0652-1

Testing Speaking Ability in the Classroom: the Semi-direct Alternative

Jerry W. Larson
Brigham Young University

ABSTRACT Although a significant amount of research has been done to improve testing techniques and scoring procedures for oral proficiency testing, only limited discussion has focused on ways to improve oral achievement tests. In the following discussion, the author describes an alternative to direct, face-to-face oral achievement testing. The technique, semi-direct testing, provides a way for classroom teachers to assess the oral ability of individual students without the commitment of class time involved in direct testing. Additional advantages of semi-direct tests are also mentioned. A study was conducted to compare achievement scores of semi-direct and direct tests of German and Spanish. An affect questionnaire was also given to each of the students to determine student attitudes toward the two testing procedures. Results of the study are presented.

Ranking high among the major issues in foreign language education in recent years is the testing of oral language skills. Subsequently, a great deal of emphasis has been placed on improving oral testing techniques and scoring procedures. Both oral proficiency and oral achievement testing[1] have received and should continue to receive attention.

Jerry W. Larson (Ph.D., University of Minnesota) is Assistant Professor of Spanish at Brigham Young University; Provo, UT.

Oral Proficiency Testing

Significant progress has been made in the area of testing oral language proficiency. For many years the Foreign Service Institute (FSI) Oral Interview was the only direct speaking test in common use in the United States. It was adopted by the Peace Corps, the Central Intelligence Agency, and several other government agencies, but it was used only to a limited extent by universities and other non-government groups.

Because the FSI technique, especially the rating scale, presented some serious problems to university and secondary school language programs (e.g. tests were expensive and time-consuming to administer to large numbers of students, and the rating scale was not precise enough to discriminate among the various levels of student ability), much work has been done either to adapt the FSI testing procedures to university and other academic language programs[2] or to devise other ways of measuring oral proficiency.

In the 1970's oral proficiency testing research and development began to increase rapidly. Journal articles and papers at professional conferences reported on proficiency-related studies dealing with topics such as test reliability and validity,[3] correlations between speaking and other types of language tests,[4] and new elicitation and scoring techniques.[5] Each of these has been useful in refining the oral testing process.

Oral Achievement Testing

Although oral proficiency testing seems to have

received the lion's share of the attention, oral achievement testing should not be left unattended. It is generally the most common type of testing being done in foreign language classrooms; yet, comparatively few articles and research studies have focused on the improvement and further development of oral achievement tests.[6] An increase in presentations and discussions regarding new techniques and ideas for improving oral achievement tests would be useful. Suggestions on how to make oral tests more feasible in classes with limited time for testing would be welcomed by many teachers.

The purpose of the following discussion is to present an alternative approach to direct oral testing, a semi-direct procedure that perhaps will make it possible for foreign language teachers with limited time to test the oral skills of individual students.

Semi-direct versus Direct Testing

Because direct methods of testing language ability demand a substantial commitment of time on the part of the teacher, other approaches to assessing students' speaking skills have been developed. One such procedure is semi-direct testing, which characterizes tests that elicit active speech by the examinee through means such as tape recordings, printed test booklets, or elicitation techniques other than direct, face-to-face interviews. Government language programs, for example, have used the Recorded Oral Proficiency Examination (ROPE)[7] as an alternative oral proficiency test when conducting an oral interview was not possible. The test parallels the FSI interview, eliciting about 15 minutes of speech during a 30-minute testing period. The ROPE, which was designed as a proficiency measure, has the added advantage of being well-suited for achievement testing too.

The benefits of having semi-direct tests for achievement testing as well as proficiency testing are of particular interest to teachers in academic institutions, for most courses are achievement-oriented. Among the advantages of semi-direct tests are test uniformity, similar administrative conditions for all examinees, efficient and flexible use of teacher time, and the ability to focus on problem areas. In addition, teachers are able to review and revise the test recording before administering it to their students, making sure that all test items are appropriate, clear, and precise.

