

EFL Female College Students and Instructors' Preferred Method of Speaking Assessment: A Perspective from Saudi Arabia

ABSTRACT

Introduction: At the College of Languages and Translation (COLT), the students take four speaking courses in the first four semesters of college. In each course, they take two interim tests and a final exam.

Aims: The study aims to find out which speaking assessment method EFL students and instructors at COLT prefer: face to face or lab testing and reasons for preferences.

Procedures: Face-to-face assessment is the most common method at COLT. The students are tested individually. Each student randomly draws one topic from a basket and engages in a conversation/interview with the teacher who asks 3-5 questions about it. The test session for all the students lasts from 8 a.m. to 1 p.m. non-stop. The second method is language lab testing used by few instructors where all the students take the test at the same time and they all answer the same set of questions in about 60 minutes. The lab test consists of several questions. The students read the printed questions, take short notes, then record their answers on an MP3 player.

Results: Most students prefer lab assessment, whereas most instructors prefer face-to-face assessment. Students reported that lab test questions are comprehensive and help them improve their speaking skills better. Testing conditions are the same. They are less anxious. They lose few marks if they miss a question. Instructors reported that face-to-face test questions are easy and cover part of the material. All students pass. Students are more anxious. Comparisons of students' test score showed that lab tests are more reliable, valid and have a better discriminating power between students who have mastered and those who have not than face-to-face tests. Recommendations for effective speaking assessment methods are given.

Keywords: *Speaking assessment, speaking tests, speaking tasks, speaking in EFL, EFL learners, lab testing, face-to-face testing*

1. INTRODUCTION

The speaking skill in the second/foreign language is difficult for students to develop and challenging for teachers to teach and assess because it is a productive **skill**

which requires students to be unique and creative in the way they express themselves orally. A review of the second language teaching and learning literature showed a plethora of studies that investigated several aspects of speaking assessment. The first line of research focused on standardized tests such as the IELTS and TOEFL [1]; [2]; [3]; [4]; [5]; and on specific-purpose speaking tests such as the Occupational English Test for assessing the language proficiency of health professionals [6]; [7]; [8] and others.

A second line of research focused on variables that affect speaking test assessment such as the effect of task type and examinee first language background on the internal structure of the test [9]; [10]; the impact of raters' language background on their judgements of the speaking performance in the College English Test-Spoken English Test (CET-SET) of China [11]; variability in raters' scoring decisions and their expertise in scoring speaking tests [12]; inter-examiner reliability in the pronunciation assessment component of speaking tests [13]; the criteria English teachers associate with general oral proficiency [14]; scoring procedures and rating scales used such as checklist vs a rubric with level descriptors [15]; divergent interpretations of the test event by the students and variation in interlocutor behaviour [16]; topical knowledge, anxiety, and integrated speaking test performance [17].

A third line of research specifically focused on the methods utilized in speaking assessment such as using interviews or conversations [16]; Oral Proficiency Interviews [18]; Dynamic Assessment based on Vygotsky's Zone of Proximal Development, in which the students receive a pre-test and self-reflection, feedback and knowledge expansion, post-test and self-reflection, post-feedback, and semi-structured interviews [19]; guiding students while listening to audio files, and having students transcribe, analyse, self-assess their own speaking samples, and record feedback on their own performance [20]; Story Retelling Speaking Tests [21]; Group-based speaking tests such as group oral discussion [22]; [23]; [24]; rehearsed speech where the students memorize scripts for their oral responses [25]; pressured vs. unpressured on-line testing conditions [1]; use of paired tasks as in the paired format of the Cambridge Speaking Tests [26]; [27]; use of individual tasks in which the students interact with an examiner and another paired task in which they interact with other student [28]; text-speaking, graph-speaking, and listening-speaking tasks [9]; integrated and independent speaking tasks [29]; and answering questions, describing a picture and giving a presentation in a language lab [10].

At the College of Languages and Translation (COLT), King Saud University, Riyadh, Saudi Arabia, 4 speaking courses are offered to students in the first four semesters of the translation program. In each speaking course, the students take 2 speaking interm tests per semester and a final exam. The most common speaking assessment method is face-to-face testing in which the students are tested individually and each student talks about a topic in front of the instructor. Another speaking assessment method is lab testing, where all the students take the test at the same time and each student answers a set of questions and records her answers using an MP2 player.

