

# SURVEY OF TESTS IN ENGLISH AS A FOREIGN LANGUAGE

ROBERT LADO

University of Michigan

## I. Introduction

This survey summarizes a larger report on tests in English as a foreign language prepared for NAFSA (National Association of Foreign Student Advisers). Those dealing with foreign students have long felt the need for information about the students' control of English before and after they come to the United States. The present survey describes and evaluates tests now available and summarizes general problems not yet solved in the field.

The need to determine how much English a student knows before he leaves his country has been felt by those who must pass on their application for study in the United States. Only with such information can a reasonable plan of study be made for him. If he does not know English well he must either study it in his country or plan to attend a school in the United States that offers special work in English. There is also need for tests to place the students at their proper level in their English classes, and to measure their progress, or to diagnose their difficulties. This variety of needs has resulted in a variety of tests, and accordingly the evaluation of each will indicate what it does or does not do. The only consistent requirement that every test must meet is that of measuring the student's control of the English language with as little interference as possible from other factors.

The tests reviewed in the survey are (1) the Educational Testing Service Examination for Foreign Students, (2) various tests from the English Language Institute of the University of Michigan, (3) a vocabulary test from the English Language Institute at Queens College, (4) the Inter-American Tests, (5) several tests from the Investigation of the Teaching of a Second Language, (6) the U. S. Department of State English

## ROBERT LADO

Proficiency Examination, and (7) the Villareal test for Spanish-speaking students. The next section presents a review and evaluation of each test individually.

### II. Reviews of Present Tests

#### (1) EDUCATIONAL TESTING SERVICE: ENGLISH EXAMINATION FOR FOREIGN STUDENTS

The English Examination for Foreign Students is designed to give educational institutions information about the student's knowledge of English before he is admitted. The Educational Testing Service provides the materials to the admitting institution at a charge of two dollars per student, and scores the objective parts of the test. In a circular of December, 1949, the Service announced that it was suspending the administration of the Examination overseas because of prevailing political and currency conditions abroad.

The examination is divided into five sections as follows: (A) Reading Comprehension, (B) Aural Comprehension, (C) Pronunciation, (D) Scientific Vocabulary, and (E) Nonverbal Reasoning. The entire test, except a composition part which the Service does not score, is of the multiple-choice, objective type.

Section A, Reading Comprehension. This section has thirty-five vocabulary items that give a word and require the student to identify its opposite from a number of choices. For example,

HOT 1-fine 2-big 3-cold 4-blue

It also has ten items in which the student chooses the word that best completes a given sentence. For example,

The color of milk is \_\_\_\_\_.  
1-green 2-red 3-white 4-blue

Twenty items force the student to choose between a correct and an incorrect grammatical form. For example,

He  $\left. \begin{array}{l} 1-is \\ 2-does \end{array} \right\}$  not coming.

## SURVEY OF TESTS IN ENGLISH AS A FOREIGN LANGUAGE

And twenty five items check comprehension of seven short reading passages. For example,

Yesterday we took an automobile ride and saw some very beautiful scenery. At this time of the year, autumn, the leaves begin to turn yellow and red, and the new freshness in the air, after the heat of summer, makes a person really feel that he is alive.

A. The ride which the speaker tells about

- 1-is one that he hopes to take
- 2-is one that he took the day before
- 3-is one that he can scarcely remember
- 4-is a beautiful dream

B. The scenery which is described is

- 1-the kind which makes one feel sad
- 2-the most beautiful in the world
- 3-not worth seeing
- 4-delightful

Section B, Aural Comprehension. The aim of this section is to test understanding of spoken English. The student listens to sixteen short passages recorded on phonograph records and has to answer 50 multiple-choice items on them. The passages and the items are of the same type as those used in reading comprehension. The passages increase in length and difficulty from approximately 33 words for one of the shortest to some 241 words for the longest and last.

This is perhaps the weakest section of the Examination for a number of reasons. First, it is a wholesale transfer of a reading technique into aural comprehension disregarding the differences involved. Sentence types, sentence length and sentence sequences as well as vocabulary are not equal in reading and in spoken material. The passages do not represent lectures as usually given by college professors much less conversational material. Also, the task of having to listen to 200 or more words, and then work through sixteen or more choices may be very valuable in reading comprehension where

## ROBERT LADO

one can recheck what was read, but in aural comprehension it lays too much emphasis on memory.

