

## Constructing Communicative Spoken Language Tests for the Language Learners in Pakistan

<sup>1</sup>*Dr. Khurram Shahzad*

<sup>2</sup>*Azhar Habib*

<sup>3</sup>*Farhana Yasmeeen*

<sup>1</sup>*Assistant professor - National University of Modern Languages, Islamabad*

<sup>2</sup>*Lecturer - National University of Modern Languages, Islamabad*

<sup>3</sup>*Assistant Professor - Lahore Garrison University, Lahore*

---

### Abstract

The assessment of speaking skills calls for a systematicity in order to assure fairness and impartiality in assessment. For the Pakistani students, English is doubly important as it ensures better grades during education and great chances of success in the job sector. In either case, oral proficiency is considered to be a fundamental requirement to access knowledge, and it supports personal and professional mobility in an ever-growing worldwide community. The usefulness of the spoken language test is based on how the test developers construct the test (Luoma, 2004). Using observations and semi-structured interviews, the data have been collected from various language institutes of Rawalpindi and Islamabad. Having analyzed the data, the researchers find many loopholes in the test, and hence the researchers use Bachman & Palmer (1996) model of spoken language test to develop the spoken test. It has three phases, which are: design stage, operationalization stage and administration stage. Test development stage includes a) construct specification that consists of the purpose of the test, brief description of the test, characteristics of the testees, trait to be judged or measured; b) test task specifications – here we should know about rubrics to be used and administration plan; c) assessment specifications – analysis of the results. Only a well-constructed communicative language test can describe the true proficiency of the learners.

**Keywords:** Assessment, speaking skills, communicative test, Pakistan

---

*Author's E-mail: kshehzad@numl.edu.pk*

### Introduction

Evaluation of speaking skills plays a very important role in the teaching and learning process. It includes the whole process of

designing, administering and marking tests as part of the curriculum. The role of assessment can be selective, for tests as instruments measure whether the learners have mastered what they are supposed to master. This role can also

bediagnostic, for it identifies the areas where learners are weaker and require further assistance. Last but not least, evaluation can have a predictive role with reference to the continuance of the learners' learning process and study-career in future. Since the curriculum comprises a blend of all these roles, evaluation has to match up with them (Airasian, 2005).

As far as English language teaching (ELT) is concerned, evaluation is also very central to the teaching / learning process. It gives meaning and substance to all that goes by the name of language teaching. The purpose of assessment is to improve students' learning and to test their performance (McNamara, 1996; Bachman, 1990). Language assessment gives a complete picture of learner's progress and performance, but, to some extent, it is wide-ranging and multifaceted as human language is. It should always be viewed as information that should be used to help improve learners' achievement. Assessment is a kind of process that is used to gather the evidence of learners' performance over a period of time (Airasian, 2005).

Contrarily, inadequate assessment is not only useless but harmful as well. It may mar the very language teaching practices and preclude the achievement of the course objectives (McNamara, 1996). Hughes (2003) is of the view that assessment helps improve instruction. It is only achieved when assessment itself is authentic and takes into consideration

the objectives of the study programme.

In fact, in a country like Pakistan which has a history of language teaching, very little attention has been paid to the notion of evaluation. Most of the time, learners are evaluated on an annual basis in a way which is nothing but mechanical and repetitive (Shahzad, 2018). This mode of assessment, over the years, has become so predictable that learners can decide before-hand what to prepare for the exam and how to get through. However, the success achieved this way is by no means an indicator of language proficiency.

Most of the assessment is done in the written form which just highlights one of the aspects of linguistic competence to the exclusion of the rest. Thus, it becomes almost impossible for the teacher to discriminate between the ability to cram certain language items or the real talent of the learner to use language in real life situations (Shahzad, 2018). To the dismay of the researchers, language assessment in Pakistan conspicuously lacks the pragmatic / practical component. This pragmatic / practical component consists of the use of oral skills, and these are the most neglected skills in ELT in Pakistan. We know that the language is primarily oral and secondarily it is written. So there is too much emphasis on the written aspect which is to put the proverbial cart before the horse. It is not just a simple violation of a sequence of skills; rather, it disturbs the whole process of learning a language.

