

OS PARÂMETROS DO “MODELO LOGÍSTICO DE 3 PARÂMETROS”, DA TEORIA DE RESPOSTA AO ITEM, COMO ELEMENTOS DA PRÁTICA AVALIATIVA DE PROFESSORES DE MATEMÁTICA DA EDUCAÇÃO BÁSICA

André Ricardo Cola

Pontifícia Universidade Católica de São Paulo, PUC-SP

andrecola@uol.com.br

Resumo:

O objetivo desse texto é desenvolver o conceito de cada um dos três parâmetros previstos no Modelo Logístico de 3 Parâmetros, da Teoria de Resposta ao Item, e relacionar com o trabalho avaliativo de professores de matemática da educação básica. Os parâmetros propostos no modelo são: discriminação do item, dificuldade do item e probabilidade de acerto casual. De posse de tais conceitos, o professor da educação básica pode utilizá-los para analisar os resultados das avaliações com foco nas questões, complementando a análise feita global e individualmente para cada aluno. Articulados com os demais processos avaliativos que ocorrem em sala de aula, os parâmetros carregam informações úteis que contribuem para uma avaliação norteadora dos processos de ensino e de aprendizagem de alunos.

Palavras-chave: avaliação educacional; avaliação em larga escala; teoria de resposta ao item

Introdução

Frente ao entendimento cada vez mais consensual de docentes, gestores e profissionais da educação de que a avaliação deve servir como instrumento norteador do processo de construção do conhecimento, esse trabalho visa apresentar o Modelo Logístico de 3 parâmetros, um dos mais utilizados para modelar dados de acordo com a Teoria de Resposta ao Item (TRI), propondo a exploração dos parâmetros desse modelo como forma de subsidiar decisões do professor da educação básica em sua prática educativa.

A Teoria de Resposta ao Item é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da habilidade (ou habilidades) do respondente. (ANDRADE, 2000, p. 7)

Tal relação é estabelecida de maneira que quanto maior a habilidade do respondente, maior será a probabilidade de acerto ao item.

Ainda segundo Andrade (2000) existem diversos modelos matemáticos propostos nos estudos disponíveis e todos eles são dependentes essencialmente de três fatores: (i) da

natureza do item, podendo ser dicotômicos ou não dicotômicos; (ii) do número de populações envolvidas, ou seja, apenas uma ou mais de uma; e (iii) do número de habilidades que é medido, podendo ser apenas uma ou mais de uma.

1. O Modelo Logístico de 3 Parâmetros – ML3

O modelo logístico de 3 parâmetros (ML3), desenvolvido por Birnbaum (1968), é o modelo mais utilizado nos sistemas de avaliação em larga escala (PISA, SAEB/Prova Brasil, SARESP). Trata-se de um modelo unidimensional, por avaliar apenas uma habilidade dos respondentes, para itens dicotômicos, ou seja, itens de múltipla escolha (corrigidos como certo ou errado) e também itens de resposta livre quando avaliados de forma dicotomizada.

O modelo logístico de 3 parâmetros considera três dados essenciais para avaliar a qualidade do item e é dado por:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}$$

com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, n$, onde:

U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente ao item i .

θ_j representa a habilidade do indivíduo j .

$P(U_{ij} = 1|\theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Função de Resposta do Item – FRI.

b_i é o parâmetro de dificuldade do item i , medido na mesma escala da habilidade.

a_i é o parâmetro de discriminação do item i , com valor proporcional à inclinação da Curva Característica do Item — CCI no ponto b_i .

c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i (também chamado de probabilidade de acerto casual).

D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

2. Os parâmetros do ML3

A fim de detalhar a utilidade do modelo para sistemas de avaliação, vamos explorar a interpretação dos três parâmetros mencionados anteriormente.

O parâmetro a de um item/questão é o parâmetro de discriminação. Refere-se à capacidade que o item possui para diferenciar os participantes que dominam dos participantes que não dominam a habilidade avaliada no item. É comum, em testes avaliativos, a presença de itens que não avaliam especificamente, ou pelo menos mais fortemente, a habilidade que inicialmente desejou-se avaliar. Tais itens possuem um índice de discriminação baixo, o que pode não indicar a verdadeira habilidade que o avaliado possui no assunto. Dessa forma, quanto maior a discriminação do item, melhor será sua capacidade de diferenciar os participantes que possuem dos participantes que não possuem a habilidade que se deseja avaliar.

