     The article initially treats the difference between measurement (quantitative) and
evaluation (qualitative) and explains the essentials for meaningful evaluation as being
clearly defined course objectives, teaching methods, and testing procedures. The
educational objectives in the "Taxonomy of Educational Objectives, Cognitive Domain"
are then considered as an acceptable guide to testing. Advantages and
disadvantages of objective and essay tests and their relevance to these educational
goals are considered. Particular testing techniques for students at different levels of
reading and writing skill are identified, and suggestions are made for the composition
and grading of essay tests. (AF)

Modern Language NOTES

A MEANINGFUL EVALUATION OF FOREIGN LANGUAGE LEARNING: READING AND WRITING

by Jacqueline Benevento, District Coordinator of Foreign Languages, Collingswood, N.J. Public schools

(Adapted from panel presentation, NJFLTA meeting, College of St. Elizabeth, March 16, 1968.)

It seems fitting for me, as the last speaker on this panel, to pause and ask you to reconsider with me the title of our topic: "A Meaningful Evaluation of Foreign Language Learning."

Let us at this point make a distinction between the terms measurement and evaluation. Suppose there are 50 items in a multiple-choice reading comprehension test. A certain pupil gets 45 items correct --- we assign a number to his achievement and say that he has a score of 45. We have measured his achievement in this particular test. At the end of the marking period, we consider his several test scores, and we add to these measurements our judgment about the quality of his participation in classroom and language laboratory, his homework, and so forth; that is, judgments in all those areas which we describe in words but do not assign numbers to. Measurement is a part of evaluation. Measurement may be thought of as a quantitative process, and evaluation a qualitative one.

We know that a test is more reliable when it is objectively scored. Our evaluation of the pupil's performance at the end of the marking period will be more reliable as it is more objective. As reliability increases, so will meaningfulness. Thus, a meaningful evaluation is one which contains, among other things, reliable quantitative measurements of his achievement.

We also know that a test is valid when it tests what has been taught. Meaningful test-making is closely intertwined with the teaching process. It is suggested that a "table of specifications" be kept as a unit is planned and taught, thus ensuring more valid tests.

For a test to be meaninful to the pupil, it must give him a chance to demonstrate what 'he has learned, and it should be returned to him as soon as possible, while interest is still high, so that it may properly be a part of the learning process.

We also know that our teaching should be guided by what we want the pupil to learn; that is how we want his behavior to be changed by the time he finishes the course, or the unit, or the class period.

The problem has been, it seems to me, how to state our long-range objectives in a common frame of reference, and how to determine where our objectives for the foreign language learner fit into general educational objectives. After we define our long-range objectives, we must then define our short-range objectives in specific terms.

To recapitulate before continuing: For evaluation to be meaningful, we must do three things, all inter-related: 1) define our objectives, both long-range and short-range, clearly and in terms of pupil behavior; 2) teach to meet these objectives; and 3) test to determine pupil achievement of these objectives.

It is at this point that the Taxonomy of Educational Objectives, Cognitive Domain comes into focus. The Taxonomy, published in 1956, is the result of work begun in 1948 to classify educational goals in a common frame of reference for greater precision, and to provide a convenient system for categorizing test items. The principle of complexity is the ordering basis for educational objectives in the cognitive domain. These objectives are: Knowledge, Comprehension, Application, Analysis, Synthesis and

Evaluation. You will note the hierarchical arrangement, with each higher level subsuming those below it. I have obtained permission of the publisher to reproduce the Condensed Version of the Taxonomy which defines and explains these categories more fully, and I urge you to pick up a copy as you leave today. This Condensed Version does not contain the test examples for each category as the Handbook does.

When we think about reading and writing in our native tongue, we usually think of entering into the higher levels of the cognitive processes, beyond the knowledge of facts and terminology. What I am inviting you to do today is to explore with me the cognitive processes which we can reasonably expect of the foreign language learner at various stages of learning the reading and writing skills, to decide in which cognitive areas you wish to measure his achievement and to write tests accordingly. (A selected bibliography on educational measurement and evaluation in general, and on foreign language testing in particular, is to be found at the end of this article.) My purpose is to stimulate your thinking about testing from a different perspective. As you will quickly see the terrain is chartered, and I can only ask you to share with me my throughts and questions as we explore together.