Second, the subject of various passages and the continuity of the several items for each passage in some cases provide clues to the answers and decrease the validity of the scores as a measure of aural comprehension. Familiarity with subjects like Henry David Thoreau, the historical incident of the United States with the Tripoli pirates, the movement of people from the farms to the cities, the present rivalry between the airlines and the railroads, etc., is bound to reflect itself in the scores even though it may not be at all related to aural comprehension.

Finally, the need to use records and a phonograph complicates the problem of giving the test abroad where it is most needed. The Educational Testing Service cannot be blamed for this difficulty, since it was the opinion of the advisory committee on the planning of the test that a section on aural comprehension be included because of the great importance of this skill.

Section C, Pronunciation. Thirty of the pronunciation items are pairs of words. The student must decide whether or not they rhyme. For example,

1. send - lend
2. friend - fiend

In the remaining thirty items the student checks the number of the syllable that receives loudest stress in a word. For example,

	1	2
accept	AC	CEPT

Sections D and E, Scientific Vocabulary and Nonverbal Reasoning are rather well done along standard lines. The Scientific Vocabulary section is to be disregarded for non-science majors and it is not really essential for science majors because the student's previous grades in science are also good indicators of future success. Admitting colleges have access to the student's records. Furthermore, scientific vocabulary

## SURVEY OF TESTS IN ENGLISH AS A FOREIGN LANGUAGE

tends to be largely cognate and as a result knowledge of it in the native language transfers easily to the foreign language.

In general the English Examination for Foreign Students of the Educational Testing Service represents a considerable amount of work and expense. Norms based on several hundred scores are available. Its two practical drawbacks are the length and difficulty of administration and its cost per student. From a technical standpoint one is forced to say that it has not made use of recent advances in linguistic science or allowed for the implications of its findings.

### (2) ENGLISH LANGUAGE INSTITUTE, UNIVERSITY OF MICHIGAN

These tests measure achievement in connection with the teaching of English as a foreign language. Two of them, however, can be used to determine a student's mastery of English before he leaves his native land. They are the Examination in Structure and the Test of Aural Comprehension. Passing mention will be made of other tests also.

Examination in Structure. This test has three equivalent forms of 150 items each. The items are of various types: multiple-choice, completion with the aid of a list, completion, re-arrangement of elements, conversion of one pattern to another, and adjustment of a given form to a given pattern. The student is allowed 60 minutes to finish the test. It can be scored objectively by an untrained person with the use of a scoring key.

The reliability of this test is .97.<sup>1</sup> Norms based on scores of 357 students toward the end of their intensive training in English are available.

---

<sup>1</sup>Reliability refers to how dependable the scores are, i.e., how nearly the same scores the students would get if they took the test again. Complete or perfect reliability would be 1.00. The procedure used is one called the chance-half method; the test is split into two halves by obtaining separate scores for the odd-numbered and the even-numbered items.

This method gives the reliability of half the test, and since the full test is more reliable than half of it, an empirical formula known as the Spearman-Brown formula is used to estimate the reliability of the full test from that of half of it. .97 is the reliability of the complete test.

## ROBERT LADO

Test of Aural Comprehension. This test also has three equivalent forms labeled A, B, and C. The student indicates his understanding of short passages of spoken English by checking multiple-choice items. The passages range in length from three to forty-seven words, and they sample specific matters of the sound system, the structure system and the words of English in actual operation. The choices on the first part of the test are sets of three pictures one of which matches what the examiner reads. In the second part the choices are of the standard type in the form of phrases in English.

The validity of the test rests in part on the fact that the items check specific matters of English vocabulary, pronunciation and structure in actual utterances. Further evidence of its validity was obtained by correlating scores on the test with the composite opinion of eight teachers of the English Language Institute who had been in daily contact with the students in class and at meals for a period of six weeks. The correlation was .80. The reliability of this sixty item test is .88.<sup>2</sup>

The test has sufficient range to test both advanced and elementary students. Of 300 students whose scores were tabulated no one made a perfect score, and the distribution of scores approaches a normal curve.

