

INTRODUCTION: PROBLEMS AND TESTS

John A. Upshur
University of Michigan

This volume is an accident. There were no preconceived notions about the nature of an audience receptive to information about foreign language testing nor, partly as a consequence, of what kinds of information would be most cordially received. Thus no author had a general plan for what should be included; no editor could decide on what published papers to collect or even on what kind of papers to solicit. Instead, this volume records what a group of people with varying interests chose to present about foreign language testing and the reactions which were made to their presentations.

The conference, of which this is a record, was also an accident. The idea for the meeting grew out of private and public expressions of dissatisfaction by a number of people who are involved in test production and who depend upon test results—dissatisfaction with the quality and range of foreign language tests available, with the uses to which tests often are put, and with the ends they are made to serve. In translating the idea to a plan, it became necessary to find representatives of language test writers, language test users and language test needers who could all be available for a meeting at the same time. Constraints imposed by schedules, finances, planning time, and ignorance on the part of the planners introduced an accidental element into the composition of the conference membership. The freedom allowed the participants in their choice of papers to present and comments to make increased the accidental quality of the proceedings.

The conference produced more than a dozen papers and more than a dozen hours of debate and discussion, and the editor who would put all of this material into a single volume has a number of decisions to make about his use of the material. For example, How much repetition should be included? Each of the participants had his own point of view, one which was quite legitimately applicable at many times to many problems. But repetitions which serve as useful reminders, provide emphasis, give a picture of the priority of concerns, and set trends in a multiply directed discussion may appear only repetitious in print. How much of the discussion that gets off the track should remain? How many of the misunderstandings and *non sequiturs* should be retained? How many of the inconclusive arguments deserve reporting? These decisions are largely

reflections of a more fundamental decision the editor must make. Is the purpose of the volume primarily to present utilitarian information to potential test writers and users, or is it to reflect as accurately as possible the "state of the art?" If the first purpose is adopted the editor selects (or rewrites) the material to provide the most orderly and economical development of what he considers to be the most practical information. On the other hand, the true state of foreign language testing today includes much unresolved argument and lack of communication, a great deal of incomplete thought, and certainly repetition—often extending beyond the boundaries of boredom. Thus to give the most valid picture of the state of the art, the editor might best present an unedited transcript of the proceedings meanwhile suppressing all his urges to cull and comment.

I have tried in editing this volume to take a middle path. All of the working papers submitted for the conference have been included even though some overlap, and even though the reader may think that not all are so informative or useful as he would expect articles in a professional journal to be. But the inclusion of all papers requires the exclusion of much discussion. I have, however, tried to retain at least some examples of repetition, misunderstandings, etc. The reader may feel frustrated when problems are introduced, discussed but left dangling, unresolved. Yet this should serve to remind him that many of the questions asked about foreign language testing still have no satisfactory answers, and also, perhaps, that many important questions have not even been posed.

Having decided on what papers to include (in this case, all papers), the editor must decide on an order in which they are to be published. Anything but an alphabetical or random order reflects the editor's evaluation of the papers or his own concept of the functions of foreign language measurement. I must, therefore, first disclaim any ordering according to personal estimates of the values of papers, and then make explicit my own concept of the function of testing so that the reader does not assume that it is a concept necessarily shared by all the contributors to this volume.

Testing is an operation performed to gain information about the value of some variable which is significant for the solution of some problem. I thus view testing as a problem solving procedure. This is not the only possible view (one may be interested in building a better mouse trap but completely uninterested in catching mice), but it is the view taken here.

One does not test without a purpose; the problem to be solved defines the purpose for testing. Further, the problem is of a special

kind: the person with the problem has several courses of action open to him—several decisions he might make. In order to make "correct" decisions, he needs information; in foreign language testing he needs information about second language related abilities (the variables) of some group of people. The abilities may be such as language dominance, language learning potential, vocabulary size, communication ability in a specified context, etc; the group of people may be a teacher's class of students in an elementary French course, Navajo school children, Japanese applicants to U.S. graduate schools, etc. The test is the means by which information is gained.

The view of testing as a problem solving procedure includes more than concern with making decisions when faced with a problem. Decisions can be evaluated; some are better than others. When alternative courses of action are open to the person with a problem, and these different courses of action each have different sets of consequences, some consequences are better than others in terms of values held by the problem-solver. A test which gives consistent information about the value of some variable is a reliable test. If the test also permits the problem solver to make decisions which have valuable consequences, the variables measured can be said to be significant for solving the particular problems, and the test can be said to be valid in that problem context. Thus the determination of the value of consequences in a problem-solving-by-test situation is validation of the test; evaluation of decisions is test validation.

Thus far I have spoken only of *a* person with *a* problem and *a* test used to provide information for its solution. But obviously there are many foreign language tests, and many people with many problems. Almost as certainly a number of different people have problems with similar dimensions, i.e., who need information about some of the same kinds of abilities of the same sort of examinees. Which of the large array of foreign language tests available will provide them with the information which will allow them to arrive at the best solution to their problems? The potential user must know or learn what tests are available, what variables they measure, whether they are suitable for his use in terms of time and facilities for administration and scoring, in terms of the people he wishes to test, whether they provide the refinement of measures he needs, how valid they are in the solution of his problem. In brief, any foreign language test is useless until potential users know of its existence, its values and its limitations; potential users remain *potential* until information about tests is available to them.