The Taxonomy, then, can help us direct our pupils beyond acquisition of facts and terminology to the higher-level intellectual skills. It should be emphasized here that knowledge is not unimportant. As weh have seen, it is the first step in the hierarchy. . .obviously, one cannot comprehend what he doesn't know; or apply, analyze, synthesize or evaluate what he doesn't comprehend. We also know that we cannot expect pupils to learn what they haven't been given an opportunity to practice. (Parenthetically, pupils also seem to study what they expect to be tested on.)

Let us now consider the two major kinds of tests: objective-type tests and essay tests, and their advantages and disadvantages, because testing will be a matter of deciding what it is we want to measure, and then choosing the kind of test to accomplish this purpose.

Objective tests include the following kinds of items: recognition items for reading tests, such as true-false, matching and multiple-choice; and recall short-answer completion items for writing tests. Advantages of objective tests includes: 1) greater smapling (a much more extensive area can be tested in a shorter period of time than can be tested with an essay test); 2) greater scorer reliability (only one answer is accepted no matter who does the scoring); 3) greater ease of scoring.

As for essay (composition) tests, they are less reliable to score, but certain procedures can be used to improve reliability. Although a good essay test cannot be quickly prepared, it does take less time to prepare than an objective-type test. The greatest advantage of an essay test is that it clearly provides the opportunity to test the pupil's writing in the light of how he chooses, organizes and presents information.

The kind of test chosen will be clearly affected by the level of learning to be tested. In the Knowledge category (as defined in the Taxonomy), objective tests are more efficient. Some publishers make such tests available to users of their textbooks. Several examples of the objective-type items are given in Making Your Own Language Tests by Dr. Nelson Brooks, which is available for you as you leave today. Numerous test items are also illustrated in the Lado and Valette books in the bibliography. These items which measure control of reading and writing in the earlier stages of learning, where the emphasis is necessarily on the Knowledge area, properly deal with structure and vocabulary.

There is no doubt that objective-type test items are easier to write for the Knowledge category. However, if we wish to test pupil achievement in the higher congnitive levels, we must learn how to write test items to measure it. Why do we not often see reading tests prepared for measuring beyond the comprehension level? Why not writing tests appropriate to the three highest levels of the cognitive domain? Does measurement of the higher levels of analysis, synthesis and evaluation lie outside the realm of foreign lang-

uage teaching? Just how creative can we expect the learner to be in a foreign tongue?

Tests for the intellectual skills must involve knowledge and skills which the pupil has previously learned, but the problem itself must be new to him. Thus, in the earlier stages, reading material for comprehension should be in the form of re-combination narratives, for example. Lado lists multiple-choice items for reading comprehension at the intermediate level dealing with interpretation and with the testing of sequence signals. "Pure" reading tests for the Comprehension and Application levels for more advanced students, so as not to involve writing, can also best be tested by skillfully constructed multiple-choice items. Since tests of this type are rather difficult to compose, the teacher may wish to use examinations available in some review workbooks.

Since writing involves reading, it is higher in complexity than comprehension. To test writing in the earlier stages is to test the knowledge of spelling and punctuation. Completion items may be used, but the difficulty is in finding items for which only one answer would be correct. A relatively easy writing test at the Application level is a dictation which the pupil has not yet studied in that form. . .perhaps a résumé of a selection assigned for extensive reading. What he has learned about spelling and punctuation would have to be applied. Dictations of this type should not be mistaken for "creative" writing. . .the vocabulary and syntax are supplied by the teacher. One-sentence answers to questions referring to passages supplied for reading comprehension could also be used. The questions would have to be very carefully constructed, and the passages not previously studied, so that memory, which would fall into the Knowledge category, would not be tested.