Because it is an aural test it is subject to the difficulties noted above for another aural test. We must have a mechanical record and play-back equipment at the examination center abroad. This difficulty is considerably reduced by the fact that the present test can be read by any educated native speaker of English without serious loss of reliability. Each item depends on a specific matter of vocabulary, structure or pronunciation which will not be seriously affected by personal differences in the dialect of the examiner. For example, the problem involved in understanding the difference between "milk chocolate" and "chocolate milk" will be substantially the same regardless of personal dialect. The test presents this problem as follows:

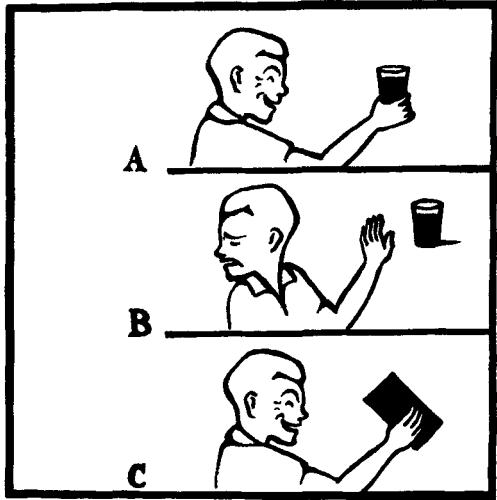
The examiner reads: "The boy likes milk chocolate."

<sup>2</sup>Chance-half method corrected by Spearman-Brown formula.

## SURVEY OF TESTS IN ENGLISH AS A FOREIGN LANGUAGE

The student selects one of the following choices:

In spite of the value of its features for placement, progress data, and diagnosis, the Test of Aural Comprehension does not fill the need for a more practical test to be used abroad before the student leaves for the United States. Even though no records or playback equipment are essentially needed, the examiner is forced to give full attention to the administration of the



test during the entire process, and it is doubtful that consular offices are willing or able to detach their personnel from regular duties for such periods of time. The scoring is, of course, no problem since the test is completely objective and the examiner simply uses a punched stencil to score it.

Norms based on scores of 300 students toward the end of their intensive work are available. Progress norms based on an eight-week intensive course are also available on the Test of Aural Comprehension and on the Examination in Structure.

Test of Pronunciation for Japanese Students. This is an objective, paper-and-pencil test of pronunciation based on a scientific linguistic description of the sound system of English compared with the sound system of Japanese. Even though the technique does not test what the student actually says but rather what he thinks he says, it shows promise of being sufficiently effective to warrant its use. It has a great practical advantage: any untrained person can give it and score it, and

## ROBERT LADO

it does not require the attention of the examiner except to hand the booklet to the student and to pick it up when the student has finished.

Test of Aural Perception for Latin American Students, and Test of Aural Perception for Japanese Students. These tests measure the student's ability to hear accurately the significant sounds and sound contrasts of English. They are based on linguistic comparisons of the native language of the students with the sound system of English. They do not depend on the association of a symbol with a particular sound as is usually the case in such tests. These tests are useful primarily to pronunciation teachers and to those who include pronunciation in their regular classes. They are not intended for use in determining a student's control of English for acceptance purposes.

### (3) ENGLISH LANGUAGE INSTITUTE, QUEENS COLLEGE

Vocabulary Test. I have examined two experimental forms of a vocabulary test being developed at Queens College. They are labeled Form XA2 and Form XB2. Each form consists of seventy-five items giving a word in a short sentence and five other words from which the student must choose the one that means the same. For example,

The man is very tall.

- A. short
- B. broad
- C. high
- D. tiny
- E. fat

This test is interesting because it gives the word being tested in a context. No norms are yet available.

### (4) INTER-AMERICAN TESTS

The Inter-American Tests were developed under the direction of Professor H. T. Manuel to provide comparable measures in English and Spanish. A general ability test is also part of the series. The language tests have A and B forms parallel in content and similar in difficulty.



## **SURVEY OF TESTS IN ENGLISH AS A FOREIGN LANGUAGE**

Inter-American Tests of Language. There are four tests in this series: (1) Reading, (2) Language Usage, (3) Vocabulary and Interpretation in the Natural Sciences, and (4) Vocabulary and Interpretation in the Social Studies. The Reading Test is available for three different levels: a primary test, an intermediate one, and an advanced test. All three have sections on vocabulary and paragraph reading comprehension. The vocabulary section is of the standard type which requires the student to identify the synonym or antonym of a given word. The reading comprehension section is of the paragraph type with multiple-choice items to check understanding of the paragraph.

