

Entenda a Teoria de Respostas ao Item (TRI), utilizada no Enem

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). Esta forma de avaliação, conhecida como a Teoria Clássica dos Testes (TCT), apresenta limitações quando se pretende comparar desempenhos de alunos submetidos a provas diferentes. Isso ocorre devido a dependência que existe entre a avaliação das características do teste e a amostra de respondentes. O índice de discriminação (grau com que o item diferencia pessoas com níveis distintos de proficiências) e a dificuldade do item dependem fundamentalmente do grupo de respondentes, ou seja, o item discrimina mais ou menos, ou é mais ou menos difícil de acordo com o grupo de respondentes. Um grupo de pessoas com menos proficiência submetido a um teste fácil pode obter como resultado um mesmo escore que um grupo de indivíduos de maior proficiência submetidos a um teste difícil. A comparação dos percentuais de acerto indicará que os grupos possuem a mesma proficiência, sendo que na verdade os grupos possuem proficiências diferentes.

Para contornar estas dificuldades e também para permitir uma medida mais apropriada da proficiência do aluno, foi desenvolvida a TRI, cujo foco principal é o item e não o teste como um todo. Dentro do contexto da TRI, as características dos itens e dos testes são estimadas independentemente das proficiências dos alunos e vice-versa.

Para exemplificar, suponha que desejemos medir a altura de uma pessoa, em metros, por meio de um questionário utilizando a TRI. Uma pergunta (item) que poderia ser feita é “Você consegue guardar sua bagagem de mão no avião sem pedir ajuda?”. Uma pessoa que responda “sim” a este item deve ter, no mínimo, 1,65m. Esta seria então a “altura” do item. Um outro item: “Você acha que se daria bem em um time de basquete?” teria uma altura em torno de 1,90m. Ao final de um conjunto de perguntas como esta, seria possível saber, com uma certa precisão, a altura do respondente.

Não existe um único conjunto de itens capazes de medir a altura. É possível obter 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,. O mesmo ocorre com o Enem que, ao invés de medir altura, mede proficiência.

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.

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. Novamente, apesar de aplicações realizadas em momentos distintos e com provas diferentes, a existência de uma escala padrão possibilita a comparabilidade de desempenhos.

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. O parâmetro de dificuldade representa a proficiência mínima que um respondente deve possuir para que sua probabilidade de acerto seja alta, ou seja, a “proficiência do item”. O parâmetro de discriminação deve ser um valor mínimo de modo a garantir que respondentes com proficiências diferentes tenham probabilidades diferentes de acerto.

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, de acordo com as necessidades. O importante é que as proficiências dos alunos submetidos a esses diferentes testes são medidas na mesma escala e, portanto, comparáveis entre si. Da mesma forma, as medidas de 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 escore 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.