

AN EVALUATION OF ORAL/AURAL TESTING IN MODERN FOREIGN LANGUAGES AT BOSTON UNIVERSITY

by Gail King and Diane Nadeau

ABSTRACT

This paper examines various forms of pure skills and skills combination laboratory testing done at Boston University during the past ten years. Suggestions for refinements and basic testing principles are discussed for each format.

I. INTRODUCTION

The original intent of this project was to research different ways of using the language laboratory facilities of the Geddes Language Center (GLC) at Boston University for language testing at the elementary, intermediate, and advanced levels. We discovered that, although a tremendous amount of literature exists about language testing in general, very little of it contains any useful information regarding lab testing per se.

Because lab testing proved to be an unresearched field, we felt that the next most logical approach of this research would be to study the oral/ aural exams already on file at GLC and, based on readings and our own expertise, to determine the strengths and weaknesses of these various test formats, and to suggest improvements where appropriate.

We found many similarities among the tests we examined. To facilitate their identification, we will refer to formats as follows:

--LISTENING COMPREHENSION-AURAL DISCRIMINATION:
students choose among multiple choice questions on grammar objectives or on comprehension of a passage.

--MEMORY RECITATION: students recite memorized material.

--DISCUSSION: students discuss preassigned topics without written support.

--LISTENING/SPEAKING,

FOR MIMICRY: students hear and repeat;

FOR GRAMMAR: students respond with appropriate grammatical objective;

FOR PASSAGE RECOGNITION: students identify and comment orally on a familiar excerpt;

FOR PERSONAL EXPRESSION: students respond orally to unfamiliar material.

- LISTENING/WRITING,
FOR PASSAGE RECOGNITION: students hear a familiar passage and identify and comment on it in writing;
FOR DICTATION: students produce the written equivalent of what they hear.
- LISTENING/WRITING/READING: students hear clues, write answers, and read their answers aloud.
- LISTENING/WRITING/READING: students read and answer written questions after listening to a clue or passage.
- LISTENING/READING: students choose among written multiple choice questions based on listening passage.
- READING/SPEAKING: students read a passage aloud.
- WRITING/READING OR ORAL COMPOSITION: students record an original composition with written support.

II. GENERAL DESIGN FACTORS

Although most of the publications consulted divide test designs according to each of the four skills (listening, reading, speaking, and writing), these four skills are rarely taught in isolation and most oral/aural tests therefore appear in hybrid form. Mimicry involves listening and speaking; pronunciation often depends on reading and speaking; dictations combine listening and writing; multiple choice can combine listening and reading; oral answers include listening or reading and speaking, to name a few combinations. It is therefore essential to determine which skill or which combination of skills one wants to test before designing an exam. With those objectives in mind, the test designer must also develop a consistent standard of correction. Whether the exam requires an objective or subjective student response, a scale which reflects the test's objective(s) is an essential part of any test. It might resemble the following:

type of response	objective tested	number of errors	grading scale
oral	pronunciation		0 1 2 3 4 5
oral	fluency		0 1 2 3 4 5
oral/written	grammar		0 1 2 3 4 5
oral/written	vocabulary		0 1 2 3 4 5
oral/written	content		0 1 2 3 4 5

This type of scale is particularly useful when students must make an oral or written response beyond true/false or multiple choice selections. Even while demonstrating comprehension of passage, students may make incorrect use of the language. We believe that faulty language skills often undermine communication and that, although native proficiency is a remote incentive to elementary and intermediate language students, linguistic exactness underlies any testing activity. It is also true that both native and nonnative instructors quickly learn a set of typical errors which native English speakers make when learning a foreign language. The instructor therefore automatically interprets and "understands" these errors. In these cases, communication is established but the language is not used properly. Would a native Frenchman who is not a language instructor, understand so readily "une femme's place" and "j'ai passe mon examen" to mean "a woman's place" and "I passed (not took) my exam"?

