

## **An Evaluation of Oral Interview and Voice thread as Elicitation Techniques to Assess EFL Learners' Speaking Skill**

**Farideh Ataeifar \***

Shiraz branch, Islamic Azad University, Shiraz, Iran

**Mohammad Sadegh Bagheri**

Shiraz branch, Islamic Azad University, Shiraz, Iran

### **Abstract**

The purpose of this paper is to evaluate and rate two selected elicitation techniques, oral interview and Voice thread, an on-line program, to assess the speaking skill of the EFL students at the Zand Higher Education Institute in Shiraz, Iran. Testing techniques were selected based on the level of the students (upper-intermediate). The speaking ability of a group of the students was evaluated using both techniques, numerical results and data were specified and analyzed in accordance to the communicative competence. As testing speaking is one of the main issues in the foreign language testing, the implication of the study reveals an overall view of the rating of the above mentioned testing techniques which further enhances the evaluation of speaking skill.

**Keywords:** Elicitation technique, Speaking skill, Voice thread, Oral interview

### **INTRODUCTION**

Testing speaking is considered to be one of the crucial points in both English teaching and testing. Moreover, it is required to make teaching consistent and at the same time motivate learners to succeed. Since speaking is a means of displaying communicative competence, many researchers approach the definition of it in varied ways. According to Savignon (1983), communicative competence is the ability to function in a truly communicative setting, a dynamic exchange in which linguistic competence must adapt itself to the total informative input, both linguistic and paralinguistic, of one or more interlocutors. So many methods focus on how to teach speaking in the target language effectively.

In contrast to the other researchers, Canale and Swain (1980) narrowed down the communicative competence into four components and made it clearer for language teachers. These components are grammatical competence, sociolinguistic competence, discourse competence, and strategic competence. Mostly, testing techniques of the speaking ability are aimed at the communicative activities, so many classroom teachers

have concentrated on promoting communicative competence in language learners by using communicative activities those which rely more on the student's ability to understand and communicate real information. (Celce-Murcia 1991) are as economical of time and effort as possible will have a positive washback effect (Hughes 2003).

The next level of teaching and testing has come about by the rapid progression and integration of technology in a time of growing global and cultural diversity affiliation. Curriculum development is being adapted to individual learning styles, encouraging collaborative teamwork, and facilitating critical thinking and problem solving through a variety of communication, visualization, and simulation technologies. Students and teacher should be familiar with and build competence around technological tools used in the classroom, online and in the world around them. The technology tools should support the increasing diversity in students' abilities, thoughts, perceptions, cultures and lived realities. By integrating technology-enhanced reflective practice into the learning process contributes to students' continued growth in both receptive and productive abilities, in the case of the current study, speaking skill.

### **Statement of the Problem**

Researchers vary in stating what testing techniques are and what objectives they should complete. However, a few studies conducted on using voice-conferencing technologies on language testing or even comparing it with other techniques. Mixed results from previous studies where researchers used both synchronous (Beauvois, 1992; Satar & Özdener, 2008) and asynchronous (McIntosh et al., 2003; Poza, 2005) types of communication have shown that some learners felt more comfortable speaking in the computer-mediated environment, while others viewed the computer-mediated environment as unhelpful in the fostering of communicative competence. In addition, the technology itself was found to impede some learners as they disliked the delay in communication and ideas (Satar & Özdener, 2008).

Therefore, this study was conducted to examine the integration of the Web 2.0 technology *Voice thread* as a new elicitation technique to determine if this asynchronous voice-conferencing tool can be an appropriate one to assess speaking skill.

### **Objective of the Study**

The purpose of this case study is to determine if the use of the Web 2.0 asynchronous voice-conferencing technology, *Voice thread*, as an elicitation technique is rated higher or lower compared with the traditional one, oral interview, to assess speaking skill.

### **Research Question**

This study was designed to answer the following research question:

- Is *Voice thread* as an elicitation technique rated higher or lower compared with the traditional one, oral interview, to assess speaking skill of EFL upper-intermediate Iranian students.

