TESTING AND EXPERIMENTAL METHODS

J.P.B. Allen and Alan Davies (eds.)

The Edinburgh Course in Applied Linguistics, Volume 4,
Oxford University Press, 1977

Volume Four of the Edinburgh Course in Applied Linguistics
has been eagerly awaited for some time, so long, indeed, that
its final arrival was bound to be an anticlimax. The quality
of the first three volumes raised expectations to such a point
that the final volume in the series would have had to be super-
lative to meet them. And unfortunately Volume 4 is not super-
latively good. However, the tardy arrival of this volume could
have been forgiven, and a less than superlative quality would
have been acceptable, if the delay had meant that the book con-
tained the latest thinking in experimental design and language
testing, giving a new slant on the latter and thereby repre-
senting a genuine advance over previous publications in the
field. This was not to be, however, and Volume 4 does not, on
the whole, add to our knowledge about testing and experimental
methods. Since the book forms part of the ECAL, this is per-
haps unfair criticism, since the main aim of the course is to
serve as a textbook in applied linguistics. As such, it is per-
haps unrealistic to expect this volume to take us beyond the
state of the art during the past ten years, and to open up new
possibilities for experimentation and testing. Nevertheless,
this was precisely the expectation of many applied linguists
and this expectation has not been met.

Volume 4 covers an essential area within the field of
applied linguistics, without which the first three volumes
were clearly incomplete, and that is the area of the instru-
ments of proof of the claims made in applied linguistics. As
the editors put it: "If applied lingustics is more than a set
of anecdotes and subjective impressions, then it needs the
strengthening of the scientific method", and that is what
Volume 4 sets out to provide. No exaggerated claims for
originality are made: "While psychological and educational
measurement in general are well-established disciplines, the
use of these concepts and techniques in applied linguistics is
relatively new: it is the purpose of this volume to give them
their rightful place at the center of the discipline." The
editors might also have mentioned that language testing and
experimentation into language use, language learning and

language proficiency require concepts and techniques that are in many ways distinct from those used in other fields of educational measurement, and that this volume is justified on that count alone.

The centrality of testing and experimentation in applied linguistics and language teaching, and thus the importance of this particular volume in the ECAL series is underlined by Davies, who, in a statement that is of particular relevance to the present state of language teaching, writes: "Our only hope of escaping from the tyranny of fashion is through submitting our guesswork to the rigour of hypothesis and experimentation." Current fads in language teaching methodology, in materials production and in syllabus design cry out for evaluation and experimental justification, and so this book is of interest to language teacher and applied linguist alike.

The book is divided into six chapters, but falls into two main areas of interest: testing and experimentation. Alan Davies, in his Introduction, points out the close connection between tests and experiments. It is suggested that a test is an experiment in miniature, and an experiment is "testing writ large". Experiments require tests, which, it is pointed out, should have been properly tried out, and the process of test construction can be seen as a process of experimentation. The distinction is drawn between test construction and test use, and the reader is warned not to use tests under construction as if they were already tests in use, e.e., in order to validate or assess experiments.

Elisabeth Ingram's chapter on Basic Concepts in Testing covers the topics of reliability and validity. The exposition of the different types of reliability is particularly clear, and the concept of validity, particularly pragmatic validity, is discussed with reference to language teaching: "In the field of language teaching, teachers are usually only too aware that ... the ability to use the language ... is what matters. The problem is how ... to make sure that the examinations and tests which the learner has to take are good indications of ability to use the target language in real situations." One important but frustratingly brief section in Ingram's chapter is the discussion of the distinction between norm-referenced tests (that is, tests which essentially compare learners with other learners, and seek to discriminate among them) and criterion-referenced tests, or tests which measure the learner

against what he has to learn, and "merely" ask whether he has
or has not attained the criterion level. Since in the criterion-
referenced tests the concept of discrimination is less relevant
than that of dichotomization, the central problem is how to dis-
tinguish those who know from those who do not. As Ingram points
out, criterion-referenced testing is in its early stages and
implies new tools for test analysis, which at present are not
available. In a final section, Ingram returns to what is almost
the leitmotiv of the book by pointing out the importance of
having good tests when carrying out experiments. The Colorado
and Pennsylvania projects are both quoted as telling examples
of how not to run experiments.

