

Testing Oral Language Proficiency of University EFL Students

M^r. Cherif Hadjira
Department of English
University of Batna
cherif.hadjira@gmail.com

Abstract

The present study aims at developing a series of objective criteria for measuring and scoring the oral proficiency of EFL students, attempting to move toward a more objective mode for scoring the oral language proficiency. To achieve this purpose, eighty students from the department of English, at Batna University, are selected based on their availability and their successful passing of conversations one, two, and three. Then, their oral proficiencies are rated against a validated and newly-developed checklist. The obtained scores are compared with the group's performance in their previous conversation sessions.

The results indicate a low correlation between the two groups of scores. It is also proved that the subjective measures are not reliable enough to indicate the students' abilities in terms of oral language proficiency.

ملخص

تهدف هذه الدراسة إلى تطوير مجموعة من المعايير الموضوعية لقياس وتقييم الكفاءة الشفهية لمجموعة من طلبة اللغة الانكليزية وذلك محاولة منا للمرور نحو وضع أكثر موضوعية في تقييم وتنقيط الإتقان الشفهي للغة. ولتحقيق هذا الغرض قمنا باختيار 80 طالبا من قسم اللغة الانكليزية في جامعة باتنة، وقد تم اختيارهم على أساس مدى تواجدهم في القسم وكذلك لتمكنهم الشفهي الناجح في ثلاثة محادثات كانوا قد قاموا بها سابقا. ثم تم تقييم كفاءتهم الشفهية باستعمال مرجعية تنقيط لبعض الخبراء في هذا الميدان، وذلك سعيا منا لتحقيق موضوعي من كفاءتهم الشفهية في اللغة.

وتمت مقارنة النتائج التي تم الحصول عليها مع نتائج الأداء الشفهي للمجموعة أثناء الدورات السابقة للمحادثة. وقد أشارت النتائج النهائية إلى وجود ارتباط نسبي بين مجموعتي النتائج. وقد ثبت أيضا أن المعايير و التدابير الذاتية ليست موثوقة ولا يمكن الاعتماد عليها بصفة كلية في تقييم و تنقيط الطلبة فيما يتعلق بالإتقان الشفهي للغة.

1-Introduction

Many language tests follow a psychological rather than linguistic theoretical framework, evidenced by the use of a single modality (such as a paper-and-pencil test that ignores spoken and oral comprehension) (Pray, 2005). Most current tests of oral proficiency have the same deficiencies, and many of the measures used by the teachers share the problem of subjectivity. This status is sustained by factors such as large classes, teachers' inadequate command of English, and the lack of easy access to support materials and facilities (Ramanathan, 2008, Sook, 2003). Therefore, due to the complicated nature of this skill, testers and language teachers should make use of reliable analyses for the purpose of objectivity. Testing is, however, necessary in foreign language learning, though language tests are no easy task for practitioners of different skills. Ur (1996) states that “when testing the oral proficiency of learners we may simply interview them and assess their responses; or use other techniques like role play, group discussion between learners, monologue, picture-description and so on”.¹

2-purpose of the Study

The purpose of the present paper is to evoke the fact that in university conversation classes there exists no clear-cut checklist or a hard and fast set of criteria for measuring the oral proficiency of students learning English as a foreign language. Various types of tests designed and administered- mostly paper and pencil listening tests, student-student, and student-teacher interviews rated without using established criteria- are not suitable to the mode. Therefore, an objective and integrated checklist is needed to measure the students' competence on the basis of their performance. To do so, the researcher appropriately modified the existing checklists to include an important

¹ - P. Ur, **A Course in Language Teaching**. (Cambridge: Cambridge University Press, 1996) p.133.

factor, "communication" which is essential for the purpose of assessing levels of oral ability, to help the test designers move from subjective teacher-made tests towards a more standardized testing of oral/aural skills. This checklist is developed as comprehensively as possible so that the researcher is able to take into account most of the required criteria in the tests for measuring oral proficiency. Sample models for developing this checklist are extracted from Farhady, Jafarpur, and Birjandi (2001), Heaton (1990), Hughes (2003), IELTS Testing Center (2000), and Underhill (1987). The most significant criteria considered in the checklist include accent, speed of response, diction, listening comprehension, communication, and fluency to name but a few.

