

SYSTEMATIC PREPARATION OF THE MULTIPLE-CHOICE LISTENING TEST

by Gilbert A. Jarvis

The preparation of foreign-language listening test items in a multiple-choice format is one of the most difficult and demanding tasks of the foreign-language educator. Ironically it is one of the tasks for which the typical teacher has little preparation, little expertise, and little time. In contrast with the professional item writer working for a test publisher, for whom the preparation of a few items constitutes an entire day's work, the classroom teacher must often create entire tests during a forty-minute preparation period or amid myriad other evening activities. The unskilled teacher, moreover, has few useful sources of information. Valette (1967) has provided an extremely useful collection of various item types and formats but has not intended the volume as a delineation of specific considerations in constructing the test items. General measurement texts discuss psychometric factors, however, extrapolation to the foreign-language situation remains a formidable task.

The purpose of this discussion is to present a relatively concise statement of a sequence of procedures for preparing the listening skill test and to indicate where media are crucial to the development of a sophisticated test. The discussion is limited to non-statistical considerations with which the practitioner must deal.

First: Determine Testing Facilities

While it is true that the listening comprehension test can be administered in the classroom, it is much more effectively administered in the language laboratory where fewer extraneous and irrelevant elements enter into the evaluation. For example, the student's score in the classroom includes, in varying and uncertain degrees, a measure of the room's acoustical properties at the student's particular location, of his proximity to the ventilating system (which may make noise), of the severity of his or a neighboring student's coughing or sneezing, of the personality or smile of a neighboring student of the opposite sex, and many similar factors, all of which are considerably diminished in the language laboratory. Without the availability of a laboratory the teacher must accept limitations inherent in the classroom and must take measures to minimize their effects. The teacher who meets several classes in different rooms must recognize that the setting *does* influence student results. The long history of research on this question indicates that comparisons between groups tested in different surroundings lack meaningfulness to the extent that any of the rooms

Multiple-Choice

deviates from a typical classroom. The teacher makes such comparisons whenever he applies the same grading standards to multiple students. The effects of the differences in facilities are minimized if the administration of the test is standardized. The crucial element in this standardization of conditions is an audio recording of the test. With a recording all groups of students hear the same rendition, a condition which is for practical purposes impossible when the teacher reads from a script in the classroom. Not only do linguistic factors such as rate, frequency and duration of pauses, intonations, and pronunciation change but also factors such as the length of time between items, the loudness with which the teacher reads the items, and whether the teacher stops to make a comment which might even include a clue for a particular group of students — all interfere with standardization of conditions. The adolescent and the young adult have keen senses of justice, and as long as we persist in grouping these learners into what we call classes and in making comparisons among them, they will recognize that it is blatantly unfair to ignore inequities in testing conditions.

The use of a recording has further advantages. While making the recording the teacher (or perhaps another speaker of the language) need not be concerned with anything other than the preparation of the tape. Matters such as timing, instructions, sequencing, and number of repetitions of items can receive appropriate attention. Furthermore, the test can be re-used, not only in other classes but at other times, if desired, with no additional preparation time.

Attention must also be given to other media which can be used. There is a certain lack of realism inherent in presenting listening items without a visual dimension. Hearing spoken language, isolated from all other context, can appear as an academic exercise rather than as the simulation of a communication situation, which it is believed to be by many educators. Language rarely occurs in such isolation. The teacher can quite easily avoid isolation by adding a visual dimension to the test. Items can be related to films, filmstrips, slides, transparencies, posters, or simple photographs cut from magazines. The visual can simply situate a single item, or it can provide a complex of events and relationships about which questions are asked. Rather than providing unnecessary clues, as some may claim, the visual restores clues which are unnaturally deleted in some purely auditory tests. Decisions about the use of visuals, in the final analysis, depend upon the instructor's evaluation of the appropriateness of simulating realistic contexts for his particular instructional setting.

Second: Define the Scope of the Test

Given adequate facilities, the instructor determines the sophisti-

Multiple-Choice

cation of the test by his skill in systematically planning the test and in writing the individual items.

The fact that the teacher wishes to give a listening test peremptorily implies that listening ability is one of the instructional objectives of the language program. The listening skill identifies the behavior to be evaluated, but the content must also be specified. What language items should the student have learned to comprehend during this part of the instructional strategy? Most practically, the content is usually stated in terms of the instructional materials used (page numbers, chapters, lessons). A detailed listing of each vocabulary item or structure would require excessive time for most practitioners.