O parâmetro b do item/questão é o parâmetro de dificuldade, ou seja, está associado ao nível de dificuldade que o item traz na habilidade avaliada. Quanto maior seu valor, mais difícil é a questão. Em uma prova/teste elaborado com uma boa qualidade, devem-se ter itens de diferentes níveis de dificuldade para avaliar adequadamente os participantes em todos os níveis de conhecimento na habilidade avaliada.

O parâmetro c de um item refere-se à probabilidade de acerto casual. Em provas de múltipla escolha, representa a probabilidade de um participante acertar a questão não dominando a habilidade exigida.

Nas avaliações em larga escala que utilizam a Teoria de Resposta ao Item em sua aplicação, esses parâmetros não são determinados por quem elaborou os itens, mas sim por meio da pré-testagem que ocorre com uma amostra da população que será submetida à avaliação. Dessa forma, os parâmetros são determinados considerando características da população respondente do teste. Em outras palavras, o pré-teste determina os bons itens para avaliar habilidades desejadas (parâmetro de discriminação); o nível de dificuldade de cada item, permitindo estabelecer uma classificação entre fáceis, médios e difíceis, por exemplo (parâmetro de dificuldade); e também a probabilidade de acerto casual para cada um dos itens presentes no teste.

A relação existente entre a probabilidade $P(U_{ij} = 1|\theta_j)$ de acerto de um item i e os parâmetros a , b e c do modelo é representada na Figura 1, por meio da Curva Característica do Item – CCI.

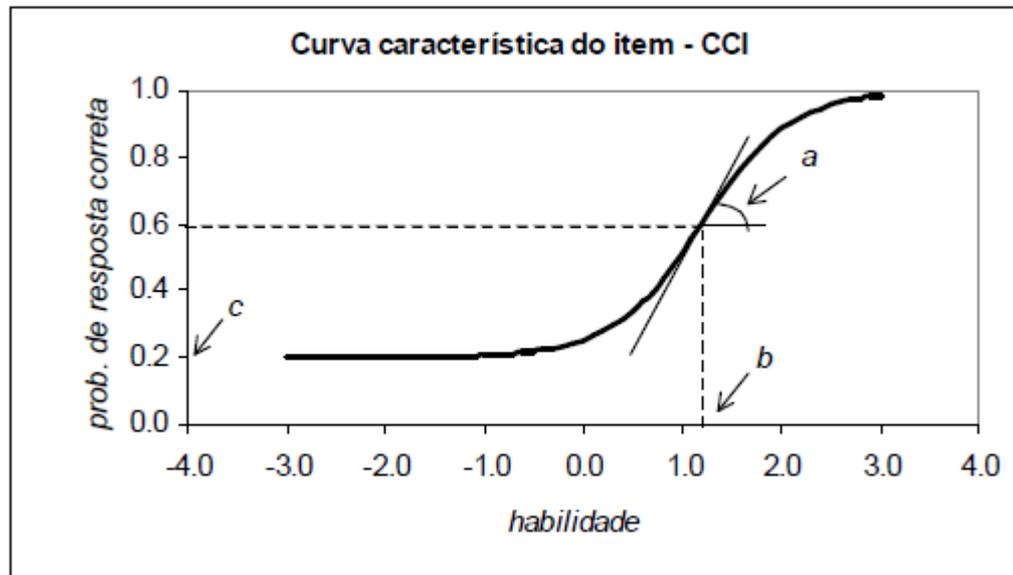


Figura 1: CCI - Curva característica do item

A escala da habilidade é uma escala *arbitrária* onde o importante são as relações de ordem existentes entre seus pontos e não necessariamente sua magnitude. O parâmetro b é medido na mesma unidade da habilidade e o parâmetro c não depende da escala, pois se trata de uma probabilidade, e como tal, assume sempre valores entre 0 e 1.

Na realidade, o parâmetro b representa a habilidade necessária para se obter uma determinada probabilidade de acerto. Assim, quanto maior o valor de b , mais difícil é o item, e vice-versa.

