Testing Oral Language Skills via the Computer

Jerry W. Larson
Brigham Young University

ABSTRACT

Although most language teachers today stress the development of oral skills in their teaching, it is very difficult for them to find time to assess these skills. This article discusses the importance of testing language students' oral skills and describes computer software that has been developed to assist in this important task. Information about various techniques that can be used to score oral achievement test performance is also presented.

KEYWORDS


INTRODUCTION

To foreign language educators in the United States, the decade of the 1980s was known for its emphasis on teaching for proficiency, particularly oral proficiency (Harlow & Caminero, 1990). As Harlow and Caminero point out, several factors contributed to this emphasis on oral skills, not the least of which was the report of the President's Commission on Foreign Languages and International Studies entitled Strength through wisdom: A critique of U.S. capability, which appeared in 1979. In addition to national interests, foreign language students themselves also expressed a strong preference for being able to communicate orally (Omaggio, 1986).

This interest in developing oral communicative competence has not dissipated in the 1990s. In fact, the emphasis seems to be increasing in some institutions. However, even though many foreign- and second-language teachers stress oral communication in their teaching, they are not being
Testing Oral Language Skills via the Computer

“pedagogically fair” because they do not test their students’ speaking skills and are actually “sending the wrong message to their students” (Gonzalez Pino, 1989). Good pedagogy dictates that if we stress oral skills in teaching, we must also test them: “If we pay lip service to the importance of oral performance, then we must evaluate that oral proficiency in some visible way” (Harlow & Caminero, 1990).

Initially, academia expressed interest in using the government’s Foreign Service Institute (FSI) oral proficiency interview procedure for assessing speaking skills. But, unfortunately, the FSI oral proficiency interview procedure presented serious problems for university and secondary school language programs. Teachers found that this type of assessment was expensive and too time-consuming to administer to large numbers of students. They also determined that the rating scale was not precise enough to discriminate among the various levels of student ability in our schools (Larson, 1984).

In an effort to avert some of the difficulties and inherent problems of the FSI oral assessment procedure, the American Council on the Teaching of Foreign Languages (ACTFL), in conjunction with the Educational Testing Service (ETS), introduced its own oral proficiency guidelines and rating procedures based on the government’s oral proficiency interview model. These new rating guidelines were adapted to the demonstrable proficiency levels of secondary- and college-level students. Beginning in 1982, numerous Oral Proficiency Assessment workshops were funded by the U.S. Office of Education to familiarize teachers with the new ACTFL/ETS guidelines and to help them implement the adapted oral assessment procedures. Unfortunately, adopting the ACTFL/ETS oral assessment procedures in many language programs is simply not possible because of the time, effort, and expense incurred in executing such a program.

As an alternative to a formal oral proficiency interview, innovative foreign language teachers have tried either to adapt the ACTFL/ETS rating scale to meet their needs (Meredith, 1990) or to devise alternative ways (or “methods”) of testing oral skills. Examples include Gonzalez Pino’s “Prochievement Tests” (Gonzalez Pino, 1989), Hendrickson’s “Speaking Prochief Achievement Tests” (Hendrickson, 1992), Brown’s “RSVP” classroom oral interview procedure (Brown, 1985), and Terry’s “Hybrid Achievement Tests” (Terry, 1986).

A CASE FOR COMPUTERIZED ORAL TESTING

Another alternative for testing oral skills has been developed at Brigham Young University (BY U). BYU has approximately 5,000 students in beginning language classes in which one of the primary goals of instruction is to facilitate the assessment of oral competence in the language being
studied. Given the pedagogical need to “test what is taught,” instructors are faced with the practical dilemma of how to test the oral skills of such a large number of students on a fairly frequent basis.

Many of the language programs at BYU have adopted a quasi oral proficiency interview format for evaluating students’ speaking abilities, which still requires an enormous commitment in terms of teacher time and student time. Since instructors have only a limited number of classroom contact hours, it is impractical—and often counterproductive—to periodically use classroom time to administer an oral interview to each student individually. Therefore, some of the teachers have made arrangements to have their students come to their office outside of class time to take oral tests. This approach, of course, requires a considerable amount of extra time on the part of the teachers.

