# Identifying the Rules for the Evaluation of English Language Speaking Skill: A Library Based Review

**Muhammad Mushtaq,** Lecturer in English, Govt Post Graduate College, Sahiwal, Pakistan. mushi121@hotmail.com
**Aamir Aziz\*,** Department of English, University of Sahiwal, Pakistan. aamiraziz@uosahiwal.edu.pk
**Muhammad Ramzan,** Lecturer in English, Government College University Faisalabad (Sahiwal Campus), Pakistan mr9535199@gmail.com
**Kalsoom Jahan,** Lecturer in English, Lahore Garrison University, Lahore, Pakistan.

**Abstract-** The main purpose of this study is to propose some principles for the testing of speaking skill. It focuses on exploring the types of speaking tests, characteristics of a good speaking test, administration of a speaking test, types and appropriateness of speaking tasks, analysis of IELTS speaking test and scoring of oral assessment. The data is collected from various sources such as web search engines, articles from different research journals and from different books as it is a library based research. This essay concludes that to develop speaking skill, appropriate measure to test the speaking skill is very important. For an assessor, it is very important to have knowledge of the criteria for the preparation of oral test. Moreover, selection of tasks to test speaking skill and scoring criteria of these tests can lead a test to a standard oral test.

**Keywords: testing of speaking skill; construction of speaking test; testing criteria; scoring of speaking test; testing tasks**

## I. INTRODUCTION

Language is a system of communication. There are two channels of communication that are written and spoken. Speaking plays an important role in the communication of any language. Rivers andTemperley (1945), in their book, A Practical Guide to the Teaching of English, opinedthat usually we spend more than 45% of our time on listening, 30% on speaking, 16% on reading and only 9% on writing. So, after listening, speaking is most important skill to learn. To become a proficient and fluent speaker, one needs to have good practice of this skill.

Language testing is the subfield of applied linguistics (Grabe, 2002). Testing of speaking skill plays a vital role to check the mastery of good communication, the main focus of language learning is to be able to communicate well because the world has become a global village. The main focus of testing of language skills is on the evaluation of first, second or any foreign languageto rate the learning of a particular language. The assessment of language is also utilized for workplace, citizenship, asylum and immigration purposes (Clapham& Corson, 1997; Inbar-Louri, 2008).The students, sometimes, can't communicate properly in speaking. The speaking activity of a test induces more problem and tension than any other component of language test. Students are examined one by one. After having waited nervously for their turn, they have to perform task in a short time. It seems appropriate that a learner could easily perform poorlyif the tester let the testing situation intimidate them. According to Heaton, (1975) testing the speaking ability is the most vital aspect of language testing. According to Kitao & Kitao (1996), listening and speaking tasks are used usually like completion of information gaps and role play in the process of testing communicative competence. According to Stephan Krashen(1978), the leaner must have self-confidence and good self-image only then he will be able to perform well in speaking skill. For this purpose, teacher can play avital role to enhance the speaking skill of the student by giving positive feedback. His role is very vital to find out such activities that can not only improve his speaking skill but also can assess his development well.

Testing of speaking skill is a very difficultto assess at all stages beyond the primary level of repetition and mimicry as it is too complex a skill to allow any trustworthy analysis to be made for the requirement of

objective testing.  It is possible for the people to generate practically all the accurate sounds but still they are not able to communicate their ideas effectively and appropriately.  People, on the other hand,can make a lot of errors both in syntax and phonology and yet they succeed in expressing their thought fairly clearly (Heaton, 1975).

Testing of speaking is not only to test how a person can utter words of foreign language but it is also the testing of how a person can convey his or her message meaningfully. That's why the notion of communicative competence is becoming more popular now a days (Bailey &Nunan, 2004; Canale, 2014; Harding, 2014; Morrow, 2018; Nation & Newton, 2008).  Role play and completion of information gaps are the most commonly used tasks for listening and speaking in the process of testing communicative competence (Bellack, Hersen, &Lamparski, 1979; Kitao & Kitao, 1996; Lynch & McNamara, 1998).

For a good speaking test, selection of different tasks has a very important place.  Selection of the test tasks should be according to the goals of learning.  Moreover, it should be communicative to the maximum. According to O'Malley's (1996), appropriate selection of the tasks for the tests and determination of evaluation criteria is a major challenge.  Similarly, Brown (2004) discussed two major challenges in speaking assessment: (1) the interaction of listening and speaking (e.g. the use of much clarification) can make it difficult to treat speaking apart, (2) the speaker's strategy to dodge certain form to convey meaning may make it difficult for test makers to design a solid elicitation technique (one that can result in the expected target form).