Preparing and Using Semi-direct Tests

Teachers today are well aware of the importance of both teaching and testing oral skills. They are also conscious of the fact that it takes a considerable amount of time to administer individual face-to-face oral tests to a number of students. The nature of semi-direct oral tests allows a teacher to administer several identical oral tests simultaneously, if language laboratory recording facilities are used. (In the absence of a language lab equipped with individual recording stations, one cassette player with one recorder can be used per testing station in order to provide identical tests to each student. Students simply place the master test tape in the player and respond onto their individual test tapes in the separate recorder.) Semi-direct, or tape-mediated, testing in this manner virtually eliminates the complaint that someone else's oral test was easier because the questions were different or that certain interferences took place during one test and not during another. In addition, once the examinees' responses have been recorded, the teacher or rater can review each response as needed when evaluating a student's performance.

In deciding whether to use semi-direct oral tests, the teacher should consider the available facilities, any specific learning objectives, and time. Thus if adequate recording facilities are not available, it would not be feasible to use this type of test. Before arranging for equipment and facilities for testing, it is also important to determine whether a semi-direct test is appropriate for assessing the skills and information being learned by the students at that time. The teacher must also determine how much time is available 1) before the test, for adequate preparation, 2) during the test, for proper administration, and 3) after the test, for fair evaluation.

The semi-direct testing format allows for testing a wide range of oral ability, from acceptable pronunciation to mastery of complex grammatical structures. Several types of test questions can be employed, including responses to simple questions, word associations, sentence repetitions, translations and interpretations, sentence recombinations, reactions to visual cues or printed situations, or free expression. The number of different types of test items is limited only by the teacher's creativity.

For preparing the semi-direct test it is recommended that the teacher write out the complete test script before recording the master tape. This allows the teacher to double-check for accuracy and completeness. It also helps to relieve the "taping anxiety" that affects many teachers. The master tape should include complete and clear instructions regarding the types of test items that follow and how and when the student is to respond. It should also contain all pauses of appropriate length for the student's responses. This will eliminate the teacher's having to start and stop the master tape

during the test. It is also advisable at the beginning of the master tape to instruct the students to record their names onto their individual test tapes, leaving a short pause for them to do so.

Scoring Semi-direct Tests

Since semi-direct tests are particularly well suited to evaluating discrete aspects of students' oral language ability, some sort of scoring form or checklist can appropriately be used by the teacher when rating each student's performance. This form should list the particular linguistic features being tested, each one to be checked as either "mastered" or "not mastered." Such a checklist will aid the teacher and the student in identifying troublesome areas in the student's acquisition of the language.

If the test items are more open-ended or somewhat global in nature, i.e. commenting upon or synthesizing a presented situation, a more general score sheet can be used. The sheet might contain numerical weightings for performance related to general fluency, grammatical accuracy, and quality of the response ("quality" involving, for example, the student's use of appropriate register and vocabulary for the situation, or trying to use more than just simple sentences and structures). The students should be informed beforehand about the various categories on the score sheet and how they will be scored. Using a scoring sheet has the additional benefit of improving the accuracy of the scoring.[8]

A Recent Study Comparing Semi-direct and Direct Tests

Oral tests have been a regular part of the beginning and intermediate German and Spanish programs at Brigham Young University for the past several years. Students have been required to take face-to-face oral tests each semester. In most cases the tests are administered by the classroom teacher. For the most part, the results have been quite satisfying; however, two concerns have been the lack of consistency among teachers in administering and scoring the tests, and the amount of time required to give individual oral tests. For this reason, an experiment was undertaken to compare the results of the customary direct oral achievement tests with results of laboratory-administered (semi-direct) tests, which would be scored by the same person.[9] It was felt that a semi-direct test offered some compensating advantages over direct tests, i.e. improved test uniformity, administration and scoring consistency, and time efficiency. It was also believed that if results of the semi-direct tests correlated highly with the results of the direct tests, student scores on the semi-direct tests could be considered indicative of probable performance on a more direct measure.

Procedures. The study compared a semi-direct testing approach to a direct testing approach. Twenty-nine intermediate German students and 20 intermediate Spanish students took two oral exams at the end of the semester: a direct, face-to-face test and a semi-direct, laboratory-administered test. The German direct test consisted of three parts totaling 61 possible points. Part A assessed the ability to pronounce German. The student was given a sheet with ten sentences to read aloud. In each sentence only one sound, or combination of sounds, was observed for scoring purposes. One point was assigned to each sound uttered correctly.