For more advanced pupils, another technique at the Application level is to ask for a brief paragraph about each of two or three pictures. Only mechanics would be graded. At an even more advanced level, the pupil would be asked to write two short compositions about two pictures, with a time limit of about 30 minutes each. Content would be graded as well as mechanics, with two separate grades. Regarding content, the teacher can make a checklist of the points of information which should be included, their logical arrangement, and the sequence signals expected to be used. The teacher

may also ask the pupil to write, in a controlled manner, a letter to a friend in France, a description of his house, an account of his vacation, etc.

Higher lever skills in writing requiring the pupil to analyze, synthesize and evaluate seem to be best tested by the true essay-type questions; that is, we ask him to express his thoughts in writing. While it is quite true that we can't expect the pupil to learn how to write unless we ask him to write, it is also true that the essay test demands a rather sophisticated control of the language. Let us now consider the higher levels in the cognitive domain, and some illustrative essay test items.

Analysis is important in advanced foreign language study when the student is expected not only to comprehend a work of literature, but also to appreciate its beauty through a study of its structure. As a beginning writing test in this area, the pupil may first be asked to write a résumé of a selection, or to outline a story. At an even more advanced level, we might ask him to determine the point of view of an author, or to write an "explication de texte" of a poem not previously studied in class.

Synthesis is the intellectual skill which is often called "creativity." The learner must create something new. We may ask him to complete an unfinished story or poem, to relate the most amusing or saddest thing that ever happened to him, and, much later, to write an original poem or short story.

The ability to evaluate involves all the behaviors which precede it in the Taxonomy. The learner must have knowledge of the criteria for evaluation, he must comprehend what he is to evaluate, he must apply the techniques of evaluation, he must be able to analyze what he is to evaluate, and he must be able to synthesize all these elements so as to write an original communication. To ask him to write a critque of a play or film would fit into this category.

Composing good essay test items is quite demanding. Some principles to be observed are: 1) do not test for specific information; 2) make clear the scope of the response; 2) allow the pupil enough time to select, organize and write; 4) have all pupils answer the same questions. Scoring essay tests is also demanding. It is a subjective process, but there are a few suggestions to make it more reliable: 1) make a list of the important points which you think should be mentioned; 2) do not look at the

pupil's name before you mark his test; 3) score all responses to one item before going on to the next one, if there is more than one question; 4) inform pupils about the method of scoring in advance; they should know if errors in spelling and punctuation will affect their marks, or if separate marks are to be given for content and mechanics. My own bias is that if there are too many errors in mechanics, the wrong testing instrument has been selected.

The problem seems to be to determing just w h e r e in the learning sequence the pupil can move out of the Knowledge category into the higher level processes. It seems to me that testing reading and writing beyond this point has remained a rather vague area for most of us. Just when will the foreign language learner be able to read and write meaningful material without being hampered by mechanics? It is clear that we cannot expect the fourth year high school pupil to write a good résumé of a story or to write an interesting letter unless we have carefully taken him through succeeding steps in his foreign language learning so that he can arrive at this skill. Since less than ten percent of those pupils who begin foreign language continue into the fourth year, small wonder that the area of the essay test has not been clearly delineated, and that those relatively few foreign language teachers who do teach fourth year classes are concerned about guidelines for testing. What I hope to do here is to offer you some of the tools, so that together we may develop these guidelines most applicable in our individual situations.

Selected Bibliography on Testing

Bloom, Benjamin S. (editor) et al. Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. David McKay Company, 1956.

Diederich, Paul. Short-cut Statistics for Teacher-made Tests. Educational Testing Service, 1960.

Ebel, Robert L. Measuring Educational Achievement. Prentice-Hall, Inc., 1965.

Lado, Robert. Language Testing: The Construction and Use of Foreign Language Tests. McGraw-Hill Book Co., 1964.

Smith, F.M. & S. Adams. Educational Measurement for the Classroom Teacher. Harper & Row, 1966.

Valette, Rebecca M. Modern Language Testing: A Handbook. Harcourt, Brace & World, Inc., 1967.