Speaking is, doubtlessly, one of the most complicated skills to assess (Chou, 2018; Correia, 2016; Guettal, 2008; Levelt, 1994; Morrow, 2018; Rychtarik, 2014; Schmidt, 1992).  It is neither desirable nor possible to separate the skill of speaking from the listening skill in many tests of oral production.  In normal speech situations, it is clear that the two skills are interdependent.  It is not possible to hold any meaningful conversation unless you understand what is being said and at the same time without making oneself understood.  This very interdependence, however, of the listening and speaking skills increases the complexity of any serious endeavor to analyze specifically what is being tested.  Therefore, the assessor is put under great pressure (Heaton, 1988). Moreover, as the spoken language is temporary, it is not possible without a tape-recorder to use such procedures as in the checking of test at leisure.  The tester of an oral production is working all the time under great pressure, making subjective judgments as swiftly as possible. Even though speech samples can be recorded during a test of speaking, by itself the tape recording is inadequate to give an correct means of checking or reassessing a score, since it cannot bring back the full context of the real situation, all of which is so necessary to any assessment of the communication that occur.

**Research Questions**
This study will focus on the following questions:
1.       What is meant by testing of speaking skill?
2.        Can a speaking test be integrated or not?
3.       What are problems in the administration of speaking test and how can it be tackled?
4.       What type of tasks can be used to test oral communication?
5.       What is the important features of computer based oral test?
6.       Which way of scoring a speaking test be more effective?


II.    METHODOLOGY

This research is library based and is qualitative as well as descriptive in nature.  That's why the material has been downloaded from online sources for this study.  Different research engines were utilized to collect data. Essays, blog posts, articles and books were consulted for this essay from Google and Google Scholars as it is the most widely used web search engines on World Wide Web.  Realizing the significance of databases due to the increasing use of it by researchers, databases such Foreign Language Assessment Directory (FLAD), MLA International Bibliography,  Journal Storage popularly Known as JSTOR, Academic Research Complete were consulted to obtain material for this essay.

To get some particular information from research journals, websites of Clarivate Analytics (Master-Journals List) and SJR (Scimago) were visited to find suitable journals about speaking skills assessment. But these databases do not provide even a single journal concerned with speaking assessment, speaking testing, speaking evaluation etc. Some journals from the fields of general linguistics and ELT were consulted. Among them include Applied Linguistics, TESOL Quarterly, Computer Assisted Language Learning, The Cambridge Guide to Second Language Assessment, Language testing etc. The reason for selecting journals from these two sources is that they are recognized being most credible and authentic sources of indexing academic research journals.

To search relevant information from different databases, web search engines and journals keyword technique was utilized. For this purpose, some keywords covering the topic of the essay were determined which included some general and some specific words or terms. Among general terms included speaking, testing, evaluation, assessment, tasks, scoring, rubric, assessment, testing, principles, criteria etc. Then, these terms were joined together with one another to make them specific to the topic of the essay like testing speaking, speaking skill test(s), testing speaking skills, principles for devising a speaking skill test, assessing speaking skills, criteria for the assessment of speaking skills, speaking test tasks, scoring rubrics etc.

These terms and keywords were typed in the search bars of different web search engines, journals and databases in two different ways. First of all, they were typed word by word in search bars and then they were retyped adding (+) like (speaking+skill+tests+principles+criteria+tasks+rubrics+scoring)

III.    DISCUSSION

## 1.    Testing of Speaking Skill

Testing of speaking skill is not an easy task for the testers to test because most of the times the oral tests are for the time being and transitory unlike writing which may be viewed again.  This is the reason that most of the times, speaking test is marked subjectively.  There is another problem that students have to face is that they due to pressure of assessment and anxiety, they usually are not able to perform well even if they are very good in communication.  Sometimes, the personality and attitude of the assessor also effects performance of the testee.

In the testing of speaking skill, there are two things that are very important i.e. coherence and cohesion. Sometimes a student may be good at grammar and structure but not at communication and vice versa. As Heaton (1975) opined that it is possible for the people to generate practically all the accurate sounds but still they are not able to communicate their ideas effectively and appropriately.  People, on the other hand, can make a lot of errors both in syntax and phonology and yet they succeed in expressing their thought fairly clearly.

## 2.    Integrated Test Tasks

For speaking test, integrated tasks can also be chosen. There are some researchers who examined integrated test speaking tasks.Lee (2006), among them, focused his attentionon the TOEFL-iBT speaking component which incorporatesthe tasks of independent speaking, reading-listening-speaking tasks and listening-speaking tasks and came to know that an increase in the quantity oftasks favorably affected the score reliability. Iwashita et al. (2008)studied the features that differentiate performance on the component of speakingof the TOEFL-iBT and came to know thatphonology, fluency and linguistic resources, all importantly affectedspoken performances evaluation. Swain et al. (2009), investigating the strategic behaviors,associated with the section of speaking of the TOEFL-iBT, proposedthat strategies of reported test-taking fell into five categories: cognitive, metacognitive,communication and affective approach. They foundthat the integrated tasks especially those who were providing both listeningand reading input, elicited more in the use of strategy than independent tasks.

