

Toward an Objective Assessment of Oral Proficiency

SHIRO SATO

An Achievement-Based Measure of Oral Proficiency

One view of language proficiency holds that it is not possible to specify any set of subskills or linguistically defined elements that constitute proficiency. It is rather described in terms of how well one can communicate regardless of how much knowledge of the target language he has acquired. But it is also true that without a fair amount of knowledge of linguistically defined elements one would most likely fail to communicate orally to the other party. This can take place even when EFL learners are helped, to a great extent, by the factor of redundancy.

In an ordinary foreign language classroom both the teacher and the learner need the kind of information that tells them how many of the stated objectives have been met during a certain period of time and what needs to be reinforced further. The teacher can, then, reflect on this information in judging the soundness of teaching methods, teaching materials, and teaching objectives. A test which provides EFL teachers with this sort of information is an achievement test. By having students practice various types of drills (oral translation, directed speech, dialog, etc.) in class the teacher expects the achievement made through such drills to ultimately lead to competence whereby a student can utilize the acquired skills or knowledge to express his ideas or views in a novel communication

context. It seems clear, though, that one's competence cannot be acquired all of a sudden but rather after a period of achievement. If this is the case (or most probable), the teacher is required to check periodically each learner's progress in various sub-elements of speaking skill (pronunciation, stress, intonation, fluency, etc.) and keep cumulative records of them. In other words, the teacher who has perhaps spent hundreds of hours teaching and evaluating his students' performance in oral drills has also observed the successes and failures of his students in their attempts to use English. Thus, the teacher's ratings based on the cumulative records of each student may very well become a sound basis for measuring their capacity to speak English. Experimental findings reported by Ingram (1968) appear to support this claim. In her experiment, student scores on well-constructed, objective, standardized oral and written tests were compared with scores of these same students on teacher-made examinations and with teacher ratings of the students' abilities in oral and written skill of English. The results of her experiment are shown below.

TABLE I
 Rank Correlations of Standardized and Teacher
 Measures of ESL¹⁾

	Teacher Exam	Teacher Ratings		
		Oral	Written	Pooled
Standardized Oral Test63	.75		
Standardized Written Test68		.76	
Total Standardized Test Scores69			.81

A look at Table I enables us recognize the superiority of a teacher's subjective ratings to the teacher's more objective tests in assessing students' proficiency of expressive skills.

A Communicative Competence-Based Measure of Oral Proficiency

The other view of proficiency tests holds that the test attempts to uncover the learner's oral competence. According to this view the test constructor disregards the particular teaching methods, materials, and syllabuses under which the learners have been taught and instead he tries to assess the degree to which he can communicate in a socio-linguistic situation at the time the test is given. The purpose of such tests lies in the fact that the test results will be used to help determine a person's qualifications for a given task. An entrance examination of a higher educational institution is a case in point. If an English and American Literature Department of a college sets the minimum requirements of oral proficiency necessary to pursue the learning of English after entering college, the department is recommended to give such a test, from which the test items are sampled to reflect the requirements. If we aim at maintaining objectivity without losing much validity, it necessarily takes a considerable amount of time, especially in the scoring. Is a case of entrance examinations where only a limited time is allowed for scoring and evaluation, this 'time factor' presses upon us as the severest stumbling block. If, on the other hand, the principal purpose of giving an oral proficiency test is to use the test results only as that of preliminary screening or as the data for future use, the following testing techniques may meet the needs.

The notion on which such testing techniques are based derives from an assumption that the learner's oral proficiency can be mea-

sured by imposing 'time measures' on the subjects. Along this line several means for achieving complete objectivity in scoring oral performance are increasingly coming into use. One such device is to count the number of simple messages which can be communicated orally in a limited amount of time (Heinberg, 1970). Palmer (1970) measures the amount of time required by each subject to convey a rather complex oral message in a communication context. Still another device which combines time measures with measures of the accuracy of oral communication was reported by Upshur (1971).