3-Background of the Study

Until now, several studies have been conducted in developing measures for evaluating

language learners' oral proficiency. Harris (1968) suggests a list of criteria for measuring oral skills, which is technically known as "Sample Oral English Rating Sheet".

Harris's sample comprises five criteria to be rated: pronunciation, grammar, vocabulary, fluency, and comprehension, each of which includes 5 levels.

The proficiency guidelines for speaking were developed in 1982 by the American Council on the Teaching of Foreign Languages (ACTFL) with the purpose of creating a criterion that could be used to identify the foreign language proficiency of speakers ranging from "no knowledge" of EFL to "total mastery" gained through widespread application. The ACTFL guidelines include: superior, advanced (high, mid, low), intermediate (high, mid, low), and novice (high, mid, low) levels.

Also, Underhill (1987) has offered a rating scale for measuring speaking skills. A rating scale, as defined by Underhill includes 1) very

limited personal conversation, 2) personal and limited social conversation, 3) basic competence for social and travel use, 4) elementary professional competence, and 5) general proficiency of all familiar and common topics.

One area of decision-making in rating scales is scoring. Farhady, Jafarpur, and Birjandi (2001) state that depending on the objective of the test, scoring may be done holistically or discretely; the former refers to an overall impression according to which the interviewee either receives excellent, good, fair, or pass/fail scores. The latter, on the other hand, rates the interviewee's performance separately on scales that relate to accent, structure, vocabulary, fluency, and comprehension. Another crucial work in this realm is a checklist developed by Hughes (2003). The checklist assigns the candidates (interviewees) to a level holistically and rates them on the six-point scale of each of: accent, grammar, vocabulary, fluency, and comprehension. The test is both given and rated by the teacher with no student self-evaluation and self-judgment about their progress.

However, more recent studies, emphasizing the interactional aspect of language, have focused on learners' awareness of the test procedures. For example, a different view of language assessment, inspired by the idea of Task-Based Instruction (TBI) is casting light on the field of foreign language testing. In task-based language assessment (TBLA) language use is observed in settings that are more realistic and complex than in discrete skills assessments, and which typically require the integration of topical, social and/or pragmatic knowledge along with knowledge of the formal elements of language (Mislevy, Steinberg & Almond, 2002). In another case, Lambert (2003), giving the tests at the end of the term to nine classes of between 26-31 first year Japanese university students majoring in electrical and mechanical engineering, predominantly male, upper elementary to pre-intermediate level, concludes that recordings of the student-student interviews would provide a clear justification for the marks awarded and it is also a good

idea to give the students a chance to think about what they would say by putting the actual test roles on the Intranet. Hughes (2003) talks about the need and relevance of testing to the EFL enterprise, saying: "I believe that the teaching profession can make three contributions to the improvement of testing: they can write better tests themselves; they can enlighten other people who are involved in testing processes; and they can put pressure on professional testers and examining boards, to improve *their* tests".¹

The significance of the study resides in the fact that Speaking is a productive skill, and hence, EFL teachers ought to give it a great importance, in practice and assessment, for the sake of the general oral mastery of language. Ur (1996) prompts this belief by saying: "I think that oral testing is worth the investment: not so much for the sake of the overall validity of the proficiency test of which it is part, as for the sake of the backwash."²

So, the current study could function as a prerequisite to interactional approaches to language testing since its main goal is to suggest a rather valid and reliable checklist as a measurement device for assessing oral proficiency.

In the light of the above studies, it should be clearly noted that the current scoring methods applied in Algerian universities are mostly impressionistic, based on experience and lack of validity and reliability; the checklist proposed can be utilized as an alternative method in order to obtain objective scores which are true representative of the students' oral communicative ability.

¹ - A. Hughes, **Testing for language teachers** (2nd ed.). (Cambridge: Cambridge University Press, 2003). P.5

² - P. Ur, **A Course in Language Teaching**.(Cambridge: Cambridge University Press,1996) .p.135

In other words, the same checklist could be used by both teachers and students in methods such as TBLA, student-student interviews, etc.