Huang and Hung (2010) made a comparison ofa speaking-only test task and a reading-to-speak test task in terms of the anxiety that each induced and theperceptions that they produced and showed that although the two tasks of oral test incurred a comparable amount of the anxiety of test, the participating learners performed an overwhelming preference for the task of reading-to speak test.

### 3. Administration of Oral Test

Administration of speaking test is a major problem because testing of speaking takes a lot of time of testers. For a tester it is very difficult to take test of a lot of learners due to the shortage of time. According to Heaton (1975), it is not difficult to understand the huge problems related to a test situation in which hundreds of students have to be examined by a few number of examiners, each learner being tested for a time of ten or fifteen minutes. Although, the utilization of language laboratories for such type tests has made it feasible in some cases to manage more reliable spoken tests to a large number of students, the authentic scoring of the tests has not been so easily solved.

### 4. Test Taker's Characteristics

According to Bachman and Palmer's (1996) model of language use, there are five characteristics of a test taker as under:

i. Language knowledge
ii. Topical knowledge,
iii. Personal characteristics,
iv. Affective schemata
v. Strategic competence

And among them there should be the interactions that constitute the essentialgroundunderlies the use of language and performance of language test.

Among these characteristics, affective schemata represent "the basis on which language users asses, consciously or unconsciously, the characteristicsof the language use task and its setting in terms of past emotionalexperiences in similar contexts" (p. 65). As Bachman and Palmer pointed out, these schemata function toeither inhibit or facilitate the access of language user to the full spectrumof language strategies and knowledge available to a test administrator ashe or she sets out to engage a language task.

### 5. Tasks for Speaking Test

There is a variety of speaking skill tasks which may provide with particular information about the testees' speaking capability. "When we are assessing speaking, we guide the examinees' talk by the tasks that we give them. These outline the content and general format of the talk to be assessed and they also provide the context for it" (Luoma, 2004: 29). Brown (2004:141) gives five types tasks of assessment. Brown's proposed categories are given below but the tasks in each category come also from the descriptions by Heaton (1988), Hughes (1989) and O'Malley (1996).

### 5.1. Imitative Tasks

Imitative tasks employ repeating a small stretch of language with major emphasis on pronunciation. This type of test is used by the tester if s/he is not interested in the competence of testee in understanding and conveying meaning or in getting involved in the conversation of interaction. The competence assessed through these tasks is that of purely phonetic, prosodic, lexical and grammatical etc.

### 5.2. Intensive Tasks

There are certain tasks that are usually included in intensive tasks:

### 5.2.1 Reading aloud

According to Heaton (1988) and Hughes (1989), reading aloud activity utilization may not be appropriate due to the reason of its variation in processing spoken input from that of written one. However, pronunciation, stress pattern and rhythm may be checked by using the technique of reading aloud. According to Brown (2004), for some other tasks that are more communicative, reading aloud can be used as companion activity by the test taker.

### 5.2.2 Directed response tasks (e.g. response to a recorded speech)

Speaking task that is the most popular due to its practicality and excessive use in lab is directed response despite its mechanical and non-communicative nature. DRT is advantageous to draw out a certain form of grammar or transform a sentence which needs minimal processing (micro skills 1-5, 8 & 10) (Brown, 2004).

### 5.2.3 Sentence or dialogue completion

Heaton (1988) warns that the tasks of this type may give illogical flow of conversation only if the dialogue completion or sentence activity is done in lab (which is what normally administered). Thus, this mode will probably be useful only for measuring test taker's micro skillgiving the right language chunks and other features of pronunciation. Though, a more responsive-type of sentence or dialogue completion, as Brown (2004) exemplifies, which may truly be free of above said shortcoming and save us from the risk of assessing

a test taker's competence as insufficient caused by misunderstanding of aural activity in processing the input. SDC assists us to measure the competence of speaking apart from its interrelatedness to listening.

### 5.2.4 Translation up to simple sentence level (interpreting-game)

As Hughes (1989) gives the description, interpreting may encompass the conductor of test acting as a native speaker of test taker's first language and interpreting the utterance into English by the test taker. It is assumed that as speaking is negotiation of intended meaning (O'Malley, 1996), interpreting-game can be utilized to measure the competence of test taker in conveying his message into the foreign language (Brown, 2004).

### 5.2.5 Picture-cued task (including simple sequence)

To elicit description, pictures are the most convenient mode to adopt(Hughes, 1989). Besides describing positions, location,comparison and order of events,a picture that have detailed information to discuss may be used to elicit the competence of test taker in telling directions, a plan and even opinions (Brown, 2004).

### 5.3 Responsive Tasks

These types of tasks are the tasks in which small dialogues are used as well as reaction or response to spoken prompts such as requests, comments and simple greetings etc. there are also some other popular responsive tasks that are given below:

### 5.3.1 Questioning and answering

According to Brown (2004), at responsive level, questions appear most probably to be referential (as compared to intensive, display question). It is required on the part of test taker in referential question that he should give such type of question that can produce meaningful language in reply/answer. Such type of questions may ask an open-ended response or a counter-question directed to the interviewer (Brown, 2004).

