

RESEÑA AL LIBRO *A Practical Handbook in Language Testing for the Second Language Teacher*. Elana Shohamy. Experimental Edition. Israel: Shoshana Goldberg, 1985, XI + 221 pp.<sup>(1)</sup>

Harold Ormsby  
C.E.L.E. - U.N.A.M.

In think it is reasonable to view all introductory language testing textbooks as attempts at communicating with those who, in the last analysis, are the most frequent and therefore really the most experienced practitioners of language testing: classroom teachers. We seem to be aware that it is the teacher who is in immediate, constant contact with students and who must make immediate, on going decisions about them; just as we seem to be aware that what language teachers need to know about testing is somehow different from what others who work in the field need to know.

Each classroom is a unique educational microcosm with very specific testing possibilities and needs. It is the teacher's job to decide how to take advantage of each classroom's possibilities and how to meet its needs. Introductory textbook writer have the obligation to know and understand these microcosms in a general way and the task of deciding what information, presented in what ways, will make the teacher's work easier and more effective. Shohamy's *Practical Handbook* is the result of this sort of understanding and decision making process.

The book is a vademecum, as its title promises. Chapters tend to be quite short and sections consist in general of only a very few pages, each clearly marked as to its specific

subject. Frequent "over-view" statements, tables, lists and an unusually large number of examples keep the force of the communication between author and reader moving.

In the Foreword there are two paragraphs that make an important contribution to understanding the book (p.v). The first of these is a Statement of the handbook's basic language learning philosophy:

Language is viewed as a means for communication, and thus the ability to produce and comprehend authentic language in each of the skills is emphasized. .. Only little attention is given to testing isolated and discrete elements such as grammar and lexicon...

The second is a straight-forward statement of what the handbook does not do:

(It) does not offer a deep theoretical view of testing and it does not deal with the built-in conflicts in the field. It does not offer thorough explanations of statistical and psychometric procedures.

The handbook is, in short, "intended for classroom language teachers" and their immediate practical needs.

The book is divided into two major parts, plus two appendices, a bibliography and finally - something very unusual and admirable in an experimental edition - a complete index of terms dealt with.

Part One (Chapters 1-9), "The Process of Producing Language Tests", starts with background information about language testing, including two well done tables on historical and current trends in language testing, several very interesting pages of students' actual views of language tests, and a rather depressingly long (but all too accurate) list of "things that are wrong with our language tests'."

In these early pages the author also sets up the dichotomy between "classroom tests"

(planned, written, administered, analysed and reported by teachers) and "external tests" (prepared and handled by staff and agencies outside the classroom) . Throughout the handbook Shohamy succeeds in keeping a tight focus on the first type, with only occasional glances at the second. This allows her, among other things, to insist on the need to test both process and product (in the classroom), without her having to get into the very sticky question of whether external tests can ever be used to test process.

The remainder of Part One goes through the making of language tests for the classroom: determining the purpose(s) of a particular test, planning it, selecting items and tasks, administration, scoring and rating students' performance, test and item analysis, and reporting results meaningfully to students, parents and the school. I will comment here on Chapters 5, 7, 8 and 9 only.

Chapter 5, "Selecting Items and Tasks", subjects each item or task type to a 10 point analysis in table form: "Preparing the items", "From a test-taker's point of view", "Authenticity", "Scoring", "Best use", "Worst use", "Advantages", "Disadvantages", "Variations" and "Criticism". For example, preparing items for multiple-choice tests is "difficult and time-consuming and it requires talent, training and experience" which leads to the conclusion that its worst use is for classroom testing (p. 38). Preparing open-ended questions, on the other hand, "is not a particularly difficult task... but questions must be directed at the most relevant information"; their best use, then, is precisely "in classroom situations", although a disadvantage is that they are "difficult to score, especially (in the case of) a longer essay question" (pp. 43-44). In this chapter Shohamy describes, analyses and exemplifies multiple choice, true/false and open-ended items as well as summary, cloze procedure and C-Tests. Full-text translation, main-point summaries and re-

porting tasks are dealt with in table form only.

Chapter 7, "Scoring and Rating" puts a heavy emphasis on the latter, giving specific detailed examples of scales for writing and speaking. This very short chapter is a refreshing departure from the more unusual "mark it right or wrong and count up" approach taken by some language testing books.

Although I will discuss some of the topics of Chapter 8, "Analysing Test Items and Tasks" later, it is worth mentioning here that the author advocates "eyeballing" a test; that is, doing a bare-bones analysis (like figuring an overall average) and then stepping back to take a long look in order to come to a cursory (but very useful) conclusion. Given the classroom teacher's time limitations and purposes, this in itself is a strong recommendation for this chapter.