It is natural that for any oral communication to take place two parties are necessary: the sender and receiver of a message. And there are some features unique to everyday conversation, such as interaction, unpredictability, and the relationship between the speakers. The testing technique which entails all these features is the interview. Thus, in judging the learner's oral performance at least the two factors should be incorporated together with the proper use of linguistic knowledge; one factor is the ability to comprehend the message and the other is the ability to make inappropriate utterance in a communication context. It must be noted, however, that no matter how valid a test of oral proficiency is made, a question of scoring oral performance objectively remains unsolved. The following is a tentative list of questions which must be explored.

1. Should we score linguistic adequacy such as grammar, lexicon, and phonology independently of situational appropriateness, or should we rather combine the scores for these aspects with that of appropriateness? If so, how are we to combine?; should situational appropriateness have a greater weight in marks over linguistic adequacy, or vice versa?

2. Are there degrees of adequacy or in appropriateness of an utterance? If so, how many degrees should there be and on what basis we can determine these degrees?

Levenston (1973) suggests three levels for this : a) fully appropriate in form and content, b) appropriate in content but not in form (even though grammatically well formed), and c) inappropriate in content.

3. No less important is how we are to interpret silence. The examinee may keep silent either when he does not comprehend what the examiner says, or when he is at a loss as to how to express himself. In either case it is highly probable that the longer the examinee keeps silent the less competent he is.

These questions must be answered carefully since they influence, to a great extent, the validity of the tests. Moreover, establishing the criteria such as these mentioned above will undoubtedly contribute to increasing the reliability of an oral proficiency test. Other suggestions which will be useful also for oral reading or imitation are made by Upshur.

- (1) that graders use an analytic scheme so that one element of speech is rated at a time: analytic scoring of speech might include separate ratings of phoneme differentiation; word stress; intonation; vocabulary choice; aspects of sentence structure such as verb tense formation, concordance, relativization; fluency, absence of 'false starts'; and so on;
- (2) to increase the number and types of questions or speech

- tasks so that random errors tend to balance out ;
- (3) to increase the number of graders so that individual rater errors tend to cancel out ;
 - (4) to train all raters with a set of standard recorded interviews or free speech tests and to have all raters review these standard types periodically²⁾.

The Cloze Technique as a Useful Device for Measuring Oral Proficiency of EFL Students

Since the cloze technique was invented by Wilson Taylor (1953) to measure the readability of prose, it has repeatedly been demonstrated, in numerous studies with native speakers of English, to correlate well with knowledge of vocabulary, reading comprehension and ability to read aloud.

An important breakthrough was brought about by Darnell (1968). In a fairly extensive study with non-native speakers he used the cloze method as a gauge of EFL proficiency and found high correlation (.82) between the cloze scores and the Test of English as a Foreign Language (Educational Testing Services ; Princeton, New Jersey). A surprising and yet interesting finding was that the cloze test, which is the written test, correlated most highly with the listening comprehension part (.73) of the various parts of TOEFL. Following this stimulating study with second language learners, Oller and Conrad (1971) tested the extent to which the cloze technique can be used among EFL learners as a yardstick for discriminating proficiency in English. The subjects used in their study were 102 foreign students to be enrolled at UCLA in 1970. They were from 38 different countries with 27 different language backgrounds. They were grouped into one of five proficiency levels : beginning, intermediate,

advanced, advanced composition, and non-native graduate students enrolled in one of the courses in TESL. In this experiment with EFL students it was found that the mean scores of the cloze test increased as the level of proficiency increased, thus suggesting a useful tool to measure proficiency of EFL students. A number of other experiments on the cloze test seem to support the idea that it can be used as a reliable, valid, and economical tool to measure overall proficiency of EFL students.

As has already been mentioned above, the interview technique presents both theoretical and practical problems which must be solved before it can be successfully used. Corder suggests a contradiction to be found in the relationships between the theoretical validity and practicality of the techniques of oral interview and free conversation.

The more ambitious we are in testing the communicative competence of a learner, the more administratively costly, subjective and unreliable results are³).