Despite the utilization of these speaking assessment methods for a long time, there is lack of research that focuses on speaking assessment for EFL college level

students in Saudi Arabia, such as speaking assessment methods that students and instructors prefer, their advantages and shortcomings, the kind of tasks speaking tests focus on; content covered by speaking tests; how speaking tests are conducted; how they are scored; how marks are assigned; validity, reliability and discriminating power of speaking tests utilized; and effectiveness of different testing procedures. Therefore, the present article aims to find out which speaking assessment method EFL female college students and instructors prefer: face-to-face or language lab testing; what are the advantages and shortcomings of each method as perceived by students and instructors; which method is more effective in measuring the speaking skill in terms of the types of tasks given on the speaking test; how many tasks the students perform on the test; how the students are tested; content sampling and breadth of coverage of the test tasks, validity, reliability discrimination power, scoring accuracy and rater reliability and accuracy. (It needs revision and rewriting)

Identifying students and instructor's speaking assessment method preferences will be based on surveys with a sample of students and speaking instructors. Students and instructor's preferences will be reported quantitatively and qualitatively. Furthermore, validity, reliability discrimination power of the speaking tests in both methods will be based on the final speaking scores of two groups of students in Speaking III who were tested using the two-assessment methods.

Results of the current study are useful for EFL college speaking instructors. It will enrich their understanding of speaking assessment methods, their effectiveness, advantages and shortcoming, reliability, validity and discrimination power issues, how they are scored, speaking test item selection, what they measure and which tasks and skills they cover. It will give them an idea about the testing and grading procedures in each assessment method. (It needs revision: problem of punctuation)

2. METHODOLOGY

4.1 Subjects

The student sample consisted of 69 EFL female in the translation program at COLT. They have experience with the lab testing method in the Speaking III course and the face-to-face speaking assessment method in the Speaking I and II courses. In addition to the Speaking courses, students at COLT take Listening, Reading, Writing, Grammar, and Vocabulary Building courses in the first 4 semesters of the translation program (20 hours per semester including the Speaking course).

The instructor sample consisted of ten female instructors who have taught speaking at COLT. They had a Ph.D. or an M.A. degree and were specialized in linguistics or teaching English as a foreign language. Six instructors have used the face-to-face assessment method, two have used the language lab assessment method only, and another two have used both speaking assessment methods. All the instructors have used the same speaking textbooks.

4.2 Materials and Tasks

The textbooks used for the four speaking courses at COLT are Interactions I & II and Mosaic I & II from McGraw Hill. The courses aim to develop students' ability to talk

about a particular topic for at least 5 minutes without preparation, with fluency, correct pronunciation, stress, intonation, grammar, choice of words and well-organized cohesive ideas. For example, in Speaking III, the students practice the following skills as described in the Mosaic I textbook: using expressions with riddles and brain teasers; using expressions to elicit the 'truth'; discussing general information; sharing unusual or surprising events; reporting surprising research; role-playing: characters with definite views; presenting rationales in active and passive voice; role-playing: eyewitness reporting; agreeing/ disagreeing with a suggestion or point of view; responding to questions; using tone of voice to express likes and dislikes; conducting interviews using tag questions; role-playing: hopes and desires; discussing personal wishes/dreams; presenting facts, feelings, opinions and reasons; discussing funny people; telling jokes; using humor in discussions; describing a scene using analogies; using notes to describe a process; discussing pros and cons of investment options; creating chronological round-robin story; discussing various forms of communication; telling stories using digressions; paraphrasing short readings; using an outline to organize thoughts for presentation. (It needs revision and rewriting)

In addition, the students practice using the following functions: Offering and requesting clarification (*does that make sense? Are you following me?*); asking for confirmation (*I do not exactly know what you mean; I am not sure I am getting this*); making generalizations; introducing information; expressing opinions; shifting focus; agreeing and disagreeing; likes and dislikes; asking for information, seeking confirmation and challenges with tag questions; expressing wishes, hopes and desires; stating reasons; and telling a joke.

1.3 Speaking Assessment Methods at COLT

4.3.1 Face-to-Face Method

In face-to-face assessment at COLT, the test consists of an oral part and a written part. In the oral part, the students are tested face to face individually. The instructor prepares a set of questions or topics for students to talk about and writes each on a card (See Appendix A). The course instructor sits in the classroom while the students are outside. She calls on the students one by one to go in and take the test. Each student randomly draws one topic from a basket, [and] engages in an interview or a conversation with instructor during which the instructor asks 3-5 questions on the topic. The topics differ from one student to another (See Appendix A). The instructor might also ask the students to describe a picture. In Levels III & IV, the exam consists of one task only in which each student gives a presentation about a topic in front of the class. She completes the conversation and/or presentation in 10 minutes or less. The test session for all the students lasts from 8 a.m. to 1 p.m. non-stop and sometimes it lasts for two days, if the number of students enrolled in the speaking courses is large. The written part is administered on a separate day and it is similar to a writing test.