More often than not, it is one of the main reasons that learners are unable to communicate in English language. On the contrary, we see a number of learners who have passed the subject of English with good marks, yet are unable to communicate in English. This situation looks more hopeless when we deeply inquire into their written competence also. We come to know that even their written competence is not due to the actual proficiency of language (Shahzad, 2019).

For the first time in the history of ELT in Pakistan, many institutes such as National University of Modern Languages (NUML), Pakistan Institute of Modern Languages (PIML), and Federal Institute of Modern Languages (FIML), etc. have come into being which recognize the pragmatic proficiency in language, i.e. oral competence. The teaching and assessment of the written component continues side by side; however, it was not done so at the cost of the spoken component as was the case before the inception of these institutes and as it still goes on in other universities of Pakistan. However, the researchers have noticed that still these institutes have miles to go before they could claim their due place in the country. The researchers' focus is on the system of evaluation and particularly the oral evaluation about which certain immediate steps should be taken to improve the exam system of these institutions. The above mentioned language institutes have been selected for the study, and the researchers believe that the results of the study

would bring innumerable benefits for the teachers and the learners alike.

### **Statement of the Problem and Rationale**

The assessment of speaking skills calls for a systematicity in order to assure fairness and impartiality in assessment. For the Pakistani learners, English is doubly important as it ensures better grades during education and great chances of success in the job sector. In either case, oral proficiency is considered to be a fundamental requirement to access knowledge, and it supports personal and professional mobility in an ever-growing worldwide community.

The study aims at investigating and assessing the methods of evaluation of oral proficiency, i.e., speaking skills used at these institutions. These institutions are known for their language teaching at multiple levels and through diverse means. Having studied the assessment processes at these institutions thoroughly, an urgent need was felt to ameliorate the assessment of oral proficiency. The evaluation methods used for measuring oral proficiency and the criteria against which this measurement is done is critically examined. In the course of the research, the researchers noted that the teachers do not go by a set, well-defined criterion as proposed by any ELT expert. This brings in the risk of going subjective and individualistic.

Given such relevance of oral

proficiency, the assessment process must be reviewed as it is very critical in the general practice of language teaching in Pakistan. The researchers investigated what the curriculum of the institutions is as far as the desired oral proficiency for the learners was concerned and what criteria were being followed during evaluation. Furthermore, the study also investigated whether the criteria being followed at the university matched up with the criteria built up by the Council of Europe (Common European Framework of Reference, CEFR) for the evaluation of oral competence. Latter, two oral tests were developed and conducted taking into consideration Bachman and Palmer's model.

### **Objectives of the Study**

The objectives of the study are as follows:

- To assess the contemporary methods used for evaluating the spoken language of language learners at different language institutes in Rawalpindi and Islamabad.
- To study the conduct of teachers during the evaluation of spoken language testing.
- To design and administer a sample oral test (SOT) with a rating scale built by Council of Europe (CEFR) to assess learners' oral proficiency in English.

### **Research Questions**

1. What is the present state of assessment of oral proficiency

practised at various institutions in Rawalpindi and Islamabad?

2. What is the need of a well-defined criterion to be used for spoken language testing?

3. How can a sample oral test be developed to restructure spoken language assessment in the light of the findings of the research?

These questions set and define the scope of this research which primarily aims at working out their answers.

### **Methodology**

This research was primarily an evaluative study, using a qualitative form of enquiry. The population for the study comprises a total number of thirty five participants (teachers). The teachers selected for the study had varied experience of teaching English language.

As regards tools, the study made use of observation sheet and semi-structured interviews. Semi-structured interviews were conducted. The interview contained both open-ended and closed-ended questions. To supplement the findings of the interviews, an observation sheet was used which aimed at studying and elucidating the techniques employed and time spent to evaluate the oral proficiency of the students, etc.

Taking into consideration the model of communicative language testing presented by Bachman (1990)

and the theoretical framework for the construction of spoken language test outlined by Bachman & Palmer (1996), the researchers used them to evaluate the spoken language tests performed at the different institutions.

### **Sample Oral Test (SOT)**

Sample oral test (SOT) was constructed taking into consideration Bachman's (1990) model of communicative language testing and test development procedure based on Bachman & Palmer's (1996) testing framework.