IV. EXISTING TESTS

The tests mentioned in this paper reflect only those which are on file at the Geddes Language Center at Boston University. Other tests, or variations of those listed below, obviously exist, but it is our intent to do an evaluative rather than a comprehensive study. We determined that it was best to divide test formats into "pure skill tests" where the objectives involved only one skill and "hybrid skills tests" which combined two or more skills. We found that very few tests or test sections isolated any one skill: it is extremely difficult to test a skill in isolation because language is an activity which depends upon both reception (listening, reading) and production (speaking, writing). Although language teaching may stress one or two of these skills, no one skill is likely to be taught exclusively. A hybrid format is also more likely to occur in an oral/aural test where listening is frequently combined with other skills.

A. PURE SKILLS TESTS

1. LISTENING. The only test on file which isolated this skill was the LISTENING COMPREHENSION - AURAL DISCRIMINATION format. The exam was in two parts. First, students heard a series of questions and had to choose from three multiple choice possibilities, also read aloud. Each question tested a specific grammatical point which students were to recognize aurally and resolve. Typical objectives tested were verb tense selection, object pronoun selection, conjunction selection, mood selection, and gender identification via pronoun substitution. In the second part, students heard short passages which were again followed by multiple choice comprehension questions.

All the questions and multiple choice answers in this test were presented orally. Students therefore received no written support which might have given them clues to the correct answers, supplementing their aural perceptions. The passages were short so that listening comprehension did not become a memory test. The questions were well designed: they tested comprehension of grammatical and structural concepts and information, rather than less important specific details. An example can best illustrate this format for passage comprehension questions. A passage contained the sentence, "Cecilia couldn't go to the party on Sunday because it was her brother Gustavo's birthday." Asking what day the party was on would have been memory testing, whereas asking why Cecilia couldn't go to the party was a truer evaluation of listening comprehension. The choices for this kind of test were often a combination of a) the correct response, b) a grammatically inappropriate response c) a factually inappropriate response. For example, possible choices given for the question, "Are you the mother?" were "a) Yes, I'm the mother." (correct response); "b) Yes, the mother is here." (factually inappropriate); and "c) Yes, she is the mother." (grammatically inappropriate). These multiple choice answers contained distractors which sounded familiar enough, because of vocabulary or phrasing, to make them viable though inappropriate options.

The multiple choice format lends itself to easy correction; however, good multiple choice questions are extremely difficult to compose. The options should neither be so specific that they test pure memory nor so general that they reflect common knowledge or facts derived from another source. To reduce guessing on this kind of exam we would include a "none of the above" option. Further, when administering a multiple choice exam such as this one, we would give clear instructions that no one write until all options have been read in order to prevent poorer students from watching the hand movements of their neighbors or of the better students in class. To prevent insufficient vocabulary from interfering with listening comprehension, design a passage around vocabulary which students have already studied, but present them with a truly original and unfamiliar passage.

The multiple choice format is not the only way to test listening comprehension in isolation. Other formats using pictures and charts can be used at the beginning, intermediate, or advanced levels. At the elementary level, Linder suggests the use of pictures for aural discrimination by matching a spoken fact or statement concerning time, schedules, food, numbers or family members to a drawing or chart. The emphasis of this listening test is mainly comprehension of isolated vocabulary, which is why the format is more appropriate at introductory levels. Valette proposes another format for pure listening: translation into English of oral statements or identification, in English, of grammatical concepts such as number, gender, and verb tense.

2. SPEAKING. The isolated speaking test is also rare, but a RECITATION format exists which required students to recite part of a poem from memory. Students were also asked, in a DISCUSSION section to speak on a previously assigned topic during a timed pause. It is likely this was memory recitation as well, since the topic was preassigned.

One may question whether these two formats really do involve "speaking" since in the one case the speaking was not original and in the other it was not spontaneous. It remains true, however, that RECITATION is an excellent format for testing pronunciation and intonation without having students rely on structure. Preassigned DISCUSSIONS are valid speaking evaluations as long as the severity of the grading is adjusted accordingly, that is, by penalizing the monotone delivery of totally memorized student recordings. In testing isolated speaking, first we must decide whether to use recording (i.e. lab) facilities or to conduct individual face to face sessions (i.e. interviews). Both have their advantages and disadvantages.