The Language Usage Test, intended for the higher grades, has a section on active vocabulary and one on expression. The active vocabulary section uses an ingenious device to test recall of words-- hence the name "active vocabulary." The student is not given complete words in the choices but only the initial letter and the number of letters in the word. In indicating his choice it is assumed that the student recalls the word with the clues given. The section on expression is also of the multiple-choice type. The student selects the expression that he considers best among those given.

The tests on Vocabulary and Interpretation in the Natural Sciences and in the Social Studies have sections on vocabulary and paragraph comprehension in those general fields of study.

These tests have been published by the Educations Testing Service after their use in connection with a report on education in Puerto Rico. They were carefully prepared and widely administered there. The reading test is as strong as any now available. There was no attempt to measure aural-oral skills in these tests.

### **(5) THE INVESTIGATION OF THE TEACHING OF A SECOND LANGUAGE**

The Investigation of the Teaching of a Second Language carried on under a grant from the Rockefeller Foundation to the University of Chicago made available a number of tests in English as a foreign language. We will mention here the tests

ROBERT LADO

on Elementary Structure, Reading Comprehension, Aural Comprehension, and Writing and Structure.<sup>3</sup>

Elementary Structure Test. This test differs from the others in that it is pitched at a very elementary level. It deals with the elements of English as usually found in a first year course abroad. A score of 90% may be interpreted as indicating a good grounding for further English study in the United States.

Reading Comprehension Test. This test is made up of passages of approximately 250 words for reading comprehension. The materials were slightly simplified below the level at which college students are asked to read their work. The topics of the passages are those likely to be found in non-technical articles on current affairs and American life, and were drawn from two books and two magazine articles. Norms based on scores of 250 students are available.<sup>4</sup>

Aural Comprehension Test. The test being reviewed was copyrighted in 1945. The examiner's part is recorded on high-fidelity records. The student listens to the record and answers 44 multiple-choice questions having three choices each. The student simply circles the letter of the choice which is correct according to the passage. There are four long passages in the test and they last from four to five minutes each, or two and a half to three pages each of double spaced typing. The topics of the passages are "Ways of Judging a Social System," "How Children Are Reared in America," "Gossip in a Small Town," and "The Baxter Family (H. H. Davies)." The last passage is a dialogue recorded by two voices.

The instructions are "Listen carefully to each passage, and take notes if you wish on the blank paper provided; but do not attempt to look at the exercises in the test booklet during the reading of any passage." When each passage has been presented, the student answers from nine to thirteen multiple-choice questions from memory or from his notes.

<sup>3</sup>General reference is made to Chapter VII, pp. 247-78 of An Investigation of Second-Language Teaching by F. B. Agard and H. B. Dunkel, Ginn and Company, 1948.

<sup>4</sup>*Ibid.*, pp. 256-8.

## SURVEY OF TESTS IN ENGLISH AS A FOREIGN LANGUAGE

A serious weakness has been discovered in this test. I took the answer sheet without having seen or heard the passages and proceeded to answer the questions which presumably were testing their comprehension. In multiple-choice items containing three choices each one would expect to guess about a third of them correctly. The scores on the four passages, however, were actually 80%, 54%, 78%, and 58%. These high scores confirmed my strong suspicion that familiarity with the areas of the passages would weigh heavily on the answers, thus invalidating the test as a measure of aural comprehension. In the last passage, the dialogue, in which there could be little transfer from previous knowledge except perhaps for insight into the way people behave, the continuity of the story made obvious some of the answers.

The reliability of the Aural Comprehension Test is given as .81, and norms based on scores of 112 students are available.<sup>5</sup>

Writing and Structure Test. This test is divided into two parts, the first part is a writing test and the second part a structure test. In the writing test the student reads a brief anecdote and then writes, without access to the text, the answers to ten simple questions on the story. He also has to write five sentences describing his family, and five sentences telling how he spent the preceding Sunday.