SOT was administered to the learners of language learning; moreover, CEFR analytical rating scale was used as a model to measure the spoken language of the testees. Sample oral tests were piloted with eight diploma level students. One SOT had five test tasks on it; it was administered to four learners. The second sample oral test was administered to two pairs of learners. It had only three test tasks. These tests were constructed to continue for 13 to 15 minutes. The researchers had CEFR analytical rating scales with them and each time they used them to mark the students. The piloting was done to check the validity and reliability of the SOTs, and certain changes were made before the final SOTs were developed. The researcher along with his partner evaluators started the sample oral tests in a cozy and comfy room having a table to put the testing material on and (a) comfortable chair(s) for the testee(s). The learners were given the oral test having instruction and the criteria which were used to measure

their oral proficiency on it. In this way, the complete procedure of the test was described to them. The test started with the first oral task that was interview which provided the testers an opportunity to ask questions on the bio-data of the testees, their likes and dislikes, hobbies, travelling, town, future plans, etc. It worked as a warm up activity and continued for three minutes. The testees felt relaxed. The next task on the test was picture description. The testees were given one minute for preparation and then asked to speak as much as they could. This activity was to be completed in three minutes. The third test task was about expressing an opinion. The testees were given an option to choose a situation or topic and express their opinion. This activity was also to be completed in three minutes and it did. The fourth task on the test was a role play. The testees were given situations to choose from and act out the role play with the interlocutor. While doing this, the main role had to be done by the testees. It also took three minutes to complete the test task.

Hence, all the eight testees were given the test which was conducted. The testers acted as interlocutor and examiner. With some testees, the one tester acted as an interlocutor and the other two acted as an examiners only and paid full attention to listen to the language produced by the testee and his or her conduct. In this way, the testers played their roles while conducting the oral test.

## **Literature Review**

Assessing speaking skills is characteristically different from those of other language skills, and at the same time has become very important in the past few decades. It is because speaking and listening skills have gained much significance in the teaching and learning process (Underhill, 1989). In fact, the ability to communicate in real-life situations is emphasized in language testing (Douglas, 2014). Unluckily, speaking skills has not been greatly priced in the language testing literature. Underhill further opines that it is all because there are some complexities and spoken language tests have always been taken as the more conventional language tests. Bachman & Palmer (1996) are of the view that it is very difficult and tiresome to prepare, administer and mark oral tests as compared to any other language tests. Thus, it can be one of the main causes for not testing speaking skills at all at various institutions all over the world. Underhill (1989) has pointed out that traditionally in language testing, 'test' was considered the most important instrument while the testers and testees were not given any significance. However, the situation is quite the opposite in spoken language tests these days. Now the candidates who take tests are considered important and what goes on between them is of utmost importance. This has not only enhanced the reliability and validity of the tests, but also provided a very natural environment to the testees, and interaction goes smoothly and naturally (Hughes, 2003).

Underhill (1989) defined an oral test as 'a test in which a person is encouraged to speak, and is then assessed on the basis of that speech'. Underhill (1989) and later on Bachman (1990) have recognized various features of speech and one of them is that it is relatively exceptional that speech would be a one-way communication system, i.e. meaning that usually we have two people taking part in communication with one person speaking and the other listening, and vice versa. Thus it is a two-way process. There are different types of criteria which further make it difficult for the testers to choose from while evaluating the performances of the testees. These are grammar, vocabulary, pronunciation, fluency, appropriateness, etc. Thus testers have to be extra-careful. While the assessment of this skill, they should develop their tools as objectively as possible.

To test spoken language proficiency in a variety of ways, a range of attempts have been made. That is way, oral testing is divided into direct and indirect methods; this division is built on whether the measurement is deemed to happen directly or indirectly.

Indirect tests are the types of tests which comprise parts where the candidate is not supposed to generate any language at all, for instance cloze tests related to writing skills. Besides, as far as oral proficiency tests are concerned, they consist of parts where the candidate has got to speak according to printed or recorded stimuli

(O' Sullivan, 2008). On the other hand, Semi-direct tests are the kind of tests that involve active speech from the testee though they are constructed by 'nonhuman' methods, for example printed test booklets or tape recordings. These tests, for instance, employ a picture to elicit a story from the testee. Direct methods incorporate methods such as role-plays and interviews.