Instructors use a rubric for scoring the interm tests and another one for scoring the final exam (See Table 1 & 2). The scoring rubric consists of 4 skills: Content and Comprehension, (ii) grammar, (iii) fluency and voice control, and (iv) pronunciation. A total of 15 marks is allocated to each question on the interm test (6 marks are allocated to content and comprehension, 3 marks are equally allocated to grammar,

fluency and voice control, and pronunciation) with a total of 75 marks for the whole Interim test. The **students mark** is then converted into a mark out of 15 plus a mark out of 10 for the written test. In the final exam rubric in Table 2, 10 marks are allocated to each question, with a total of 50 marks for the whole final exam. For each question, she circles a score for each skill that reflects a student's performance (1 for poor, 3 for average and 6 for perfect). Scoring rubrics are not shown to the students. No comments or feedback are given to the students on their performance (strengths and weaknesses) during the test.

Table 1: Interim Test Scoring Rubric (Face-to-Face Assessment Method)

Test Qs	Content & Comprehension	Grammar	Fluency & voice control	Pronunciation	Total 15/75 marks
	6 marks Poor excellent	3 marks Poor excellent	3 marks Poor excellent	3 marks Poor excellent	
Q1	1 2 3 4 5 6	1 2 3	1 2 3	1 2 3	
Q2	1 2 3 4 5 6	1 2 3	1 2 3	1 2 3	
Q3	1 2 3 4 5 6	1 2 3	1 2 3	1 2 3	
Q4	1 2 3 4 5 6	1 2 3	1 2 3	1 2 3	
Q5	1 2 3 4 5 6	1 2 3	1 2 3	1 2 3	

Table 2: Interim Final Exam Scoring Rubric (Face-to-Face Assessment Method)

Test Qs	Content & Comprehension	Grammar	Fluency & voice control	Pronunciation	Total 50 marks
	10 marks Poor excellent	10 marks Poor excellent	10 marks Poor excellent	10 marks Poor excellent	
Q1	0 ½ 1 1½ 2	0 ½ 1 1½ 2	0 ½ 1 1½ 2	0 ½ 1 1½ 2	
Q2	0 ½ 1 1½ 2	0 ½ 1 1½ 2	0 ½ 1 1½ 2	0 ½ 1 1½ 2	
Q3	0 ½ 1 1½ 2	0 ½ 1 1½ 2	0 ½ 1 1½ 2	0 ½ 1 1½ 2	
Q4	0 ½ 1 1½	0 ½ 1 1½	0 ½ 1 1½	0 ½ 1 1½	

	2	2	2	2	
Q5	0 ½ 1 1½	0 ½ 1 1½	0 ½ 1 1½	0 ½ 1 1½	
	2	2	2	2	

4.3.2 Language Lab Assessment Method

In the language lab assessment method, all the students take the tests as a group at the same time, but they answer the questions individually. They complete 8-10 questions covering different themes, skills and language functions printed on a piece of paper. The test tasks are similar to those performed in class. To prevent cheating, two versions of the test questions are prepared (same questions, but in a different order). The students write their names on the question paper. A week in advance, the instructor asks the students to bring an MP3 player, practice using the MP3 player at home, test it at the beginning of the test session. Each student sits in her cubicle with her headsets on, reads the questions, thinks about them for a while, then records her responses to the questions in any order provided that she mentions the question number at the beginning of the answer. To minimize noise, the students keep their headsets on and record answers in a soft voice. Students are not allowed to write down answers to the questions and read them into the MP3 player. They are not allowed to re-listen to their answers, delete and re-record answers. At the end of the test session, the instructor collects the MP3 devices and test question papers of the students. The test session takes about 60 minutes.

When the students are tested in the language lab, the instructor listens to each student's MP3 at her own convenience at home or at the office. While listening to the recordings and scoring the students' responses, she writes comments on strengths and weaknesses for each student.

To assign marks, the lab instructor uses a scoring rubric as in Table 3 that consists of 3 skills: (i) idea generation or content; (ii) grammar and vocabulary; (iii) pronunciation and fluency. 50% of the mark is allocated to ideas (content); 30% to grammar and vocabulary, and 20% for pronunciation and fluency, i.e., 50% for the content (ideas) and 50% for the form (grammar, vocabulary, pronunciation, and fluency). For each skill category, a continuum that shows marks for the excellent, average, and poor performance are given in the cells. If the test is comprised of 10 questions, then 10 marks are allocated to each and the total marks is then converted into 25 for the Interm test, or 50 for the final.