### 5.3.2 Paraphrasing

There can be written or aural input in oral Paraphrasing but aural input is more preferable than written input. A paraphrase as an assessment of speaking skill should be administered with caution because instead of assessing speaking skill of the learners, short term memory and listening comprehension would mistakenly be judged.

### 5.4. Interactive Tasks

Interactive tasks are the tasks that are based on larger dialogue on interactional and transactional discussion. These tasks are as given below:

### 5.4.1 Interview

Interviews are widely utilized, even with their limitations e.g. the behavior of interviewers may affect testees' performance because of familiarity (O'Sullivan, 2002) and personality (Berry, 2007). The interviewers may also bias the interaction by controlling and leading the discussion (Lazaraton, 1996; Young &Milanovic,1992). There are different types of interviews among which are face-to-face, one-on-one or two-on-one. Each of these types have its own advantages and disadvantages. Among these types, two-on-one interview may be time saving and scheduling and provides interaction that is authentic between two test takers, though it can be risky due to one test taker's domination over the other. According to Canale (1984), to conduct an oral proficiency test there are four main steps that needs to be followed: (i) Warm up: a small talk about identity, origin and the like; (ii) Level-Check: description without any break, wh-questions, reading passages aloud, short guided role plays, and explanation how to make or do something etc.; (iii) Probe: questions related to some field of life and (iv) Wind-down: easier questions containing test taker's feeling about the interview.

Like Canale, Hughes (1989) put forwarded eleven rules to conduct an interview: (i) construct tests of speaking as long as possible, (ii) include as large a specified substance sample as it is possible in available time, (iii) prepare the spoken test with great care, (iv) maximum 'fresh starts' should be provided to the test takers, (v) select the interviews with care and coach them too, (vi) have help from some other tester, (vii) choose such tasks and topics that can't create any difficulty for the students in their mother tongue, (viii) always conduct these interview in a room that is peaceful, (ix) get maximum information from the test taker (x) the interviewer should talk as minimum as possible

How to score an open ended response is a great challenge. To ensure reliability, creating a workable, consistent scoring system has be one of the great challenges in designing an interview as methods to assess speaking (Brown, 2004). According to O'Malley (1996), there are at least two solution to the problem of how to score an interview that are analytic and holistic scoring rubric. There is another alternative according to him that is rescoring of performance with the help of recorded sound on the tape.

### 5.4.2 Drama tasks

A division of drama like tasks has been made by O'Malley (1996) in sub-types i.e. simulation, role play and improvisations. Each varies in scripting and preparation. Improvisation presents very little opportunity for test taker to organize the situation and may spur creativity in language use. A slightly longer time is provided in role play for the test taker and he can manage what to say. Meanwhile, planning and decision making is required in simulation. A real world socio drama may be involved in simulation which is the zenith of speaking competence. An unpredictable response may be caused in drama like task such as interview. That's why some care should be taken to handle this type of task like that of interview.

### 5.4.3 Discussion and conversations

As far as discussions and conversations are concerned, Brown (2004) gives somewhat similar complications in terms of response predictability hence, consistency of the scoring to the tasks of interview and drama. Testers look like to select tasks of this type as informal assessment to produce and observe testee's performance in: (i) starting, sustaining and ending a topic, (ii) controlling, getting attention and interrupting, (iii) questioning, paraphrasing and clarifying, (iv) comprehension signaling (such as nodding), (v) using appropriate patterns of intonation, (vi) using kinesics, facial expression, body language and eye contact, (vii) being formal, polite and some other sociolinguistic situation.

### 5.4.4 Games

There are a lot of games for this purpose that can't be counted but practically all games which can prompt spoken language objectively can be utilized as task of informal assessment for speaking. Brow (2004) warns us that the use of games for speaking test may go beyond assessment and adds that a specific perspective requires to be carried out in order to keep it in line with the principles of assessment. Brown (2004) mentioned some games such as, crossword puzzle, tinker toy, predetermined direction map, information gap that can all fall under information gap activities by O'Malley's (1996) standpoint as he elaborates that information gap are the activities in which a learner is given with information that the other (e.g. his pair) does not know but need to. An activity of information gap contains collecting complete information of building restructure, find the difference between two pictures or sequence a picture into order. To score an information gap activity, O'Malley (1996: 83) suggests test maker to consider the speaker's "accuracy and clarity of the description as well as on the reconstruction." (O'Malley, 1996: 83)

### 5.5 Extensive Tasks (Monologue)

The monologues are types of tasks in which speaker has to use language for long time and need a lot of preparations. The terms are self-explanatory and some may in fact have some characteristics with some types of already explained only with longer and broader scope of the use of language.