### **Framework to Developing an Oral Proficiency Test**

The usefulness of the spoken language test is based on how the test developers construct the test (Luoma, 2004). Bachman & Palmer (1996) have given three stages of test development. These three stages are: design stage, operationalization stage and administration stage.

Test Development stage includes a) construct specification that consists of the purpose of the test, brief description of the test, characteristics of the testees, trait to be judged or measured; b) test task specifications - here we should know about rubrics to be used and administration plan; c) assessment specifications - analysis of the results (Bachman & Palmer, 1996).

As far as the operationalization stage is concerned, it is further categorized into two stages. The first stage deals with the test task specifications development, explaining the purpose of the each test task separately which is part of the test, the trait the test is supposed to measure, the setting, allotted time to each test task,

test input features and the marking key. The second stage relates to the blueprint of the test. It means that 'how test tasks will be organized to form actual tests' (Bachman & Palmer, 1996, p. 90). In other words, the blueprint is basically the whole structure of the test; it contains the number of test tasks, their specifications, their purposes, and criteria to be measured and the marks allotted to each test task.

The third and last stage of test development is known as administration stage. It includes the testees who will be administered the test, collection of test results and the analysis of the results.

### **Data Analysis**

#### **Analysis of the Interviews**

In this part of the paper, the researchers would present the analysis and later on the findings of the interviews conducted with twenty eight teachers, fifteen male teachers and thirteen female teachers. The researcher would use the techniques of thematic analysis to highlight the findings and present discussion on them. All interviewees had varied experience of teaching.

Interviews were of seven to ten minutes duration. The interviews consisted of seven to eight questions. The interviews provided the researchers with valuable information concerning the assessment of speaking skills in different institutions. Some of the questions were:

- How do you assess the oral proficiency of the language learners?
- How many test tasks do you give to the learners to assess their oral proficiency?
- How many evaluators are you while assessing the speaking skills?
- Do you use any criteria to assess the oral proficiency of the learners?
- Do you use any rating scales or bands to assess the oral proficiency of the learners?

From the data collected through these questions, many themes emerged from the interviews. These themes along with the discussion are as follows:

### **Interpretive & Thematic Analysis of the Interview**

The following themes emerged from the interviews.

#### **Elicitation Techniques**

First of all, topics and proverbs or quotations were given to the learners to assess their oral proficiency; moreover, only one test task was given to them. Hence we can say that when topics or quotations were assigned to them, they did not match up with the contention made by Hughes (2003), McNamara (2000) and Bachman (1990) that elicitation techniques should be based on real-life situations which would help the learners to expose their actual ability to use the language in real communication.

Secondly, the interviewees told the researchers that four to five students were called in the classroom for oral assessment simultaneously. They were assigned topics and if the learners found any difficulty to speak on, the assessors sometimes changed the topic and sometimes did not. And if they changed the topic after all, it was done only once.

Hence, it means that if the second time the topic is difficult, it is not changed at all. The candidates do not have any other choice; rather they have to express themselves on the topic whether they like it or not or they are able to speak on the topic or not, and this way they are graded or we can say their spoken language was tested. Moreover, one of them told the researchers that if they found that the topic was difficult they asked easy questions; if the topic was easy, they asked difficult questions. The researchers fail to understand what these difficult and easy questions mean.

#### **1.8.2.2 Time Allotted to the Learners**

According to the interviewees, they gave learners four to five minutes for preparation. After that, they called the first examinee to 'come and speak on the topic' while the other examinees waited for their turn. In this way, they were given more time for preparation of the topic than the others. They give each learner four to five minutes on average to speak on the topic. When the researchers asked, 'how much time do you give to the learners for their oral assessment'? One of the interviewees told the researchers that even 'one

minute is enough to assess' the spoken language of the learners'.

Furthermore, as one of the interviewees said that they asked questions to the examinee 'related to the topic' so that 'we can understand' whether the learner has 'comprehended the proper idea of the topic' or not. The researchers cannot make this out whether they want to observe and see learners perform in real-life situations or want to check their knowledge about the topic.