Alan Davies' chapter on The Construction of Language Tests
points out the relation of different linguistic theories to
language testing, and emphasises the need to integrate teach-
ing with testing. The latter is not an end in itself, but mere-
ly an aid in the taking of intelligent decisions about learners
and learning. An extensive discussion of validity follows, in
which it is pointed out that the validity of achievement tests -
with which most teachers are concerned - is directly dependent
on the syllabus being used, so that there is a limit to the
extent to which the ultimate validity of achievement tests can
be improved. An interesting relationship emerges in the dis-
cussion of the validity of proficiency tests between such
validity and ESP in the broad sense, through the question which
is common to both areas: What is the learner learning which
language for? Thus "work-sampling tests" - sampling the
demands made on a learner - are akin to materials development
and language testing based on a needs analysis. This some-
what theoretical section is followed by a section on item-
writing which emphasises that the teacher should ask him/her-
self "Which specific language skills do I want to test here?"
and "How can I get at this skill, i.e., by what techniques or
device can I be sure of validity?". The item writing section
is not intended to be comprehensive, and is not. It does not
offer anything new, and the practical section which ends the
chapter is of less usefulness than other such practical sec-
tions. It is hard to believe that even the keenest item
writer will persevere in writing 25 items of a type similar
to the model:

    "Some taste... soup" = some soup that has no taste
                        (ANS=tasteless)"

What is really needed is not an indication of how to fill in the familiar matrix of levels and skills with item types (this is more than adequately done already by Heaton, Harris and others) but a presentation of testing techniques, with a discussion of their suitability for a series of different purposes in different types of test (proficiency, diagnostic, etc.) or in a variety of different syllabuses (notional, structural, ESP) in the case of achievement tests. What are the implications of a notional/functional syllabus for testing techniques? Can we get at "communicative competence" through multiple-choice techniques, or the pseudo-random cloze?

The second section of the book, Chapters 4 to 6, are by Ruth Clark and deal with the design, analysis and interpretation of experiments. The emphasis in Chapter 4 is on the importance of clear operational definitions in experimental design, to avoid the collection of uninterpretable or simply useless data. The reader is given many examples to make unfamiliar and possibly daunting principles and concepts clear. The pitfalls of measurement in research are pointed out with the aim of alerting the reader, not to the impossibility of research in the social sciences, but "to the lamentable naivete of much research in these fields". Yet again, the main concern is how to assure the reliability and validity of experimental observation. This chapter also deals breifly with two important issues: how to reduce bias in controlled experimentation, and how to allow effects to show themselves which do not show up in natural situations because of so many other factors which obscure them. The chapter ends with some very useful practical work on experimental design.

Chapter 5 on Procedures and Computations continues the discussion of four experiments introduced in the previous chapter and deals with the various ways of summarising most appropriately the data obtained. The chapter contains a very useful account of the mechanics of establishing frequency distributions, ranks and the like, based, like so much else in ECAL, on materials tried and tested on several generations of students of applied linguistics in Edinburgh.

Chapter 6 gives a clear and sensible introduction to the concepts necessary for the understanding of statistical inference and provides some excellent exercises for practical work as well as a useful bibliography.

Fianlly both Chapter 6 and Appendix I provide some useful directions on how to calculate certain basic statistics.

One area of applied linguistic concern missing from this book, as well as from similar books, is the design of surveys and questionnaires, both of which are important, if not essential, tools in much language-related research. This may in part be due to the fact that the book has only three authors, although it contains six chapters, and the interests of these authors are less in the area of surveys than in that of testing.

Nevertheless, Volume 4 remains a worthwhile addition to the ECAL series. If it offers nothing new to the language tester and researcher, this is perhaps becuase it is a textbook intended for the training of applied linguists, rather than a reference book, or an original contribution to the field. Like the other three volumes, it must be read as a textbook, with the intention of learning new concepts and techniques, rather than be consulted as a reference book. Although most of the information contained in this volume already exists elsewhere, it does not exist as a coherent whole, so that Volume 4 represents a useful compilation of topics relevant to language testing and experimental methods.

J. Charles Alderson
Research and Development Unit
Centro de Enseñanza de Lenguas
Extranjeras
Universidad Nacional Autonoma
de Mexico