To alleviate the substantial extra demand on language teachers’ time, we felt it would be possible to devise a way for language students to come to the Language Resource Center to test their oral skills, without requiring the teachers’ presence. In order to implement this form of testing, we first developed a series of semi-direct oral exams administered via audio cassette tapes. (These tests were somewhat akin to the Simulated Oral Proficiency Interview exams (SOPI) developed by the Center for Applied Linguistics in Washington, D.C. [see Stansfield, 1989]. Our exams were not standardized nor nationally normed; they were oral achievement tests based on our own curricula.) Having students take oral tests in the lab on tape allowed teachers to spend more classroom time interacting with students. However, due to the linear nature of cassette recordings, scoring the test tapes still required a substantial amount of teacher time.

As is customary with oral tests, the initial questions on our taped tests were designed to “put the students at ease” during the test rather than to elicit highly discriminating information. Teachers necessarily lost large amounts of time listening to a series of questions at the beginning of the tape that did not significantly help determine the students’ level of speaking ability. In fact, searching out specific responses on an individual student’s audio tape sometimes required nearly as much time as it took to listen to the entire recording. We therefore wanted to devise a way to facilitate and expedite test evaluation as well as test administration. The solution to this problem has come in the form of computerized oral testing.

**BENEFITS ASSOCIATED WITH COMPUTERIZED ORAL TESTING**

There are several notable benefits associated with computerized oral testing. First, the quality of voice recordings using digital sound through the computer is superior to that of analog tape recordings. This superior...
quality makes it easier for the teacher to discriminate between phonetically similar sounds that could ultimately cause confusion in communication. Since the procedure for recording responses into the computer is no more challenging than recording onto a cassette tape, students who are a little timid around technology should not feel any more anxiety than when recording responses into a regular tape recorder.

A second advantage of computerized oral testing over face-to-face interview assessment is that all students examined receive an identical test: all students receive the same questions in exactly the same way; all students have the same time to respond to individual questions; and students are not able to "manipulate" the tester to their advantage. In addition to enjoying the same administrative advantages as the SOPI, computerized oral tests can include various kinds of response prompts (e.g., text, audio, graphics, motion video, or a combination of these inputs).

Finally, besides the benefit of test uniformity mentioned above, another significant advantage of computerized oral testing is the access to student responses for evaluation purposes. The students' responses can be stored on the computer's hard disk—or on a network server—and accessed almost instantaneously for evaluation at a later time by the teacher.

FEASIBILITY OF COMPUTERIZED ORAL TESTING

Many language learning centers are currently equipped with multimedia computers, including sound cards, for student use. These computers are capable of recording students' responses to questions or situations that can be presented in a variety of formats (e.g., digitized audio, video, animations, diagrams, photographs, or simple text) via the computer. Basically, all that has been lacking is a computer program that will facilitate computerized oral language testing. During the past couple of years, staff members of the Humanities Research Center at BYU have been developing software, called Oral Testing Software (OTS), that facilitates computerized oral testing. The OTS package has now been piloted by several hundred language students, and evaluative comments from students and teachers have been very favorable.

BYU'S ORAL TESTING SOFTWARE (OTS)

The OTS program is operational on both Windows and Macintosh platforms. The software is designed to accomplish three purposes: (a) to facilitate the preparation of computerized oral language tests, (b) to administer these speaking tests, and (c) to expedite the assessment of examinees' responses to the questions on these oral tests.
Test Preparation Module

The OTS program is designed to assist teachers without a lot of computer experience in creating computerized oral tests. Teachers simply follow a step-by-step procedure outlined in the Test Preparation Module. This module utilizes a template that guides test creators (i.e., teachers) through the various steps in constructing a test that can be administered via the computer. Teachers may return to any of these steps to edit any portion of the test as needed.

The first step in the test creation process is to specify the number of the test item, that is, in what point in the test order this item is to appear. Individual items may appear in any order desired. (During this stage, it is also possible to delete previously created items that are no longer wanted.) After indicating the number of the item in the test, teachers may, if they wish, include introductory information for the item, which can be presented via a variety of digital sources.

Next, teachers choose what type of prompt or elicitation technique they wish to employ for the item. Again, any source of input mentioned above can be used. Teachers also indicate how much time, if any, will be allowed for students to prepare to answer the question. Some oral response items may require several seconds for students to organize their thoughts before recording their answer. If response preparation time is indicated, the OTS will cause the computer to pause for the designated time and then initiate the recording mode for the students to record their response.

The Test Preparation template also allows test creators to determine the length of time to be allowed for students to respond to a given question. During test administration, once the designated time has expired, the computer notifies examinees that time for responding is over and then presents the next item.

Once a test item has been created, the Test Preparation Module allows teachers to preview that item before inserting it into the test proper. The final step is to compile the entire set of items into a working test. To accomplish this task, teachers simply click on the designated button, and OTS combines the items into a test file for subsequent administration.