Table 3: Interm and Final Exam Scoring Rubric for Each Question (Language Lab Assessment Method)

Test Ques.	Ques. Marks	Idea generation (content) 50%	Grammar and vocabulary 30%	Pronunciation and fluency 20%	Total Marks 100
		Poor excellent	Poor excellent	Poor excellent	
Q1	10	1 2 3 4 5	1 2 3	1 2	

Q2	10	1	2	3	4	5	1	2	3	1	2	
Q3	10	1	2	3	4	5	1	2	3	1	2	
Q4	10	1	2	3	4	5	1	2	3	1	2	
Q5	10	1	2	3	4	5	1	2	3	1	2	
Q6	10	1	2	3	4	5	1	2	3	1	2	
Q7	10	1	2	3	4	5	1	2	3	1	2	
Q8	10	1	2	3	4	5	1	2	3	1	2	
Q9	10	1	2	3	4	5	1	2	3	1	2	
Q10	10	1	2	3	4	5	1	2	3	1	2	

3. DATA COLLECTION AND ANALYSIS

The 69 students responded to a survey with open-ended questions about their preferences as follows: Which speaking assessment procedure do you prefer and Why? Compare test anxiety in both assessment methods; the mark you got on both; the kind of feedback you get from the instructor in both methods; the difficulty level of the questions in both methods; Which assessment method covers more speaking material and tasks practiced in class? How effective is each method in improving your speaking skill? (It needs revision: Punctuation)

Similarly, speaking instructors in the sample responded to a survey with the following open-ended questions: Which speaking assessment method do you prefer: face to face or language lab and why? Compare the face to face and lab assessment methods in terms of speaking skills and tasks covered by the test; item difficulty level and discrimination power in both; test anxiety in students; time required for administering the test in both methods; cheating in both methods; how tests are scored; scoring time and scoring accuracy in both methods; percentage of passing students in both methods; and feedback given to students in both methods. (It needs revision)

Students and instructors' preferences and views are reported in the results section quantitatively and qualitatively.

Moreover, to calculate the reliability coefficient and discrimination power of tests in the face to face and lab assessment methods, the course grades for students who have completed the speaking III course over 2 semesters were obtained from the Registration Department at KSU. The internal consistency reliability coefficient, and the percentage of students who got A, B, C, D, and F grades were computed for 120 students who were took the face-to-face test in the Fall semester, and for 69 students who took the lab test in the spring semester.

4. RESULTS AND DISCUSSION

6.1 Students' Views

85% of the students who had experience with face-to-face and lab assessment preferred the lab assessment method. The students reported that taking the exam in the language lab reduces their anxiety as they have some time to think about the answer, organize their ideas and are not being watched over by the instructor. In face-to-face assessment, the feel tense while waiting for their turn to take the test. In lab assessment, they are given 8-10 questions and the overall test mark (50 marks)

is divided by 8 or 10 questions. If the student misses a question, she will lose few marks, but in face-to-face assessment they will lose more marks as the whole exam consists of one or two topics. When they are given 8-10 questions to answer, different parts of the speaking material and skills and tasks are covered. But when one topic is given face to face, it covers only part of the material, skills or tasks practiced in class. The lab test questions are usually of different difficulty levels; **thus they** match students' ability levels. Teachers can give written feedback to students on strengths and weaknesses and areas of improvement because they grade the test at home not on the spot as in face-to-face assessment. This feedback is beneficial for them and enables them to work on their weaknesses. The students reported:

Najat: *Because the lab test has many questions that cover all types of skills and tasks that we have practiced in class, we study harder and practice more before we take the test.*

Dalal: *I feel more relaxed when I take the speaking test in the lab because I have more time to think and organize my ideas. I talk to myself and do not have to look at the teacher who is looking at me while talking face to face.*

Noura: *Because the lab test focuses on many details in the **course**. I now can speak fluently using correct grammar and pronunciation and can easily generate ideas.*

Nadia: *The face-to-face test is easy because I talk about one topic and answer a couple of questions and finish in less than 10 minutes. Knowing that the test is going to be easy, and most students pass, I do not work hard during the semester and do not practice outside the classroom unless I have to prepare like a presentation or a debate.*

Hanan: *When we take the speaking test in the lab, the teacher gives us back the test papers and MP3 **players, I know** what marks I got for each question, on which question I did well and which one I did not.*

Furthermore, the students indicated that face-to-face assessment is not fair to the students who take the test early. Those who take the test later have an advantage of knowing the test topics from students who have been tested. In lab assessment, all the students answer the same questions at the same **time and** there is no such leakage of test questions. The students added that their speaking skill improves better in the case of lab assessment than face-to-face **assessment as** the lab tests are more demanding and require a lot of studying, practice and reviewing. Face-to-face tests are easy and do not require much effort in studying and reviewing.