4-Hypothesis of the Study

For the sake of arriving at an objective decision, this study pursued to provide an answer or rather check the following main hypothesis:

1. Which measure, subjective or objective, provides a more valid and reliable estimate of the oral proficiency of the EFL learners?

5-Methodology

5.1-Participants

Subjects in the present study are 80 students selected from the department of English-2nd Year class- at Batna University.

The rationale for their selection is their availability and the fact that the participants had already passed three conversation courses successfully and they are also taking conversation four at the time of the study. Twenty five percent of the participants are male (n=20) and the rest are female (n= 60), ranging from 20 to 27 years old.

5.2-Instrumentation

One instrument utilized in the process of the present study is the proposed checklist including a series of standards and criteria for measuring oral communicative abilities of EFL students on an academic level. Another instrument is the IELTS format of interview (a speaking test) in which the interviewees are asked to answer general and personal questions about their homes and families, jobs, studies, interests, and a range of similar topic areas in about five minutes. The other instrument utilized in the present study is an audio recorder for recording the interviews.

5.3-Procedure

In order to validate the newly designed checklist, that is, to determine the extent to which the checklist measures what it is supposed to measure, a pilot study is conducted. Ten students were randomly selected and rated using both the new checklist and the one designed by Hughes (2003) to determine the criterion-related validity of the new checklist. The correlation coefficient obtained between the two series is 0.968 indicating that the new checklist is valid.

By the end of the semester, the subjects are asked to speak for about two minutes on a particular topic for which they are given almost two minutes to think about. All of the selected subjects are interviewed and rated against the new checklist by two raters, first by the researcher and then by a teacher of English. Each interview session is held in the presence of the researcher, the classroom teacher (both as interviewers) and one of the subjects (as the interviewee). Each interview commences with a set of simple questions and then proceeds to more challenging ones, and before each session, the subjects are asked to explain and write down brief notes on the sources and material which they practice in conversation courses and also the methods applied during the courses and the final examinations. This is done as a warm up activity to decrease the psychological stress and to ensure that the same mode and channel is being used to score the oral proficiency of the subjects in the previous courses. In order to enhance the reliability of the scores, rating activities are carried out first by the researcher and then by an inter-rater, and agreement is reached on each student's score.

All the subjects' scores in conversation courses are collected from the Department of English, at Batna University, and their average scores are calculated. After gathering the required data, the next step is to rate and score each interview based upon the developed checklist with the aim of attaining more reliable and objective scores. The correlation coefficient determines whether or not there is a possible relationship amongst these series of scores.

6-Data Analysis

The interviews obtained during the study are assessed through listening to the recordings and the performance of each interviewee is rated on the basis of the criteria indicated in the developed checklist; first by the researcher and next by an inter-rater. After calculating the average of the scores given by two raters, two series of scores are attained - the average scores in interview and the average scores in conversation courses.

By using the Microsoft Excel software (2003 version) and calculating the variables, the correlation obtained was 0.0045 which indicates that the correlation between the two series of scores is substantially low. This proves the hypothesis of the study that the previous ratings are wholly implemented in a subjective manner compared to the ratings made against the newly developed checklist including the objective criteria.

Minimum and maximum values are higher for the subjective rating than for the objective ones which might indicate that instructors in conversation courses are more generous, and so these scores do not represent the true oral abilities of the subjects. Also the mean of the students' scores in conversation courses is 15.87 while the corresponding mean score in interview is 11.77. However, the difference between the standard deviation of both groups is not meaningful and indicates that the use of standard criteria for scoring oral proficiency causes the scores of the students to fall off in a similar manner, i.e. the subjects who receive higher scores among others by subjective scoring measures also receive the higher range of scores by the objective measures although their range of scores lowered meaningfully in objective scoring. The median of the scores in conversation courses is 16 which shows that half of the scores are higher and half of them are lower than 16. The median of the objective scores is 12.