Test Administration Module

After a test has been created and loaded onto the student computers (or network server), the OTS program administers the test to students individually. If teachers designate during the test preparation process that students must be in the examinee registry to take the test, the program verifies the eligibility of individual students. If students are not listed in the
Testing Oral Language Skills via the Computer

registry, they cannot take the test. As the students sit down to take the test, the Test Administration Module first checks to make sure the volume in their headset is adequate for them to hear satisfactorily whatever instructions or questions are to be presented. The program also verifies that their voice is recording properly in the designated storage area (e.g., on the computer's hard disk, on the network server, or on some other designated storage medium such as a Zip or Jaz cartridge).

Once the volume and recording checks have been completed, test items are administered to the examinees in the order and manner specified by the teacher when the test was created. As students take the test, they will see a graphic representation of how much time remains for them to prepare their answer (if such time was allowed during the test creation phase) and how much time they have to complete their response. When examinees have finished responding to a given item, they may move on to the next item by clicking on the STOP button which stops the recording process and takes them to the next question. (This feature provides for a less frustrating testing experience since the examinees are not forced to wait for time to elapse to move on.)

During the administration of the test, examinees generally will not have to perform any tasks other than speaking into their microphone or clicking on buttons to advance through the test. The software automatically presents all audio, video, graphics, and other question prompts. The students are not required to type or execute complicated computer commands, thus minimizing test anxiety.

Test Results Assessment Module

Assessing student responses in OTS is accomplished through the aid of the “Test Assessment Module.” This module facilitates access to students' responses, which are listed in a Results File compiled separately for each student. Using the assessment program, it is possible for teachers to listen to specific responses of individual students independent of any other responses. From the Results File, teachers simply select the specific student and response(s) they wish to listen to, and the computer plays the response(s) instantly. The teachers do not have to spend time fast forwarding or rewinding a tape to find the desired response. Teachers can listen to as little or as much of each response as necessary to assess the student's performance properly. This also makes it possible to listen only to those responses that give useful information about the student's speaking ability. It is no longer necessary to spend time listening to warm-up items which often do not contribute meaningful or discriminating assessment information. Thus, the overall time needed to assess a student's performance is greatly reduced. Additionally, if the students' responses have
been recorded onto a network server accessible to teachers in their office, response evaluation can be completed without having to go to the Language Resource Center. If networking is not available, it is possible to save students’ responses onto Zip or Jaz cartridges for later assessment. The Assessment Module also includes a notepad on which teachers can type any assessment notes they would like to pass on to the students or to keep for later reference when discussing the students’ performances or assigning grades. On the notepad, the date of the test and the times it began and ended are also listed. All of this information can be saved on the computer or printed for hard copy storage.

PROFICIENCY ASSESSMENT

In spite of the fact that accessing students’ oral responses can now be done relatively easily via the computer, it is still necessary to employ some scoring procedure in order to give meaningful input to students regarding their performance. Over the years, various techniques have been used to assess responses of oral tests. Assessing global oral proficiency was a vital concern for U.S. government agencies and led to the formulation of the FSI oral proficiency scales which specified 11 major ranges of proficiency. Using these scales, agencies were able to describe in a meaningful way the speaking ability of their employees or candidates for employment. However, these proficiency scales were a little too broad to show relatively small increments of improvement in oral ability, and were, therefore, not as useful as needed in academic settings. This need gave rise to an adaptation of the FSI scale known as the Interagency Language Roundtable (ILR) scale which later was further refined by ACTFL and ETS. These new scale guidelines were an expansion of the lower levels of the FSI scale, providing nine proficiency descriptions for the FSI levels 0-2. Although more discriminating than their parent FSI rating scales, these guidelines are still considered by some teachers to be too imprecise for assessing oral performance progress in classroom situations.

Achievement and Progress Assessment

A number of procedures have been employed to score oral achievement and progress tests. The technique one chooses to use would depend upon the objectives of one’s language program. If, for example, teachers stress pronunciation development, they would want to use assessment criteria that include this aspect of language development. If they believe competence in certain components of the language are more important than oth-
ers, they would use a scoring procedure that allows them to weight those components.

The oral assessment techniques described below have been discussed in various publications in the field of foreign language education. The information presented here is not intended to be an exhaustive, all-inclusive listing but represents samples of the kinds of procedures that educators have tried with success. The first techniques described below are “unweighted” procedures in which each component of the language is given equal weight in the assessment. Weighted procedures are then discussed.