## Null Hypothesis

The following null hypothesis was provided for the research question:

*H<sub>0</sub>*: Oral interview and *Voice thread* are rated the same as two elicitation techniques to assess speaking skill of EFL upper-intermediate Iranian students.

## Significance of the Study

The results of this study provides foreign language test designers, teachers, and curriculum planners with data on the effectiveness of two elicitation techniques, asynchronous voice-conferencing and oral interview, to asses one of the important English productive abilities, speaking. It also makes foreign language educators familiar with a new elicitation technique, *Voice thread*, and let them know how to use it in an online environment. Moreover, the results of the research were beneficial to get and view the testing procedures. In addition, learners were positively influenced by testing speaking because their speaking results have been improved by the end.

## REVIEW LITERATURE

The varieties of evaluation differ according to the purpose of our tests. For example, to test speaking ability, the performance evaluation plays a significant role. The performance is tested under certain circumstances and Luoma (2004) explains it is often tested in live interaction, where the test discourse is not entirely predictable, just as no two conversations are ever exactly the same even if they are about the same topic and the speakers have the same roles and aims. As a matter of fact, there are different types of evaluation: a teacher's evaluation, self-evaluation, or peer-evaluation. According to Underhill (1987), in order to use self-evaluation method, we should be teaching the learner to know how he is getting on independently of the teacher. And similarly in the peer-evaluation, learners must be taught to evaluate their peers. Furthermore, some techniques will be easier to self- correct than others.

Mostly, the evaluation is measured through scores, similarly marks or grades. And Brown (2004) claims that grades must be the most-talked-about topic in anyone's school years. In addition, grades still play a leading role among teachers and learners. "It's ironic – when our study life – is inserted into [marks] only... (Brown 2004, p. 282). In particular, scores must show criteria as well. "If we are to interpret the score on a given test as an indicator of an individual's ability, that score must be both reliable and valid" (Bachmann 1996, p. 24). On the other hand, test takers should not see that we are writing notes or marks while they are being tested. Heaton reports: "Never mark in front of a student" (Heaton 1990, p. 67). Due to the fact, test takers can be stressed out or their performance can be reduced.

Technology and the internet can also be used to asses students' language ability and their development of specific language skills. Moreover, many teachers and testers are looking for novel ways as elicitation techniques with the current initiative to improve the oral proficiency of language learners (Pufahl & Rhodes, 2011). Technological

resources may help teachers provide an environment that will decrease anxiety levels and be perceived as less threatening to learners (Crookall & Oxford, 1991; Pufahl & Rhodes, 2011; Ravenscroft, 2009). Researchers found that studies incorporating text-based computer-mediated communication have decreased participants' anxiety levels due to less pressure to formulate responses under a time constraint (McIntosh, Braul, & Choe, 2003).

## **RESEARCH METHOD**

### **Participants**

The participants of the study were EFL students of the Zand Higher Education Institute in Shiraz, Iran. They were in their second year, 2016, of studying English Translation. The study took place during the first nine weeks of the college term. The class consisted of 15 students, with 7 males and 8 females. To reduce the rater-effect, the teacher was the researcher herself. She had about 11 years of experience teaching at higher education institutes and universities. Her experience was strongly founded in communicative approach to language teaching and testing.

### **Setting**

All the steps of the study were conducted in their own classroom at the Zand Higher Education Institute in Shiraz, Iran.

### **Criteria to choose testing techniques**

The purpose of a test is one of the main criteria in selecting the appropriate testing techniques. "If the student has a purpose in speaking, he immediately finds himself in a situation in which what he says and how he says it have significance" (Brown & Yule 1983, pp. 117-118). This means that, if a speaker delivers information to a listener, who is unaware of this information, then the speaker is more motivated to deliver the message (Brown & Yule 1983, p. 117). Despite this fact, we should know "...what it is you want to test" (Brown 2004, p. 49) when considering the purpose of a test.