Part B consisted of ten "structured interview" questions. For each question one point was subtracted for every error or repetition. An additional point was subtracted if the answer was inappropriately short or not in sentence form.

Part C required students to respond to picture cues, giving either questions or statements depending upon the content of the pictures. Two points were given for each error-free response; one point if the grammar was incorrect but the response was understandable; no credit if the response was inappropriate or not intelligible.

The German semi-direct test consisted of eight structured interview-type questions that were prerecorded on a cassette tape. The tests were administered in the language laboratory. Each student recorded his or her responses onto individual blank tapes while listening to the questions on the master tape being played from the laboratory console. The same scoring procedures used for scoring the structured interview section of the direct test were used in scoring the semi-direct test questions. At the beginning of the semi-direct test recording, instructions were given to the students informing them that each question would be given twice, after which there would be a pause for them to respond to the question. They were also informed that proportionally more time would be allowed for answering the more difficult questions. In addition, students were told that they would be graded on the quality (i.e. use of complex structures, appropriate vocabulary, etc.) as well as on the accuracy of their responses.

The Spanish direct and semi-direct tests each consisted of the same structured interview-type items. The items ranged in difficulty from elementary questions, such as "Where are you from?" to more

grammatical and conceptually difficult questions for intermediate students, such as "What would you do if you were given $100?" All of the tests were scored using the same scoring sheet, which focuses on four areas: comprehension, fluency, grammatical accuracy, and quality of response. Credit varied according to the difficulty of the item. The same preliminary instructions on the prerecorded German semi-direct test tape were also given on the Spanish test tape.

The tests were administered a week apart. To control for rater differences, the same persons in German and in Spanish scored both sets of exams. The two test scores of each student were correlated using the Pearson product-moment formula to determine the extent of correlation between the two instruments.[10]

In addition, all of the students were asked to complete a questionnaire dealing with their attitudes toward various aspects of each of the two tests. (A sample of the questionnaire is included in the Appendix.) Student responses were recorded on a 10-point (0 to 9) Likert scale. The mean and standard deviations for each variable within each item on the questionnaire were calculated and t-tests were performed in order to determine whether the responses differed significantly.

Results. The correlation between the direct and semi-direct tests of German was .69; the correlation coefficient between the Spanish tests was .89. The higher Spanish correlation could possibly be due to a test-retest effect, since the two Spanish tests were very similar, while the German tests were different. Table I lists the descriptive statistics, correlation coefficients (r-values), and levels of probability (p-values) of the tests.

Further examination of the data shows that of the 29 students taking the German tests, 27 scored higher on the direct test, while only 2 scored higher on the semi-direct measure. Of the 20 students in Spanish, 15 scored higher on the direct test, and 5 scored higher on the semi-direct test. However, a rank-order comparison of the students reveals that they tend to be ranked in the same order on both types of tests (German, Rho = .67, $p = .001$; Spanish, Rho = .91, $p < .001$). Comparative rankings are found in Tables II and III.

Responses of the students regarding their feelings about the two test types differed for the two language groups (see Table IV). Data collected from the German students showed that they did not have a strong preference for either of the two tests in terms of their own *performance,* test *fairness,* and item *difficulty.* Although statistics were not significant on these items, students tended to rate the semi-direct test more positively when evaluating *pleasantness, frustration,* and *personal preference.* Significant statistics revealed that the direct test provoked more *nervousness,* and the semi-direct test allowed for more *comfort.* Of the 29 students who took the two tests, 17 preferred the semi-direct test and 12 the direct interview.

The Spanish questionnaire yielded significant results showing student preference for the direct test in terms of their *performance, pleasantness* of test-taking, *fairness* of the items, *frustration* level, *difficulty,* and *personal preference.* Students experienced *nervousness* on both tests, and they felt somewhat more *comfortable* during the direct test. All 20 of the Spanish students taking both tests personally preferred the direct interview.