The great problem in using this kind of test is that of scoring it. No objective measure of its reliability is given but even if personnel can be trained to score it with any degree of uniformity, we cannot hope to train personnel in the field to score it. Norms based on scores of 197 students are available.<sup>6</sup>

The structure part of this test is in multiple-choice form. It has ninety-five items. The student is required to select the correct word or phrase out of the three offered. These items form part of a continuous narrative, the description by a foreign student of his arrival in the United States. The structure part is reported to be highly reliable and a better predictor of

---

<sup>5</sup>Ibid., pp. 260-278.

<sup>6</sup>Ibid., p. 265.

## ROBERT LADO

writing ability than the writing section itself.<sup>7</sup> The reliability of the structure part is given as .96.<sup>8</sup> Norms based on the scores of 197 foreign students nearing completion of their special English work are available.<sup>9</sup>

### (6) U. S. DEPARTMENT OF STATE, ENGLISH PROFICIENCY EXAMINATION<sup>10</sup>

The purpose of the English Proficiency Examination is to test the control of English by the foreign student before he leaves his country. A team of two persons is needed to give the test, one a native speaker of English, the other a native speaker of the language of the student. It takes about one hour to give the examination, exclusive of the personal interview part.

The test is divided into four parts. The first consists of eighty true-false statements which are read aloud to the student. Each statement is read twice, and the student makes his response by checking either "True" or "False" on his answer sheet. Each statement deals with facts that are known to the student so that his problem is only to understand the language itself.

The second part of the examination also tests aural comprehension, this time of two anecdotes read aloud to the student. Each anecdote is read twice in a natural style as with the true-false statements, and the student writes short answers to ten questions asked orally after each anecdote. The

---

<sup>7</sup>Ibid., p. 264.

<sup>8</sup>Ibid., p. 278.

<sup>9</sup>Ibid., p. 265.

<sup>10</sup>This description is based on a copy of the examination and of the instructions for its administration supplied by the U. S. Department of State. The request for a copy was addressed to Dr. F. J. Colligan, Chief, Division of Exchange of Persons, and was answered by Miss Marita Houlihan, Training Programs Branch of that Division.

I am also indebted in this section to Mrs. Margaret Emmons, my associate in the Sub-Committee on Evaluation of the English Section of NAFSA. Her description of the test is available in her booklet entitled Orientation and English Instruction for Students from Other Lands, Bulletin 1950, No. 8, Federal Security Agency, pp. 2-3.

## SURVEY OF TESTS IN ENGLISH AS A FOREIGN LANGUAGE

answers are to be graded according to whether or not they show that the anecdotes and the questions were understood.

In the third part the student writes a short composition on a subject familiar to him, and in the last part he is interviewed to determine his ability to speak English.

The grading of the test is admittedly the most difficult problem in its use. There is no problem, however, in grading the true-false section. The difficulties begin with the short written answers to the questions about the anecdotes, they increase with the written composition, and they are probably greatest in grading the interview to determine the student's ability to speak English.

Even though the test does not require a mechanical recording, the problem of individual variation among examiners is still unsolved. In the true-false statements this variation is probably not significant, but in the anecdotes it might well be. Memory also might be too heavy a factor in the anecdotes, since answering ten oral questions which constitute problems in themselves requires a somewhat severe memory strain.

The fact that the test requires the full attention of two persons increases the difficulty of its administration abroad. And finally, the subjective nature of much of the scoring requires trained personnel at the testing centers abroad.

No norms are yet available on this examination. The Department of State considers it still in the experimental stage.

### (7) VILLAREAL TEST OF AURAL COMPREHENSION FOR NATIVE SPEAKERS OF SPANISH

The first full volume devoted entirely to the measuring of aural comprehension by Spanish-speaking students is a dissertation by Jesse J. Villareal. He developed two equivalent forms of a test of aural comprehension consisting each of ten selections and five dialogues, a total of fifteen passages. These passages are recorded on phonograph records. The students listen to the passages and indicate their responses by means of seventy-five multiple-choice answers having five choices each. The questions and choices on the answer sheets are in Spanish for the Spanish-speaking students and in English also.

## ROBERT LADO

The reliability as reported by its author is .80. The validity<sup>11</sup> is based on the criteria of preparation of the test items and on correlations of test scores with ratings by "qualified observers" and by the students themselves. The correlation of test scores with ratings by qualified observers was .78 and with ratings by the students themselves, .31.

