
REVIEWS

The *TESOL Quarterly* welcomes evaluative reviews of publications relevant to TESOL professionals. In addition to textbooks and reference materials, these include computer and video software, testing instruments, and other forms of nonprint materials.

Edited by **HEIDI RIGGENBACH**
University of Washington

Recent Publications on Statistics, Language Testing and Quantitative Research Methods: II

Fundamental Considerations in Language Testing

Lyle Bachman. Oxford: Oxford University Press, 1990. Pp. xi + 408.

Assessing Achievement in the Learner-Centred Curriculum

Geoff Brindley. Sydney: National Centre for English Language Teaching and Research, 1989. Pp. vii + 183.

Testing for Language Teachers

Arthur Hughes. Cambridge: Cambridge University Press, 1989. Pp. vii + 176.

■ In the first part of this article (*TESOL Quarterly*, Vol. 23, No. 1), I reviewed four books, those by Brown (1988); Butler (1985); Henning (1987); and Woods, Fletcher, and Hughes (1986). I stressed the importance of the ability of applied linguists and language teachers to read the literature of their field evaluatively, without being intimidated by its quantitative aspects. I also stressed the importance for language teachers of reaching sound testing decisions that respond to specific educational contexts, either with existing tests or with teacher-developed exams. In this second section, I review three more volumes on testing.

Bachman's large work, *Fundamental Considerations in Language Testing*, begins with the statement that "this is not a 'nuts and bolts' text on how to write language tests" (p. 1); the reader will find this true. What, then, is it? Bachman describes it as a "discussion of fundamental issues that must be addressed at the start of any language testing effort" and an exploration of "some of the problems raised by [the fact that] language is both the instrument and the object of measurement" (p. 1). Bachman identifies three fundamental considerations: the contexts determining the uses of language tests; the nature of language abilities; and the nature of

measurement, focusing throughout on the measurement of language *proficiency*.

I begin with the last chapter because this is, for me, the most interesting and satisfying. It is also the chapter where Bachman speaks directly to his audience and identifies, as I do not think he does until that point, precisely who his audience is. Here Bachman discusses “authentic language” tests (p. 300), and develops a distinction between what he calls “real-life” (RL) and “interactional/ability” (IA) tests (pp. 301-302), which is immediately convincing and useful. Around this contrast Bachman builds a wide-ranging and valuable discussion of problems and directions in language testing, located firmly within the discourse community of language-testing researchers. The persistent problems and future directions he discusses, and the terms in which he considers them, are not those of teachers who must develop or select language tests for their own students. This chapter clarifies for me the uneasy feeling I had throughout the book, that of needing to identify the intended audience.

There is much of value in Bachman’s book for practicing teachers and students in TESOL courses. I suspect, however, that teachers, reading alone, will find the book hard going indeed, and that master’s degree students will be grateful for a well-prepared instructor to guide them through the text, in particular helping them identify concrete cases, problems, and data sets that can bring into perspective the theoretical issues discussed by Bachman. This need for supplementary example is reminiscent of Henning’s volume discussed in the first part of this review; however, the book is unlike Henning’s in most regards. Bachman allows himself full range to consider, at length, the fundamental issues that concern him, and is unabashedly willing to dismiss issues of less interest with but brief comment. This is a personal book, making no attempt at complete coverage, of everything a graduate student might want to know about language testing. While this approach makes the book interesting for the committed reader, it may present a problem for anyone designing a one-semester course in language testing who aims to cover specific ground in the way recommended in the first part of this review.

Chapter 1 is not, for me, the book’s strongest. The brief sections on second language acquisition research, language ability, and measurement theory leave me wondering what central point I should be carrying away to apply to my reading later in the book. It seems that several concerns are raised, only to be left hanging, while we wonder what position Bachman holds, or what he assumes us to have inferred from the discussion. A figure borrowed from

Chapter 6 attempts, it seems, to link the several short discussions of areas that especially interest Bachman to some sort of philosophical base or overall plan, but the reader without a solid grounding in these areas will find the discussion obscure. Placing the overview of the book earlier might have helped to clarify this chapter. In fact, the book is cohesive, but to learn this it is necessary to read all of it. This is not, in my view, a book that is made to dip into, in the way that many of us have become accustomed to doing with sourcebooks for testing courses.