Third: Assure the Representativeness of the Test

In measurement terminology the test must have "content validity," meaning that the test items must be representative of the entire content as defined in the preceding step. A test is always a sample of the total content. It is easy to see, therefore, that a sample should always represent the whole from which it was "drawn."

Common violations include testing a portion of the content to a disproportionate extent. Lack of representativeness occurs, for example, when a structure, expression, or word, which may indeed have occurred in the lesson, appears with much greater or lesser frequency on the test. A particular idiomatic expression used once in a dialogue and receiving no special subsequent emphasis violates the representativeness when it appears in five of fifteen listening test items.

A listening test including only discrimination items, where the student must merely distinguish between pairs of sounds, is not valid if in the instructional strategy emphasis was solely upon listening vocabulary. If the listening test contains vocabulary which appeared in a text passage utilized only in reading skill development, it lacks validity. The test items must, therefore, be apportioned to the various segments, both in the content dimension and in the behavioral dimension, of the instructional strategy.

Fourth: Determine Item Format

The format of the individual items may vary considerably. The two most commonly used formats, both in foreign languages and in other areas, are adequate for all listening skill tests. They may be schematically represented as:

- Type I* xxx xxxxx xx x xxxxxxxx xxx . . .
Completion
- a. xxxxx
 - b. xxxx
 - c. xxx xxxxx
 - d. xxxxxxxx

Multiple-Choice

- Type II* xxxxxx xx x xxxxxx xxxxxx xxx xxxx xxxxxx?
Rejoiner a. xxx xx xxxxxx xxxxxxxx.
 b. xxxx xxx xxxxxx xxx.
 c. xx xxxxxxxx xx xxx xxxx.
 d. xxxx xx x xxx xxxxxxx.

The ubiquitous use of these formats assures minimal student confusion, provided that the test maker follows the logic inherent in each of the format types. In Type I items both the stem and the answer are parts of the same utterance. In other words, if the utterance were occurring in an actual communication situation, the same speaker would be uttering both the stem and the answer. In contrast, Type II items inherently imply two speakers. The correct answer is the most appropriate retort to the stem. Frequently, therefore, the stem is a question. It may, however, be a statement which elicits another utterance. (E. g. "It's really a nice day today," could evoke the response "Yes, how about going on a picnic.")

Type I items require that attention be given to avoiding either unnecessary repetition or extraneous clues. Any item which is written in such a way that the same word begins each alternative should be modified in the interest of efficiency by moving that word to the end of the stem. If, for example, the stem of Sample Item I below ended with the words "have to," each alternative would inefficiently begin with "look at a."

Sample Item I
(assuming English as
a foreign language)

- If you forget the date, you have to look at a . . .
- a. compass
 - b. clock
 - c. cantaloup
 - *d. calendar

An extraneous clue would be introduced into the above item if one of the wrong answers (called "distractors") were "album." The student could eliminate such a distractor merely by knowing that the word "a" is not used before a vowel sound — knowledge quite irrelevant to that which the item is intended to measure. The clue is, in a sense, a violation of the validity of the item.

It is also at this stage that the test maker must decide whether or not items are to be based upon the content of a passage or upon a visual such as a short film. If not, each item will be independent, and comprehension of it will be sufficient to answer it. The test maker

Multiple-Choice

must also determine whether the test will measure solely listening skill or whether the items will require an additional skill. If listening-reading hybrid items (oral stem, written answers) are to be used, for example, the instructor should consider projecting the written answers from an overhead projector rather than duplicating them for each of the students. The projection permits the perpetual use of a single answer card or sheet on which the student marks only the letter of the correct answer. Even if the test is solely of listening skill, the overhead projector can be used to indicate the number of the item. The teacher can indicate item number as the test tape is playing. Auditorily numbering the item in the foreign language does provide additional, desirable practice in hearing the language; however, any item which is missed because of misunderstanding of the numbering clearly lacks validity. It measures only comprehension of numbers, not any content of the item. Certitude that every student comprehends all numbers becomes then a prerequisite to auditory numbering in the language.

Fifth: Write Item Stem and Correct Answer

Determining precisely what an item measures is facilitated by writing the stem and the correct response prior to creating the distractors. A useful criterion for evaluating the clarity of the item results from asking the question "What doesn't the student 'know,' if he wrongly answers the item?" If this question cannot be answered precisely by either a single statement or a very short series of either-or statements, the item may well be testing non-linguistic factors, probably abilities in logic or reasoning. Such items would probably be answered incorrectly by some students if presented in the native language.

