

Criterion-Referenced Language Testing, de
James Dean Brown e Thom Hudson.
Cambridge University Press, 2002. 320p.

Ricardo Augusto de Souza
Universidade Federal de Minas Gerais

A obra *Criterion-referenced Language Testing*, escrita em co-autoria por James Dean Brown e Thom Hudson, aborda aspectos conceituais e discute dimensões práticas relativas à avaliação referenciada no critério no contexto do ensino de línguas estrangeiras. Trata-se de uma obra lançada pela editora *Cambridge University Press* no mercado editorial em 2002, como título da série *Cambridge Applied Linguistics*, que conta como editores Michael H. Long e Jack C. Richards. O livro é suplementado por um sítio de apoio na Internet, contendo resumos dos principais pontos de cada capítulo, listagem de termos-chave e exercícios, cujo endereço eletrônico é www.cambridge.org.uk/elt/crlt/.

Os autores, ambos professores da Universidade do Havaí em Manoa, definem os testes referenciados em critérios (TRC) como procedimentos avaliativos cuja meta é averiguar evidências de capacidade de desempenho oriundo da aquisição de um conjunto de habilidades e conhecimentos específicos ou da capacidade de desempenho de tarefas relevantes, das quais o instrumento de avaliação constitui um mecanismo de amostragem e um instrumento de mensuração. Trata-se, portanto, de uma estratégia de avaliação adequada a situações de ensino de línguas nas quais se busca a verificação da aprendizagem de conteúdos ou capacidades previstas em parâmetros curriculares, assim como a situações de avaliação formal nas quais almeja-se a verificação de habilidades lingüístico-comunicativas correlatas à atuação potencial em domínios ou contextos semelhantes aos representados no teste. Ainda, a leitura da obra corrobora a compreensão de que a avaliação lingüística referenciada no critério compreende um conjunto de abordagens diferenciado dos testes

referenciados em normas (TRN),¹ pois nesta última busca-se a comparação entre os desempenhos individuais dentro de uma corte de sujeitos submetidos ao mesmo instrumento de avaliação formal, com vistas à distribuição normal dos candidatos, baseada na diferenciação de sua proficiência geral.

A obra é desenvolvida ao longo de sete capítulos que seguem a dois prefácios (p. ix-xvi), o primeiro assinado pelos editores da série *Cambridge Applied Linguistics* e o segundo escrito pelos dois autores. De maneira geral, pode-se perceber a existência de dois grandes blocos temáticos em torno dos quais os sete capítulos estão organizados. O primeiro bloco contempla fundamentos conceituais da avaliação formal e da avaliação lingüística, assim como critérios, procedimentos e metodologias para o processo de avaliação, contemplando desde o diagnóstico de necessidades de avaliação formal até a administração de testes e o retorno de resultados. O segundo bloco trabalha com aspectos técnicos e fortemente embasados em métodos quantitativos específicos à teoria de medida.² Os aspectos discutidos no segundo bloco enfocam tanto a verificação da qualidade e eficácia de itens para testes quanto o exame e operacionalização das noções de confiabilidade e validade. Os três primeiros capítulos e também o sétimo e último capítulo podem ser vistos como articulados ao primeiro bloco temático. Os capítulos quatro, cinco e seis, por sua vez, podem ser tomados como constituindo o segundo bloco de temas.

O primeiro capítulo, *Alternate paradigms*, apresenta detalhadamente os TRN e os TRC como paradigmas de avaliação lingüística. Os autores enumeram definições relevantes para os dois modelos e expõem uma comparação sistemática de ambos. O capítulo é concluído com considerações sobre a adequação dos TRC como instrumentos de

¹ O uso nesta resenha das expressões “testes referenciados em critérios” e “testes referenciados em normas” como equivalentes a *criterion-referenced testing* e *norm-referenced testing*, respectivamente, baseia-se nas ocorrências de tais expressões em Depresbiteris (1999).