### 5.5.1 Speech (oral report/presentation)

To present reports, papers, or models in settings of education is a common practice. That's why, reports and oral presentations can be utilized to assess the skill of speaking holistically or analytically. But it should be kept in mind that it is best to use for advanced or intermediate levels that are focusing on delivery and contents (Brown, 2004).

### 5.5.2 Picture-cued story telling

Just like the limited version, the main concern of using pictures or the series of pictures at this level to make it a stimulus for purpose to produce longer stories. For this, a series of six picture with an appropriate amount of details about the setting and the character will be adequate to test (Brown, 2004).

### 5.5.3 Retelling a story, news event

Retelling a story or a news event can also be a good task to assess oral skill because retelling a story is different from paraphrasingand takes longer stretch of discourse with varied, preferably narrative, genre. The main emphasis is usually on meaningfulness of the relationship of events within the story, fluency and interaction to audience (Brown, 2004).

### 5.5.4 Translation (extended prose)

A paragraph or more than a paragraph mostly in written form is given in the native language of the test taker that is to be read prior to the interpretation of text with ease in actual testingis used in this type of task. The text can include a procedure, dialogue, synopsis complex directions, or a play script. Caution should be made concerning with this type of task because this particular type requires a skill not intended for every speaker of a language. Therefore, if this type is to be used a degree of confidence should be made sure (Brown, 2004).

## 6.    Computer Based Tests

According to Brown (1989) the best way to proceed is that if a program has expertise and resources, making placement tests during a program appear as best practices. Bernstein, Van Moere, and Cheng (2010)

recommended two major categories of computer based tests to measure speaking skills i.e. automated and semi-direct tests. The later makes use of computer to distribute tests and collect testees' response yet; the rating part is carried by human raters with the help of predefined rubrics. Compared to the semi direct speaking skill tests, automated tests make use of computer not only for test development as well as administration, but it also utilizes computers in scoring process. Examples of computer based speaking tests contain the Computerized Oral Proficiency Interview (COPI; Malabonga& Kenyon, 1999), TOEFL iBT (Zechner, Higgins, Xi, & Williamson, 2009), and Versant English (Pearson, 2011).

Fully automated-computerized tests have many advantages (Eyckmans, Van de Velde, Van Hout, & Boers, 2007; Qian, 2009; Van Moere, 2010) e.g. these tests do not need on-site testers. For, these tests can be conducted at different places by test takers from different regions. Secondly, the same test may be given to many test takers. Thus, it economizes administration costs.A number of studies by Davidson and Coombe (2012), Louma (2004), Neri, Cucchiarini, Stik and Boves (2000), Pennington (1999) etc. have also highlighted the significance of the assessment of speaking skill through computer based procedure.

## 7.    Important Features of IELTS Speaking Test

The IELTS speaking test has three parts and takes between 11 to 14 minutes. In the first section, (4 to 6 minutes) the candidates are asked to talk about themselves, their interests and to answer the questions on familiar areas of topic. In the second section, 3 to 4 minutes, the candidates discuss about the proposed topic on a cue card. Between one and one and a half minutes, the candidate must speak with a few questions of examiner at the end. The candidate has the opportunity for 4 to 5 minutes to discuss issues of a more abstract nature in the third section. These topics or issues are thematically interlinked inthe part two. For example, if part two question requires for a description of a favorite teacher, part three will be based on a discussion of issues that are related to education. If part two is related with an interesting place or holiday, then part three will also be concerned to travel or tourism, and so on.

The IELTS test scores are arranged in bands from0 to 9, and 9 being the highest. There are basically four components; listening, speaking, reading, and writing. Listening and reading are marked clerically and their particular bands are divided into two. It is possible,for example, to achieve a score of say, 4.5, 5.0, 5.5, 6.0, 6.5, and so on. While the testing of speaking and writing are marked by examiner and are documented as whole bands only: 4.0, 5.0, 6.0, etc. In practice, this is meant that a learner has the opportunity to get better by one or more whole bands in speaking and writing and as a consequence of it, work in these areas  can have a high level of pay-off. This presents students a big advantage to perform well in the section of speaking and with limited available time for IELTS preparation, hard work on the speaking part has a high level of possible return.

If the students want to answer examiner's questions, they should follow the following framework. There are two considerations. Firstly, the criteria of marking suggest that hesitation will be punished and as anyone knows who has attended an interview that it is easy to slip into incomprehensibility if you are not able to assemble an argument rather quickly. Secondly, the test of IELTS clearly functions in an academic context, so a student could expect that calling up an academic perspective would be appropriate. In other words, by approaching the questions given by the examiner from a theoretical perspective, a student could justifiably think that they were contributing properly within the parameters of the discussion.

A set of (IELTS type) questions is given to the students and then they are asked to think about how to answer them. Such as; (a) what is the advantage of tourism?, (b) do you believe that the use of technology in education at school level is a good idea? (c) do you believe that men and women should be regarded equal at the place of work?, (d) do you believe it is essential to have nice friends? (e) are you confident about the future of your planet?