### **Criteria**

When the researchers asked about the criteria used to measure the oral proficiency of the learners, some of the interviewees talked about only one or two linguistic competencies like grammar and pronunciation whereas the others talked about syntax, vocabulary and pronunciation, and some of them had a vague idea about the criteria. Moreover, all the interviewees maintained that they had all the criteria in their mind so much so they talked about 'confidence' as one criterion to measure the oral competence of the examinees which the researchers did not find as one of the components of communicative competence. Furthermore, the interviewees did not talk about the other communicative competencies like sociolinguistic competence, pragmatic competence and organizational competence as elaborated by Canale & Swain (1980), Bachman (1990) and later on Bachman & Palmer (1996), for in real-life

situations the learners are supposed to interact with other people and not to deliver speeches, sermons or presentations.

### **Rating Scales or Bands**

Another theme that emerged from the interviews was about rating scales or bands. All of the interviewees maintained that they did not use any sort of rating scales, holistic or analytic, nor bands, nor did the institutions provide them with any scales of their own. So it seems that their process of marking the learners is very subjective and unnatural.

### **No. of Assessors**

The interviewees also pointed out that they were mostly two assessors while evaluating the spoken language of the learners. Both of them asked questions or interacted with the examinee that is also against the norm set by Underhill (1989) and Bachman & Palmer (1996) who view that one examiner should act as an interlocutor and the other should pay full attention to the language produced by the examinee. The interviewees told the researchers that sometimes when there is shortage of teachers, 'one senior teacher is appointed' who conducts the evaluation of the learners'. Furthermore, one of the interviewees told the researchers that 'if there are more students, one examiner asks questions' and the other does the writing of the names, marks, etc. and

while doing so, he is able to judge the learners as well.

Taking into consideration the themes which emerged from the interviews, we can say that the whole process of conducting the oral test or measuring the oral proficiency of the learners does not match up with the theoretical framework described by Bachman & Palmer (1996).

Moreover, according to Underhill (1989), at least three test tasks should be used to assess the oral competence of the students; almost 13 to 15 minutes should be spent to assess the spoken language of the learners. Not only will they produce more language but also the test itself will become valid and reliable. The test constructors should develop the test in such a way that it should offer the real-life situations to the learners.

### **Analysis of the Observations**

The researchers used observation as data collection technique to find out exactly how the oral proficiency test was conducted at the institutions. The researchers used non-participant technique of observation. The researchers observed that topics were handed over to the learners on a piece of paper which had neither the instructions for them, nor the test tasks to be used to elicit the language for assessment, nor the criteria used to assess the oral proficiency, nor the expected time for the completion of the test tasks, nor levels, etc.

Rather the assessors called a group of five to six learners at a time, giving them a topic and some time for preparation. After 5 minutes, the test takers, who were two in number, started calling the candidates one by one. While they were conducting the oral test with one learner, not only did the others try to talk with other learners so that they could muster up some points concerning their topic, but also they got more time for preparation.

Second, as far as the operationalization stage of the theoretical framework developed by Bachman & Palmer (1996) was concerned, the researchers observed that the examiners did not have any official document which could describe the levels of language proficiency with descriptors. They did not have any rating scale to measure the language of the examinees, or any document which could have described the language areas to be tested. Moreover, they did not have any document which could have explained to them the number of language test tasks to be used to assess the communicative competence of the learners. However, the assessors did have a sheet which contained three columns where they were supposed to write the name, roll number and the marks awarded to the learners.

Third, the researchers observed that the examiners conducted the oral test in their own way. They made use of only one test task which was sometimes converted to a question and answer sort of task, and that is what some of the teachers pointed out in the interviews

that they used two test tasks. They gave topics to the learners since they had a number of topics with them. Sometimes, they gave the same topics to different learners. So when they faced problems and did not have anything to say on the topic, they made request to the examiners to get the topic changed which was accepted by them.

Fourth, as far as the administration stage of the test was concerned, the researchers observed that some of the learners when asked to speak on the topic could not express themselves, for it was basically a proverb. Now the examiner started explaining the proverb which should have been done when the examiner gave the topic to the learner. Moreover, sometimes they raised some questions pertaining to the topic and sometimes they did not do so. During the administration stage, the researchers also observed that out of two examiners, sometimes one examiner received phone calls and did messaging. Last but not least, the examiners did not use any rating scale - holistic or analytic - or bands to grade or mark the learners, and the one who did messaging or received a phone call relied on the other scorer and copied the marks of the other examiner on the sheet given to him or her.