Administering the speaking test in the lab is quickly accomplished and the conditions are generally more uniform than during the live interview. Drawbacks include mechanical breakdown, scheduling make-up sessions for students who miss the exam, and the mechanically time consuming correction process. However, it is obvious that if the speaking time is limited to one or two minutes, corrections would not be any more demanding than correcting a written paragraph. Checking precise pronunciation on a taped RECITATION is more methodical than during a live interview since uncertain sounds can be replayed. We suggest testing different speech patterns by assigning a variety of genres for RECITATION - narration, poetry dialogues, and underlining the precise pronunciation objectives aimed at on a copy of the passage for each student as you listen.

The personal interview method may appear, at first, more appealing to students, particularly if they are not used to regularly assigned recording drills. Yet with speaking the only skill tested, the live interview remains an unnatural, one-way conversation. The instructor only initiates the students' oral presentations. It is quite difficult for the instructor to appear responsive while remaining neutral and not interfering. The only advantage for the live interview for DISCUSSION or RECITATION is that the instructor can aid the hesitant student with added visual and verbal cues. This type of prompting by the instructor is contrary to the uniform possible in a recorded speaking test where students receive no special assistance. Despite this drawback, the live interview is a logical choice when the goal of the speaking test is vocabulary: pictures or the verbal cue

“how does one say (word) in (target language)?” are appropriate during the live interview. Design a scale/score sheet which lists the areas being tested; vocabulary, pronunciation, contest, fluency, etc. Mark and grade it during or immediately following the students discussion.

3. **WRITING AND READING.** Obviously, no oral/aural test can examine these skills in isolation. They are, however, often combined with speaking or listening skills as we will show below.

B. HYBRID SKILLS

As we have indicated, it is logical that there are more examples of hybrid exams on file at GLC: the receptive/productive aspects of language are rarely isolated in language teaching. In addition, the majority of the language students at Boston University register for the four-skills track, rather than any other approach, therefore most exams reflect this student distribution.

1. **LISTENING/SPEAKING.** The LISTENING/SPEAKING combination is one of the most popular of the test formats for the simple reason that it makes the most effective use of the oral/aural skills within framework of the audiolingual exams on file in many languages. Since it is impossible to list all of the items on all of the different language exams on file, we will group them into four basic test formats: MIMICRY, GRAMMAR, PASSAGE RECOGNITION, and PERSONAL EXPRESSION.

Pronunciation and retention are the best applications of the LISTENING/SPEAKING for MIMICRY format. At the elementary level, mimicry aimed at the pronunciation of particularly difficult or unfamiliar sounds in the target language. Valette recommends testing one phoneme per enunciation in order to standardize corrections. At the upper levels, retention was added to mimicry by presenting longer sentences for repetition. Again, use a marked copy for corrections.

LISTENING/SPEAKING for GRAMMAR asked students to respond orally with grammatical transformations. Often students had done a similar type of exercise in written form, either in class or on homework assignments; but oral testing is more difficult since students do not have the leisure to ponder and debate the correct answer. The tests included typical grammatical objectives: pronoun substitution, tense transformation, mood selection, negatives, interrogatives, indirect discourse, plurals, and relative pronoun.

LISTENING/SPEAKING for PASSAGE RECOGNITION was tested in three ways. In the first, an excerpt which students had heard or studied before was read and students were to answer orally five questions relating to the passage. In the second, the passage and questions were played through once for comprehension, and then repeated a second time with pauses for questions. In the third, an excerpt was played from a passage they had heard or studied before

and they were asked to identify and situate the excerpt. The excerpt was then replayed and students were given a short time to think about it again and then to revise what they had said the first time.