After that the students are prompted to think about the questions from a range of perspectives like cultural, economic, historical, linguistic, philosophical, political, psychological, and social. Then, the students are paired up and assigned to use these perspectives in their discussions for at least one minute. Aim of this practice is to facilitate testees to feel at ease in discussion, and use the said perspectives a spur for their ideas. It is possible, at this point, for the students to feel growth of confidence. After that at an appropriate time, the students are asked to interview their class-teacher, who then provides them with some model answers. Such as:

Well! From a social viewpoint I believe that tourism may contribute to develop understanding at international level and promote contact among people from different nations. On the economic level, there are a lot of advantages for the leisure business as well as hotels with encouraging effects on employment. Philosophically, it may make our planet a more tolerant place,

### 7.1. Comparison of IELTS and TOEFL Tests

IELTS and TOEFL are the two most popular tests in Pakistan. Though both have same purpose that is to assess the proficiency of candidate in foreign language but both adopt different pattern for different language skills. Same is the case in speaking skill too. Current practices of testing of speaking skill have been observed in IELTS and TOEFL as given by are discussed in the table below:

### 7.2 Comparison of current practices in IELTS and TOEFL

| TEST | IELTS | TOEFL |
|---|---|---|
| **Mode** | Informal and face to face with an examiner. It is recorded for the candidate. s/he can ask for a remark if requires | |
| **Duration** | 11-14 minutes. Time is controlled by the evaluator. However, it is neither less nor more than 11-14 minutes | 20 minutes/4:30 hrs |
| **Test Division** | **3 Parts**<br>1. Question-Answering 4-5 minutes. 12 questions are asked based on three topics i.e. yourself, your country and your life<br>2. It is called "The Long Turn". It is about a face to face talk with the examiner for 3-4 minutes. Topic card is given to for 1 min to prepare for 1-2 min talk about yourself/smth happened in your life.<br>3. discussion for 4-5 minutes with examiner on the topic chosen in part-2 | **3 Sections**<br>1. Independent Speaking: on personal ideas, opinion or experience. 15 seconds are given for preparation and 45 seconds for the response. There are 2 tasks for this section<br>2. Speaking in response to what you have listened/read. It integrates speaking with reading and listening. **Time in Seconds:** a. Listening: upto 60 b. Reading: 45 c. Preparation: 30 d. Response: 60<br>1. speaking in response to listening only. It is integrated with listening. Students are answer questions in response to an audio recording **Time in Seconds:** 1. Listening: 60-120 2. Preparation: 30 3. Response: 60 |
| **Marking Criteria** | 1. Fluency & Cohesion 25%<br>2. Vocab 25%<br>3. Grammar 25%<br>4.Pronunciation 25% | **0-4 Score/Scale marking**<br>1. Weak: 0-1<br>2. Limited: 1.5-2<br>3. Fair: 2.5-3<br>4. Good: 3.5-4 |
| **Topics/Contents** | Work, study, home, hometown, art, birthday, childhood, clothes, computer, daily routine, dictionaries, evenings, family & Friends, flowers, foods, going out, happiness, hobbies, internet, leisure, music, neighbours&neighbourhood, newspaper, pets, reading, shopping, sport, TV, weather | |
| **Tasks** | 1. Question/Answers<br>2. Face to face talk<br>3. Discussion | **Tasks for Sec. 1**<br>1. choose topic from one option<br>2. choose from two options<br>**Tasks for Sec. 2**<br>1. read a short passage<br>2. listen to an audio and answer |

Identifying the Rules for the Evaluation of English Language Speaking Skill: A Library Based Review

| | | question based on what you have read and listened | |
| | | **Tasks for Sec. 3** 1. Explain a problem 2. Explain a lecture . | |

*Comparison of IELTS and TOEFL*

## 8.    Debates on analytical and holistic scoring rubrics

The discussion between analytical as well as holistic scoring rubric for speaking assessment has been well debated (Bachman, 1988; Douglas & Smith, 1997; Fulcher, 1997; Underhill, 1987; Weir, 1990). Holistic scoring is frequently preferred to analytical for verbal assessment tests to assess the general communicative efficiency of the testees (Weir, 1990). Plus, holistic scoring assures efficient scoring and makes score reporting process easy, and minimizes cognitive burden on the evaluators.

On the other hand, holistic scoring poses some problems i.e. (a) the relative value of sub-features of scoring rubric are implied and may also be distinctive. Constituents of the holistic rubrics may be valued differentially by different raters relying on their experiences, backgrounds and perceptions of how a strength or weakness affects the communicative quality in a specific assessment framework (Brown, 1995). Another difficulty with holistic scoring process is concerned with the explanation of scores. According to Weir (1990), the usual performance accounts of all holistic score levels may not work for testees with different performances on all components. In TAST, a single holistic score may be enough for decisions making about examinees' weaknesses and strengths.