The chapters that deal with measurement are excellent, although the incorporation of work with data into the text, or the provision of concrete examples for the theoretical arguments and explanations Bachman is presenting, would help a language-testing initiate make better sense of the discussion. But, as I have indicated above, I do not believe this is a book for the uninitiated. Chapters 2 and 3 cover more or less typical ground for a book on language testing, but, again, Bachman's individual concerns are evident throughout. The sections on "Steps in Measurement" in Chapter 2 and "Research Uses of Tests" in Chapter 3 are especially welcome. Bachman does not, however, attempt to introduce or teach any actual statistical techniques for working with quantitative data. The book holds true to the promise of its title in focusing on considerations, not applications.

Bachman takes (in Chapter 6) a wider view of reliability and threats to it than do many more traditional books on language testing; he provides a clear and quite detailed explanation of classical measurement theory, generalizability theory, and item response theory, while providing statistical formulae for concepts under consideration. In discussing validity (in Chapter 7), he follows recent practice in stressing construct validation. While he discusses correlational, factor analytic, and multitrait multimethod approaches to construct validation, he does not provide either the statistical formulae or any concrete examples of these.

In Chapters 4 and 5, Bachman introduces and develops in detail his framework for communicative language ability and "test method facets" (p. 115), which has already been influential in language testing through dissemination and discussion at testing meetings, and in its application in the Cambridge-TOEFL Comparability Study (Bachman, Davidson, & Foulkes, 1990). In large part the framework is presented descriptively and in language that appears to imply an a priori reality for the model. In fact, it could well provide researchers in language testing with researchable issues beyond the end of the century, and indeed Bachman himself has suggested that something of the sort should happen

(Bachman & Clark, 1987). Bachman's framework is theoretically interesting in a number of ways, but my own feeling is that it would have received the attention it deserves more easily if presented in a separate forum. Its presence in this more broadly construed text confirms for me that this is the book Bachman wanted to write, and that it must be read on that basis.

Brindley's *Assessing Achievement in the Learner-Centred Curriculum* was prepared as a report to the Australian Adult Migrant Education Program (AMEP), an Australian government program serving over 130,000 learners, primarily adult ESL learners, each year. Brindley makes no claims that this book could or should be used as a textbook for language-testing courses, and thus my review of it here may seem unfair. Nevertheless, Brindley has provided us with a powerfully informative example of language testing research and language tests as pedagogical problem solving in action. In the report, Brindley investigates the kinds of information on learner achievement that are needed at all levels of the AMEP, and the ways that fine-grained assessment of learners' language gains over courses of instruction can be carried out. As part of this, he looks at assessment methods in other contexts and considers their applicability or adaptability to the needs of the AMEP.

Brindley provides a fascinating look at teachers' responses to a survey of their views on the importance of assessment in their work. He found that placement into classes figured highest, with diagnostic information for use in course planning second. Looking at assessment methods favored by teachers, he found that informal methods ranked highest (with observation followed by recycling of work ranked most highly) and informal discussions with learners second; standardized testing ranked lowest. Teachers were very interested in learner self-assessment, although most of them felt underequipped in applying this. Interestingly, self-assessment is scarcely dealt with at all in the books I have reviewed here and in Part 1, although it is becoming an increasingly important movement, particularly in Europe and Australia.

Self-assessment is just one form of assessment found appropriate within the learner-centered educational philosophy held by the AMEP. The book contains the best discussion of learner profiling I have encountered, along with reference, although limited, to the *folio* of learner work that parallels the developments in portfolio assessment currently enjoying popularity in U.S. composition assessment. While Brindley's discussion of accountability may be rather obscure to those of us outside AMEP, it does demonstrate that while some of our language testing problems can be solved

with reference to a wider perspective of literature, research, and experience, others must perforce be solved more introspectively.