The student who cannot answer Sample Item I may not know the word "forget," but it is more likely that he does not know "date" or "calendar." Structurally the sentence is not complex. The expression "have to" is important but can probably be approximated from the context. Thus, the item does seem to have adequate clarity in terms of what it measures. The information is crucial if the teacher is truly to function as a facilitator of learning.

Sixth: Write the Distractors

Once the stem and correct answer are written for an item, adding the distractors is a formidable, time-consuming task. How many distractors should there be? Good logical and statistical arguments can be made for three, four, or five alternatives. Whatever the number, all the distractors require a measure of plausibility. If the purpose

Multiple-Choice

of the test is to discriminate or distinguish among students, all the distractors must attract some students. Moreover, if no one chooses a particular distractor, it has no *raison d'être* and should be exchanged for a more attractive one.

Distractor plausibility is achieved by the existence of a homogeneity among the alternatives. The distractor must in some way have a resemblance to the correct answer. In other words, there must exist an element which binds the alternatives together. It can be achieved in several ways.

A conceptual similarity, or similarity in meaning, can yield homogeneity. Consider the item:

"It's getting very hot in this apartment; please

Sample Item II

- open the "
- a. cupboard
 - b. closet
 - *c. window
 - d. oven

All the distractors are not only objects which can be opened, but they are also objects usually found in an apartment. Homogeneity would be lacking if one of the distractors were "hair" or "tree," neither of which have meaningful link to the others.

Phonetic similarity may also provide homogeneity among the alternatives of the listening item. The distractor plausibility comes from a similarity in sound to the correct answer. It must be recognized that this type of item tests a different kind of listening from that tested by meaningful similarity. Psychologically the student must make discriminations rather than associations. An item of this type could be made from Sample Item II by changing the distractors to:

- a. winter
- b. windy
- c. window
- d. winding

It is, of course, very possible to write an item which combines types of homogeneity. One should recognize, however, that some precision in interpreting what the student does not know may be lost. Moreover, the teacher must be alert against accidentally providing clues in the combination of types of homogeneity. If, for example, several items have two phonetically similar alternatives and two meaningfully homogeneous alternatives, and the correct answer is always (or never) one of the phonetically similar alternatives, the student

Multiple-Choice

with insight can utilize this information to eliminate alternatives.

Concern must also be given to avoidance of other unintentional clues. The correct answer should appear equally in all positions, in random sequence, and should not be the longest alternative a disproportionate percentage of the time. (There is a tendency among the uninitiated to make the correct answer the longest one.) General agreement exists, moreover, that incorrect (unauthentic) language should not appear in the distractors.

Seventh: After Administering Provide Feedback

Knowledge of results, always an important learning variable, is often overlooked in the FL listening test when, in fact, it is extremely easy to provide students with feedback. In the simplest form the instructor can merely re-play the test tape, explaining and repeating as necessary. The tape assures the rendition's being identical to that of the testing situation. The feedback can be enhanced, according to current psychological findings, if the student is presented with a written representation while listening to the tape. The visual representation can be accomplished through use of the overhead projector or by giving students written copies. If the test copy or answer sheet has been prepared in a conventional duplication manner, the instructor simply prepares a master copy and duplicates the oral items on the back side of the student papers, thus using no additional paper.

As more and more schools provide tape playback facilities in a "Learning Center" or "Listening Center," the teacher should consider placing used listening test tapes in the center for student practice. At times a closet with a tape recorder suffices. Occasionally a choice has to be made between student practice with the tape and re-use of it as a test instrument.

Summary

Following these steps should assure the instructor of a relatively refined test. If test analysis by computer is available, the instructor can procure statistical information which can aid him in determining the success of the test, after it has been given. Rarely will the statistical data (with which few practitioners are unfortunately conversant) disparage the quality of a test which has been systematically prepared and for which a tape recording has been utilized.

Summarily, the procedures for systematic preparation of the listening test in multiple-choice format are:

1. Specify the facilities: the language laboratory, if possible, a tape recording at the very least, and any projection materials.
2. Define the scope of the test. What content is to be tested?

Multiple-Choice

3. Ensure the content validity by making certain the items are representative of the entire content.
4. Determine item format(s).
5. Begin writing items by writing stem and correct response.
6. Write distractors considering plausibility and potential inadvertent clues.
7. After recording and administering test provide adequate feedback to the learners.

BIBLIOGRAPHY

Vallete, R. M. *Modern Language Testing*. New York: Harcourt, Brace, and World, 1967.

ABOUT THE AUTHOR:

Mr. Jarvis is an Assistant Professor of Humanities-Education at the Ohio State University, Columbus, Ohio.