**Unweighted Assessment Procedures**

**The San Antonio Independent School District (SAISD) Oral Proficiency Assessment Guidelines**

The SAISD test is a tape-mediated oral exam. The scoring procedure, outlined and described by Manley (1995), was designed to test only three content areas (self and family, school, and leisure activities) but could be used to score responses in other topic areas as well. Using a “Grading Grid,” teachers check a value of 0-5 for responses to each item. The values correspond to the following criteria: 0 = no response, 1 = response unclear or not comprehensible, 2 = response is a word, list, phrase, or simple sentence with a mistake, 3 = response is a simple, short sentence without a mistake, 4 = response is a complex sentence with a mistake, and 5 = response includes complex sentences without mistakes. A bonus point is awarded for exceptional quality of response.

**Winthrop College Model**

Joseph Zdenek of Winthrop College recommends scoring high school students' responses on a scale of 0-3 or 0-5 (see Zdenek, 1989). The 0-3 scale focuses on pronunciation, fluency, and structure, assigning three points maximum for each area. The 0-5 scale evaluates comprehension, grammar, vocabulary, fluency, and accent.

Zdenek allows for inadvertent poor performance on a given question by granting an additional question worth up to 10 points. The extra points are added to the original score points, and the total is then divided by 250 to get the absolute score.

**Student Oral Assessment Redefined (SOAR)**

Linda Paulus of Mundelein High School proposes evaluating students'
oral performance in three domains: strategic, sociolinguistic, and discourse competencies (see Paulus, 1998). In this scoring procedure, language errors do not factor into students' scores, unless they interfere with communication of the message. The scoring criteria for this approach are (a) quality and extent of communication (Discourse Competency), (b) verbal and nonverbal communication strategies (Strategic Competency), and (c) culturally appropriate language and behavior (Sociolinguistic Competency).

**Classroom Oral Interview Procedure (RSCVP)**

This procedure, described by James Brown of Ball State University, is a modified OPI (see Brown, 1985). The exam takes place in stages. Stage One is similar to the warm-up stage of the OPI, consisting of routine inquiries about the examinee, the weather, etc. Stage Two presents questions and information based on what students have been studying in their course. Stage Three includes “divergent exchanges” designed to go beyond the students’ “prepared storehouse of responses.”

The focus of the evaluation is on Response, Structure, Vocabulary, and Pronunciation (RSVP). (Response is basically equivalent to fluency, and Structure refers to grammatical precision.) Each category is scored from 0-3, where 3 = “better than most” responses, 2 = a “force mode” or assumed competence for that level, 1 = “less well than most,” and 0 = “unredeemed silence or inaudible mumble.”

**Keene State College Model**

Providing another variation on the FSI model, Helen Frink of Keene State College has reduced the number of competency categories to four: poor, fair, good, excellent (see Frink, 1982). Each of the categories has an assigned point value (poor = 10, fair = 14, good = 17, excellent = 20). Frink discourages assigning values between the assigned values, claiming that little, in the long run, is gained by assigning, for example, a 16 instead of 17. She claims that such a practice will negate the ease and speed with which the scoring form can be completed.

**Weighted Assessment Procedures**

**University of Virginia Simplified Scoring Procedure for High School Students**

Juan Gutiérrez of the University of Virginia suggests using a simplified scoring approach for novice- and intermediate-level high school students which places more emphasis on vocabulary and grammar by assigning greater value to those two components (see Gutiérrez, 1987). His rating scale for each response is as follows. (Specific criteria are not explained...
Testing Oral Language Skills via the Computer

for each of the numerical ratings.)

Pronunciation 1 2 3 4 5 x 4 =
Vocabulary 1 2 3 4 5 x 7 =
Grammar 1 2 3 4 5 x 6 =
Fluency 1 2 3 4 5 x 3 =

BOYLAN’S MONOLOGUES AND CONVERSATIONAL EXCHANGES MODEL

For oral testing at the beginning and intermediate levels, Patricia Boylan uses a two-part oral test featuring monologues and conversational exchanges (see Boylan cited in Omaggio, 1986). The monologues are based on situations listed on topic cards prepared by the teacher from recent course material. The students speak briefly about the theme and then answer questions from the teacher about the topic. The next part of the oral test consists of the students’ asking questions of the teacher based on topics from the cards or on a role-play situation. The responses are graded according to the weighted scale shown below. The various components’ weightings are: vocabulary, 19%; fluency, 17%; structure, 16%; comprehensibility, 34%; and listening comprehension, 14%.

