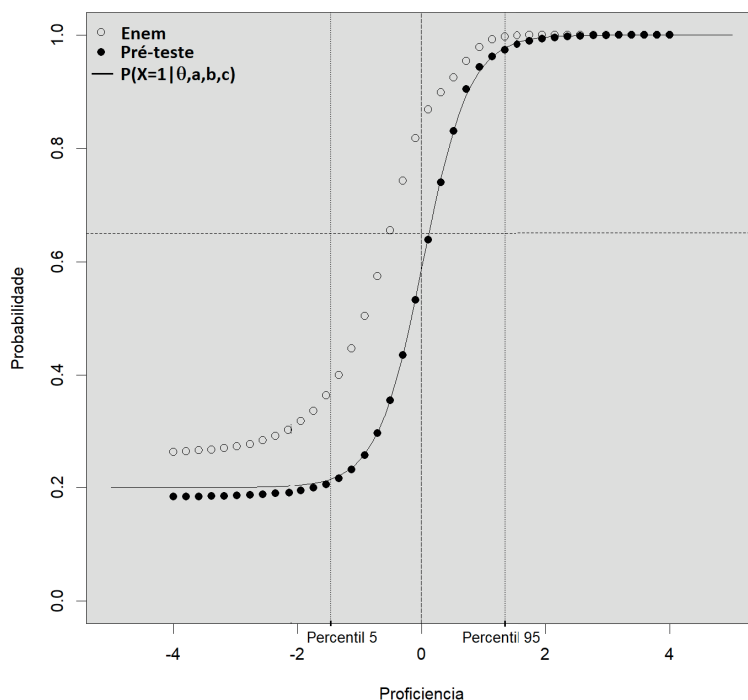


FIGURA 2
GRÁFICO DE ITEM FICTÍCIO QUE APRESENTA DIF

Fonte: Elaborada pela CGIM/Daeb/Inep.

Nota-se que o item apresentou um comportamento diferente entre o pré-teste e o exame para grupos que possuíam a mesma proficiência na escala, o que caracteriza um item que possui DIF.

Durante as análises, 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.

CÁLCULO DAS PROFICIÊNCIAS

Obtidos os parâmetros calibrados na escala do Enem, 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), mediante a seguinte fórmula:

$$E(\theta|\mathbf{x}, \boldsymbol{\eta}) = \frac{\int_{\mathbb{R}} \theta L(\mathbf{x}|\boldsymbol{\eta}) f(\theta) d\theta}{\int_{\mathbb{R}} L(\mathbf{x}|\boldsymbol{\eta}) f(\theta) d\theta}$$

Na qual:

- $f(\theta)$ é a função de probabilidade *a priori*.
- $L(\mathbf{x}|\boldsymbol{\eta})$ é uma função matemática associada ao padrão de respostas (\mathbf{x}) e aos parâmetros dos itens $(\boldsymbol{\eta})$, 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 simples, pois depende da complexidade da função que está sendo usada e, no caso do Enem, 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.

As estimativas das notas são calculadas utilizando-se pontos de quadratura para aproximar a distribuição *a priori* das habilidades de cada respondente. No Enem, utilizam-se 40 pontos de quadratura e *a priori* com distribuição normal. Os parâmetros resultantes devem então passar por uma transformação linear de modo a estarem na escala original do Enem 2009 (0,1).

ESCALA

As proficiências são calculadas em uma escala (0,1 – média 0 e desvio-padrão 1). Para melhor entendimento das notas dos participantes, foi arbitrado que para a divulgação dos resultados seria mais adequada uma distribuição de proficiências com média 500 e desvio-padrão 100, ou seja, uma escala (500,100).

Genericamente, utiliza-se a seguinte fórmula de transformação das proficiências para a escala (500,100):

$$\theta_{(500,100)} = 100 \theta_{(0,1)} + 500$$

em que $\theta_{(0,1)}$ é a proficiência na escala antiga, ou seja, na escala (0,1).

Os parâmetros dos itens também são transformados para a escala (500,100), da seguinte forma:

$$b_{(500,100)} = 100b_{(0,1)} + 500$$

$$a_{(500,100)} = \frac{a_{(0,1)}}{100}$$

em que $b_{(0,1)}$ e $a_{(0,1)}$ são os parâmetros dos itens na escala (0,1). O parâmetro “c”, por representar uma probabilidade, não sofre alteração.