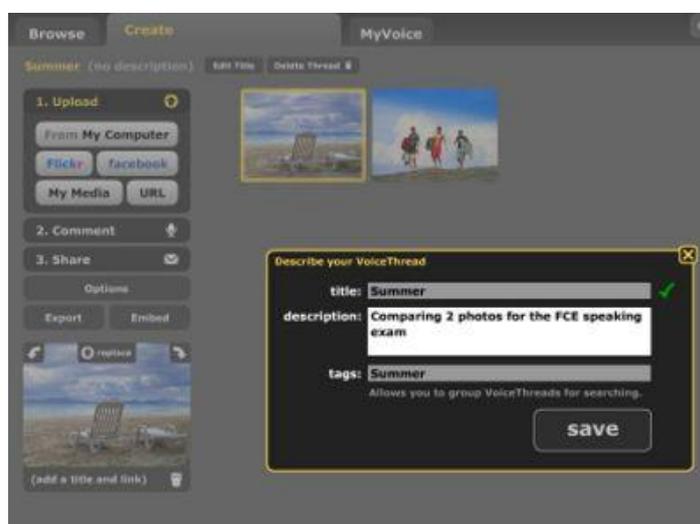
Another consideration to choose the testing technique is the context that should be chosen, so that learners can speak based on that context (Luoma 2004). It is easier for students who do not have to think about the input when tested in spoken interaction. The level of learners as related to their age was another criterion for the selection of tasks. Learners in a certain age group are interested in certain topics and the context for testing tasks should be relevant to their age. Above all, the level of learner's language competence should be specific to the tasks as well. Learners are to be tested on the basis of their acknowledged level of language competence. Therefore, the testing techniques are adjusted to learner's language competence.

All in all, "The broad aim of all these techniques is to encourage learners to speak by giving them something to speak about..." (Underhill 1987, p. 40). Similarly, Fulcher (2003) believed that a goal to design a test task is to elicit enough speech to allow a

rating to take place; moreover, he suggested that techniques will elicit behavior which is not only a reliable and valid indicator of the ability in which we are interested but it can also be reliably scored. To achieve the above goal in the case of the current study, two elicitation techniques were chosen by the researcher: oral interview and on- line *Voice thread*.

### ***Voice thread***

The group used *Voice thread* to record and forward their voice messages to their teacher who was the researcher herself. It is an asynchronous voice conferencing technology that allows learners to communicate by posting voice recordings to a web using cell phones or microphones to record their voices from a computer. The technology allows students to post comments in English around an image, a video or a sound clip, or a question or series of questions provided by the teacher. A screen shot of *Voice thread* can be viewed below.



**Figure 1.** *Voice thread* Screenshot

It is a Web 2.0 technology that is free to use. Students created their own accounts and were required to have a username and a password to access the account. *Voice thread* is great for getting students use and develop their speaking skills in a fun and creative way. It is also very teacher-friendly. It has privacy controls, comment moderation, and a system of identities that allows all students in a class to use it conveniently.

### **Evaluation of Testing Techniques**

As a matter of fact, there is nothing like the best technique or the best test. When evaluating or assessing the learner's performance, different principles must be taken into account. O'Malley and Pierce (1996) conclude that:

Oral language assessment can take various forms depending on your purpose for assessment, student's level of language proficiency and the purposes for which students use oral language in the classroom. (O'Malley & Pierce 1996, p. 69). Without the purpose of a test, it can't be evaluated. To find out whether the testing technique was

prepared ahead, there can be a construction of the preparation projects used in the research. Also, this can be evaluated with regard to the testing techniques.

Next step in testing the speaking ability is to design a test specification where the tester can include skills to be assessed or scoring procedures used. In this case, it is also helpful to evaluate whether test specifications were completely understood by the tester.

### **Rating Scale**

Fulcher (2003) believe that rating scales are important in tests of speaking because they operationalize the construct that the test is supposed to measure, whether the description of the construct is thick, or thin as in the scaling of descriptors.

Luoma (2004) distinguishes two kinds of scales: holistic and analytic scales. Holistic scales express an overall impression of an examinee's ability in one score, but analytic scales express the test taker's strengths and weaknesses and contain a number of criteria. O' Mally and Pierce (1996) state the rating scales must be aimed at the learners and teachers should assign the rating that most closely fits the student's actual performance. Underhill (1987) adds that the rating scale should be kept as simple as possible. Thus, testers can more concentrate on the evaluation of performance when knowing the description of rating scales by memory. They do not have to worry about the complex description and can evaluate it clearly.