One possible solution to this is the conversational cloze test, which is made from a transcript of ordinary conversation. We can assume that the conversational cloze test measures oral competence indirectly if its scores correlate consistently with those of more elaborate oral tests. The superiority of cloze tests to other oral tests is great because the former is rather easy to construct, administer, and score. All of these features attached to the cloze test make them quite reliable. Brown (1983) asked 30 students of English from various nations to take part in two five-minute interviews and to take two conversational cloze tests. The performance of these students

at the interviews was scored for effectiveness, correctness, and range and accuracy of vocabulary. The two conversational cloze tests consisted of a total of thirty-eight items and were constructed from the transcripts of two simulated, authentic conversations where every seventh word was deleted. The rank correlation between the scores of the students on the two conversational tests and their scores on the interview tests was 0.8. This high correlation gives some evidence that conversational cloze tests may be a good, indirect way of measuring one's oral ability.

With a similar purpose as Brown's in mind, Hughes (1981) attempted to clarify the differences in effectiveness to gauge oral competence between the conversational cloze and the ordinary cloze (its test material is taken from prose). The subjects of his study were 64 foreign students with various linguistic backgrounds. The scores of both types of cloze tests were correlated with the teachers' ratings of the students' ability to converse in English. A total of six teachers were asked to rank their students in order according to (a) oral comprehension, (b) oral production, and (c) communicative competence. These teachers were also asked to grade each student's performance on pronunciation, grammar, vocabulary, fluency, and comprehension. The results showed that both of the two conversational cloze tests were superior to the two prose-based cloze tests as indicators of teachers' judgements of their students' oral ability.

It was shown in the two experiments on conversational cloze tests that the tests are highly valid as measures of one's oral proficiency. Furthermore, the tests are easy to make, administer, and score, the features of which lead to high reliability and practicality. While, at this stage, any conclusive statement cannot be made, the conversational cloze procedure appears to be a reasonably prom-

ising device to assess one's oral ability.

NOTES

- 1) Elizabeth Ingram, "Testing in the Context of a Language Learning Experiment," *Problems in Foreign Language Testing*, eds. J.A. Upshur and Julia Fata. Language Learning, Special Issue Number 3 (1968), 147-161.
- 2) John A. Upshur, "Objective Evaluation of Oral Proficiency in the ESOL Classroom," *TESOL Quarterly*, 5 (March, 1971) 47-59.
- 3) S.P. Corder, *Introducing Applied Linguistics*, London : Penguin, 1973, p. 364.

REFERENCES

- Briere, Eugene J.; Clausing, Gerhard ; Senko, Donna ; and Purcell, Edward. "A Look at Cloze Testing Across Languages and Levels." *Modern Language Journal*, 62 (January-February, 1978) 23-26.
- Brown, David. "Conversational Cloze Tests and Conversational Ability." *ELT Journal*, 37/2 (April, 1983) 158-161.
- Darnell, D.K. "The Development of an English Language Proficiency Test of Foreign Students Using a Clozetrophy Procedure." *ERIC ED* 024-039, 1968.
- Hughes, Arthur. "Conversational Cloze as a Measure of Oral Ability." *ELT Journal* 35/2 (April, 1981) 161-167.
- Levenston, E.A. "Aspects of Testing the Oral Proficiency of Adult Immigrants to Canada." *Papers on Language Testing*. Edited by Leslie Palmer and Bernard Spolsky. Washington D.C.: Teachers of English to Speakers of Other Languages, 1975.
- Oller, John W., Jr. and Conrad, Christine A. "The Cloze Technique and ESL Proficiency." *Language Learning*, 21 (December, 1971) 183-195.
- Pilliner, A.E.G. "Subjective and Objective Testing." *Language Testing Symposium* Edited by Alan Davies. London : Oxford University Press, 1968.
- Soudek, Miluse and Soudek, Lev I. "Cloze after Thirty Years: New Uses in Language Teaching." *ELT Journal*, 37/4 (October, 1983) 335-340.
- Spolsky, Bernard. "Language Testing-The Problem of Validation." *TESOL Quarterly*, 2 (June, 1968) 88-94.
- Taylor, Wilson C. "Recent Developments in the Use of Cloze Procedure." *Journalism Quarterly*, 33 (Winter, 1956) 42-48.