Os resultados de proficiências estimadas pelas equipes, separadamente, são comparados. Para confirmação se os resultados estão corretos, o protocolo de conferência exige que sejam idênticos até a quinta casa decimal, entre todas as equipes envolvidas. Os resultados são divulgados apenas quando há 100% de concordância nos valores obtidos pelas equipes técnicas. Essa tripla conferência juntamente com o rigor científico dos métodos estatísticos utilizados garantem a confiabilidade dos resultados.



CONSIDERAÇÕES FINAIS

Conforme demonstrado ao longo do documento, o cálculo das proficiências por meio da TRI exige o uso de estatísticas avançadas e de *software* específico. Apesar de complexo, o método de cálculo é objetivo e não realiza distinção entre os participantes. A proficiência está relacionada ao número de acertos, aos parâmetros dos itens e ao padrão de respostas. Logo, não se realizam comparações tendo como base a quantidade de acertos de diferentes indivíduos. Além disso, não é adequada a realização de comparações tendo a quantidade de acertos e proficiência de diferentes áreas de conhecimento, já que cada área possui sua própria escala.

Ressalta-se, ainda, que as proficiências na TRI são estimadas em uma escala métrica que não possui mínimo e máximo preestabelecidos. Esses valores variam de acordo com as características do conjunto dos itens que compõem cada prova, a cada edição de aplicação. E, a partir do processo de equalização, como relatado neste documento, as diferentes aplicações de cada área são comparáveis por terem seus itens alocados em uma mesma escala.

Conclui-se, também, que 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 com base em simples ponderações. Por serem informações sensíveis às atividades realizadas pelo Inep, os parâmetros dos itens do Enem são divulgados cinco anos depois dos resultados de uma determinada edição. Para tanto, o Inep coloca à disposição o Serviço de Atendimento ao Pesquisador (SAP) (Portaria Inep nº 467/2014) nos termos do entendimento e da decisão da Controladoria-Geral da União (CGU) (Parecer nº 2245 de 10/06/2016, Referência Processo nº 23480.005140/2016-90).



REFERÊNCIAS

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

BAKER, F. B.; KIM, S. H. (Ed.). *Item response theory: parameter estimation techniques*. 2nd. ed. New York: Marcel Dekker, 2004.

BIRNBAUM, A. Some latent trait models and their models ant their use in inferring an examinee's ability. In: LORD, F. M.; NOVICK, M. R. (Ed.). *Statistical theories of mental test scores*. Reading: Addison-Wesley, 1968. p. 397-479.

BOCK, R. D.; GIBBONS, R.; MURAKI, E. Full-information item factor analysis. *Applied Psychological Measurement*, [S. l.], v. 12, n. 3, p. 261-280, Sept. 1988.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Nota técnica: procedimento de cálculo das notas do Enem*. Brasília, DF, 2011. Disponível em: <https://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_procedimento_de_calculo_das_notas_enem_2.pdf>. Acesso em: 6 jul. 2021.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Nota técnica: Teoria de Resposta ao Item*. Brasília, DF, 2012. Disponível em: <https://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_tri_enem_18012012.pdf>. Acesso em: 6 jul. 2021.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Entenda a sua nota no Enem: guia do participante*. Brasília, DF, 2012. Disponível em: <https://download.inep.gov.br/educacao_basica/enem/guia_participante/2013/guia_do_participante_notas.pdf>. Acesso em: 6 jul. 2021.

FUNDAÇÃO CESGRANRIO. *Relatório técnico da apuração das provas objetivas*. Rio de Janeiro, 2020. Documento não publicado.

HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. *Fundamentals of item response theory*. Newbury Park: Sage Publications, 1991.

KLEIN, R. 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, jul./set. 2003.

LAWLEY, D. N. On problems connected with item selection and test construction. *The Royal Society of Edinburgh*, Edinburgh, v. 61A, p. 273-287, 1943.

LORD, F. M. *A theory of test scores*. [New York: Psychometric Society], 1952. (Psychometric Monograph, 7).

PASQUALI, L. *Psicometria: teoria e aplicações*. Brasília: Editora Universidade de Brasília, 1997.

PASQUALI, L. *Psicometria: teoria dos testes na psicologia e na educação*. Petrópolis: Vozes, 2004.

STROUD, A. H.; SECREST, D. *Gaussian Quadrature Formulas*. Englewood Cliffs: Prentice-Hall, 1966.

YEN, W. M.; FITZPATRICK, A. R. Item Response Theory. In: BRENNAN, R. L. (Ed.). *Educational Measurement*. Westport: Praeger Publishers, 2006. p. 111-153.



(CC) BY-NC

VENDA PROIBIDA

