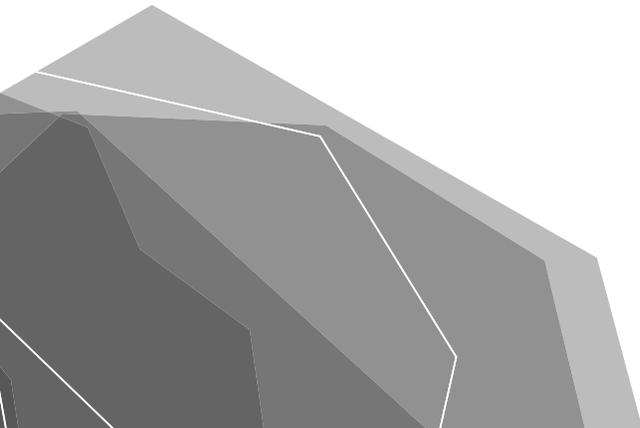


**RE
SE
NHA**

review



CONSTRUINDO TESTES: COMO ELABORAR E VALIDAR ITENS DE MÚLTIPLA ESCOLHA

DEVELOPING TESTS: HOW TO CREATE AND VALIDATE MULTIPLE CHOICE ITEMS

CONSTRUYENDO TESTS: CÓMO ELABORAR Y VALIDAR ÍTENS DE OPCIÓN
MÚLTIPLE

Alessandra Ramos de Oliveira Harden

Haladyna, Thomas M. *Developing and validating multiple-choice test items.*

Publicado pela primeira vez em 1994 e ainda, para nosso azar, sem tradução para o português, *Developing and validating multiple-choice test items*, de Thomas Haladyna pode ser considerado como leitura fundamental para qualquer um que trabalhe com avaliação educacional. De fato, o autor já esclarece, logo na introdução, que escreveu para “aqueles seriamente interessados na elaboração de itens para a avaliação de desempenho” (p. viii).¹ Esses “interessados” são colocados por Haladyna em dois grupos: o primeiro é formado por estudantes universitários das áreas de mensuração educacional, que têm na leitura do livro a oportunidade de compreender melhor as duas fases da elaboração de testes, que seriam a elaboração ou construção propriamente dita do teste e a validação das respostas dadas. O segundo grupo é constituído por profissionais que trabalham diretamente com a elaboração de itens, para quem certamente Haladyna apresenta amplo material para ajudá-los a aprimorar sua prática. Ou seja, esse é um livro também para todos os professores.

Sem dúvida, o que dá o tom ao livro é a familiaridade do autor com o tema e a propriedade com a qual lida com as muitas questões ligadas à avaliação de desempenho. Isso não é surpresa para quem sabe da vasta experiência do

¹ Todas as traduções de trechos do livro aqui citados são de minha autoria.

autor. Professor emérito da Arizona State University, Haladyna tem uma longa lista de publicações que demonstram sua dedicação à psicologia educacional, à psicometria e a outros tópicos ligados à elaboração de testes (para aplicação tanto em larga escala quanto para turmas pequenas). O saber adquirido com os anos de trabalho é mencionado como uma das justificativas para a terceira edição de *Developing and validating multiple-choice test items*:

[...] meus mais de 30 anos de experiência com planejamento, aplicação e avaliação de sistemas de testagem, e com ensino no nível fundamental, na graduação e na pós-graduação me ajudaram a entender melhor o processo de elaboração de itens, as vantagens trazidas por itens bem construídos e a importância da validação das respostas dadas (p. viii).

Essa vivência de Haladyna em ambientes ligados ao ensino se manifesta não apenas na riqueza de informação e no cuidado com o detalhe, mas também na qualidade didática do próprio texto. As quase 300 páginas são escritas de maneira clara e organizada e formam capítulos bem concatenados, ao longo dos quais Haladyna vai gradual e pacientemente guiando o leitor pelo universo das avaliações com base em itens de múltipla escolha (ME). Pouco a pouco, o conteúdo denso é apresentado, e até tópicos mais espinhosos começam a fazer mais sentido. A receita do sucesso do professor: ele define até mesmo os conceitos mais básicos e adiciona ao texto tabelas e quadros sinópticos, além de presentear o leitor com pequenas introduções e sumários em cada capítulo.

É assim, portanto, que está estruturado cada um dos onze capítulos do livro, que são precedidos por uma introdução. Nela, Haladyna explica que a terceira edição é uma versão revisada e melhorada das outras duas edições e expõe os motivos para tanto: o contínuo interesse dos leitores, a necessidade de ampliar a pesquisa sobre a construção de itens, o desenvolvimento na construção de itens de ME e a já mencionada convicção do autor nos benefícios de itens bem feitos e na relevância da validação. Encontram-se também nesse prefácio breves considerações sobre o que Haladyna considera limitações do livro, sobre o então atual estado dos testes de ME, além da indicação do tipo de leitor a quem o livro é dirigido, conforme citado acima, e uma apresentação do modo como o livro está organizado, com os capítulos divididos em quatro partes.