6.2 Instructors' Views

Unlike the students, 80% of the instructors surveyed preferred the face-to-face assessment for **several reason**. Although face-to-face tests take a long time to administer to all the students, it is easier to prepare and score. Depending on the speaking course the students are enrolled in, each student answers 3-5 questions on a topic in a conversation/interview with the instructor, describes a picture, or gives a

presentation they give in class about a topic they choose and prepare. In the latter case, the instructor does not prepare any test tasks. The conversations/interviews given on the test are realistic. The instructor can see the student's facial expressions, gestures, personality (shy or self-confident) and ability to extemporate. When the student picks a topic from the basket, she does not have time to think about it. But if she hesitates or cannot generate ideas, the instructor prompts her by asking questions or requesting clarification of certain points. It is easier to grade face-to-face tests using the rubrics in Table 1. But in Table 2, it is time-consuming to add fractions in the rubric and there is a possibility of making mistakes while adding them up. The test tasks are easy for all the students. As a result, all of the majority pass the course and only those who missed many classes fail.

On the other hand, the instructors mentioned several shortcomings. For example, the test questions do not cover the whole material, skills or tasks covered in class or the textbook. The students feel nervous, and anxious during the test because they engage in a conversation/interview with the instructor or give a presentation in front of their classmates. Moreover, the test conditions for the students in face-to-face assessment are different. Teacher fatigue may affect her ability to concentrate as the test session lasts 5 hours or more. Performance of students who take the test in the morning may not be like that of students who take the test in the afternoon and who wait for their turn to be tested for a long time. Students who take the test later have an advantage of knowing the test topics given to students who were tested earlier. This will give them an opportunity to prepare something and even expect which topics might be given to them, especially in the case of large classes when some students take the test the next day. In large classes, the instructor might give the same topics to several students because it is not possible to prepare 60, 70 or even 80 different topics. There is no chance to give feedback to each student in face-to-face assessment as there is no time. The student will only see her grade without knowing how the grade was assigned. Some instructors wrote:

Suad: *Because speaking skill is difficult to assess and requires a lot of time, I give few, easy questions to finish giving and scoring it in shortest time possible.*

Fatima: *I give few, easy questions to help the students pass and to avoid their complaints because the speaking skill is difficult.*

Ghadah: *When the students fail, they will have to take the course over, I will see them again in my class the following semester and will have more students in my speaking class than when all students pass. That is why I give easy questions that the students can answer and hence pass.*

By contrast, 20% of the instructors prefer the lab method because it has numerous advantages. The lab test takes less time (about 60 minutes) to administer than face-to-face assessment which takes at least 5 hours. Each student can answer between 8-10 questions or tasks. All students answer the same set of questions. Unlike face-to-face assessment, there is no discrepancy in question difficulty level and amount of details required. All students take the lab test at the same time and under the same conditions. There is no way for any student to cheat or know the types of test questions. The tasks/questions/topics given cover a wide range of skills practiced in class, some of which are easy, some are difficult, and some are of average difficulty.

The total exam mark is divided over 8-10 questions rather than one or two questions in face-to-face assessment. If a student cannot answer a question, she will lose few marks only. But in face-to-face assessment, she will lose the whole exam mark if there is one question and 50% of the mark if the test has 2 topics. Each student receives feedback on her performance as the instructor writes down comments for each student as the instructor scores answers at her own convenience. Since the lab test has many questions with many details, it sorts out students according to their mastery of the speaking skills tested. You will find **students who** get an A, B, C, D and even F; unlike face-to-face assessment where students talk about one easy topic, and most students pass with high grade and no failures. The scoring rubric used for scoring lab tests is more balanced as it allocates 50% of the question mark to content (ideas) and 50% to the form (vocabulary, grammar, pronunciation, fluency...etc).

Instructors who use lab assessment mentioned some shortcomings. For example, lab assessment lacks real-life interaction with another speaker as the student talks to herself while recording her answers. Anyway, topics on a lab test do not require a conversation or interview with another person. Unlike face-to-face assessment in which the instructor assigns each student a mark on the spot, scoring lab tests is time consuming. But the instructor can grade the recordings at her own convenience. Time spent on scoring lab test is worth it because the instructor can write feedback, can re-check answers, and compare different students' answers. Instructors commented:

Majda: It is true that scoring the student's recordings is time consuming, but I grade at my own convenience. I am not under pressure as in face-to-face testing where I have to finish testing a large number of students in a day and I test students non-stop, which is very tiring. In the end, the total time spent on administering and scoring the test in both methods is the same.