² Além da obra ora resenhada, acreditamos que outras introduções aos fundamentos da teoria da medida adequadas ao leitor sem treinamento aprofundado em matemática ou estatística poderão ser encontradas no capítulo 2 de Bachman (1990) e no capítulo 2 de Pasquali (2003).

avaliação compatíveis com as necessidades de contextos específicos de aprendizagem e uso de línguas estrangeiras e também compatíveis com teorizações acerca da competência comunicativa.

O capítulo dois, intitulado *Curriculum-related testing*, discute as contingências da avaliação lingüística nos contextos escolares e de programas de ensino de línguas estrangeiras. Os autores procuram discutir as situações nas quais, no cenário da instrução formal, tornam-se necessários o emprego de instrumentos de avaliação lingüística e as funções que a avaliação pode desempenhar, tais como diagnóstico inicial na análise de necessidades, diagnóstico final de aprendizagem e nivelamento. Ainda, Brown e Hudson defendem a maior adequação dos TRC, comparativamente aos TRN, para estes contextos. Ao término do capítulo, os autores expandem a noção de avaliação referenciada no critério para além do emprego de testes propriamente ditos, demonstrando a compatibilidade entre vários dos princípios subjacentes aos TRN e estratégias alternativas de avaliação, tais como portfólios e projetos.

O capítulo intitulado *Criterion-referenced test items* é o terceiro na obra de Brown e Hudson. Nesse capítulo, os autores apresentam uma tipologia dos itens para avaliação lingüística compatível com os TRC e várias linhas de orientação para sua construção e análise. São contemplados fatores textuais da apresentação dos itens, apresentação essa que é corretamente enquadrada como evento discursivo, assim como considerações sobre diversas modalidades de formatação. O capítulo é encerrado com indicações sobre procedimentos de desenvolvimento de um instrumento de avaliação do tipo TRC, indicações essas que reiteram a perspectiva de grande adequação desse modelo a objetivos de programas instrucionais e contextos específicos de uso de línguas discutida no capítulo anterior. Os autores argumentam que tal adequação se dá justamente em função da possibilidade de que itens de um TRC sejam desenvolvidos com vistas à obtenção de altos níveis de compatibilidade com objetivos avaliativos específicos.

Os três primeiros capítulos são freqüentemente ilustrados por exemplos e relatos de casos colhidos da experiência profissional dos autores no ensino do inglês como segunda língua. Em nosso entender, tal recurso a casos reais é um dos principais méritos da obra, pois ele realiza uma aproximação entre elaboração conceitual e considerações sobre contingências vivenciadas na atuação profissional dos autores que

poderá ser facilmente transferida para contextos similares e certamente familiares a um número grande de professores de línguas estrangeiras.

O quarto capítulo é intitulado “*Basic descriptive and item statistics for criterion-referenced tests*”. Nele, os autores buscam demonstrar como a construção de um instrumento do tipo TRC eficaz pode beneficiar-se de análises quantitativas para a averiguação da adequação dos itens que comporão o teste. Inicialmente, são apresentados conceitos e procedimentos fundamentais de estatística descritiva de testes educacionais, como média, mediana, moda, estimativas de facilidade e estimativas de discriminação de itens. A seguir, os autores conduzem-nos ao entendimento do valor da análise estatística para a verificação da eficiência de itens individuais para os objetivos de uma dada avaliação formal. Brown e Hudson defendem a aplicabilidade de tais conceitos e procedimentos aos objetivos dos TRC, uma vez que sua utilidade pode parecer mais evidente para os objetivos dos TRN. Os autores apresentam procedimentos estatísticos específicos que melhor vêm ao encontro dos objetivos dos TRC.

O capítulo quatro culmina com uma apresentação da Teoria da Resposta ao Item e dos desenvolvimentos por ela trazidos à construção e à análise de bancos de itens para a avaliação educacional. Segundo Pasquali (2003), a contribuição da Teoria da Resposta ao Item à psicometria contemporânea vincula-se ao fato de que ela contempla a noção de traço latente (em contraponto a traço comportamental observável), adotando como axiomas a noção de que o desempenho em uma tarefa seja causado por traços latentes e a noção de que a relação entre o sucesso no desempenho de uma tarefa e o conjunto dos traços latentes a ele subjacentes pode ser descrita matematicamente. Em convergência com tais argumentos, Brown e Hudson demonstram a utilidade dos métodos estatísticos ligados à Teoria da Resposta ao Item no estabelecimento de estimativas da probabilidade de que o acerto em um dado item seja decorrente do construto da habilidade/capacidade a ele subjacente, ou seja, a um traço latente como, por exemplo, a capacidade de recuperação de referentes em situações de substituição lexical em um teste de leitura em língua estrangeira.