The bulk of Brindley's book is devoted to a detailed description of, and argument for, various criterion-referenced achievement assessment instruments, a wealth of which are presented within the middle chapters of the book, in a way that is both fully contextualized and firmly based theoretically. Contained in this book is enough project material for several language testing classes to explore in practice many of the issues raised in a more abstract way in Bachman's book. This volume should prove useful supplementary reading for any language testing course, (Teachers seeking another book to fill this supplementary niche might also examine Weir's *Communicative Language Testing* [1988].)

Hughes' *Testing for Language Teachers* starts with a clear specification of its audience and purpose: The book is intended "to help language teachers write better tests" (p. ix). The text is consistent with this purpose in its tone, level of complexity, and limits on its assumptions of prior knowledge and experience. I note, however, that there is almost no reference to statistics throughout the book, relegating statistical information that does appear to an appendix. I believe that teachers can handle more statistics than Hughes appears to give them credit for, and that indeed they must. How can teachers design effective tests if they have no tools with which to evaluate their efforts? With almost every section of this book, specialists in language testing will want more; but for short in-service courses and for practitioners using the book in their home contexts to provide guidance as they actually solve their testing problems, what Hughes offers is probably enough, except in limits he places on statistical tools. I suspect, however, that even those using the book in short courses will need to supplement the book with some manipulation of quantitative data.

Hughes' explanations of basic concepts are brief but clear and pertinent. In Chapter 3, I note in particular a helpful discussion of the differences between kinds of language tests, with welcome attention to achievement versus proficiency, and norm-referenced versus criterion-referenced tests, and a welcome dismissal of the objective/subjective testing myth in a few lines. The further readings are quite well chosen to offer more detail in each area but, in most cases, without becoming too technical. But the very limited extent of the discussions does cause me some concerns: As an example, Chapter 4 deals with validity. A sound grasp of validity is especially important for a practitioner audience, both because of its classroom impact and because many teachers will have previously

studied language testing, if they have studied it at all, in an older paradigm where validity was underemphasized relative to reliability. I would have liked to see more attention to the argument for an emphasis on validity in testing and test design, and some discussion of the operational implications of such an emphasis. Similarly, in Chapter 5 on reliability, the techniques discussed are fine, but given the stress placed in the book on teacher-made tests and the recurring advice (with which I agree) to avoid multiple-choice testing, I would have expected a different balance. It seems curious in this chapter to place the focus of concern away from the kinds of criterion-referenced testing techniques Hughes has declared himself to favor.

For me, the weakest chapter of the book is perhaps the most important one: Chapter 7 on “Stages of Test Construction.” There is so little detail in the main text that too much weight may be given to the examples, neither of which does much to illustrate how the principles Hughes has expounded can be put into operation. Further, no mention is made of a stage *after* the test has been administered: Practitioners need to learn how to make use of test results, and how to evaluate those results to plan for improved testing in the future.

The remaining chapters on test techniques and the testing of specific language skills contain useful advice and examples. The final chapter, on test administration, will be a real help to teachers attempting serious testing in their classes or schools for the first time.

Despite the limitations I have described, this is an excellent small book, probably the first serious British contender as a replacement for the classic *Writing English Language Tests* (Heaton, 1975), a book that is still used widely for courses in language testing for teachers. Teachers looking at the Hughes book should probably also look at Heaton, and at *Testing Language Ability in the Classroom* (Cohen, 1980), and *Techniques in Testing* (Madsen, 1983).