Conclusions and Recommendations

Results of the analysis of the data indicate several differences between the two language groups with respect to outcome as well as to attitudes toward the two testing procedures. The most notable disparity, however, is with respect to attitudes toward test type. Three reasons for this can be proposed. The first is tester personality: several comments from the German students after the tests were completed seemed to confirm that the person administering the German direct tests was somewhat impatient and at times a little brusque with the students. A second major factor that is believed to have influenced the affect responses is the fact that the German direct test counted toward the final grade for the course, whereas the direct test for the Spanish students did not. Finally, the Spanish direct test was very different from the German direct test, and this could have affected both student test-type preference and performance. Instead of a general preference for direct, face-to-face tests as found in other research by Shohamy,[11] the perceived difficulty of each test seemed to be related directly to test type preference, for whatever reason.

Although the rank orders of scores for both the direct and semi-direct tests within a language were similar, and both the German and Spanish tests yielded correlation coefficients significantly different from zero, the results were not sufficient to conclude that semi-direct tests can replace direct tests and give consistently the same outcome with respect to speaking achievement. Teachers should be cautious about comparing scores from direct tests with scores from semi-direct measures. This does not mean, however, that a teacher is not justified in using semi-direct measures for testing

discrete elements of oral language. For teachers who do not have sufficient time to administer individual face-to-face tests, semi-direct tests would appear to be a viable alternative, especially since the teacher is able to score the test tapes at a later time.

This study could be profitably replicated using a greater number of students in each group. Both the direct and semi-direct tests should be of equal importance to the students and be similar—but not identical—in composition. Several administrations of each type of test should be given throughout the semester before having the students complete the affect questionnaire. Additionally, data should be organized in such a manner as to determine whether test type preference appears to be related to individual student achievement.[12]

Table I

Means, Standard Deviations, and Correlation Coefficients of Direct and Semi-Direct Test Scores by Language

Language:	German		Spanish	
Test type:	Direct	Semi-Direct	Direct	Semi-Direct
Possible:	61.0	27.0	119.0	119.0
Means:	45.9	14.8	91.7	85.6
S.D.	5.6	4.9	19.4	20.7
r		.69		.89
p		< .001		< .001
N	29	29	20	20

Probability level accepted as significant = .01

Table II

Comparative Student Rankings for Direct and Semi-Direct Tests
(German)

Student	Direct	Semi-Direct	Student	Direct	Semi-Direct
A	1	2	P	15.5	24
B	2	5	Q	17.5	14.5
C	4	1	R	17.5	18
D	4	5	S	20.5	5
E	4	14.5	T	20.5	9.5
F	6	18	U	20.5	14.5
G	8	7.5	V	20.5	21.5
H	8	9.5	W	23	26.5
I	8	11.5	X	24.5	14.5
J	11	3	Y	24.5	25
K	11	7.5	Z	26.5	18
L	11	11.5	AA	26.5	21.5
M	13	21.5	BB	28	28
N	14	26.5	CC	29	29
O	15.5	21.5			

Table III

Comparative Student Rankings for Direct and Semi-Direct Tests
(Spanish)

Student	Direct	Semi-Direct	Student	Direct	Semi-Direct
A	1	2	K	10.5	8
B	2	4	L	12	15
C	3	1	M	13.5	11
D	4	9	N	13.5	10
E	5	6	O	15	14
F	6.5	7	P	16	17
G	6.5	3	Q	17.5	16
H	8	13	R	17.5	18
I	9	5	S	19	20
J	10.5	8	T	20	19

0652-7

Table IV

Means, Standard Deviations, and T-test Results
for Affect Questionnaires by Language*