A aplicação de métodos da Teoria da Resposta ao Item, segundo Brown e Hudson, possibilita o desenvolvimento de bancos de itens para testes que podem ser selecionados em função de características

específicas que sejam adequadas ao nível de habilidade esperado em uma dada situação, sendo, portanto, claramente compatível com os propósitos da TRC. Igualmente, os autores apresentam aplicações de tais métodos para o estabelecimento preciso de vínculos entre a dificuldade de itens e os níveis de habilidade esperados em situações de avaliação de desempenho nas quais atuam múltiplos juízes, tais como entrevistas orais e avaliações de produção textual. A exposição da Teoria da Resposta ao Item no capítulo quatro talvez seja aquela que mais exige do leitor pouco familiarizado com Teoria da Medida na obra de Brown e Hudson. Contudo, acreditamos que o maior valor dessa exposição não esteja em sua utilidade como texto didático introdutório aos avanços recentes em métodos estatísticos, mas, sim, em apontar para o lingüista aplicado ou professor de línguas interessado em avaliação lingüística o quanto tais avanços podem contribuir para seu objeto de interesse, no mínimo encorajando-o a buscar o fortalecimento de possíveis diálogos profissionais com os especialistas da área de medidas educacionais e de psicometria.

O capítulo cinco, intitulado “*Reliability, dependability, and unidimensionality*”, trata da questão da obtenção de estimativas de consistência de um dado conjunto de escores de um instrumento de avaliação. No capítulo, procedimentos para estimativa da confiabilidade apropriados em metodologias de avaliação que instanciam direcionamentos típicos da TRN, da TRC e das aplicações da Teoria da Resposta ao Item aos testes educacionais são discutidos. Os autores demonstram que as estimativas de fidedignidade apresentam diferenças em cada uma dessas situações metodológicas, e associam os termos *reliability*, *dependability* e *fit*, referentes a modelos de conceituação da confiabilidade adequados aos quadros da TRN, TRC e Teoria da Resposta ao Item, respectivamente. Os autores detalham procedimentos estatísticos compatíveis com as três conceituações. Na TRN, em que há a pressuposição de variação de escores e de distribuição normal dos mesmos, ocorre prevalência de procedimentos correlacionais. Na TRC, a confiabilidade é contemplada por procedimentos de mensuração que asseguram a confiabilidade dos escores como indícios de diferenciação entre aptos e inaptos. No contexto das aplicações da Teoria da Resposta ao Item, a noção de confiabilidade é associada ao nível de conexão entre um escore e o traço latente inferido pela habilidade de execução do item.

O sexto capítulo é intitulado “*Validity of criterion-referenced tests*”. Tal como indicado no título, nesse capítulo os autores tratam da questão da validade na avaliação referenciada na norma. A noção de validade de um teste é usualmente compreendida como a acuidade com a qual um dado teste realiza a mensuração do fator que ele pressupõe medir. O capítulo cobre inicialmente as aplicações aos itens de um instrumento de TRC da noção de validade de conteúdo, que é associada com a representatividade de um dado item em relação ao domínio ou situação na qual se dá o emprego de uma habilidade, sendo que esse domínio ou situação é externo ao teste. No tocante à validade de conteúdo, duas abordagens de sua operacionalização, a abordagem orientada por argumentos teóricos e a abordagem orientada por parecer especializado, são examinadas. Passa-se então à problematização da noção de validade de construto, noção essa associada à representatividade dos itens de um teste em relação aos construtos psicológicos subjacentes à habilidade cuja verificação é almejada, assim como ao detalhamento de abordagens para sua operacionalização e estudo, com a exposição de procedimentos de tratamento estatístico compatíveis. O capítulo é concluído com um tratamento da questão acerca do vínculo entre validade de conteúdo e validade de construto, assim como sobre suas implicações na utilização de testes como instrumentos de tomada de decisão. Todos os tópicos abordados no capítulo seis são ilustrados com relatos detalhados de casos da experiência profissional dos autores.