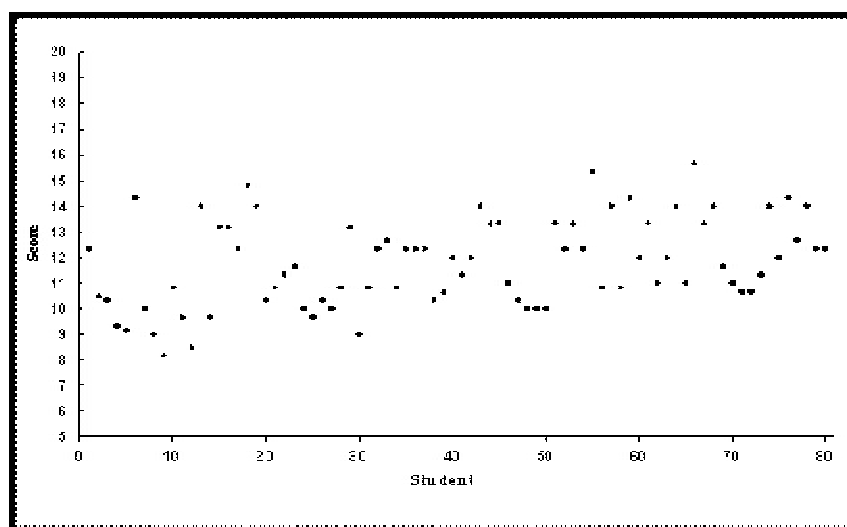
Table1 illustrates the average scores of conversation courses assigned to students by their instructors through traditional subjective means of testing and scoring oral communicative abilities. 70 percent of the scores are in the range of 15 to 20, and the rest of the scores fluctuate between 12 and 15.

Table1. Descriptive statistics on objective and subjective measures

Statistical Evaluations	Subjective Scores	Objective Scores
Population	80	80
Min. Value	12.33	8.17
Max. Value	19.33	15.67
Range	7	7.5
Mean	15.87	11.77
Standard Deviation	1.70	1.70
Standard Error	0.19	0.19
Median	16	12
Sum	1,270.1	942.38
Sum of Squares	20,395.27	11,331.72
Variance	2.92	2.92

Figure1: represents the average scores given by the two raters to the same groups of students based on the standard criteria listed in the designed checklist (See Appendix for a sample checklist). The distribution of these scores is lower than the course scores with 50 percent of the scores between 12 and 16 and the rest between 8 and 12.

Figure 1: *The average scores of subjects based on the standard checklist*

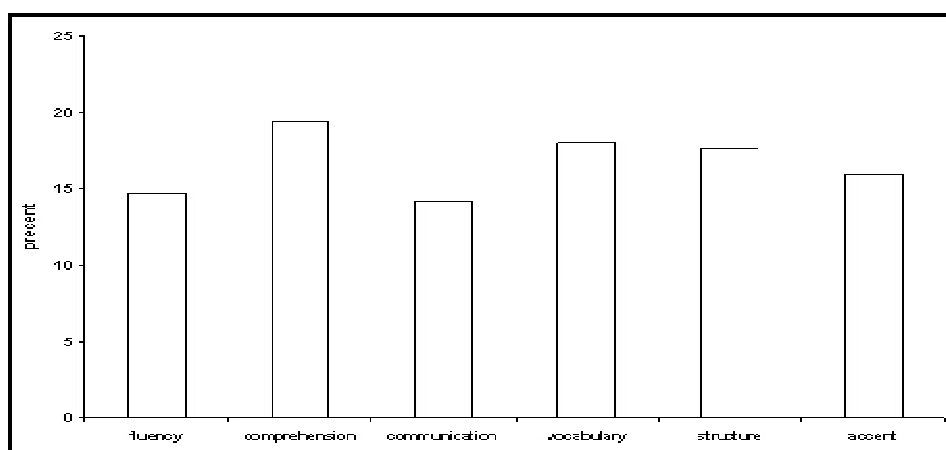


This distribution of scores, indicating the variation in the students' oral abilities, shows that the actual abilities of the students are far below than those given by their EFL teachers.

In order to determine the contribution of each scale on the objective scores and the performance of the subjects, the scores in various scales are specified in terms of six scales. Although the general performance of the subjects is weak, the figure shows the strengths and weaknesses of the subjects in different sub skills of the speaking skill.

The checklist contains 6 scales namely, fluency, comprehension, communication, vocabulary, structure and accent, each of which includes 5 levels of proficiency. The performance of the subjects on each scale is then independently calculated.