Problems, variables, techniques and validity are logically distinct "parts" of testing as a problem solving procedure, and in

principle an author could write a paper which would deal, albeit abstractly, with only one of these parts. In practice, however, this is seldom done in writing about foreign language testing. One does not generally write about techniques or operations without consideration of variables those techniques should measure; one does not generally write about variables without consideration of their relevance to some other behavior; one does not generally write about problems with no thoughts as to their solutions; one does not generally write about validity except in the context of a problem and the means of accumulating information for its solution; and one does not write and publish anything about testing without, in a sense, adding to a kind of information storage system.

Although none of the papers in this volume deals exclusively and abstractly with single "parts" of the problem solving by testing procedure, each has its own emphasis. They are ordered according to their emphasis on problems, variables, techniques, validation, and information about tests. The first paper, Hopkins' "Language Testing of North American Indians," provides an historical context of testing that group—if it can be properly referred to as a single group. It suggests that some early testing may not have been directed towards solving any particular Indian problems, that those efforts may have been largely wasted. The most pressing current problem, to the solution of which second language information is needed, is curriculum development: placement of students, the identification of student problems, and determination of student progress.

Brière's paper touches on a problem, variables and techniques. Beginning with a problem common to every foreign language teacher—pacing and control of his own teaching—and with a prescribed set of variables—defined by the text and course aims—he shows how instruments are developed to help solve the problem. Thus Brière offers an overview of the problem solving approach to testing. White's paper on the "Second Language Testing Requirements of the Canadian Public Service Commission" shows a vast range of problems, including that of a teacher's control of his own instruction, but also going beyond to complex problems with extremely important consequences, such as personnel selection and awarding of bonuses. White invites suggestions about the nature of variables which will be significant for solving the problems of the Canadian Public Service Commission.

Variables to be tested are not "drawn from thin air" by a budding test writer. They are generally "elements" or more "elementary processes" of theories about the kinds of behavior which would provide *good* or *bad* solutions to a test user's problems. If,

for example, a businessman's problem is to select a "good secretary," the variables he is interested in are those related to *good* and *bad* performance of secretaries. A "theory" or "model" of secretarial behavior might include such "elements" or "elementary processes" as typing speed, spelling ability, regularity of attendance, etc. These then become the variables in a testing program for the selection of secretaries.

The variables of most foreign, or second language testing are still "elements" of structural linguistic theories of language. But this set of variables is inadequate for reasonably good solution of many problems faced by those concerned with second language behavior of bilinguals. Whether this inadequacy is due more to the limited compass or to inherent shortcomings in the linguistic theories from which the variables are drawn was not at issue in the conference reported here. All of the participants were, however, conscious of the severe limitations of current language testing practice, and recognized that what is being measured is too often inadequate or irrelevant.

Two of the conference papers present models of language behavior. From an elaboration of the Ervin and Osgood concept of compound and coordinate bilingualism Jakobovits has developed a psycholinguistic model of bilingual language function. The variables implied by the model's "elements" and processing operations are reflected in some twenty-seven tests, less than half of which are also based upon elements of a structural linguistic theory.

Cooper's "Elaborated Language Testing Model" makes provision for adding "communicative competence" to "linguistic competence." Rejecting the notion of a language as a single code applicable in all social contexts, Cooper makes explicit that acceptable and effective language behavior presupposes not only an implicit knowledge of linguistic rules, but also another set of selection rules which determines the variety of language to be used in a given social context, i.e., which, when applied to linguistic rules, determines what is said, when, to whom and where.

The three papers which follow (Pimsleur, Spolsky *et al.*, and Plaister) also reflect foreign language testing variables. The variables, although complex, are conceived of as essentially unidimensional, in contrast to the multidimensional models of Jakobovits and Cooper. There is also in these papers a much greater emphasis on instrumentation, on the development of tests to provide reliable and "pure" measures of specified variables. In addition to providing a comprehensive picture of the development of a testing instrument, Spolsky *et al.* indicate in their conclusion how a new test for the measurement for one variable may embody useful techniques for measuring other variables as well.

Valette, Lewis and Ingram are all concerned with, among other things, test validity. Valette's paper addresses the problem of establishing criterion measures for foreign language learning. In the absence of external criteria the language teacher can never be certain of the values and efficiency of his own course. That is, if students learn just what they have been taught, their teachers must be concerned also to know the value of that learning. Lewis' paper, although covering many aspects of foreign language testing, is here placed with "validation" papers because the author makes explicit how a subset of variables are selected to be measured in a testing program when one of a larger set of problems is to be solved—validity involves the significance of variables for problem solving. Ingram presents the persistent problem (unresolved in this volume) of evaluating the results of educational experiments when data are test scores. When teaching materials or instruction produce measurable changes in students' performance, do these changes indicate only that the students have been subjected to "experimental manipulation," or do they mean that a desired goal has been achieved? If no changes are observed, is this apparent stability in performance a true indication of no change, or is it due to an instrument which does not record changes which have occurred?

Mackey presents arguments for the establishment of a test information repository and retrieval system, and suggests the kind of information that should be included. Savard explicates how such a system is already being established at the International Center for Research on Bilingualism at Laval University.

Palmer's conference summary, with the discussion it provoked, is especially useful in pointing out the many problems in foreign language testing which received short shrift or were entirely ignored at the conference. There was great concern over misapplication of tests and test information, especially when that information contributes to decisions about the careers of examinees. The caveat that test users cannot wait for "perfect tests" notwithstanding, much of the misapplication problem is attributable to a paucity of tests which measure significant variables and which are suitable for large bilingual populations heretofore largely ignored by test writers.

I must apologize to authors whose intentions have been misconstrued or misrepresented. Apologies are due also to participants whose "prize comments" have been omitted or whose dialogue may have been unfairly altered during the process of transcription and edition. For all of these errors the editor must assume full responsibility.