
REVIEW

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 POLLY ULICHNY
University of Massachusetts at Boston

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

Understanding Research in Second Language Learning: A Teacher's Guide to Statistics and Research Design

James Dean Brown. Cambridge: Cambridge University Press, 1988. Pp. xi + 219.

Statistics in Linguistics

Christopher Butler. Oxford: Oxford University Press, 1985. Pp. x + 214.

A Guide to Language Testing: Development, Evaluation, Research

Grant Henning. Cambridge, MA: Newbury House, 1987. Pp. vii + 198.

Statistics in Language Studies

Anthony Woods, Paul Fletcher, and Arthur Hughes. Cambridge: Cambridge University Press, 1986. Pp. xii + 322.

■ Lazaraton, Riggenbach, and Ediger (1987) conducted a survey of active professionals (defined as those who had made conference presentations) in TESOL and second language research and found that 67% of them were dissatisfied with the amount of preparation they had had in research design and statistics. Evidently these relatively advanced members of our profession were justified in their hesitation, since some were unable to interpret correctly a table of analysis of variance (ANOVA) results or to decide correctly between ANOVA and *t* tests for multiple comparisons across groups. One thing all were agreed upon was the necessity of training in research design and statistics for all students of TESOL and applied linguistics.

Research methods used in second language learning and acquisition have expanded a good deal in this decade: Qualitative,

and most particularly ethnographic, research has become common, and our view of what constitutes “data” has widened well beyond summary test scores. We have also embraced a new kind of measurement theory, latent trait theory, and studies that appear in our journals are increasingly likely to use the Rasch Model of latent trait measurement in particular.

I believe that every future applied linguist or language teacher should be able to read the literature that affects what and how they teach and to evaluate it, without being intimidated by, or forced to take on trust the accuracy of, the data analyses. Equally, I believe that every future language teacher or applied linguist should be able to observe and analyze a language education situation and make a determination of what tests are needed: progress checks, placement into levels, detailed diagnosis, proficiency measures, and so on. They should then be able to look at the tests available and to make an evaluation to determine whether the tests are appropriate to the needs of the context. If available tests are not appropriate, future language teachers or applied linguists should be able to put together their own test, pilot it, and improve it based on the pilot data.

Obviously, students will not leave a single-semester course able to do all these things, but they should be firmly on their way, with the tools in hand to travel the rest of the way alone. When I look at books in the general area of statistics, language testing, and quantitative research, then, I do so with this personal agenda in mind. In addition, given the apparent diffidence with which most people approach this area, the books had better be accurate, well illustrated, well provided with practice material and answers, carefully sequenced, and user friendly.

Of the four books reviewed here, Brown’s *Understanding Research in Second Language Learning* is the only one that does not aim for “productive competence” in its users. Designed for language teachers with no background in statistics, the book is “oriented toward the consumer, rather than the producer, of statistical studies” (p. x). There is a clear emphasis throughout the book on reading and interpreting second language research studies. Brown stays with that aim consistently throughout the book, offering users a number of chances to read and critique statistical studies. I am impressed by the modest and honest claims that Brown makes for the book: He does not claim that users will be able to *do* research after working through the book but, rather, attempts to explain (a) the basic terms of the statistics field; (b) how tables, charts, and graphs work; (c) the appropriate use of research designs; (d) the logic that underlies the use of statistics; and (e) an approach to critiquing and assessing statistical research.

I see these goals as important and realistic ones: I agree with Brown when he says, “not all teachers have the time and interest to do statistical research. But I believe it is irresponsible to ignore such research just because you do not have the relatively simple tools for understanding it” (p. xi). Too many graduate students do just that or, worse, read the beginnings and endings of research articles and extract the conclusions without considering the data analyses on which those conclusions are based. Perhaps that is because our field has been lacking teachers of Brown’s warmth and common sense, teachers who are able to make the numbers seem less terrifying.

Brown offers a way to enter a discourse community that is particularly intimidating to many teachers; he offers a means of moving from “outsider” to “insider,” at least at the level of understanding statistical studies. We are taken gently and thoroughly through basic concepts, including a discussion of “extraneous variables” that is the most down-to-earth and helpful I have seen in any book. We are walked through several real research studies, with exercises and commentary: The step-by-step guide to hypothesis testing, which is applied in several chapters, is especially helpful. (The only formula that was omitted from the book that I wish had been included is the Pearson correlation, which is central to this field and often discussed by Brown.) In general, I think that once teachers/graduate students find themselves able to read research with comfort, the prospect of conducting quantitatively based research will not seem at all threatening.

The other three books all aim to teach users enough statistics to be able to apply them in practice. Butler’s *Statistics in Linguistics* and Woods, Fletcher, and Hughes’s *Statistics in Language Studies* both focus on teaching statistical concepts to linguistics students (both include applied linguists and classroom researchers within their definition). However, Woods et al. go further in seeking to enable their students to evaluate the research literature and “to make them aware of such [statistical] methods so that they can recognise the potential application to their own work, and to supply them with the information necessary to engage in efficient discussion with a statistician” (p. xii).