Ahlam: In lab testing, I have students' recordings and I can listen to them again to double check that I did not overrate or underrate responses to any question. I can compare students' responses, recheck comments, and modify the marks I have assigned using the scoring rubric.

Afnan: Lab testing results in better mastery of the speaking tasks as the students answer a variety of questions that cover the whole material practiced in class and contained in the textbook.

1.4 Reliability, Validity, and Discrimination Power of Both Assessment Methods

Statistical analysis of the students' speaking test scores obtained from the Registration Department showed that face-to-face assessment is less reliable than lab assessment. The internal consistency reliability coefficient is .41 for the face-to-face test scores compared to .71 for the lab test scores. Some factors that affect test reliability is the number of questions on a test, number of skills tested, the difficulty level and variability of the tasks covered by the test questions. Taking these factors into consideration, the lab test is more reliable.

Another intervening variable that affects reliability of face-to-face tests is that the students do not take the test under the same conditions. The **topics the students talk about are** not identical in theme and difficulty levels and the times they take the test during the day (fatigue and alertness issues). Those who take the test later know the topics other students spoke about earlier (leak out of test topics).

In addition, face-to-face assessment is less valid due to the inclusion of a written component as part of a speaking test. This way instructors are testing the writing skill, not the speaking skill.

Moreover, scoring responses and assigning marks in face-to-face testing is less reliable because the instructor has to make a decision regarding the students' response in a short time. There might be variability in assigning marks to different students especially those who take the test early and those who take the test late. Once the students leave, the instructor cannot double check and compare their performance with that of other students. This was confirmed in a study by [12] which indicated that there is variability in raters' scoring decisions and that rater's level of expertise affects scores. The time taken to reach a scoring decision was found to be associated with the reliability and accuracy of scores.

An additional characteristic of a good test is that it should discriminate between those who have mastered and those who have not mastered the skills under study [30]; [31]). It should also have a high discrimination **power especially** because students at COLT are going to take 6 interpreting courses in the subsequent levels. In this respect, the lab test is more discriminating because the students respond to numerous questions that vary in task, theme, and difficulty level. The questions are representative of the skills and tasks taught in the textbook. The lab test sorts out students into excellent, very good, good, below average and those who have not mastered the speaking skills. By contrast, face-to-face testing is less discriminating because the students speak about one easy topic only. End of course grades in Table 4 show no failing students, 3.5% passed with a D, 13% passed with a C, 43.5% passed with a B, and 40% passed with an A in the face-to-face assessment. On the contrary, Table 4 shows that lab testing is more discriminating as 32% of the students failed the course, 25% got a D, 13% got a C, 18.5% got a B and 10% got an A.

Table 4 Distribution of Students' Letter Grades in Speaking III for Face-to face and Lab Assessment

	A+	A	B+	B	C+	C	D+	D	F
Face to face Assessment	17.5%	22.5%	27.5%	16%	10%	3%	1%	2.5%	0%
Lab testing assessment	1.5%	8.5%	8.5%	10%	6%	7%	8.5%	17%	32%

N (Face to face) = 120

N (lab) = 69

The preference for language lab assessment by the students and some instructors in the present study is supported by findings of a study by [10] in which Taiwanese

college students answered questions, described a picture, gave a presentation in the language lab and responded to a survey on an audiotape. [10] results showed that the students performed better on the speaking task of answering questions by exhibiting more fluency and complexity.

5. RECOMMENDATIONS AND CONCLUSION

To improve the quality of speaking tests at COLT, no matter what the assessment method is, the study recommends the following:

- 1) Since the test length affects test reliability, face-to-face tests should require the students to talk about at least 3 topics on the Interim test and at least 5 topics on the final. If conducted in the language lab, interim tests and finals should contain twice as many questions as those on a face-to-face test.
- 2) Since face-to-face testing is time consuming and the instructor has to test a large number of students individually, the testing time will be reduced if two or more speaking instructors collaborate in administering a speaking test to 30% or 40% of the students per assessor. In this case assessors should discuss the assessment method in advance and use a unified scoring rubric. Collaborative testing will also help increase the number of questions/topics on a speaking test and make the test more reliable.
- 3) In all testing modes, test instructions must specify the type and number of details and type of task that must be taken into consideration while talking about a topic, emphasize correct sentence structure, pronunciation, stress, intonation, **fluency and** so on.
- 4) In selecting speaking test topics, the topics should be comparable to but not a repeat or identical to what the students have practiced in class.
- 5) In scoring speaking tests, a scoring rubric must be prepared. 50% of the total test mark should be allocated to the content (ideas, organization, cohesiveness), 50% to the form (grammar, pronunciation, articulation of phonemes, stress, and intonation and fluency). In scoring content (ideas), marks should be allocated to each detail. In grading grammar, pronunciation, intonation, stress, **fluency...**etc., holistic grading is used. If what the student has said is off point, no marks are given, even if her grammar and pronunciation are good. In a study with Taiwanese college students, [15] found that the rating method had a noticeable impact on how teachers judge students' performance and interpret their scores.
- 6) No matter which testing mode is used, students should be given feedback on their performance. Most Taiwanese students in [15] study considered feedback helpful in improving their speaking ability.
- 7) Raising speaking instructors' awareness of the purpose of speaking assessment, and different testing modes, individual, paired, group face-to-face testing, lab testing is necessary. Training speaking instructors in selecting test questions, administering face-to-face and lab tests, and scoring them in a fair and reliable manner as described earlier is recommended. [12] indicated that training raters of