Um ponto que chama atenção ainda nesse trecho preambular é que o professor Haladyna critica a ênfase dada ao uso da ME, sugerindo que possa ter sido demasiada. No entanto, ele reconhece o papel crucial desse tipo de item tanto nos Estados Unidos, país cuja realidade guia suas considerações, quanto em outras partes do mundo. De fato, é inegável o amplo e generalizado uso da ME, que vem sendo empregada com objetivos diversos, como seleção de estudantes e de profissionais, certificação, concessão de créditos de cursos, atribuição de notas e, claro, avaliação da aprendizagem. Acertadamente, o autor não se limita a nenhum desses usos em particular, mas, ao contrário, declara que uma das premissas de seu livro é o reconhecimento do valor de itens de ME em sala de aula, em avaliações educacionais em larga escala ou em testes de proficiência profissional.

A primeira seção do livro, intitulada “A foundation for multiple-choice testing” (Fundamentos dos testes de ME), é dedicada a apresentar e contextualizar os testes que são o objeto de interesse de Haladyna na obra. São três os capítulos aí agrupados. No primeiro, “The importance of item development for validity”, o autor apresenta conceitos básicos para que os leitores possam compreender a extrema relevância da validade na construção de itens. Entende-se validade como “um processo lógico no qual definimos o que estamos medindo, criamos as medidas para fazê-lo, coletamos dados e avaliamos esses dados no que se refere à validade de interpretação de um escore de teste e seu uso subsequente” (p. 3). O mais fundamental dos conceitos é o de item, ou *test item*, e o autor se preocupa a logo defini-lo e caracterizá-lo:

[...] um item é a unidade básica observável de qualquer teste que geralmente contém uma afirmação que provoca ou exige uma resposta por parte do avaliando. Essa resposta recebe um valor numérico específico, quase sempre 1 quando está correta e 0 quando está errada, mas pode também ser alocada em uma escala de valores indo do baixo ao alto (p. 3).

O capítulo segue com esclarecimentos acerca de noções essenciais sobre as fases de construção de itens de ME. O leitor encontra aí uma lista das tarefas que cabem aos responsáveis pelo instrumento de avaliação, uma espécie de *checklist* das etapas a serem seguidas: “1 – Faça um planejamento para a elaboração dos itens”; “2 – Crie um cronograma para a elaboração dos itens;” e assim por diante (p. 14). Há também uma tabela com as fases do processo de validação dos itens, o qual Haladyna divide em formulação, explicação e validação (p. 18).

O segundo capítulo — “Content and Cognitive Processes” — está ligado à identificação do que se deseja medir com um instrumento de avaliação, ou seja, à definição dos conteúdos disciplinares e dos processos cognitivos que devem guiar a elaboração dos itens. Há uma taxonomia segundo a qual a aprendizagem pode ser compreendida (conhecimento, competências e habilidades), e o texto é enriquecido com copiosos exemplos.

No capítulo três, de título “Item formats”, o foco é a diferença entre tipos (ou formatos) de itens de teste, a qual reside no fato de o construto medido ser abstrato ou concreto. O aprendizado definido em termos concretos é designado também como de baixa inferência e aquele definido em termos abstratos é de alta inferência. Haladyna propõe distinções importantes entre os dois tipos de formatos em uma tabela cuja clareza e praticidade a tornam muito útil. Mais adiante no capítulo, a atenção se volta à avaliação do formato de item com base na validade, e o autor expõe seis argumentos de validade que servem para verificar a propriedade do uso dos formatos de ME em contextos específicos.

“Developing MC test items” (Desenvolvimento de itens de ME) é o nome da segunda seção do livro. Nela estão contidos quatro capítulos que formam um verdadeiro guia prático para a construção de itens. O capítulo quatro (“Item formats”) consiste em um estudo de oito formatos de ME, com indicações de conteúdos e processos cognitivos para cuja mensuração cada um deles é mais recomendável e discussão de vantagens e desvantagens de seu uso. Novamente, o autor adiciona muitos exemplos.