Category		German		Spanish	
		Direct	Semi-Direct	Direct	Semi-Direct
Performance	mean	4.63	5.33	6.90	3.90
	s.d.	1.79	2.19	1.21	2.38
	t	1.35		5.02	
	p	.18		< .001	
Pleasantness	mean	5.27	6.37	6.90	3.75
	s.d.	2.20	2.16	1.97	2.65
	t	1.95		4.26	
	p	.05		< .001	
Fairness	mean	6.93	7.20	8.40	3.40
	s.d.	2.13	1.97	1.05	2.39
	t	.49		8.56	
	p	.62		< .001	
Comfort	mean	5.00	6.83	6.60	4.70
	s.d.	2.27	2.10	2.56	2.75
	t	3.23		2.26	
	p	< .01		.028	
Frustration	mean	4.17	5.47	6.95	3.30
	s.d.	2.38	2.56	2.65	3.15
	t	2.03		3.97	
	p	.04		< .01	
Nervousness	mean	4.03	6.47	5.70	4.85
	s.d.	2.67	1.98	2.83	3.12
	t	4.00		.90	
	p	< .001		.62	
Difficulty	mean	5.07	5.63	6.25	3.35
	s.d.	1.80	1.63	1.94	2.58
	t	1.27		4.01	
	p	.20		< .001	
Preference	mean	5.63	6.47	7.20	2.65
	s.d.	2.37	1.41	1.32	2.58
	t	1.65		7.20	
	p	.10		< .001	

German N = 29
Spanish N = 20
Probability level accepted as significant = .01
*A sample questionnaire is shown in the Appendix.

NOTES

[1]The term *proficiency* as used in this paper refers to an acquired competence in the language regardless of how or where that competence was acquired. *Achievement* relates to acquisition of specific linguistic features that are associated with a particular course of instruction.

[2]Working together, ACTFL and the Educational Testing Service have modified the FSI rating scales, making them suitable for testing language proficiency in academic institutions. A set of *Provisional Guidelines* has been published and is available through the American Council on the Teaching of Foreign Languages (ACTFL); P.O. Box 408; Hastings-on-Hudson, NY 10706.

[3]For examples of research in this area, see A.S. Palmer, P.J.M. Groot, and G.A. Trosper, eds. *The Validation of Oral Proficiency Tests: Selected Papers for the Colloquium on the Validation of Oral Proficiency Tests* (Washington, DC: TESOL, 1981); R.T. Clifford, "Reliability and Validity of Oral Proficiency Ratings and Convergent/Discriminant Validity of Language Aspects of Spoken German Using the *MLA Cooperative Foreign Language Proficiency Tests: German (Speaking)* and on Oral Interview Procedure," Diss. University of Minnesota, 1977; L.F. Bachman and A.S. Palmer, "The Construct Validation of Oral Proficiency Tests," *TESL Studies,* 3 (1980), 1-20.

[4]See D. Hosley and K. Meredith, "Inter- and Intra-Test Correlates of the TOEFL," *TESOL Quarterly,* 13 (1979), 209-17; J.W. Oller, Jr., "Cloze Tests of Second Language Proficiency and What They Measure," *Language Learning,* 23 (1973), 105-18; J.W. Oller, Jr., "Dictation as a Test of ESL Proficiency," in *Teaching English as a Second Language: A Book of Readings,* ed. H.B. Allen and R.N. Campbell (New York: McGraw-Hill, 1972), pp. 346-54.

[5]See R.L. Jones, "The Oral Interview of the Foreign Service Institute," in *Advances in Language Testing: Series 1, Some Major Tests,* ed. B. Spolsky (Arlington, VA: Center for Applied Linguistics, 1979), pp. 104-15; P. Lowe, Jr., *Handbook of Question Types and Their Use in LLC Oral Proficiency Tests* (Washington, DC: CIA Language School, 1976).

[6]A review of recent foreign language publications shows little discussion about or emphasis on oral achievement testing *per se.* Some articles that have focused on oral achievement testing include R. Ravid, "Presentation of Procedures for Development of a Second Language Achievement Test," *Foreign Language Annals,* 16 (1983), 201-05; J.W. Larson, "Skills Correlations: A Study of Three Final Examinations," *The Modern Language Journal,* 67 (1983), 228-34; H.H. Frink, "Oral Testing for First-Year Language Classes," *Foreign Language Annals,* 15 (1982), 281-87; Clark describes some of the trends and current developments in oral achievement testing, but his discussion is more related to proficiency testing. See J.L.D. Clark, "Language Testing: Past and Current Status—Directions for the Future," *The Modern Language Journal,* 67 (1983), 431-43.