speaking tests resulted in increased inter-rater correlation and agreement, and improved correlation and agreement with established reference scores.

Finally, speaking instructors must always remember that they are not preparing the students enrolled in speaking courses at COLT to pass the courses only, but the students need to acquire advanced speaking skills that will help them in the interpreting courses that they will later during the translation program and in interpreting skills that they need to use as interpreters after graduation. (It needs revision: problem of structure)

REFERENCES

- [1] Vahid Panahzadeh; Bitā Asadi. On the impacts of pressured vs. unpressured online task planning on EFL students' oral production in classroom and testing contexts. *Eurasian Journal of Applied Linguistics*. 2019; 5(3), 341-352.
- [2] Jing Wei; Lorena Llosa. Investigating Differences between American and Indian Raters in Assessing TOEFL iBT Speaking Tasks. *Language Assessment Quarterly*. 2015; 12(3), 283-304.
- [3] Timothy L. Farnsworth. An investigation into the validity of the TOEFL iBT speaking test for international teaching assistant certification. *Language Assessment Quarterly*. 2013; 10(3), 274-291.
- [4] Shahzad Karim; Naushaba Haq. An assessment of IELTS speaking test. *International Journal of Evaluation and Research in Education*. 2015; 3(3), 152-157.
- [5] Christina Judy Fernandez. Behind a spoken performance: Test takers' strategic reactions in a simulated part 3 of the IELTS speaking test. *Language Testing in Asia*, 8(Article 18). 2018.
- [6] Sally O'Hagan; John Pill; Ying Zhang. Extending the scope of speaking assessment criteria in a specific-purpose language test: Operationalizing a health professional perspective. *Language Testing*. 2016; 33(2), 195-216.
- [7] Rosemary Wette. English proficiency tests and communication skills training for overseas-qualified health professionals in Australia and New Zealand. *Language Assessment Quarterly*. 2018; (2), 200-210.
- [8] Elizabeth Manias; Tim McNamara. Standard setting in specific-purpose language testing: What can a qualitative study add? *Language Testing*. 2016; 33(2), 235-249.
- [9] Xun Yan; Lixia Cheng; April Ginther. Factor analysis for fairness: Examining the impact of task type and examinee L1 background on scores of an ITA speaking test. *Language Testing*. 2019; 36(2), 207-234.

- [10] Huei-Chun Teng. A study of task types for L2 speaking assessment. Annual Meeting of the International Society for Language Studies (ISLS) (Honolulu, HI, Apr 2-4). 2007.
- [11] Ying Zhang; Catherine Elder. Investigating native and non-native English-speaking teacher raters' judgements of oral proficiency in the college English test-spoken English test (CET-SET). *Assessment in Education: Principles, Policy & Practice*. 2014; 21(3), 306-325.
- [12] Lawrence Edward Davis. Rater expertise in a second language speaking assessment: The influence of training and experience. Ph.D. Dissertation, University of Hawai'i at Manoa. ERIC ED554751. 2012.
- [13] Michael D. Carey; Robert H. Mannell; Peter K. Dunn. Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*. 2011; 28(2)201-219.
- [15] Mu-hsuan Chou. Teacher interpretation of test scores and feedback to students in EFL classrooms: A comparison of two rating methods. *Higher Education Studies*. 2013; 3(2), 86-95.
- [14] Takanori Sato. The contribution of test-takers' speech content to scores on an English oral proficiency test. *Language Testing*. 2012; 29(2), 223-241.
- [16] James Simpson. Differing expectations in the assessment of the speaking skills of ESOL learners. *Linguistics and Education: An International Research Journal*. 2006; 17(1), 40-55.
- [17] Heng-Tsung Danny Huang. Modeling the relationships among topical knowledge, anxiety, and integrated speaking test performance: A Structural equation modeling approach. Ph.D. Dissertation, The University of Texas at Austin. ERIC ED520180. 2010.
- [18] Shelley Staples; Geoffrey T. Laflair,; Jesse Egbert. Comparing language use in oral proficiency interviews to target domains: Conversational, academic, and professional discourse. *Modern Language Journal*. 2017; 101(1), 194-213.
- [19] Fahmi; Bambang Widi Pratolo; Nurul Amalia Zahrani. Dynamic assessment effect on speaking performance of Indonesian EFL learners. *International Journal of Evaluation and Research in Education*. 2020; 9(3), 778-790.
- [20] Shu-Chen Huang. Understanding learners' self-assessment and self-feedback on their foreign language speaking performance. *Assessment & Evaluation in Higher Education*. 2016; 41(6), 803-820.
- [21] Akiyo Hirai; Rie Koizumi. Development of a practical speaking test with a positive impact on learning using a story retelling technique. *Language Assessment Quarterly*. 2009; 6(2),151-167.

