



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

NOTA TÉCNICA

Assunto: Teoria de Resposta ao Item

A decisão de implementar no Exame Nacional do Ensino Médio (ENEM) a Teoria de Resposta ao Item (TRI) teve duas finalidades principais: (1) permitir a comparabilidade dos resultados entre os anos e (2) permitir a aplicação do Exame várias vezes ao ano.

A comparação dos resultados entre avaliações é possível na medida em que, com a TRI, uma escala métrica é estabelecida. Assim como existem escalas padrões para mensurar comprimento (metro) e temperatura (Celsius), com a TRI desenvolve-se uma escala padrão de conhecimento. As provas, nas avaliações educacionais, são instrumentos de medida do conhecimento, comumente denominado de traço latente. Por sua natureza, os conhecimentos adquiridos pelos estudantes não podem ser mensurados diretamente, mas é possível utilizar instrumentos de medida que buscam mensurá-los indiretamente. Essa é a fundamentação da Teoria da Medida e é ela que embasa a construção também de instrumentos psicológicos que buscam medir: inteligência, depressão, personalidade etc.

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 ENCCEJA, Prova Brasil e 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). Este exame surgiu em 1964 e é amplamente utilizado em todo o mundo. Desde o ano de sua origem, este exame já avaliou mais de 25 milhões de alunos e tem sido administrado por mais de 4.500 centros em 165 países

do mundo. No TOEFL, os alunos marcam o horário em um dos centros credenciados e podem realizá-lo várias vezes ao ano. A prova é adaptativa, realizada no computador, e cada candidato responde a um conjunto de itens (questões) diferentes. Apesar de em cada aplicação o candidato receber uma prova distinta, todos os resultados são comparáveis e considerados isonômicos.

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. O exame é aplicado sete vezes ao ano, em outubro, novembro, dezembro, janeiro, março (ou abril), maio e junho. Estudantes de outros países também podem prestá-lo, caso estejam interessados em ingressar em uma das universidades que aceita os resultados do SAT. Novamente, apesar de aplicações realizadas em momentos distintos e com provas diferentes, a existência de uma escala padrão possibilita a comparabilidade de desempenhos.

Um pouco mais sobre TRI

Usualmente, quando desejamos medir a proficiência de um aluno em determinada área do conhecimento, fazemos uso do escore (número de acertos) do aluno em um teste com um determinado número de itens (questões). Os itens são analisados através de parâmetros denominados de discriminação e de dificuldade, que dependem fundamentalmente do grau de dificuldade do teste, como também depende o escore do aluno, e do grupo de respondentes. Comparações entre desempenhos de alunos submetidos a diferentes provas são difíceis de serem realizadas. Esta é a Teoria Clássica dos Testes (TCT).

Para contornar estas dificuldades, e também para permitir uma medida mais apropriada da proficiência do aluno, foi desenvolvida a TRI, cujo foco principal, como bem diz o seu nome, é o item e não o teste como um todo. Dentro do contexto da TRI, a medida de proficiência de um aluno não depende dos itens apresentados a ele e os parâmetros de discriminação e de dificuldade do item não dependem do grupo de respondentes. Em outras palavras, um item mede determinado conhecimento, independentemente de quem o está respondendo, e a proficiência de um aluno não depende dos itens que estão sendo apresentados a ele.

A TRI é um conjunto de modelos que relacionam a probabilidade de um aluno apresentar uma determinada resposta a um item, com sua proficiência e características (parâmetros) do item. O modelo utilizado no ENEM é o modelo logístico de três parâmetros que, além dos parâmetros de discriminação e de dificuldade, também faz uso de um parâmetro para controlar o acerto casual. Este último parâmetro tem um papel bastante importante nas avaliações com itens de múltipla escolha, caso do ENEM.

O modelo logístico da TRI parte do princípio de que quanto maior a proficiência do respondente, maior a sua probabilidade de acerto, traço latente acumulativo. O seu parâmetro de dificuldade é medido na mesma escala da proficiência, fato este que permite a comparabilidade entre resultados de diferentes testes e a construção e interpretação de escalas de proficiência, como a escala nacional de proficiência do SAEB construída pelo INEP/MEC para Matemática e Língua Portuguesa (www.inep.gov.br). Uma outra leitura para esse parâmetro, a qual nos parece mais apropriada, é dizer que ele representa a proficiência mínima que um respondente deve possuir para que sua probabilidade de acerto seja alta, ou seja, ele poderia ser chamado de “proficiência do item”.

Para exemplificar, vamos supor que desejemos medir a altura de uma pessoa, em metros, por meio de um questionário. Com um conjunto de perguntas seríamos capazes de definir com uma boa precisão a altura da pessoa. Uma pergunta (item) que poderia ser feita é “Você consegue guardar a bagagem no porta-malas do avião?”. Podemos imaginar que uma pessoa para responder sim para este item deve ter pelo menos 1,65m. Esta seria então a “altura” do item. Um outro item seria: “Você acha que se daria bem em um time de basquete?”. A altura deste item seria bem maior do que 1,65m. Ao final de um conjunto de respostas, seria possível saber a altura do respondente. O importante a ressaltar aqui é que não existe um único conjunto de questões capazes de medir a altura, ou seja, é possível medir a altura de maneira **isonômica** a partir de **provas** diferentes, ou seja, a partir de um conjunto diferente de questões. Itens de mesma “altura” serão respondidos de maneira igual por pessoas de mesma altura. Por isso, pode-se dizer que os resultados são comparáveis. Esta é a grande inovação da Teoria de Resposta ao Item utilizada no Enem, só que ao invés de medir altura, o Enem mede a proficiência de estudantes do Ensino Médio.

O parâmetro de discriminação, como já diz o seu nome, deve ser um valor mínimo de modo a garantir que respondentes com proficiências diferentes tenham

probabilidades diferentes de acerto. O grau de informação do item está diretamente relacionado com esse parâmetro.

A elaboração de uma boa prova exige o conhecimento dos parâmetros dos itens. Isto é conseguido através de pré-testagens de itens em amostras apropriadas de alunos nas quais estimamos os parâmetros dos itens em uma mesma escala de proficiência. Deste modo, posicionamos os itens em uma escala de acordo com o nível de proficiência que eles exigem.

O conjunto desses itens passa a formar um banco de itens na escala de proficiência desejada e a partir dele podemos construir um ou mais testes com graus de dificuldade apropriados para atender os objetivos de uma ou mais avaliações. O importante é que as proficiências de alunos submetidos a esses diferentes testes são medidas na mesma escala e, portanto, comparáveis entre si. Da mesma forma, as medidas que se obtêm da proficiência de um aluno submetido a dois testes construídos com itens desse banco serão iguais.

Por último, vale a pena ressaltar que em avaliações onde o acerto casual é possível, caso do ENEM, a medida de proficiência da TRI leva em conta não só o número de acertos, mas também o padrão de respostas do aluno. Em outras palavras, dois alunos com o mesmo score podem receber da TRI diferentes valores de proficiência. Receberá maior proficiência aquele aluno que apresentar respostas aos itens de forma mais coerente com o construto que está sendo medido.

Dalton Francisco de Andrade
Coordenador-Geral de Informações e Indicadores Educacionais

Camila Akemi Karino
Coordenadora-Geral de Instrumentos e Medidas