To evaluate communicative competence (CC) and learners' speaking performance, rating scales can be designed based on learners' needs depending on specific conditions. In the case of the present study, the rating scale consisted of five criteria: grammar accuracy, pronunciation, vocabulary, fluency, and interactive communication. First, grammar accuracy is specific to the mastery of word order, tense and pronouns. "When testing speaking it is similarly important to decide what types of errors be penalized in the scoring, and what errors should be ignored or treated with leniency" (Fulcher 2003, pp. 27-28). Pronunciation is evaluated by finding out whether learners understand foreign sounds and are able to pronounce words which do not occur in their native language. As Fulcher (2003) reports: "The speaker must first decide what to say, be able to articulate and create the physical sounds that carry meaning" (Fulcher 2003, p. 25).

The criterion of vocabulary basically tests if learners have achieved the vocabulary which they have been learning and practicing throughout their learning process. The evaluation of this sub-skill is limited regards to the learner's knowledge; however, it can be further developed by the strategies learned and used during the learner's performance.

Next criterion, fluency, tests whether learners can connect their speech and express what they wanted to say. Fulcher (2003) described fluency as: "If speech is going to be fluent, the process of planning what to say, retrieving the necessary grammar and vocabulary, and speaking, needs to be automatic. It is when speech becomes more noticeably automatic that we describe a speaker as being 'fluent'" (Fulcher 2003, p. 30).

In order to test the fluency, it is concerned whether the fluency does not disturb the learner's performance or whether the pauses are not too long.

In the interactive communication, the researcher mostly concentrated to evaluate whether learners understand each other or the interlocutor. Further, the criterion concerns whether learners can experience the ability of taking turns and the ability to grasp the control over the conversation or simply listen to the conversation. "Firstly, a second language learner must be a good listener if he or she also wishes to be a good speaker" (Fulcher 2003, p. 35). When concentrating on the evaluation of the performance according to the rating scales, the researchers should be careful not to misjudge the overall performance (Johnson 2001, p. 307). Hughes advises that "...great care must be taken to ignore personal qualities of the candidates that are irrelevant to an assessment of their language ability" (Hughes 2003, pp. 133-134). Therefore, the researcher should try to avoid the test taker's appearance and to focus only on his/her language ability.

## Procedure

Students participated in the study over a nine week period. The researcher interviewed the students individually in a face-to-face situation. Each learner was given two pictures from two different places to describe at first. Further, the test taker was asked to choose one of the places in where s/he would prefer to spend her/his holiday. The next task was to explain why s/he would choose the place or why not. The technique was selected according to learners' positive attitudes to the pictures, and the topic was relevant to their interests. By this time, learners had been practicing vocabulary related to holidays. Other skills, which were assessed such as expressing likes and dislikes and students' ideas corresponded to the overall skills of picture comparison technique.

The language was natural and comprehensive for the learners. The technique in the first part can rarely be found in the real-life situations; therefore, it was not authentic. For a test taker, it would lose the point of completing the task when the tester knew the picture, as there was no information gap. However, in the second task, the learner had an opportunity to express what s/he liked or where s/he preferred to go on holidays. This technique was appealing for a test taker and the tester as well.

It is recommended to give learners time to look at the pictures. Also the content of the pictures should be neither too comprehensive nor too simple. Nevertheless, learners were not guided by any written instructions. The technique was mainly designed to the oral interview technique and the use of written instructions was avoided. Fulcher (2003) suggests that pictures are used as a springboard to wider discussions of topical issues. There are also a lot of possibilities to combine other techniques with pictures to elicit data. The student's description was audio-recorded to be evaluated later based on the five criteria of the rating scale of the study.

To reduce time and practice effect, the second elicitation technique was used just a week later. The teacher, the researcher, and the student installed *Voice thread* and made

their own accounts in advance. Since the teacher was the researcher herself, teacher training was not required. However, the teacher trained the students on the particular application, *Voice thread*, they were assigned to use to record their voice messages. The student was assigned to perform the same task, picture description, audio-recorded his/her performance in his/her account and forward it to the teacher for scoring and later evaluation based on the same criteria of the same rating scale as the first step of the test.