O capítulo cinco tem o título “Guidelines for developing MC items” e é realmente isso: uma coletânea de 31 diretrizes dirigidas a elaboradores de itens de ME. Abrange desde cuidados gerais a serem observados com relação a conteúdo até aspectos da escrita de enunciados e distratores (por exemplo: cada item deve refletir somente um tópico do conteúdo e somente um processo cognitivo; os distratores devem ser plausíveis etc.). Além disso, o texto traz orientações específicas para cada um dos formatos de itens apresentados no capítulo anterior. Trata-se de material precioso especialmente para os que atuam na revisão técnica e linguística de itens. Sem dúvida, pode ajudar a diminuir o grande número de itens descartados em qualquer processo de elaboração e validação de instrumentos de avaliação.

No capítulo seguinte, o autor demonstra a aplicação do conteúdo tratado até aí, pois expõe o leitor a uma série de exemplos de itens, discutindo-os com o objetivo de evidenciar a variedade de formatos que os itens de ME podem assumir, dependendo da área de conhecimento, do que se deseja medir e, claro, da criatividade dos elaboradores envolvidos.

Já o capítulo sete (“Item generation”) é motivado pela necessidade contínua que se tem de gerar itens cada vez melhores. Diante desse desafio, Haladyna argumenta, “qualquer estratégia que aumente a qualidade dos itens e a velocidade de produção é bem-vinda” (p. 148). Nessa linha, o professor comenta brevemente o futuro das teorias de geração de itens para em seguida descrever, exemplificar e avaliar técnicas utilizadas para ajudar elaboradores e diminuir o risco de descarte.

A terceira seção está voltada para a validação de itens. Intitulada “Validity evidence arising from item development and item response validation” (Evidência de validade com base na elaboração do item e na validação da resposta ao item), é composta por três capítulos que versam sobre aspectos do processo de validação da resposta dada ao item.

No capítulo oito (“Validity evidence coming from item development procedures”), o autor recorre mais uma vez a uma tabela, em que lista seis regras ou princípios referentes à qualidade dos itens. Após isso, divide o capítulo em duas subseções: uma dedicada a questões gerais, em que aborda tópicos como a definição de conteúdo, a elaboração de guias para escrita de itens, a seleção e o treinamento de elaboradores, e a segurança; e outra, em que se concentra na revisão dos itens. Nesta, são elencadas oito tipos de revisão que devem ser executadas antes da pré-testagem do material.

No capítulo nove, “Validity evidence coming from statistical study of item responses”, Haladyna dá início à discussão do tratamento estatístico das respostas aos itens, revisando aspectos da Teoria Clássica dos Testes e da Teoria de Resposta ao Item. Para tanto, utiliza três abordagens “diferentes, mas complementares” (p. 202) da resposta ao item: a tabular, a gráfica e a estatística. Por fim, propõe orientações para a avaliação de itens com base na combinação do nível de dificuldade com o de discriminação atribuídos a cada item.

No capítulo dez, a preocupação é avaliar quatro pontos da análise de respostas ao item que podem colocar em risco a validade. Esses problemas são abordados individualmente em subcapítulos, em que o autor indica e discute material bibliográfico.

Na última seção da obra, “The future of item development and item response validation” (O futuro da elaboração de itens e da validação da resposta ao item), formada pelo capítulo onze (“New directions in item writing and item response validation”), o autor tece comentários sobre os temas que, na sua opinião, mais podem influenciar o futuro das grandes áreas que são objeto do livro: as políticas educacionais, a noção de validade, o desenvolvimento da psicologia cognitiva e a forma como é definido o resultado da escolaridade e da capacitação profissional. Ele discute ainda novas teorias e direcionamentos referentes à elaboração de itens.

É inegável que *Developing and validating multiple-choice test items* tem muitas e importantes qualidades. A longa lista de referências bibliográficas é mais um ponto que justifica sua classificação como essencial. Há obviamente algumas falhas que devem ser observadas e com as quais os leitores devem tomar cuidado, mas que estão mais ligadas à passagem do tempo e ao desenvolvimento da tecnologia e das pesquisas da área do que a alguma propriedade intrínseca do texto. Se a publicação foi feita em 2004, a leitura exige a complementação das informações com textos mais recentes, especialmente quanto ao emprego de tecnologia digital para a elaboração e aplicação de instrumentos de avaliação e para o tratamento de dados estatísticos. Além disso, o crescimento, no Brasil, das pesquisas voltadas às áreas focadas no livro e o aprimoramento por que passaram nossos sistemas de avaliação também tornam imprescindível o cuidado do leitor para que os ensinamentos do professor Haladyna possam ser relacionados à nossa realidade atual.

Alessandra Ramos de Oliveira Harden

Doutora em Estudos Hispânicos e Lusófonos pela University College Dublin, Irlanda
Professora da Universidade de Brasília
oliveira.ales@gmail.com