PASSAGE RECOGNITION at first seems a logical type of exercise and the popularity of the format at our intermediate levels seems to confirm its validity. Students have heard the material before; they have learned appropriate vocabulary; and they have discussed plot in class and in written answers and compositions. Students perform well on the exercise. It is necessary to remember, however, that these are not true evaluations of listening comprehension: students are not asked to comprehend but to recall. Once they mentally locate the passage they can "tune out" and still record coherent answers concerning the plot. Testing listening with material students have studied primarily tests recall and plot memory, not comprehension. Weighing grammar, vocabulary and fluency more severely than the actual passage identification, in fact, making these the real test objectives, restores some validity to PASSAGE RECOGNITION as a test of listening/speaking. Thus, familiar materials become the pretext for, not the aim of, speaking. A score/scale sheet determines the objectives and standardizes correction.

We feel that a truer test of listening/speaking consists of an original passage which implements the vocabulary and grammatical structures which the students have studied. At the beginning levels, short questions follow the passage. Instructors choose whether to grant mental preparation time or to require an immediate response. At the advanced levels, the test might require longer, more creative replies from students. Students should be given a working glossary before listening to an original passage if it contains unfamiliar vocabulary. If they are free to play back the passage a reasonable number of times, depending on the duration of the lab exam, the listening situation duplicates the students first experience listening to any new passage. Unfamiliarity with the passage gives a fairer evaluation of students' listening skill, just as the more spontaneous recording, which does not stem from previous class discussion, is fairer demonstration of students' speaking skills.

We found tests of LISTENING/SPEAKING for PERSONAL EXPRESSIONS at the beginning, intermediate, and advanced levels. A first year exam consisted of a series of personal questions on age, tastes, and facts. These differ from the grammar questions because they are not part of a larger and systematic exercise designed for a specific grammatical objective. Upon consulting with some of our colleagues, we discovered three interesting and, at times, surprising results. First, at the beginning levels, it was felt that this type of question best introduced a longer oral/aural exam. It reassured, oriented, and encouraged students; and it got students thinking in the target language. Second, a rather amusing result of such personal

questions was that students often remained inflexible in their thinking and were therefore unable to formulate a logical and grammatically correct answer. For example, if the question asked was "What did you have for breakfast?", students often resorted to inappropriate vocabulary to express what they really had eaten for breakfast rather than composing sentences which, while perhaps untrue, were grammatically correct and used vocabulary of which they were certain. Similarly when asked "Is there a cathedral in Boston?" many students were unable to answer because they misjudged the intent of the question. They mistook affirmative/negative formulation for factual exactness. Students should be forewarned to answer logically and correctly rather than factually on this kind of question. Third, elementary students, on the whole, performed better on simple, direct questions at their level of language than did second year students. There appeared to be a degree of regression in speaking.

At the advanced level, an interesting test design combined LISTENING/SPEAKING with PERSONAL EXPRESSION. Students heard a one-sided dialog on tape with blanks occurring after every statement. Students played the tape as often as necessary to get the gist of the conversation and to record logical counter replies. We do not know whether students could write out the responses they wanted to record or not. If listening and speaking were the primary objectives, we assume pencil and paper were not allowed.

2. LISTENING/WRITING. In a test of listening and writing, students actually compose phrases or sentences or provide vocabulary rather than simply making true/false or multiple choice notations, as they would on a pure listening test. LISTENING/WRITING was tested by PASSAGE RECOGNITION and DICTATION.

Instead of recording answers as described in the listening/speaking section above, students wrote out appropriate answers. The same comments concerning the use of familiar materials apply here as well. Written PASSAGE RECOGNITION tests recall much more than comprehension; the test objectives therefore remain written grammar, spelling, and sentence structure. If students are asked to write about materials they have studied only orally/aurally, we would prepare a vocabulary/spelling list. The list could be distributed either before or during the test. Cognates are often used in listening passages but, since cognates often have a slightly different spelling, such a glossary helps avoid typical spelling errors.