Baixos valores de a indicam que o item tem pouco poder de discriminação (alunos com habilidades bastante diferentes têm aproximadamente a mesma probabilidade de responder corretamente ao item) e valores muito altos indicam itens com curvas características muito “íngremes”, que discriminam os alunos basicamente em dois grupos: os que possuem habilidades abaixo do valor do parâmetro b e os que possuem habilidades acima do valor do parâmetro b .

Uma medida bastante utilizada em conjunto com a CCI é a *função de informação do item*, que permite analisar quanto um item contém de informação para a medida de habilidade. Tal função exprime a importância que têm os três parâmetros sobre o montante de informação do item. A informação é maior:

- (i) quando b_i se aproxima de θ_j ;
- (ii) quanto maior for o a_i ;
- (iii) e quanto mais c_i se aproximar de 0.

Segundo o modelo em questão, a informação fornecida pelo teste (conjunto de itens) é simplesmente a soma das informações fornecidas por cada item que compõe o mesmo.

3. A avaliação e as decisões do professor

O Modelo Logístico de 3 Parâmetros – ML3, apresentado nesse texto e amplamente utilizado em avaliações de larga escala, é adequado sobretudo quando deseja-se comparar o desempenho de diferentes populações de respondentes de um teste/prova realizada em larga escala. Utilizar o modelo em um contexto escolar, restrito a uma única sala de aula, não é recomendado por conta do número reduzido de respondentes e, na prática, o seu uso é também inviabilizado pela complexidade dos métodos matemáticos envolvidos (ANDRADE, 2000, p. 135). Diante disso, a proposta desse trabalho é aproximar os conceitos dos parâmetros a , b e c , discutidos no item 2 desse texto, ao trabalho de avaliação realizado pelo professor em sala de aula.

Considerando que o impacto da avaliação no processo de ensino/aprendizagem não é desprezível, não se pode desconsiderar o aluno e seus erros, para que a avaliação seja um meio de estudo e análise dos fenômenos de ensino e de aprendizagem (ALMOULOU, 2007, p. 104).

Segundo De Ketele apud Almouloud (2007, p. 104),

Avaliar consiste em recolher um conjunto de informações pertinentes, válidas e fiéis para examinar a adequação entre esse conjunto de informações e um conjunto de critérios pertinentes na escolha. Ou seja, a avaliação é o conjunto de procedimentos e de processos de coleta, de tratamento e de comunicação de informações feitos para tomar decisões.

Por meios dos parâmetros do ML3 o professor tem a possibilidade de tomar decisões que irão impactar em sua prática avaliativa. Nos próximos parágrafos, é explicitado como o professor pode fazer uso desses parâmetros para melhorar seus instrumentos de avaliação. A ordem escolhida para apresentar os parâmetros deu-se pela complexidade de sua utilização no cotidiano do professor.

No caso do parâmetro c , que se refere à probabilidade de acerto casual, o professor, ao elaborar questões de múltipla escolha, por exemplo, deve analisar qual a probabilidade de acerto casual seus alunos terão ao responder as questões. Essa probabilidade varia de acordo com o número de alternativas de resposta apresentado pela questão. Quando o professor escolhe, por exemplo, elaborar 5 alternativas de resposta ao invés de 3, está alterando a probabilidade de aproximadamente 33,3% para 20% de acerto casual, ou seja,

os alunos que não adquiriram a habilidade avaliada passam a ter 20% de chance de acertarem a questão, contra os 33,3% em questões com 3 alternativas. Dessa forma, fica evidente que o professor deve decidir quantas alternativas fixará para cada questão, levando em conta seus objetivos com um dado instrumento de avaliação, considerando a probabilidade de acerto casual.

Quanto ao parâmetro b , de dificuldade do item, ao elaborar o instrumento de avaliação contendo questões de múltipla escolha, é comum que os professores atribuam níveis de complexidade aos itens tomando como referência a sua própria experiência como docente ou então questões presentes em livros didáticos e outros materiais que o professor usa como referência e que trazem níveis de dificuldades declarados, por exemplo, como fácil, médio ou difícil. Contudo, após realizar a avaliação, por meio das respostas dos alunos, o professor pode determinar, por exemplo, quais foram as questões fáceis, médias e difíceis baseando-se nas respostas dos alunos e não em um critério subjetivo, como a experiência do próprio professor. Considerando todas as respostas dadas a uma determinada questão, é possível determinar o índice de acerto, por exemplo, calculando-se a razão entre o número de acertos em relação ao total de respostas. Com isso, o professor terá informações precisas sobre quais foram as questões que mais os alunos acertaram e as que menos acertaram, podendo considerá-las, por exemplo, respectivamente, como fáceis e difíceis. De posse de tais dados, fica clarificado que o professor está analisando as repostas dadas às questões e não o resultado da prova como um todo. Tal fato evidencia uma das principais características da Teoria de Resposta ao Item, ou seja, os itens/questão são seus elementos centrais de análise (ANDRADE, 2000, p. 3).