Chapter 9, "Reporting Test Results", demands that the reported results be meaningful in both a lay and technical sense. The author rejects raw scores and simple-minded percentages in favor of percentiles, z-, t- and grade-level scores, and profiles. As Shohamy points out, these are not particularly difficult for teachers to compute or, more importantly, for parents and students to understand.

Part Two (Chapters 10-13) "Testing Specific Skills", deals, as one would expect, with testing reading, listening, speaking and writing. Each chapter is a compendium of specific examples of test tasks, including discussion, and excellent suggestions and guidelines. A great deal of very useful information is packed into tables that go over the nature of each test or task, its purpose, the procedure itself, variations, scoring or rating techniques and appropriate uses.

In Chapters 10 and 12 Shohamy gives example after example of test tasks: notes, advertisements, short letters, forms, newspaper

articles and short narratives for reading tests, interviews of various sorts, roleplay, group discussions, reporting and picture descriptions for speaking tests. In Chapter 11, the author does about all one can do about listening tests without having recordings included with the book. And in Chapter 13, on tests of writing, she gives a fair number of ideas for tasks, exemplifies a couple of them and then goes into the important question of rating written production. One very positive aspect of this chapter are the eight different photocopied examples of actual student production (bad handwriting and all) that are offered as rating exercises.

Appendix A, "Statistical Computations", deals very briefly with the concept (but not the mathematics) of correlations and gives the formulas for standard deviation, internal consistency (reliability) and standard error of measurement. Appendix B, "Testing Specific Language Aspects", deals very briefly with classical sorts of multiple choice grammar and vocabulary tests.

The bibliography is divided into three parts: other introductory language testing books, articles on language testing research, and general educational testing books. Besides serving as the list of references for the handbook, this bibliography makes for, I think, a good starting point for further reading.

For more than 25 years now, our profession has been looking for ways of communicating directly and effectively with classroom teachers. Although our attitudes and traditions have perhaps made this process much more complicated than it ever had to be, we have indeed come closer and closer to achieving that goal. Just like many of its predecessors this handbook can be taken as a step in the right direction --this step perhaps a bit more "right" than others. Nonetheless, I think there is room for some questions.

While the use of rating scales is an ex-

cellent idea for classroom testing, I cannot help but wonder how a non-native (or non-near-native) teacher can be reasonably sure he is applying reasonable expectations and criteria to his students' performance. There are always some disagreements among native teachers as to, for instance, what errors truly "interfere with communication" or what "almost effortless self-expression might be. Where and how can the non-native teacher find some degree of confidence in his own judgements? I cannot say that I can propose a really effective solution to this problem. Perhaps just mentioning it, giving a teachers' insecurity the dignity of recognition, would be enough. I hope Shohamy will address it in the next edition of her book.

This leaves me with two questions that are more technical. The first is the question of expecting variance in classroom tests; the second, a related question about the applicability of the concept of reliability to classroom testing. Shohamy herself opens the question about variance when she says (p.77):

However, it should be remembered that not always are we interested in discriminating (among students). In a criterion-referenced test we would like all students to master the item. (*Italics hers*).

And again: (p. 80):

The desired level of difficulty for a criterion-referenced/mastery test is not based on the ability of the item to discriminate between high and low achievers... Instead, the difficulty of each test item is determined by the specific learning outcome it is designed to measure,

I take no issue with either of these statements. In fact, I would like to see them raised to the level of general principles of classroom testing.

I find nothing in the classroom situation,

and particularly in achievement/mastery testing in the classroom, that makes variance theoretically inevitable. Furthermore, what variance that does crop up can often be best explained in ways that make it extraneous to the test itself and of much wider importance to the teacher as professional. If one takes this stance, such things as distribution graphs, the mean, standard deviation and correlations have to be understood and presented in very different ways, if at all.

Reliability as traditionally understood has to do with the relationship across observations between mathematically explicable and mathematically inexplicable variance. I would argue that in the classroom a clinical view of explicability is much more valid and useful than this statistical one. Variance, rather than being an ontological axiom whose vagaries are defined in a conventional way, should be a cause for practical concern and lead to a much more real way of understanding and evaluating language teaching/learning. Variance and variability in general and statistical reliability in particular should, I think, be relegated to the areas of external and experimental testing.

I recommend this book very highly, especially to teachers-of-teachers and, of course, to their students. Its organization is very comfortable and seems to be designed to encourage discussions that can lead novices to an understanding of what classroom language testing is about.

1. Privately published. Available from Shoshana Goldberg, Ben Gurion St. 11/24, Raanana 43300, Israel.