

Entenda o método TRI de avaliação do conhecimento (adotado no Enem)

Roberto Lobo

25 de fevereiro de 2015

Avaliar o conhecimento dos estudantes de forma objetiva e segura sempre foi um grande desafio para os professores mais comprometidos. Nenhum método respondia exatamente ao que esses professores desejavam.

A primeira decisão para assegurar maior objetividade foi a introdução dos testes de múltipla escolha em substituição aos discursivos, por dependerem menos da análise subjetiva e da individualidade dos corretores e transformava claramente acertos e erros em uma escala crescente de mensuração das competências avaliadas.

No entanto, havia sérios problemas nesse método. Somente para citar dois, um prova mais difícil reduzia globalmente o índice de acertos, fazendo com que fosse difícil avaliar a evolução de uma turma ao longo do tempo ou da eficácia de políticas educacionais, uma vez que não havia como comparar provas diferentes de diferentes épocas. Como exemplo podemos citar os exames de suficiência para exercício profissional cuja dificuldade ao variar de ano para ano pode gerar injustiças em razão de diferentes graus de dificuldade das provas aplicadas.

Outro seria a possibilidade de que acertos ocorressem em função de escolhas aleatórias, permitindo que alunos fossem promovidos na disciplina, ou vencendo disputa de vagas em razão de notas obtidas sem a garantia do conhecimento correspondente. Para diminuir os erros absolutos resultantes da variação das dificuldades dos testes propostos a partir de questões diferentes, alguns sistemas de avaliação adotaram a nota relativa, como fez o **Provão do MEC**, por exemplo, na década de 1990.

A TRI é utilizada no Enem desde 2009

As notas eram normalizadas, ficando, por exemplo, os 10% melhores com nota A ou 5, os 30% seguintes com nota B ou 3, 40% seguintes com C ou 3, 30% com D ou 2 e finalmente os 10% de piores resultados com E ou 1. Esse tipo de avaliação, muito comum nos EUA, evitava que testes mais difíceis em seu conjunto prejudicassem os estudantes, fazendo com que todo conjunto de resultados de diferentes anos tivesse distribuição semelhante de notas, qualquer que fosse sua heterogeneidade, embora não fosse capaz de medir a absoluta dos estudantes, nem a evolução da competência geral do sistema de ensino por meio de um melhor resultado alcançado com o correr do tempo. Também não resolveu o problema de eventuais resultados distorcidos em razão da prova objetiva.

A grande ambição sempre foi a de conseguir e medir, de fato, a competência dos estudantes – não somente a relativa, mas a absoluta, o que evitaria os problemas mencionados anteriormente, como o “chute” nas respostas. **Há metodologias que desde a década de 1950 usam um modelo para tentar estabelecer uma avaliação da real competência dos estudantes em certos tópicos ou disciplinas - e que continuam em evolução até hoje - sendo conhecidas como a Teoria da Resposta ao Item (TRI)**. Muita gente sabe que a TRI (IRT em inglês) é cada vez mais utilizada, mas poucos entendem de fato como funciona.

Essa metodologia denominada TRI foi introduzida, na década de 1980, nos exames do Scholastic Aptitude Test, o SAT, que mede a competência de estudantes secundários nos EUA e tem servido como um dos critérios de ingresso no ensino superior. A metodologia também é adotado no Graduate Record Examinaton, o GRE, para estudantes candidatos à pós-graduação.

O exame da OCDE, o Programme for International Student Assessment, Pisa, passou a utilizar a metodologia a partir de 2000. **No Brasil, o Enem a utiliza desde 2009.**

Para quantificar competências, em uma classificação numérica, a TRI se baseia em uma teoria e uma hipótese: a de que é possível colocar a competência de forma ordenada em um eixo que vai do menor para o maior nível e, correspondentemente, que as questões de múltipla escolha também podem ser colocadas sobre um eixo semelhante da mais fácil à mais difícil. Permite, assim, fazer uma estrita correspondência entre esses dois eixos.

Havendo uma correspondência entre os eixos, espera-se que um estudante com certa competência tenderá a responder corretamente às questões com dificuldade abaixo de sua competência e errar nas que estão acima desta competência. Assim não importa quantas questões o estudante acertou, mas até que nível de dificuldade ele acertou.

Por isso, independente da especificidade das questões, desde que a cada uma seja atribuída sua classificação de dificuldade, as provas teriam a capacidade de situar os estudantes nesse eixo de competência, independentemente das peculiaridades de cada prova. **Um exemplo simples seria a medida da altura de um indivíduo.**

Imaginemos que temos dez varas não identificadas escolhidas entre as varas de tamanho 150, 155, 160, 165 e assim por diante - de cinco em cinco centímetros - até 1,95 metro. Um indivíduo de 1,82 metro será maior que sete varas e menor do que três.

Porém, se medirmos esse mesmo indivíduo com outras dez varas não identificadas de tamanhos 170, 175, etc, até 2,15, ele será maior do que 3 e menor do que 7. Se essas medidas correspondessem às questões de uma “prova convencional”, ele teria tirado 7 na primeira prova e 3 na segunda, se tomarmos como nota o número de varas que ele é maior.

No entanto, se soubéssemos a quanto corresponde a altura de cada vara, saberíamos medir sua altura de forma única e objetiva, pois sua altura ficaria entre 1,80 e 1,85 nos dois casos, um resultado mais próximo da realidade e fixo para todas as vezes em que a altura for medida. Temos nesse caso um uma imprecisão de apenas 0,5, ou 1 ponto na avaliação de 0 a 10. Se aumentarmos o número de varas para diferenças menores, ou seja, com mais varas, a imprecisão será cada vez menor.

No caso de testes que pretendem avaliar conhecimento, as questões são preparadas e submetidas a um grande número de respondentes e estatisticamente classificadas em uma ordem crescente de dificuldade, como no caso da vara.

O problema é que a competência não tem um critério físico exato como a altura de um indivíduo. No caso da avaliação de conhecimentos, é preciso introduzir uma probabilidade de acertos e erros - porque há um pequeno grau de variação entre as respostas de alunos com o mesmo grau de conhecimento - por meio de um estudo estatístico mais sofisticado. Mas o importante é que o princípio de medição é o mesmo. Também o “chute” pode ser estimado e compensado.

Sem dúvida, a **TRI é um método sofisticado de preparação e análise, muito mais poderoso** e preciso do que os clássicos processos de avaliação estudantil que vêm sendo também amplamente usado em outras áreas, como a Psicologia.