What pleases me most about these three books is their shared emphasis on the ethical imperative of language testing. In teaching a course in language testing, at any level, it is critical to emphasize test uses and the power of test-based decisions. Teachers and applied linguists—anyone who is likely to read testing research literature and draw pedagogical or bureaucratic conclusions from it, or to evaluate existing tests, or to develop new tests—must be provided with ways to consider what tests *mean*, in terms of the methods used, the abilities elicited, the societal and individual consequences that ensue, and so on. Bachman’s book provides a

strong theoretical foundation for doing this; in Chapter 7 he provides a highly relevant overview of ethical considerations in language testing. Bachman does not directly engage the specific issues, instead he provides a summary overview, rightly indicating that this is an area where language testing needs to focus much more effort. Brindley does this implicitly at every point in his book, and explicitly at the many points where he discusses accountability and collaborative assessment. The AMEP assessments serve as an example to testers and teachers alike, of what can be achieved by careful thought to what learners, teachers, and other involved parties need from a language assessment, and how testing experts can work with them to achieve it. The Hughes book begins with a discussion of *backwash* (also known as *washback*), or the effects of testing, beneficial and otherwise, on curriculum. Hughes argues that good tests are a powerful tool for curricular improvement. He also argues that teachers, students, and institutions (schools/colleges) need tests, and thus teachers must take responsibility for better testing. Hughes reiterates these views throughout his book and, despite its limitations, the book goes a pleasing distance in attempting to illustrate how this perspective can be turned into good testing practice.

Surveying these three books and the four previously reviewed, I realize that comparing books in this area is difficult because each is different in both approach and goals. In order to begin to compare textbooks in this area, questions about audience become critical. Lazaraton, Riggenbach, and Ediger (1987) found that 47% of their respondents, all of whom had established some researcher credentials in TESOL, had taken only one “tools” course. It seems most likely, then, that readers of these books also will have taken, or will be taking, only one course in language testing, or statistics, or research design. I have not located, among those I have reviewed, the perfect coursebook for a single-semester course in language testing. Another book eagerly awaited by language testers is Hatch and Lazaraton’s *The Research Manual: Design and Statistics for Applied Linguistics*, expected soon. This book promises to fill a major gap, but again will not be all things to each of us. The perfect coursebook, for me, would have Henning’s or Hughes’ coverage of key practical concepts, Brindley’s learner focus and wealth of examples, Bachman’s level of theoretical analysis of issues facing language testers, Woods et al’s (or Hatch & Lazaraton’s) level of statistical explanatory power, and Brown’s attention to interpretation of the research literature. Is this possible in one volume? And if it were, could it be taught or learned in a single semester? Realistically, teachers and teacher-trainers must make difficult

decisions about inclusion and exclusion, about ethos and values, about the kinds of graduates one wants a program to produce, and must select from the increasing array of materials in statistics, language testing, and quantitative research methods in order to match our requirements. The choices are better than they have been at any time in the past.

ACKNOWLEDGMENTS

I am grateful to Andrew Cohen, Dan Douglas, Larry Selinker, and Bernard Spolsky for comments that have assisted me in my review. All opinions, of course, remain my own.

REFERENCES

- Bachman, L., & Clark, J. (1987). The measurement of foreign/second language proficiency. *Annals of the American Association of Political and Social Sciences*, 490, 20-33.
- Bachman, L., Davidson, F., & Foulkes, J. (1990, March). *The Cambridge-TOEFL Comparability Study*. Panel presented at the 12th Annual Language Testing Research Colloquium, San Francisco, CA.
- Cohen, A. D. (1980). *Testing language ability in the classroom*. Rowley, MA: Newbury House.
- Hatch, E., & Lazaraton, A. (in press). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House/Harper & Row.
- Heaton, J. B. (1975). *Writing English language tests*. London: Longman.
- Lazaraton, A., Riggensbach, H., & Ediger, A. (1987). Forming a discipline: Applied linguists' literacy in research methodology and statistics. *TESOL Quarterly*, 21, 263-277.
- Madsen, H. (1983). *Techniques in testing*. New York: Oxford University Press.
- Weir, C. (1988). *Communicative language testing*. Exeter: University of Exeter.

LIZ HAMP-LYONS

University of Michigan

University of Colorado at Denver (as of fall 1990)