

TESTING COMMUNICATIVE PERFORMANCE: AN INTERIM STUDY

Brendan J. Carroll

Pergamom Press. 1980

This book is part of Pergamom's series "Language Teaching Methodology" and, like others in the series, it sets a level sophisticated enough to appeal to academics and researchers while still being accessible to the interested and informed classroom teacher. This particular volume, however, presents more difficulties than most in that it tends to rely on a general knowledge of the field of language testing -- terms such as "correlation matrix", "norm-" and "criterion-referencing", etc. are not explained -- and on a knowledge of communicative needs analysis and syllabus design in general, and more specifically, a familiarity with J. L. Munby's Communicative Syllabus Design (Cambridge, 1978), whose model Carroll follows in the specification of test content. Even though Testing Communicative Performance is not a simple testing manual, it is certainly worth the extra effort to read, in that it offers a valuable contribution and a useful guide to anyone involved in the very difficult task of evaluating language use.

Carroll divides his book into five chapters: "The assessment of communicative performance", an introduction and a statement of general principles; "The design of communicative tests", a specification of test content and a description of the procedures for item construction; "The development of communicative tests", a description of the actual writing of items and tests; and "Communicative testing literature", a brief review of some current books in the field, some empirical and descriptive data from Carroll's own project and the conclusion. He includes, as well, four appendices which amplify or illustrate points made in the text.

In the first chapter, after underlining the importance of the communicative needs of the participants as the ruling principle of course design and the primacy of use over usage in communicative programs, he points out the obvious desire for authenticity, or that tests "should be real-life, interactive communicative operations" (p.11). He admits, however, that

this is, perhaps the most difficult goal to achieve. Authenticity alone is not enough to ensure adequate assessment. "Test tasks must be drawn from a range of interrelated systems, from consciously systematized functions, skills and linguistic categories, brought together in the performance of an authentic communicative operation; performances whose value is judged by reference to the satisfactory resolution of patterns of communication, function, skill and language in an effective system of language measurement." (p. 13).

There are other characteristics besides authenticity to consider. These are: Relevance, or how relevant is the point being tested to meeting communicative needs? Acceptability, or will the users accept its content and format? Comparability, or can the test scores obtained at different times or with different groups be compared? and Economy, or do the tests give us the information we need with the least expenditure of time, effort and resources? Carroll summarizes these four in the useful acronym RACE. RACE and authenticity are, then, the goals to be strived for.

Chapter Two gives the principles of test design, which consist of three processes: the description of the participant(s), the analysis of communicative needs and the specification of test content. The first two processes follow closely the processes described by Munby (1978), although Carroll further categorizes in terms of the four macro-skill areas of writing, speaking, listening and reading. For the specification of test content, he gives a resumé of ten parameters and procedures.

In Chapter Three, he applies the procedures suggested to two cases: the first and most fully developed, that of a Nigerian student wishing to study a business course in England; the second, a course in social-survival English for students wishing to spend some time in Britain. The chapter is most valuable in that it gives a concrete example of testing micro-skills such as transcoding written information by completing a graph (Munby's skill 52) in an integrative format. Other non-reading-related aspects are less fully treated, although there are some tantalizing suggestions. This chapter also describes a rating scale for describing performance levels. The nine bands of the scale range from "non-user", band 1, to "expert", band 9. Carroll suggests that combining the results of these criterion-based rating scales with the

less descriptive results of objective tests may produce a score which is more reliable, yet which maintains the advantage of the more valid rating scales. Although there are obvious theoretical pitfalls to his system, it may be extremely useful in the long run.

Another useful suggestion is the use of a two-tier testing program which is described in Chapter Four. The first tier consists of a more general usage-based phase, although still with a notional/functional orientation. The second consists of a more specific set of testing tasks which incorporate the specific events, activities, skills and functions involved in the communicative needs of the participant.

Although Chapter Five does not follow the progression exactly, it is full of interesting information. The short review of texts on, or related to communicative testing gives succinct summaries of some of the ideas Carroll finds relevant. The experimental data he cites is not impressive, as he himself admits, although it is at least suggestive. His discussion of measurement criteria, however, is especially useful and perhaps should be read not at the very end, but in conjunction with the first chapter. His conclusions are not endings but beginnings and, as such, suggest further research. He calls for such things as a closer relationship between testing and learning, involvement of the learner in self-assessment, better banding systems based on communicative activities, more comparable measures even at an international level, new approaches to test design, the "demythologizing" of elaborate, meaningless statistical displays, better definitions of performance criteria and other, more technical discussions.

In spite of the relative difficulties of approaching this rather demanding book, I feel that the effort is more than worthwhile. It is a valuable contribution to the field, in that it touches not only theoretical issues but also presents a practical working model for designing, developing and operating communicative tests and testing programs.

Diana Jenkins de Ormsby

CELE/UNAM