O parâmetro a , de discriminação do item, pode ser determinado separando-se os respondentes em dois grupos. Para facilitar a escrita desse texto, denominaremos G1 o grupo composto pelos alunos de melhor rendimento em dado teste e G2 o grupo composto por alunos de pior rendimento no mesmo teste. Podemos utilizar como métrica, por exemplo, 15% do total de alunos compondo G1 e outros 15% compondo G2. A discriminação de um item é calculada da seguinte forma: dado um item, determinam-se as razões, em G1 e em G2, entre o número de respostas certas em relação ao total de respostas; a diferença entre essas razões é a discriminação do item. Quanto menor for essa diferença, menos discriminativo será o item. Quanto maior for a diferença obtida, melhor será o item para avaliar quem realmente possui a habilidade avaliada e quem realmente não a possui.

Os resultados obtidos para os parâmetros b e c podem servir para: (i) o professor analisar os problemas de aprendizagem presentes nas respostas das questões em cada um dos níveis de dificuldade; (ii) o professor verificar se sua primeira análise sobre a complexidade das questões se concretizou ou foi refutada pelas respostas dos alunos; (iii) atribuir o parâmetro de dificuldade para o item elaborado e utilizá-lo em futuros instrumentos de avaliação, tendo o parâmetro determinado por uma população de alunos que possuirá características semelhantes, como iguais série escolar, bimestre, conteúdos curriculares; e (iv) atribuir o parâmetro de discriminação do item, podendo reelaborar ou desconsiderar os itens pouco discriminativos nas próximas avaliações aplicadas a uma população semelhante e priorizar a aplicação de itens com alto índice de discriminação.

Vale ressaltar que, nos três casos mostrados, o foco está no item/questão elaborada, não se referindo à avaliação inteira realizada pelo indivíduo. É importante, ainda, esclarecer que o professor não estaria utilizando a Teoria de Resposta ao Item em seu trabalho avaliativo feito em sala de aula, mas sim estaria aplicando os mesmos conceitos dos parâmetros do ML3, aqui constituídos como mais três elementos de informações pertinentes ao processo avaliativo, como já destacado nesse texto por De Ketele apud Almouloud (2007, p. 104).

4. Considerações finais

Os parâmetros do modelo logístico de 3 parâmetros carregam informações úteis que, em conjunto com o sistema avaliativo adotado pelo professor, podem contribuir para uma avaliação norteadora dos processos de ensino e de aprendizagem de alunos.

Por meio da perspectiva apresentada, podemos afirmar que os parâmetros de discriminação, dificuldade e acerto casual, fornecem ao professor a possibilidade de analisar as respostas dos alunos com foco nas questões. Tal possibilidade não implica em abandonar o olhar para o resultado individual e global da prova, mas sim complementar essa análise. De posse dos parâmetros de dificuldade e discriminação, aliados a demais processos avaliativos que ocorrem em sala de aula, o professor pode promover decisões que retroalimentarão novas informações que, por sua vez, promoverão o tratamento da ação educativa em conformidade com os objetivos a que ela se propõe.

Nesse texto, foram apresentados os resultados parciais de um projeto de pesquisa de mestrado que tem como finalidade investigar sistemas de avaliação em educação

matemática, por meio de recursos diversos que estão disponíveis para o professor da educação básica.

5. Agradecimentos

Agradeço à CAPES, pelo incentivo a mim proporcionado por meio da bolsa de estudo; à grande colaboradora desse trabalho Profa. Dra. Deise Aparecida Peralta.

6. Referências bibliográficas

ALMOULOUD, Saddo Ag. *Fundamentos da Didática da Matemática*. Paraná: UFPR, 2007.

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

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