### Scoring

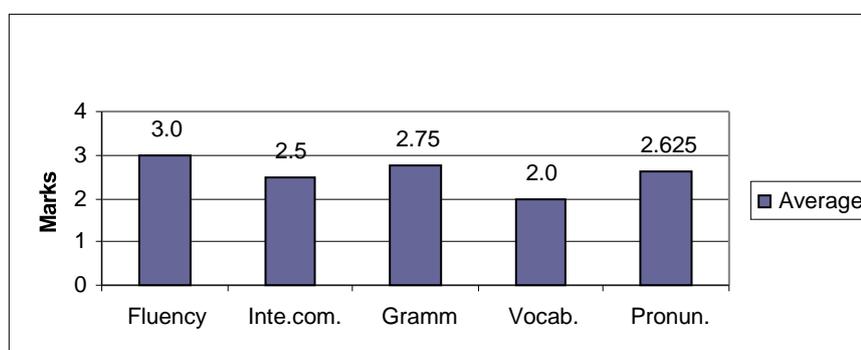
According to the test specification, the main assessed skill was to compare two different pictures. At the beginning, learners were assigned to compare the pictures by describing them. Therefore, the comparative task has been changed into the descriptive task. The rating scale was illustrated in Table 1.

**Table 1.** Rating Scale

Score	Vocabulary	Pronunciation	Grammar Errors	Fluency	Interactive communication
1	Broad & general	no errors	0-2	Smooth	Normal
2	Adequate	Occasional, ineffective mispronunciation	Few	Less hesitant	Repetition
3	Limited	Occasional, effective mispronunciation	occasional	More hesitant	Rephrasing
4	Word choice	Frequent errors, heavy accent	Frequent	Slow	Repetition & rephrasing
5	Inadequate	unrecognizable pronunciation	Totally inaccurate	Stuttered	Total failure

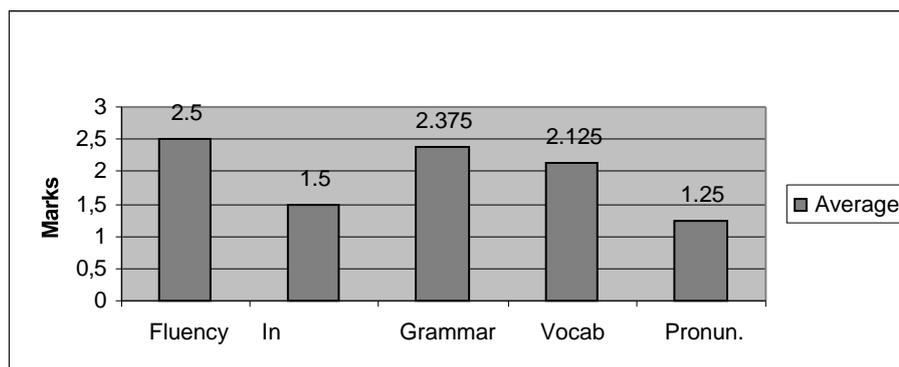
### Data Analysis

After scoring learners' speaking performance based on the five criteria of the analytic scale displayed in Table 1, grammar accuracy, pronunciation, vocabulary, fluency, and interactive communication, the average of the criteria was calculated for each elicitation technique and was demonstrated in the following figures:



**Figure 1.** The average of criteria in the face-to-face elicitation technique

The discourse component was very strong because the tasks interrelated among each other and there was strong lexical cohesion; however, the average of fluency which belongs under this category was very low (3.0). To consider the average of all criteria, the grammatical competence of the technique was seen as the relevant part of the testing. The average of grammatical competence became a leader among the other competences. It was 2.46 which specifically: the test takers were mostly successful with pronunciation (2.0) and then, followed by vocabulary success (2.63) and the final position of the success belonged to the grammatical accuracy (2.75). To compare the two techniques, the scores of the second group who were using *Voice thread* were also calculated, then the average of all five criteria was illustrated in Figure 2.