Butler’s book is a basic introduction (he does not include “more advanced techniques” such as multiple correlation and regression and ANOVA) to the kinds of quantitative methods applicable for investigations into language, for “making sense of the data” (p. vii), and he aims to explain the reasoning behind the choice of statistical methods as well as the methods themselves. In this I think he succeeds only some of the time. Butler’s method is to work through the technique and then provide an example, which results in pages

with several formulas punctuated by formal, difficult text. The chapter on correlation is an exception, beginning as it does with concepts and moving gently to formulas. Much of the time, the treatment of a statistical technique is like that in many of the basic statistics texts that have been tried and rejected by teachers and researchers in language fields: dry, dense, distant, and devoid of examples or a reality base.

In contrast, Woods et al. keep at the forefront of their discussion the kinds of reality linguists and others in fields such as TESOL and second language acquisition share. Each discussion of a new topic moves, after a brief location of the technique in its place in the world of statistics, to a language situation. A problem or question is posed, and the technique is suggested as a means to investigate it. Thus, at each stage the student knows why it is worth learning about the technique and the kinds of situations in which the technique can be used.

This book is longer than any of the others, which shows in the fuller treatment given to most issues, although sometimes a key concept for readers and designers of language research (e.g., the Pearson product-moment correlation) is given a surprisingly brief discussion. Woods et al. also include several more advanced techniques such as ANOVA, multiple regression, multivariate analysis, and principal components analysis, which they distinguish from and prefer over factor analysis. Several chapters deserve special mention: Chapter 1, "Why do linguists need statistics?" covers ground like that in Brown and noticeably lacking in Butler and Henning; chapter 2, "Tables and graphs," introduces a wide range of visual representations while covering some basic concepts; and chapter 8, "Testing hypotheses about population values," takes us logically through confidence intervals, Type I and II errors, the concept of a test statistic, classical hypothesis testing, and statistical tests of hypotheses. In general, I find the book well organized, thorough, and written at an appropriate level, rarely unduly dense or dry, and never condescending.

Henning's *A Guide to Language Testing* is, as its title suggests, unlike the other books reviewed here in that its focus is on language test design and not on research design. Its main audience is "teachers and teachers-in-training who are preparing to develop tests, maintain testing programs, or conduct research in the field of language pedagogy" (p. vii). Henning cautions users that "the book progresses rapidly. . . . Familiarity with the rudiments of statistical concepts such as correlation, regression, frequency distributions, and hypothesis testing will be useful. . . . A working knowledge of elementary algebra is essential" (p. vii). This is, then, a more

advanced book than any of the others reviewed here. Henning also makes it clear that the book needs to be supplemented with sample tests, articles on testing, and student projects.

These built-in assumptions may explain why I find the organization and development of the book hard to follow and why I feel that the book often moves straight to tools, without any scene-setting of why the tools are useful and without any data to ground them. These assumptions may also account for the number of times I noted that familiarity with a concept is taken for granted in one chapter but the same concept is then taught in a later chapter (for example, on page 19 the formula for a z-score is given, using $X = \text{raw score}$, $M = \text{mean}$, and $s = \text{standard deviation}$, but mean and standard deviation are not taught until p. 39 and p. 40, respectively).

This is a very uneven book: Henning is at his strongest when discussing very practical testing concerns, such as item analysis, and wholly statistical aspects of test evaluation; the chapter on "Language Test Reliability," for instance, is very good. The final three chapters of the book deal in turn with latent trait measurement, item banking and computer-adaptive testing, and program evaluation. These are all important areas not covered in any of the other books reviewed here. However, the treatment of these topics, although well done, is so brief that I wonder if they should not have been saved for a follow-up book, where they could have been fully discussed with multiple examples and practice activities. As an instructor, I think it would be excessive to include these areas, as well as the material in the earlier chapters, in a one-semester course.

Looking at the four books together, I note some areas of common ground and some areas in which the material diverges widely. There is also some failure to agree on basic questions in quantitative methods. This is illustrated by the following two examples, the first of which deals with interval and ordinal data. Brown tells us (pp. 23-24) that data cannot be converted upward: Ordinal data can never become interval data. Henning, on the other hand, tells us, "Interval scales are usually obtained by the transformation or normalization of ordinal scales" (p. 18). Woods et al. do not treat this issue, as far as I can see, but Butler comments, "Investigators of . . . linguistic phenomena commonly assume a higher level of measurement than . . . their data warrant" (p. 12). The second example concerns *t* tests. Henning does not cover this topic, but Butler does, without referring to constraints on their repeated use, and the same is true of Woods et al. Brown, on the other hand, stresses the need to avoid repeated use of *t* tests on the same data sets (p. 170). It would seem,

then, that readers may be forgiven for not being fully in control of some “basic concepts” of quantitative methods, when their informants do not provide them with consistent information.

In the second part of this review, which will appear in the next issue, I will examine several more books and make some summary comments about what we have and what we need in terms of materials in this area.

REFERENCE

Lazaraton, A., Rigggenbach, H., & Ediger, A. (1987). Forming a discipline: Applied linguists' literacy in research methodology and statistics. *TESOL Quarterly*, 21, 263-277.

LIZ HAMP-LYONS

University of Michigan