Score reports, based on holistic scoring process, might provide with limited information about testees performance. Whereas those based on analytic scoring capture more diverse profiles.

For, analytic scores can ensure diagnostic information about the testees with diverse profiles (Bachman &Savignon, 1986), providing the opportunity of producing a combined 'holistic' score with suitable. A score report produced from the analytic scoring process has thus the ability to work for admission and placement purposes. Bachman and Savignon (1986) say that the ability of speaking is a multi-componential feature therefore its rubrics need to be developed in terms of components to mirror current models of communicative language ability.

Analytic scoring approach has its problems also. Potential scoring in-consistency due to cognitive burden on scorers, complexity in determining the directions in analytic rubrics and getting scorers well attuned are most commonly mentioned ones (Douglas & Smith, 1997; Underhill, 1987).

Underhill (1987) observes the complexity which raters come across while evaluating the performance of testees on different criteria. Douglas and Smith (1997) propose that holistic scoring on the Test of Spoken English (TSE) permits for more reliable scoring. In addition, scorers may experience more contented while working with a specific amount of 'fuzziness' through holistic rubric.

IV.    CONCLUSION

To sum up the whole discussion, it is a fact that a language is learnt to communicate and most of communication is through speaking.  So, speaking is very important skill to convey one's message to the listener and this skill can only be learnt and practiced well if the testing of speaking is effective.  The administration of a speaking test is a great responsibility for a tester because a good tester provides such environment to the participants that not only reduces their anxiety but also encourage them to perform well. Selection of the tasks for the oral assessment should be in such a way that can assess the students' abilities effectively.  Scoring of the oral test should be in such a way that there should be no chance of subjectivity or bias. In short, oral test should be prepared to check the speaking skill effectively.

REFERENCES:

1.  Bachman, L., & Palmer, A. S. (1996). Language testing in practice. Oxford, England:
2.  Oxford University Press.
3.  Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests.
4.  Language Testing, 27, 355–377.

5. Bailey, K. M. (2003). Speaking. In D. Nunan, *Practical English language teaching* (pp. 47-66). Singapore: Singapore.
6. Bailey, K. M., &Nunan, D. (2004). *Practical English language teaching: Speaking.* McGraw-Hill Companies.
7. Bellack, A. S., Hersen, M., &Lamparski, D. (1979). Role-play tests for assessing social skills:
8. Are they valid? Are they useful?. *Journal of Consulting and Clinical Psychology*, *47*(2), 335.
9. Berry, V. (2007). Personality differences and oral test performance. Frankfurt, Germany: Peter Lang.
10. Brown, J. D. (1989). Improving ESL placement tests using two perspectives. TESOL Quarterly, 23, 65–83.
11. Brown, H. D. (2004). *Language Assessment: Principles and classroom practices.* White Plains, NY: Pearson Education.
12. Brown, H. H. (2007). *Teaching by principles: An interactive approach to language pedagogy* (3rd Ed.). White Plains, NY: Pearson Publications.
13. Chou, M. H. (2018). Speaking Anxiety and Strategy Use for Learning English as a Foreign
14. Language in Full and Partial English-Medium Instruction Contexts. *TESOL Quarterly*, *52*(3), 611-633.
15. Clapham, C., & Corson, D. (1997). *Encyclopedia of language and education. Vol. 7, Language testing and assessment*. Dordrecht, Netherlands: Kluwer Academic.
16. Correia, R. C. (2016). Assessing Speaking Proficiency: A Challenge for the Portuguese EFL Teacher. *e-TEALS*, *7*(1), 87-107.
17. Davidson, P., &Coombe, C. (2012). Computerized language assessment. *The Cambridge Guide to Second Language Assessment*, 267-273.
18. Eyckmans, J., Van de Velde, H., Van Hout, R., & Boers, F. (2007). Learners' response behaviour in yes/no vocabulary tests. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), Modeling and assessing vocabulary knowledge (pp. 59–76). Cambridge, UK: Cambridge University Press.
19. Fulcher, G., & Reiter, R. M. (2003). Task difficulty in speaking tests. *Language testing*, *20*(3), 321-344.
20. Grabe, W. (2002). *Applied linguistics: An emerging discipline for the twenty-first century*. DOI: 10.1093/oxfordhb/9780195384253.013.0002
21. Guettal, I. (2008). *Introducing storytelling activities as a basic technique to develop students'*
22. *oral performance an experimental approach case study: Second year students of English at Batna University* (doctoral dissertation, Universite Mohamed Khider).
23. Harding, L. (2014). Communicative language testing: Current issues and future research. *Language assessment quarterly*, *11*(2), 186-197.
24. Heaton, J. B. (1975). *Writing English language tests: A practical guide for teachers of English as a second or foreign language*. Longman Publishing Group.
25. Heaton, J. B. (1988). *Writing English language tests (new edition).* London: Longman.
26. Huang, H.-T., & Hung, S.-T. (2010). Examining the practice of a reading-to-speak test task: Anxiety and experience of EFL students. Asia Pacific Education Review, 11, 235–242. doi:10.1007/s12564-010-9072-6.
27. Hughes, A. (1989). *Testing for language teachers.* Cambridge: Cambridge University Press.
28. IELTS speaking part 1 topics & questions. IELTS Liz. Retrieved on December 30, 2018 from http://ieltsliz.com/ielts-speaking-part-1-topics/
29. IELTS speaking part 2 topics & questions. IELTS Liz. Retrieved on December 30, 2018 from http://ieltsliz.com/ielts-speaking-part-2-topics/
30. IELTS speaking part 2 topics & questions. IELTS Liz. Retrieved on December 30, 2018 from http://ieltsliz.com/ielts-speaking-part-3-topics-2/
31. Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on
32. language assessment courses. *Language Testing*, *25*(3), 385-402.
33. Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second
34. language speaking proficiency: How distinct? Applied Linguistics, 29, 24–49. doi:10.1093/applin/amm017.
35. Kitao, S. K., & Kitao, K. (1996). Testing speaking (Report No.TM025215). (ERIC Document Reproduction Service No. ED398261)