**Figure 2.** The average of criteria in the *Voice thread* technique

According to the rating scale average and to the criteria belonging to the grammatical competence, the component was considered as strong one since the average was estimated at 1.916. Even though, the grammatical accuracy criterion reached only 2.375. Other criteria namely: vocabulary (2.125) and pronunciation (1.25) enhanced learners to attain better overall marks.

## RESULTS AND DISCUSSION

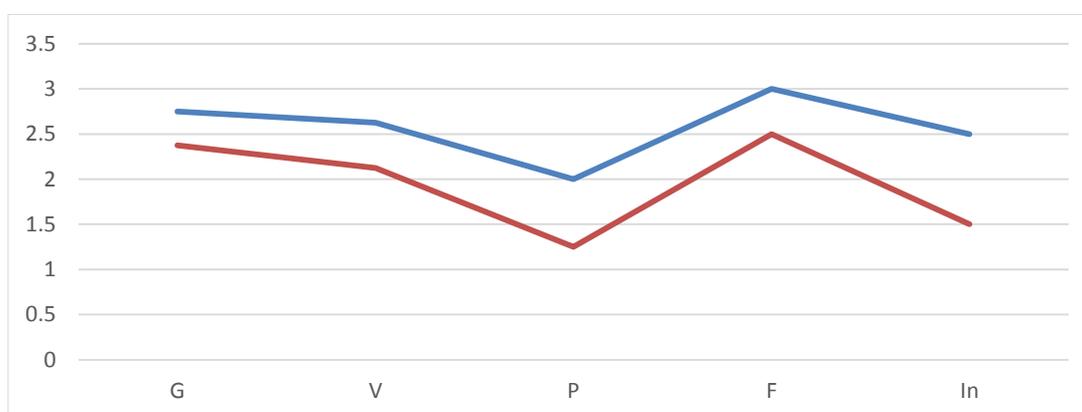
In the face-to-face interview technique, the strategic competence was not fully used because test takers used mostly the avoidance strategies to reduce the structures with the aim of transferring the message. The average of interactive communication was 2.5 and it is not the strongest component of the CC. Even though, different strategies mentioned were used among test takers, they still did not increase better results to the overall testing of the picture comparison technique. On the other hand, test takers used repetitions and self-corrections as well.

However, when not looking at the averages of the criteria or at the test takers marks, the researcher perceived the learners' strengths to be in the grammatical competence. Consequently, the results out of the criteria to the grammatical competence were analyzed as the strongest as well. Besides, learners were talking about the topics they knew in relation to the pictures. Then, the strategic competence was a strong element in this area. The technique was not found to be difficult from the standpoint of a teacher and a learner even though the ratings did not correspond to it. Also, some learners said that the technique was simple and interesting. From the point of view of a teacher, the

technique was practical, but in a large class like the current setting, an extra interviewer would be recommended since interviewing, scoring, and rating the five criteria simultaneously could be not only tiring but also error-prone.

Sociolinguistic competence became very valuable in the *Voice thread* technique. Learners were tested based on an interesting topic presented in the pictures which enabled learners to talk about the topic in a free and informal manner. However, the learners were required to record their voice messages to be evaluated by the researcher later. The sociolinguistic competence approached the learners' strengths and reached an average 1.5 which was the best result besides the pronunciation measurement.

On the contrary, the discourse competence was evaluated as the weakest component of CC, since the fluency reached only 2.5 marks average. To see the average of all criteria, and compare the techniques, look at Figure 3.



**Figure 3.** The comparison of the average of all criteria of two elicitation techniques

## CONCLUSION

Through Littlewood's (1981) division of communicative activities, the testing techniques designed based on the need of the information gap activities were mostly focused on functional communicative activities. The test takers view both techniques as appropriate and useful. In the researcher's opinion, the tests and techniques were positively perceived by learners; thus, the tests contained positive face validity. Moreover, the evaluation of selected testing techniques from the viewpoint of a teacher and a learner reached in this paper can be helpful for beginner teachers who can get positive ideas and attitudes toward testing speaking. The research of the evaluated testing techniques also had a good impact on learners who realized their strong and weak sides of their language communicative competence.