Very few LISTENING/WRITING tests are on file at GLC, perhaps because instructors do not bother to record their oral passages. They simply dictate or read the passage and questions in class. DICTATION stresses the sound-symbol relationship, spelling, and grammatical precision. It is most effective when instructors compose original dictations incorporating vocabulary with the grammar being

studied, such as combining vocabulary on travel with the imperfect tense in a descriptive paragraph. Students are told which grammar points and vocabulary to prepare. We find that it is not productive to dictate verbatim from passages students have memorized, nor is it useful to assign students to listen to recorded dictations using unfamiliar and random vocabulary, tenses, or grammatical rules. We were amazed the first time we assigned this type of dictation and found that some students could spell extremely well without having any idea of what they had written when asked to translate it. We find dictation most useful as a tool for students to use in preparing for a vocabulary test where they are expected to produce the target language equivalent of an English cue. The next step is for students to study the English equivalents of certain vocabulary items. We compose an original dictation which incorporates the assigned vocabulary and students take the passage in dictation and finally translate specific parts of it. The translation makes students aware of verb tenses, structural differences, and vocabulary. It is then easier for students to make the next step which is to study target language equivalents of English words and phrases.

3. LISTENING/WRITING/READING. An oral dimension was added to written PASSAGE RECOGNITION and DICTATION by having students record their written responses. This compromise format is more complicated than LISTENING/WRITING yet less demanding than LISTENING/SPEAKING because students have written support for their speaking. It does, however, differ greatly from the LISTENING/WRITING test since only the recorded answers are graded. It is most effective in those cases where the writing skill is stressed more than speaking.

4. LISTENING/READING/WRITING. One LISTENING/WRITING test provided students with written support in the form of a question. An original passage, using recently acquired vocabulary, was recorded on individual tapes. Students could play it back as often as necessary. They wrote answers to questions they had read about the passage they heard. The questions were designed to facilitate listening comprehension. In this case, the written reply was graded. The test is a simpler version of the LISTENING/WRITING design because the questions were provided in written form.

5. LISTENING/READING. One listening test design provided students with a series of multiple choice questions in written form, rather than relying strictly on their aural skills. The instructor created an original passage incorporating vocabulary students had learned. The pre-recorded passage was played in class. Because the passage was quite long (8 minutes), the students received copies of the multiple choice questions. There are two ways of administering such a test. In one, students answer questions in succession as the passage progresses. In another, the passage is played through once, without students having questions. Questions

are then distributed to students and the passage is replayed for question answering. In the latter situation, students have a fairly good idea of the "plot" during the second listening and they can proceed from question to question without being distracted by either the tape or the written material. Since retention is not being tested here, it becomes possible to ask for more specific details than would be appropriate in a pure listening test.

6. **READING/SPEAKING.** Students were asked to read aloud a passage they had never seen before. The passage was short but contained an excellent selection of vowels, nasals, and target language sounds. Its dialog format allowed more intonation variation than straight narration would permit. The instructors set up a general guideline concerning the objectives of this oral component: a) vowels and nasals, b) consonants, and c) fluency, elisions, mute endings, and intonation. Students had never seen the passage before the exam but they had been introduced to all of the vocabulary, verbs and structures it contained.

Despite the scoring guidelines, instructors were free to curve their scale since no numerical chart listed points per error. We suggest that it would have been better to prepare a marked copy of the passage underlining which vowels, nasals, elisions, etc. would count against students' grades if wrongly pronounced. Reading such a passage tests pronunciation, rhythm and intonation. The test objective is one step above pure mimicry since no model is provided.

7. **WRITING/READING.** The natural progression from **READING/SPEAKING** as described above is the **WRITING/READING** combination, or **ORAL COMPOSITION**. The two objectives of this test are the ability to write coherently and correctly and the ability to read one's own written composition with the proper rhythm, intonation, and pronunciation. It can be given during a class held in the lab or assigned to be done outside of class. The in-class lab exam approach guarantees that students are truly reading their own work, as is not the case in the outside-of-class lab assignment. This format is less stressful for students than the speaking test while still incorporating an oral component; it is more creative and personal than reading pre-existing materials; and it is more active and involved than pure essay writing.

IV. CONCLUSION

Every test design meets certain needs and includes certain objectives. In the process of this evaluation we have discovered that there are these key steps in designing oral/aural tests: 1) determine test objective(s), that is, which skill(s) you want to test; 2) determine the format by which students are to demonstrate the skill; and 3) prepare a grading scale/score sheet to measure the test objective(s). These may seem obvious since they pertain to all forms of testing; but it seems useful to emphasize them because, in designing oral/aural

tests, it is very difficult to isolate specific skills without careful planning.

