

Book Review:

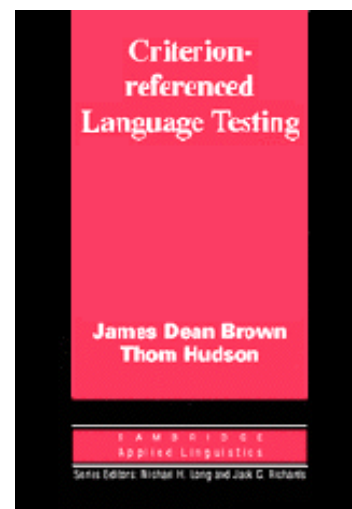
**Criterion-Referenced Language Testing
(Cambridge Applied Linguistics Series)**

by J. D. Brown & Thom Hudson (2002)

ISBN: 0-521-00083-1, hard: 0-521-80628-3

Cambridge, UK / New York:

Cambridge University Press (311 pages)



As the name implies, this book focuses on criterion-referenced language tests (CRTs). Hudson and Brown provide a theoretical and practical overview of what CRTs are, how they should be used, and how to design them. The main strength of this volume is in the way theoretical concepts are combined with actual test samples and data. Another positive feature is its online chapter summaries, vocabulary lists, review questions, and exercises. A salient weakness of this work is its scope: it is extremely broad and a large portion of the material is not directly about CRTs. Also, whereas some parts are for novices, others are for statistically savvy readers. A brief summary of each chapter follows.

Chapter 1. *Alternative Paradigms*

After contrasting several definitions of CRTs, the authors note how much ambiguity exists about what CRTs actually are. One source of confusion is the criterion used for CRTs: whereas some are based on a notion of a universal competence, others focus on specific curricula or course material. Another source of confusion arises from the diverse jargon in the testing field. Readers might wonder how CRTs differ from other types of assessment tools such as universe designed tests (Osborne, 1968), mastery-referenced tests (Berk, 1980), objectives-referenced tests (Nitko, 1984), domain-referenced tests (McCormick & James, 1988), or LSP tests (Douglas, 2000). For simplicity, these are regarded as variant forms of CRTs.

After mentioning how CRTs differ from norm-referenced tests (NRTs), some theoretical and practical perplexities in language testing today are highlighted. The lack of a broadly accepted model of communicative competence has resulted in widespread disagreement about CRT construction. Until deeper consensus about communicative competence is reached, many key points regarding communicative testing will spark debate.

Chapter 2. Curriculum-Related Testing

Exploring the testing-curricula interface, Brown and Hudson emphasize how assessment should occur at each phase of curricular design. The authors suggest that well-designed CRTs can ". . . provide the glue connecting the components of curriculum development to one another" (p. 36). The need to constantly monitor tests and see how well they fit curricular and student goals is underscored. Moreover, the value in distinguishing between instructional objectives (which should be precisely defined) and performance objectives (which are more general) is also stressed. This chapter concludes with a discussion of the role of washback and feedback. Well-designed tests should create positive washback and encourage feedback from multiple parties.

Chapter 3. Criteria-Reference Test Items

After formulating guidelines for test items based on Grice's maxims, distinctions between selected-response, constructed-response, and personal-response items are elucidated. Advice for creating test questions in each of these formats is mentioned. The need to develop test descriptors which are neither too narrow and restrictive nor too broad and vague is also underlined and practical advice about how to avoid linguistic and format confoundings (items which are unclear or poorly written) is expounded. Too many amateur tests are riddled with such confoundings.

The chapter concludes by describing ways to analyze item quality and content. Decisions about item quality and content draw heavily subjective judgments and a broad range of persons be involved in the decision making process to reduce the effects of subjectivity.

Chapter 4. Basic Descriptive and Item Statistics for CRTs

Some key statistical concepts needed to understand CRTs are discussed and various ways to interpret CRT and NRT test score distributions are outlined. A discussion of item analysis, item facility, and discrimination indices then follows and the importance of evaluating CRT test items through each of these statistical means is carefully explained. The chapter concludes by briefly discussing the usefulness of three different item response theory (IRT) models. After mentioning the respective merits of each model, practical ways IRTs can be used to generate equivalent forms through item banking are outlined.

"classroom teachers interested in grounding their testing practices more deeply in theory will find this work worthwhile."

Chapter 5. *Reliability, Dependability, and Unidimensionality*

Three aspects of test consistency are described in depth: reliability, dependability, and fit. After discussing how these concepts apply to NRTs, CRTs, and IRT, ways to enhance test consistency are outlined. A wide number of technical concepts are explained, and the authors suggest how each reliability estimate approach provides a partial clue as to how well constructed a test might be.

The most interesting part of this chapter is the discussion of how threshold-loss methods and G theory approaches to calculating dependability differ. The former methods use various formulas to ascertain the role of random chance in ascertaining examinees performance. The latter methods achieve this aim through various calculations involving phi (F) and confidence intervals.

An overview of three consistency issues in IRT concludes this chapter. Since large sampling sizes are required for two of the IRT methods outlined, some of this discussion is of limited utility to classroom teachers. Nonetheless, readers can start to appreciate some of the complexities involved in IRT.

Chapter 6. *Validity of CRTs*

The nature of test validity in general and construct validity in particular defines considered is the focus of this chapter. The implications of Messick's (1988) expanded views of validity, which note that validity does not lie in any test per se, but in expert judgments about its domain relevance and representativeness are explained. The authors caution that content relevance, representativeness, and technical quality should also bear weight in assessing construct validity. This chapter concludes with a discussion of frameworks to assess testing consequences. Negative aspects of washback are mentioned, along with ways to make CRT validity decisions and set more appropriate test standards.

Chapter 7. *Administering, giving feedback, and reporting on CRTs*

The final chapter of this book mentions some practical tips about how to get teachers to work on CRT items, ways to encourage CRT feedback, and how to report CRT results. A particularly interesting part of this chapter is a discussion of ways to handle students who intentionally perform poorly on tests. Noting how students might want to mask their true abilities under some conditions, the need to provide incentives for high performers was underscored.

Chapter 7 concludes with a discussion of ways to interpret gain scores. The need to promote the principles of well-constructed criterion-referenced testing among colleagues is also mentioned.

Conclusion

This 320-page volume is most likely to appeal to at least two types of readers. Its organization and online study questions make it well-suited to one-semester undergraduate courses on testing. And its thorough index also makes it a handy reference work for teachers working with institutional CRTs. Though parts of this text are apt to seem tedious if read from cover to cover, classroom teachers interested in grounding their testing practices more deeply in theory will find this work worthwhile.

- Reviewed by Tim Newfields

Works Cited

Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge, UK: Cambridge University Press.

McCormick, R. & James, M. (1988). *Curriculum evaluation in schools* (2nd ed.). Beckenham, UK: Croom Helm.

Nitko, A. J. (1984). Review of the book *A technology for test-item writing*. *Journal of Educational Measurement*, 14, 3-19.

Osborne, H. G. (1968). *Item sampling for achievement testing*. *Educational and Psychological Measurement*, 28, 95-104.

HTML: http://jalt.org/test/new_3.htm / PDF: <http://jalt.org/test/PDF/Newfields3.pdf>