36. Krashen, S. (1978). Individual variation in the use of the monitor. *Second language acquisition research: Issues and implications*, 175-183.
37. Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: the case of CASE. Language Testing, 13, 151–172.
38. Lee, Y. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. Language Testing, 23, 131–166. doi:10. 1191/0265532206lt325oa.
39. Levelt, W. J. (1994). The skill of speaking. *International Perspectives on Psychological Science*, *1*, 89-103.
40. Luoma, S. (2004). *Assessing speaking.* New York, Cambridge University Press.
41. Luoma, S. (2004). Assessing speaking. New York, NY: Cambridge University Press.
42. O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. Language Testing, 19(3), 277–295.
43. Lynch, B. K., & McNamara, T. F. (1998). Using G-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing*, *15*(2), 158-180.
44. Malabonga, V., & Kenyon, D..,M. (1999). Multimedia computer technology and performance-based language testing: a demonstration of the ComputerizedOral Proficiency Instrument (COPI). In M. B. Olsen (Ed.), Computer mediated language assessment and evaluation in natural language processing: Proceedingsof a symposium sponsored by the association for computational linguistics and international association of language learning technology (pp. 16- 23). New Brunswick, NJ: Association for Computational Linguistics.
45. Morrow, C. K. (2018). Communicative language testing. *The TESOL Encyclopedia of English Language Teaching*, 1-7.
46. Nation, I. S., & Newton, J. (2008). *Teaching ESL/EFL listening and speaking*. Routledge.
47. Neri, A., Cucchiarini, C., Strik, H., &Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, *15*(5), 441-467.
48. O'Malley, J. M., & Pierce, L. V. (1996). *Authentic Assessment for English Language Learner: Practical Approaches for Teachers.* White Plains, NY: Addison Wesley.
49. Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning*, *12*(5), 427-440.
50. Pearson. (2011). Versant English test: Test description and validation summary. Palo Alto, CA: Pearson Knowledge Technologies. Retrieved from http://www.versanttest.com/technology/VersantEnglishTestValidation.pdf.
51. Rychtarik, S. (2014). *Designing criteria to assess speaking skills* (doctoral dissertation).
52. Retrieved from https://www.academia.edu/7719276/Designing_Criteria_to_Assess_Speaking_Skills
53. Qian, D. (2009). Comparing direct and semi-direct modes for speaking assessment: affective effects on test takers. Language Assessment Quarterly, 6(2),113–125.
54. Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in second language acquisition*, *14*(4), 357-385.
55. Swain, M., Huang, L.-S., Barkaoui, K., Brooks, L., &Lapkin, S. (2009). The speaking section of the TOEFL iBT (SSTiBT): Test-takers' reported strategic behaviors (TOEFL
56. iBT Research Report No. iBT-10). Princeton, NJ: Educational Testing Service.
57. Test content. *TOEFL iBT.* Retrieved December 29, 2018 from https://www.ets.org/toefl/ibt/about/content/TOEFL prep online guides and tips. Retrieved December 29, 2018 from https://www.prepscholar.com/toefl/blog/toefl-speaking-scoring-system/
58. Van Moere, A. (2010). Automated spoken language testing: test construction and scoring model development. In L. Araújo (Ed.), Computer-based assessment of foreign language speaking skills (pp. 84–99). Luxembourg: Publications Office of the European Union.
59. Young, R., &Milanovic, M. (1992). Discourse variation in oral proficiency interviews. Studies in Second Language Acquisition, 14, 403–424.
60. Zechner, K., Higgins, D., Xi, X., & Williamson, D. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. Speech Communication, 51, 883–895.