The aim of this evaluation of existing Geddes Language Center test formats and the further suggestions about them is to stimulate more thought and to encourage new ideas in oral/aural testing. It is not meant as a definitive statement but as a point of departure for new creativity and application of further innovations.

Article by Diane Nadeau and Gail King, Geddes Language Center, College of Liberal Arts, Boston University, Boston, MA 02215

BIBLIOGRAPHY

Birkmaier, Emma Marie, ed., *The Britannica Review of Foreign Language Education*, vol. 1. Chicago: Encyclopedia Britannica, 1968.

Clark, John L. D., "Measurement Implications of Recent Trends in Foreign Language Teaching," in *Foreign Language Education: A Reappraisal*. Edited by Dale L. Lange and Charles J. James. *The ACTFL Review of Foreign Language Education*, vol. 4. Skokie, IL.: National Textbook, 1972.

Lange, Dale L., ed., *The Britannica Review of Foreign Language Education*, vol. 2. Chicago: Encyclopedia Britannica, 1970.

Linder, Cathy, *Oral Communication Testing*. Skokie, IL.: National Textbook, 1977.

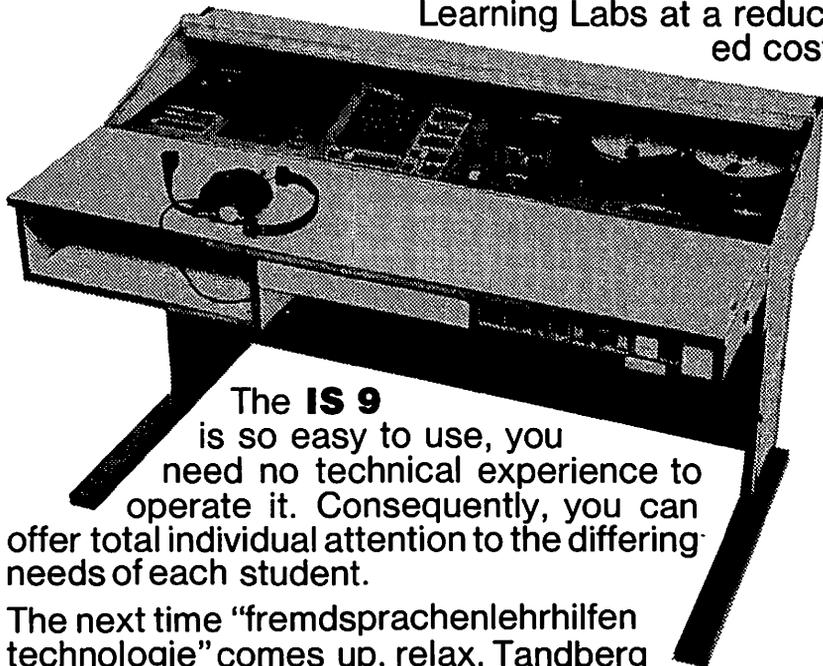
Rude, Ben D., "A Technique for Language Lab Testing" in *Language Learning*. vol. 17, 1967, 151-53.

Valette, Rebecca M., *Modern Language Testing*, 2d ed. New York: Harcourt Brace Javanovich, 1977.

TANDBERG® MAKES Fremdsprachenlehrhilfentechnologie EASY!

Difficult words become easy with the **Tandberg IS 9** Learning Laboratory.

Microprocessor technology has allowed greater freedom and flexibility to produce the most versatile Learning Labs at a reduced cost.



The **IS 9** is so easy to use, you need no technical experience to operate it. Consequently, you can offer total individual attention to the differing needs of each student.

The next time "fremdsprachenlehrhilfen technologie" comes up, relax, Tandberg makes it easy to say, "foreign language teaching aids technology."

We've worked hard, so you won't have to.

TANDBERG OF AMERICA, INC.

Labriola Court

Armonk, New York 10504

(914) 273-9150