Figure 2: *The percentages of scales of the checklist contributed to the total score*



The performance of the subjects in the areas of comprehension, vocabulary, and structure is fairly better compared to that of fluency, communication, and accent (Figure 2).

7-Discussion

One point to discuss here is that teachers' scoring the students' oral proficiency subjectively is neither reliable nor valid, and so the given scores cannot present the true ability of the subjects in oral language proficiency. By analyzing oral language proficiency in terms of a number of scales and calculating the learners' ability in terms of their performance on the scales, the researcher could now validly judge the learners' oral language proficiency. The general performance of the subjects, however, is weak in the ratings carried out, but their performance in the individual scales of the checklist is varied. That is, in certain scales they perform successfully but in others they do not.

Results show that the performance of the subjects in linguistic components is better than their performance in communicative aspects. Fluency is one of the key factors in assessing the oral language proficiency. Most of the subjects in the present study are hesitant and

their oral performances are discontinuous. Another scale on the checklist is comprehension in which the subjects show a better performance than in the other scales. In most cases they understand the question or/and the gist but are not able enough to manage the discussion. We suggest that the comprehension skill of the subjects should be assigned a higher priority in the development of the English teaching curricula.

On the scale of communication, the subjects have the weakest performance indicating the greater attention they need to pay to this aspect of their communicative competence. Although the performance of the subjects in vocabulary and grammar scales is better, there are other problems such as lack of complete accuracy that should be considered by the EFL teachers. As for acceptable and intelligible accent, interviewees show a weak performance in this scale which may be indicative of the EFL learners' ignorance of this part of language.

Analyses show that the mean score of the subjects in objective scoring is approximately four points lower than the mean score in subjective scoring. It might be that subjective scoring is implemented based on personal judgments and also the scores are allotted to the overall speaking skill of the subjects, and therefore, the range of scores is high. On the other hand, in the objective scoring, in light of the standards and criteria, the communication skill as a whole is broken into six distinct sub-skills. The scores obtained for each sub-skill are summed up in order to represent the total score given to each subject in terms of their comprehension ability, and so, the range of scores was meaningfully lower. Through objective scoring, weak and strong sub-skills of the subjects' speaking skill can be assessed enabling EFL teachers to remove the deficiencies and reinforce the stronger points.

8-Conclusion

In summary, the point to be taken into account is the lack of attention and application of specific standards to score learners' oral

productive skill. In the EFL setting, there are many teachers who score the learners' speaking ability subjectively without applying any criteria and they often show generosity in scoring; consequently, the obtained results will be a series of unreliable and invalid scores which are not truly representative of the learners' actual ability. (However, there may be few language teachers who, after a long time of experience, use their intuition to score the learners' performances subjectively. They are an exception though.) Therefore, in order to obtain better results including more reliable and objective scores in testing speaking, it is essential to utilize a series of criteria to score oral language proficiency. As Pray mentions, "Oral-language assessments must measure the essential elements of knowing a language, not just lexical knowledge. This includes the ability to produce new utterances and recombine forms to represent ideas, events, and objects on an abstract level, to produce forms of the language they have never heard before, and to demonstrate mastery over the general functions of language such as syntax, morphology, semantics, and pragmatics".¹ One concern of teachers is how to prepare reliable tests for measuring oral proficiency of the students and score their performance. To have a more reliable estimate of the students' oral language ability, using a checklist, will be very helpful. It will eliminate all those sources that threaten the stability of the test scores. The checklist can act as a blueprint to teachers who wish to assess their students' oral proficiency. It reminds them of the macro-skills as well as the specifications or micro-skills that should be included in testing oral proficiency.

9-Delimitations

Despite the promising results, this study suffered from a few problems. One shortcoming was related to our population which was

¹ - Pray. L, "How well do commonly used language instruments measure English oral-language proficiency?" *Bilingual Research Journal*, 29(2), (2005). P.405.

predominantly female; the results of the present study, therefore, might not be generalizable to the male population. Moreover, speaking skills, though emphasized, are overshadowed by other skills due to the lack of healthy environments to adequately practice or apply oral/aural skills. This results in a series of problems especially while conducting the interviews during the research.

It is also understandable that it is very difficult to design tests that get learners to improvise speech in the foreign language.