[7]For more detailed information about the use of the Recorded Oral Proficiency Examination (ROPE), see P. Lowe, Jr. and R.T. Clifford, "Developing an Indirect Measure of Overall Oral Proficiency," in *Measuring Spoken Language Proficiency,* ed. J.R. Frith (Washington, DC: Georgetown Univ. Press, 1980), pp. 31-39.

[8]R.M. Valette, "Evaluating the Second-Language Learning Program," in *Learning a Second Language: Seventy-ninth Yearbook of the National Society for the Study of Education, Part II,* ed. F.M. Grittner (Chicago: Univ. of Chicago Press, 1980), p. 160.

[9]Randall l. Jones, Professor of German at Brigham Young University, assisted in the study, supervising the testing and data collection of the German students.

[10]A Pearson product-moment formula is a statistical formula to determine correlation. An *r* value above .80 is considered more meaningful than an *r* value below .80. Thus the Spanish *r* value in Table I is more meaningful than the German *r* value.

[11]Shohamy reports that according to her findings, a face-to-face test (the oral interview in her study) is preferred over a number of other alternatives. See E. Shohamy, "Students' Attitudes toward Tests: Affective Considerations in Testing." Paper presented at the Fourteenth Annual TESOL Convention, San Francisco, 1980.

[12]Scott and Madsen found in their study with ESL students that there does appear to be a correlation between student ability and test type preference. See M.L. Scott and H.S. Madsen, "The Influence of Retesting on Test Affect," in *Issues in Language Testing Research,* ed. J.W. Oller, Jr. (Rowley, MA: Newbury House, 1983), pp. 270-79.

Appendix

Sample Affect Questionnaire

You have recently taken two similar oral tests. One was administered via tape recorder in the language laboratory and the other was administered in a face-to-face interview situation. We would like you to answer the following questions concerning your feelings about both types of tests. Please read each question carefully and answer to the best of your ability by circling the appropriate number.

1. How *well* do you feel you did on each of the tests?
   a. face-to-face— not at all well 0 1 2 3 4 5 6 7 8 9 very well
   b. language lab— not at all well 0 1 2 3 4 5 6 7 8 9 very well
2. How *pleasant* did you find the experience of taking each test?
   a. face-to-face— very unpleasant 0 1 2 3 4 5 6 7 8 9 very pleasant
   b. language lab— very unpleasant 0 1 2 3 4 5 6 7 8 9 very pleasant
3. How *fair* did you feel each of the tests was?
   a. face-to-face— very unfair 0 1 2 3 4 5 6 7 8 9 very fair
   b. language lab— very unfair 0 1 2 3 4 5 6 7 8 9 very fair
4. How *comfortable* did you feel when you took each of the tests?
   a. face-to-face— very uncomfortable 0 1 2 3 4 5 6 7 8 9 very comfortable
   b. language lab— very uncomfortable 0 1 2 3 4 5 6 7 8 9 very comfortable
5. How *frustrated* did you feel as you were taking each test?
   a. face-to-face— very frustrated 0 1 2 3 4 5 6 7 8 9 not at all frustrated
   b. language lab— very frustrated 0 1 2 3 4 5 6 7 8 9 not at all frustrated
6. How *nervous* were you as you were taking each test?
   a. face-to-face— very nervous 0 1 2 3 4 5 6 7 8 9 not at all nervous
   b. language lab— very nervous 0 1 2 3 4 5 6 7 8 9 not at all nervous
7. How *difficult* did you find each of the tests?
   a. face-to-face— very difficult 0 1 2 3 4 5 6 7 8 9 very easy
   b. language lab— very difficult 0 1 2 3 4 5 6 7 8 9 very easy
8. How did you personally *like* each of the tests?
   a. face-to-face— did not like at all 0 1 2 3 4 5 6 7 8 9 liked very much
   b. language lab— did not like at all 0 1 2 3 4 5 6 7 8 9 liked very much
9. Which of the two tests did you prefer?
   face-to-face _____
   language lab _____