In short, if we take the theoretical framework developed by Bachman & Palmer (1996) as a model and match up the oral test designed, developed and administered at the institutions, we find many things missing from the test. For example, we find:

- the purpose of the oral test was not

explicitly described;

- there were not any TLU tasks;
- there was not any official document which could have detailed the instructions about the levels of language proficiency of the learners, and the instructions concerning language ability the examiners were supposed to measure;
- there were not any guidelines for the testers on the number of test tasks to be used to assess the spoken language of the learners;
- there were not any test task specifications; and
- there was not any rating scale or bands used to grade or mark the students.

Hence, we can say that the test did not match up with the theoretical framework developed by Bachman & Palmer (1996).

### **Analysis of the Sample Oral Tests**

One of the research questions was 'How can the process of spoken language assessment be restructured and improved upon in the light of the findings of the research?' So taking this question into consideration, the researchers designed, developed, piloted and later on administered the sample oral tests. Since there were two types of SOTs, one was conducted with a pair of learners and the other was conducted with the single learner at a time in order to measure their oral proficiency. There were eight students in number who volunteered to participate in the SOTs. The one SOT was consisted of three test tasks and the

other SOT of four test tasks. As far as the marking or grading of the learners was concerned, the CEFR scale was used to do so.

The first test task on both the sample oral tests was interview which proved to be a very suitable task. It helped the researchers to explore the syntactical structures and range of vocabulary used by the learners to express their ability to use the language in real-life situations. The second test task on one SOT was picture description and the examinees described the picture. The third test task was expressing an opinion and the fourth was role-play. In all of these test tasks, examinees' communicative competence like syntax, vocabulary, pronunciation, fluency, coherence, interaction, etc against the CEFR scale was measured in the natural environment and they were graded soon after the completion of the test.

The second sample oral test was also conducted with four learners who volunteered to participate in it. They were administered the test in pairs. In the test, both the examinees had to accomplish the test tasks together and the examiners fully concentrated on their language. During the test, their oral language competence like grammar, vocabulary, pronunciation, fluency, content and coherence was completely exploited with the help of test tasks and measured on the CEFR scale soon after the test was completed. The whole process of conducting the SOT was recorded. SOT scores are presented in the table below.

Here the researchers would explain the structure and contents of the table and then the findings would be discussed. However, the researchers would not enlist the names or any personal data of the learners. Each learner has been given a number like 1, 2 or 3 and letters 'M' and 'F' have been used along with the numbers to indicate whether the examinee is male or female. Since there were two examiners, two tables are given below to display the marks awarded to the students.

## Award List

**Table 1 Marking Key**

Student No.	Range		Accuracy		Fluency		Interaction		Coherence		Total Out of 50
	Marks	Level	Marks	Level	Marks	Level	Marks	Level	Marks	Level	
	5x3		5x3		5x1		5x2		5x1		
01F	4	B2+	4	B2+	3	B2	3	B2+	4	B2	37
02F	4	B2	4	B2+	4	B2	3	B2	3	B2	37
03F	4	B2	4	B2	3	B2+	3	B2	4	B2	37
04F	4	B2	4	B2	3	B2	3	B3	3	B2	36
05M	4	B2+	4	B2	3	B2+	3	B2	3	B2	36
06F	4	B2	4	B2+	3	B2	3	B2	3	B2	36
07M	4	B2	3	B2	3	B2	3	B2	4	B2	34
08F	3	B2	3	B2	3	B2	3	B2	4	B2	31

*Constructing Communicative Spoken Language  
Tests for the Language Learners in Pakistan*