Moreover, when answers to a test are written, assessors can check them carefully at their leisure time; but speech flits past, and is very difficult to judge quickly, objectively and reliably. Recordings can be made; but this is liable to be prohibitively expensive and time-consuming. Ultimately, even if you agree on criteria for assessment, some testers will be stricter in applying them, others more lenient. It will be difficult to get reliable, consistent and objective assessment.

Bibliography

- 1-American Council on the Teaching of Foreign Languages. (1982). *ACTFL provisional proficiency guidelines*. Yonkers: ACTFL.
- 2-Farhady, H., Jafarpur, A., & Birjandi, P. (2001). *Testing language skill: From theory to practice* (9th ed.).Tehran: SAMT Publication.
- 3-Harris, D. P. (1986). *Testing English as a second language*. New York: McGraw Hill.
- 4-Heaton, J. B. (1990). *Writing English language tests* (3rd ed.). New York: Longman.
- 5-Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- 6-Lambert, I. (2003). Recording speaking tests for oral assessment. *The Internet TESL Journal*, IX(4). Retrieved 23 March 2008 from <http://iteslj.org/Articles/Lambert-SpeakingTests.html>.

- 7-Llosa, L. (2007). Validating a standards-based classroom assessment of English proficiency: A multitrait-multimethod approach. *Language Testing*, 24(4), 489–515.
- 8-Mislevy, R.J., Steinberg, L.S., Almond, R.G. (2002). Design and analysis in task-based language assessment. *Language Testing*, 19(4), 477-496.
- 9-Pray, L. (2005). How well do commonly used language instruments measure English oral-language proficiency? *Bilingual Research Journal*, 29(2), 387-409.
- 10-Ramanathan, H. (2008). Testing of English in India: A developing concept. *Language Testing*, 25(1), 111-126.
- 11-Sook, K. H. (2003). The types of speaking assessment tasks used by Korean Junior Secondary school English teachers. *Asian EFL Journal*, Retrieved 23 March 2008.
- 12- Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.
- 13- Ur, P. (1996). *A Course in Language Teaching: Practice and Theory*. Cambridge: Cambridge University Press.

Appendix

The Sample Checklist for Measuring Communicative Abilities:

Extracted from Farhady et al. (2001), Harris (1968), and Hughes (2003)

Scale I- Fluency:

- 5- Speaks fluently.
- 4- Speaks with near-native like fluency, pauses and hesitations do not interfere with comprehension
- 3- Speaks with occasional hesitations.
- 2- Speaks hesitantly and slowly because of rephrasing and searching for words.
- 1- Speaks in single word and short patterns, unable to make connected sentences.

Scale II- Comprehension:

- 5- Understands academic discourse without difficulty.
- 4- Understands most spoken language except for very colloquial speech.

3- Understands academic discourse with repetitions, rephrasing, and clarification.

2- Understands simple sentences, words; requires repetitions, slower than normal speech.

1- Understands very little or no English.

Scale III- Communication:

5- Communicates competently in social academic settings.

4- Speaks fluently in a social academic setting, errors do not interfere with meaning.

3- Initiates and sustains conversation, exhibits self- confidence in social situations.

2- Begins to communicate for personal and survival needs.

1- Almost unable to communicate.

Scale IV- Vocabulary:

5- Uses extensive vocabulary in any domain appropriately.

4- Uses varied vocabulary to discuss general topics and in special interests.

3- Uses academic vocabulary, some word usage inappropriate, slightly damages the message.

2- Uses limited vocabulary, constant use of one word.

1- Inadequate basic vocabulary.

Scale V- Structure:

5- Masters a variety of grammatical structures, almost no error.

4- Occasional grammatical errors but no problem with understanding.

3- Uses some complex sentences but lacks control over irregular forms.

2- Uses predominantly present tense verbs, constant errors interfere with understanding.

1- Severe errors make understanding completely impossible.

Scale VI- Accent:

5- Acceptable pronunciation, with few traces of foreign accent.

4- Speaks with few phonemic errors, but almost intelligible pronunciation.

3- Occasional errors necessitate attentive listening.

2- Constant phonemic errors make understanding extremely hard.

1- Severe problems make understanding almost impossible.