The usual objection to the use of large passages for aural comprehension is in order with this test. The average length of the passages is 249 words; the first passage has 488 words; memory more than aural comprehension must often be the deciding factor in giving the right answer. This test was not easy even for 104 native speakers of English who made scores between 164 and 207 out of a possible 212. Apparently none of the native speakers made a perfect score.

### III. Summary Evaluation

None of the above tests fully meets the need for a linguistically valid English examination that may be administered abroad by untrained personnel who are busy with other duties. Those tests that come close to sound construction require the full attention of the examiner for varying lengths of time, or they require the use of records and phonographs that may or may not be available. Since considerable time and money have been spent on the preparation of these tests without attaining the desired goals, it seems appropriate to look at the difficulties in the hope of finding a practical solution. Following are some of the major obstacles and their possible solutions.

1. Perhaps the greatest obstacle to progress in these tests is the lack of a clear understanding of what English language is. As a result the approach to test making, except in the case of vocabulary, is a haphazard, almost mystic attempt to capture in a paragraph or a story what a foreign student

---

<sup>11</sup>"Validity" refers to whether or not a test measures what it claims to measure--aural comprehension in this case. A test of intelligence, for example, would not be a "valid" test in aural comprehension. A test is said to be valid according to the criteria of preparation when the items were constructed for the purpose of measuring in this case aural comprehension. This claim must be self-evident or it must be supported by adequate proof.

## SURVEY OF TESTS IN ENGLISH AS A FOREIGN LANGUAGE

has to know to attend our colleges. To overcome this difficulty we must first realize that what we need to test is how well the student uses and reacts to the meaningful signals that constitute the English language. The matter of how good a student he is, or how well he is going to do in engineering mechanics, for example, can be determined by his previous academic work in his own language; for that purpose he submits his academic records to the admitting institution. Once we concentrate on testing his control of the signals of the language itself we will make progress provided also that we base our tests on a scientific linguistic analysis of English.

2. Failure to make distinctions in the linguistic problems found by the various groups of foreign students is a weakness of many of the above tests, and yet the task of preparing separate tests for all language backgrounds is so enormous that we may never hope to have such tests except for a limited few languages. A practical solution to this problem may be that of keeping separate norms for the various national groups of students that take the tests.

3. A final roadblock to progress on a practical test for foreign students has been and is the need to test at least one aural skill, aural comprehension, which requires the use of records and phonographs. We do not have enough information on the correlation of aural comprehension with other language skills that can be measured more easily than aural comprehension. There is some hope that substitutions can be made without serious loss of validity.

### IV. BIBLIOGRAPHY OF TESTS DESCRIBED

1. Educational Testing Service, P. O. Box 592, Princeton, New Jersey.  
English Examination for Foreign Students. (Norms)
2. English Language Institute, University of Michigan, Ann Arbor, Michigan.  
Examination in Structure. (Norms) Forms A, B, and C.  
Test of Aural Comprehension. (Norms) Forms A, B, and C.  
Test of Pronunciation for Japanese Students.

ROBERT LADO

- Test of Aural Perception for Japanese Students.
3. English Language Institute, Queens College, New York.  
Vocabulary Test.
  4. Inter-American Tests, Educational Testing Service or Professor H. T. Manuel, University of Texas, Austin 12, Texas.  
General Ability Test. (Norms) Primary, Intermediate, Advanced.  
Reading Tests. (Norms) Forms A and B. Primary, Intermediate, Advanced.  
Language Usage. (Norms) Forms A and B.  
Vocabulary and Interpretation in the Natural Sciences. (Norms) Forms A and B.  
Vocabulary and Interpretation in the Social Studies. (Norms) Forms A and B.
  5. The Investigation of the Teaching of a Second Language, The University of Chicago, Chicago, Illinois.  
Elementary Structure Test. (Norms)  
Writing and Structure Test.  
Structure Test. (Norms)  
Writing Test. (Norms)  
Reading Comprehension Test. (Norms)  
Aural Comprehension. (Norms)
  6. U. S. Department of State. Write to Dr. Francis J. Colli-gan, Chief, Division of Exchange of Persons, Dept. of State, Washington 25, D. C.  
The English Proficiency Examination.
  7. Villareal Test. Professor Jesse J. Villareal, University of Texas, Austin 12, Texas.  
Test of Aural Comprehension for Native Speakers of Spanish.