Part I: Monologue (34%)

<table>
<thead>
<tr>
<th>Fluency</th>
<th>1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td>1 2 3</td>
</tr>
<tr>
<td>Structure</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14</td>
</tr>
</tbody>
</table>

Section Sub-total /34

Part II: Answering Questions on Monologue (26%)

<table>
<thead>
<tr>
<th>Fluency</th>
<th>1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td>1 2 3</td>
</tr>
<tr>
<td>Structure</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>1 2 3 4 5 6</td>
</tr>
</tbody>
</table>

Section Sub-total /26

Part III: Answering Questions on Monologue (26%)

<table>
<thead>
<tr>
<th>Fluency</th>
<th>1 2 3 4 5 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>Structure</td>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
</tbody>
</table>

Section Sub-total /40

Test Total /100
After eight weeks of instruction, students in first semester French classes at the University of Illinois are given oral achievement tests based on situations listed on conversation cards (see Omaggio, 1986). The cards serve as stimuli for conversation and are not to be used as a translation task. The interview is taped and scored at a later time. Four components of the language are assessed: pronunciation, vocabulary, grammar, and fluency. The score sheet guides the teacher to assign a global grade of A through E for each of the language components. A numerical conversion table indicates the point value each letter grade represents (A = 4.5-5.0, B = 4.0-4.4, C = 3.5-3.9, D = 3.0-3.4, and E = 2.0-2.9). The numerical value obtained for each component grade is then multiplied by its respective weight: pronunciation = 4, vocabulary = 7, grammaticality = 6, and fluency = 3. (The weightings are based on research that indicates that speaking skills at this level reflect primarily knowledge of vocabulary and grammar (see Higgs & Clifford, 1982).

A scoring procedure used in the Foreign Language Achievement Tests Series (FLATS) program at Brigham Young University weighs each question according to its overall difficulty (i.e., the more difficult the question, the more points possible) as well as the language components being examined for that question. The language components examined are (a) fluency—ease of response, overall smoothness, continuity, and naturalness of speech, (b) pronunciation—correct pronunciation of words, phonemes, inflections, etc. in their respective environments, (c) grammar—linguistic correctness of response, for example, correct tense, syntax, agreement, etc., (d) quality—appropriateness and adequacy of response, for example, use of complex structures where appropriate, precise vocabulary, a thorough, complete response, and communication of relevant information. Test raters are given instruction regarding the use of the rating scales before using them.

The questions under consideration are listed on the score sheet next to their respective rating scale. An example of a FLATS score sheet appears below.
## Testing Oral Language Skills via the Computer

### Question #
#### Points Possible

**Tell a little about yourself.**
(lowest difficulty level)

<table>
<thead>
<tr>
<th>Fluency</th>
<th>Pronunciation</th>
<th>Grammar</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Total: __

**Relate several things you did this morning from the time you woke up until now.**
(medium difficulty level)

<table>
<thead>
<tr>
<th>Fluency</th>
<th>Pronunciation</th>
<th>Grammar</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Total: __

**Explain how to ride a bicycle.**
(high difficulty level)

<table>
<thead>
<tr>
<th>Fluency</th>
<th>Pronunciation</th>
<th>Grammar</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Total: __

GRAND TOTAL: _____
CONCLUSION

The general consensus in foreign- and second-language education is that oral skill development is a high priority, indeed in many cases, the top priority. If, in fact, speaking is emphasized, it should also be tested periodically. However, assessing oral skills requires a significant commitment of time and energy on the part of language teachers. In an effort to mitigate this testing burden, testing software has been developed that allows teachers to construct computerized oral tests, to administer them, and to assess students' responses with relative ease. Using this kind of software in conjunction with an appropriate scoring technique, teachers can assess their students' oral performance on a relatively frequent basis with a minimal loss of classroom time.

REFERENCES


Testing Oral Language Skills via the Computer


AUTHOR'S ADDRESS

Jerry W. Larson
Humanities Research Center
3060 JKHB
Brigham Young University
Provo, UT 84602
Phone: 801/378-6529
Fax: 801/378-7313
E-mail: Jerry_Larson@byu.edu

AUTHOR'S BIODATA

Jerry W. Larson obtained his Ph.D. in 1977 in Foreign Language Education at the University of Minnesota and his M.A. in 1974, in Spanish Pedagogy at Brigham Young University. He is currently Director of the Humanities Research Center and Professor of Spanish at Brigham Young University. His research interests include applications of technology in foreign language instruction and testing.