Tal como apontado acima, os capítulos quatro, cinco e seis são aqueles com teor mais explicitamente técnico da obra, sendo claramente apoiados na apresentação de procedimentos estatísticos específicos. Na maioria das circunstâncias nas quais os autores se detêm em exposições de mecanismos matemáticos, tais exposições são mediadas pela apresentação de exemplos de suas aplicações, o que nos leva a acreditar que tornará tais passagens niveladas com as necessidades de um leitor sem plena familiaridade com a Teoria da Medida ou em estatística. Trata-se de um estilo de apresentação que reitera a competência e a clareza didática na apresentação de métodos quantitativos que guardam interesse para o profissional de ensino/aprendizagem de línguas que caracterizam outros trabalhos do primeiro autor, tais como Brown (1988) e Brown (2001). Contudo, enfatizamos que, em nosso entender, o principal mérito da obra não reside no que ela ensina de estatística, mas,

sim, no que ela revela acerca das potencialidades frutíferas da aproximação informada de seus leitores com esse campo.

O sétimo e último capítulo da obra, intitulado *Administering, giving feedback, and reporting on criterion-referenced tests*, encerra a visão panorâmica dos TRC proposta pelos autores, retomando a tônica do que sugerimos acima como sendo o primeiro bloco temático. Especificamente nele, os autores apresentam considerações gerais de ordem prática que objetivam o oferecimento de sugestões aos professores e coordenadores de programas de ensino de línguas que necessitem ou queiram envolver-se com a avaliação lingüística nos moldes dos TRC. Mantendo o recurso a relatos de experiências profissionais dos autores, que caracteriza toda a obra, são apresentadas recomendações para o gerenciamento de equipes de trabalho para a construção colaborativa de testes e para situações de retorno e devolução de processos avaliativos.

É nosso entender que o livro *Criterion-referenced Language Testing* é uma obra altamente bem-sucedida na apresentação minuciosa dos meandros de um tipo de avaliação que vem ao encontro das necessidades de elaboração e seleção de instrumentos de verificação de aprendizagem e/ou diagnóstico de competências que certamente são familiares a um grande número de professores e coordenadores de programas de ensino de línguas estrangeiras. Trata-se ainda de um trabalho que tem como um de seus méritos mais significativos sua eficiência exemplar na reunião de direcionamentos e sugestões de ordem prática com uma apresentação não superficial de dimensões técnicas e de natureza conceitual. Assim, a obra poderá levar docentes e estudantes em preparação para carreiras no ensino de línguas tanto a conhecer possíveis soluções para a prática da avaliação quanto a reconhecer o estudo da avaliação lingüística como um campo intelectualmente desafiador e instigante. Julgamos, portanto, que a obra é um livro indispensável ao acervo de bibliotecas que dão suporte a cursos de graduação e pós-graduação nos quais se ensinam os fundamentos da avaliação lingüística ou de bibliotecas de instituições em que atuam equipes profissionais responsáveis pela elaboração de instrumentos de avaliação e nas quais altos padrões de qualidade são almejados para esses produtos.

Referências Bibliográficas

BACHMAN, L. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press, 1990.

BROWN, J. D. *Understanding Research in Second Language Learning*. Cambridge: Cambridge University Press, 1988.

_____. *Using Surveys in Language Programs*. Cambridge: Cambridge University Press, 2001.

DEPRESBITERIS, L. *Avaliação educacional em três atos*. São Paulo: Editora SENAC São Paulo, 1999.

PASQUALI, L. *Psicometria – Teoria dos testes na Psicologia e na Educação*. Petrópolis: Editora Vozes, 2003.