As their voices were recorded and evaluated later while they were using *Voice thread*, unlike in the face-to-face oral interview, there were performances in which it was difficult to stop the test takers' discussion because they got very excited about using the new on-line technique. Concerning the scoring procedures, it was structured and completed after the re-listening the audio-recorded performances. It was a long process

of transcribing the performances and the evaluation of all individual test takers according to the rating scale criteria.

All in all, the broad purpose of all elicitation techniques is to give the learners some hints, and encourage them to speak. Furthermore, good techniques are those which elicit not only enough speech to allow a rating to take place but also behaviors which can be reliable and valid indicators of the ability in which the researcher was interested. It cannot be concluded just based on the result of the present study that a technique is better than the other, neither can the result be generalized to the other testing contexts. However, both face-to-face and *Voice thread* were rated as two good and successful elicitation techniques for the particular EFL learners, the participants of the study, and satisfied the needs of both the tester and the test takers.

### LIMITATIONS OF THE STUDY

Working with on-line applications has its own limitations, particularly if the application has to be used in a crowded institute, the setting of the study, which was equipped with an internet service that was not fast enough. The students' devices were disconnected while they were recording and forwarding their voice messages. So, they had to shift to their mobile phone nets. Not only was it time-consuming, but it was stressful in some cases; especially with low-achievers.

### SUGGESTIONS FOR FURTHER RESEARCH

The evaluation of testing techniques of the speaking skill is still a "hot issue" that needs to be examined, re-proved and practiced. As mentioned, communication and speaking are important features of learning the target language and teachers should be consistent in testing the speaking skill.

### REFERENCES

- Bachman, Lyle, F. (1996). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Brown, D. (2004). *Language assessment: Principles and classroom practices*. London: Longman.
- Brown, G., & Yule, G. (1983). *Teaching the spoken language: An approach based on the analysis of conversational English*. Cambridge: CUP.
- Beauvois, M. H. (1992). Computer-assisted classroom discussion in the foreign language classroom: Conversation in slow motion. *Foreign Language Annals*, 25, 455-464.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Celce-Muria, M. (1991). *Teaching English as a second or foreign language* Waldsworth: Heinle and Heinle Publishers.
- Crookall, D., & Oxford, R. (1991). Dealing with anxiety: Some practical activities for language learners and teacher trainees. In E.K. Horwitz & D.J. Young (Eds.),

- Language anxiety: From theory and practice to classroom implications* (pp. 141-150).  
Englewood Cliffs, NJ: Prentice Hall.
- Fulcher, G. (2003). *Testing second language speaking*. London: Pearson Education Limited.
- Heaton, J. B. (1990). *Classroom testing*. London: Longman Group UK Limited.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: CUP.
- Johnson, K. (2001). *An introduction to foreign language learning and teaching*. London: Pearson Education Limited.
- Littlewood, W. (1981). *Communicative language teaching*. Cambridge: Cambridge University Press.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: CUP.
- McIntosh, S., Braul, B., & Choe, T. (2003). A case study in asynchronous voice conferencing for language instruction. *Educational Media International*, 40, 63-74.
- O'Malley, J.M., and Pierce, L.V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. Boston: Addison-Wesley Publishing Company.
- Pufahl, I. & Rhodes, NC. (2011). Foreign language instruction in US schools: Results of national survey of elementary and secondary schools. *Foreign Language Annuals*, 2, 258-288.
- Satar, H., & Özdener, N. (2008). The effects of synchronous CMC on speaking proficiency and anxiety: Text versus voice chat. *Modern Language Journal*, 92, 595-613.  
doi:10.1111/j.15404781.2008.00789
- Savignon, Sandra J. (1983). *Communicative competence: Theory and classroom practice*. Boston: Addison- Wesley Publishing Company.
- Seliger, H. W, and Shohamy, E. (1989). *Second language research methods*. Oxford: Oxford Universtity Press.
- Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques*. Cambridge: CUP.