**Table 2**

Student No.	Range		Accuracy		Fluency		Interaction		Coherence		Total out of 50
	Marks 5x3	Level	Marks 5x3	Level	Marks 5x1	Level	Marks 5x2	Level	Marks 5x1	Level	
01F	3	B2	4	B2	3	B2	3	B2	3	B2+	33
02F	3	B2	4	B2	4	B2	3	B2	3	B2	34
03F	4	B2	3	B2	4	B2	4	B2+	4	B2	37
04F	4	B2+	4	B2	4	B2	4	B2	4	B2	40
06M	4	B2	3	B1+	4	B2	4	B2	4	B2	37
O6F	4	B2+	4	B2	4	B2	4	B2+	4	B2	40
07M	4	B2	3	B2+	4	B2+	4	B2	4	B2	37
08F	3	B2	3	B1+	3	B1+	3	B2	3	B2	30

The tables above contain 12 columns. The first column has the given identification of the examinee. The other five columns on the table have the components of the communicative competence such as: range, accuracy, fluency, interaction and coherence, and all these five columns have the column of level attached to it on the right. The last column contains the total marks given to the examinee. The final marks were awarded by the examiner after the marks of the five columns containing components of communicative competence were added. The column of level attached to each scale column shows where the examinee stands according to the CEFR scale. Moreover, the number awarded to each category shows the particular score

given within that level scale. But one should remember that 5 was the maximum score that an examinee could get against 5 levels or component descriptors described by the CEFR.

Most of the examinees fell in the categories of B2+, B2 and B1+ which show that they are the independent users of the language and have to spend more time and energy on speaking skills to become expert users of the language. They had some problems in grammar and pronunciation but as a whole the interaction did not break down. Three (3) was the pass score on each category as all the categories contained five marks on the CEFR scale (Council of Europe, 2001). Then the formula given

by Underhill (1989) was applied and range and accuracy were multiplied by 3, fluency was multiplied by 1, interaction was multiplied by 2, and coherence was multiplied by 1. Thus each assessor on the panel had 50 marks. This is how the SOTs were conducted with eight volunteer learners.

### **Findings of the Study**

Here the researchers would like to enumerate the findings of the study based on the data which is just discussed above. These findings will help answer the research question of the study.

The data analysis shows that:

- the purpose of the oral test was not explicitly described;
- there were not any TLU tasks;
- there was only one elicitation technique used to measure the oral proficiency of the learners;
- there was not any official document which could have detailed the instructions about the levels of language proficiency of the learners, and the instructions concerning language ability the examiners were supposed to measure;
- there were not any guidelines for the testers on the number of test tasks to be used to assess the spoken language of the learners;
- there were not any test task specifications; and
- there was not any rating scale or bands used to grade or mark the students.

### **Conclusion**

Hence on the basis of data analysis, the researchers conclude that the spoken language test performed at the language institutions is invalid and inauthentic. It has to be well planned, has to be administered appropriately and analyzed according to the scale or band used to evaluate the learners. By acting upon such guidelines mentioned above, one can meet the objectives of the spoken language testing appropriately. For testing spoken language, a well-defined criterion is very important. It helps guide the testing practitioners; it makes the test authentic and valid. It helps not only in making valid judgement about learners' spoken language, but also it helps the teachers in further improving and diagnosing the problems of the learners regarding oral proficiency.

### **References**

- Airasian, W. P. (2000). *Classroom Assessment: concepts and applications* (4th ed.). Boston: McGraw Hill Companies.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Canale, M. & Swain, M. (1980). *Theoretical bases of communicative approaches to second language teaching and*

- testing. *Applied Linguistics*, 1, 1-47.
- CEFR. (2001). *Common European framework of reference*. Cambridge: Cambridge University Press.
- Douglas, D. (2014). *Understanding language testing*. New York: Routledge.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- O' Sullivan, B. (2008). Notes on assessing speaking accessed on 08, 07, 2010 at 10:00pm at [www.lrc.cornell.edu/events/past/2008-2009/papers08/osull1.pdf](http://www.lrc.cornell.edu/events/past/2008-2009/papers08/osull1.pdf)
- Shahzad, K. (2018). *Analyzing English language teaching and testing practices in developing discourse competence in essay writing*. An unpublished thesis. International Islamic University, Islamabad.
- Shahzad, K. (2018a). Spoken language testing practices in National University of Modern Languages. *Journal of Research in Social Sciences - JRSS*, 6(1). 291-314.
- Underhill, N. (1989). *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.