- [22] Michael Stephenson. Setting the group agenda: Negotiating deontic rights through directives in a task-based, oral, L2, group assessment. *Classroom Discourse*. 2020;11(4), 337-365.
- [23] David A. Leaper; James R. Brawn. Detecting development of speaking proficiency with a group oral test: A quantitative analysis. *Language Testing*. 2019; 36(2), 181-206.
- [24] Özlem Yalçın-Çolakoglu; Merve Selçuk. Assessing individual and group oral exams: Scoring criteria and rater interaction. *Advances in Language and Literary Studies*. 2019; 10(1), 147-153.
- [26] Gwyneth Gates; Troy L. Cox; Teresa Reber Bell; William Eggington. Line, please? An analysis of the rehearsed speech characteristics of native Korean speakers on the English oral proficiency interview--computer (OPIc). *Language Testing in Asia*, Article 10. 2020
- [26] Julie Norton. The paired format in the Cambridge speaking tests. *ELT Journal*. 2005; 59(4), 287-297.
- [27] Anne Lazaraton. Process and outcome in paired oral assessment. *ELT Journal*. 2006; 60(3), 287-289.
- [28] Lindsay Brooks. Interacting in pairs in a test of oral. Proficiency: Co-constructing a better performance. *Language Testing*. 2009; 26(3), 341-366.
- [29] Khaled Barkaoui; Lindsay Brooks; Merrill Swain; Sharon Lapkin. Test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics*. 2013; 34(3), 304-324.
- [30] Reima Al-Jarf. Issues in Assessing the Speaking Skill in EFL. *The International Conference on Language Testing and Assessment*. Guangzhou, China. November 27-30, 2015
- [31] Reima Al-Jarf. Linguistic and measurement considerations in Translation tests. *13th World Congress of the Association Internationale de Lingistique Appliquee (AILA)*. Singapore, December 16-21, 2002.

APPENDIX

APPENDIX A: SAMPLE TESTS IN FACE-TO-FACE ASSESSMENT

TEST 1: Each student answers 1 question only

- 1) Tell about your family. How many members are there? Are you the oldest/middle/youngest child? Where does your family live? Do you have any responsibilities at home? Do you have any family traditions/habits that you do regularly? What are they? What do you like/dislike about these traditions/habits and why?

- 2) Tell about your school life. Where did you go for elementary, junior high and secondary school? What was the most interesting part of your school life? What was the worst part of your school life? What was the hardest thing to learn? What was the funniest thing to learn?
- 3) Tell about your collage life. Are you happy about your major? Is it like what you have expected? What is your most wonderful experience till now? What is your favorite course? When do you expect to graduate? What are your plans after graduation?
- 4) What are your plans for your career? What do you like to be? Where do you like to work? why? Do you like to study for an M.A. or Ph.D. and why? Do you think that studying for five years is enough and why?
- 5) Tell about yourself. What kinds of movies do you like? why? What are your favorite hobbies? why? What kinds of sports do you like to watch/practice yourself? why? What is your favorite season? why? What is your favorite food? why?
- 6) What are you looking forward to in your social life? When do you like to get married? Do you like to have a big, fancy wedding or simple one? why? Who do you like to be married to? Who are you going to invite? When do you like to have children? What names do you like to give them?

TEST 2: The same question is used with each situation. Each student responds to one situation only.

Q: What would you do and react in the following situation:

- *You are alone in a classroom the first day of school. Then another student comes in.*
- *You are a professor. A new student comes into your office to meet you.*
- *You are in your office. A new worker enters and seems lost and confused.*
- *You are in a friend's party. You see a new person standing next to you.*
- *You meet a friend after being apart for a long time.*
- *You meet a family member after being apart for a long time.*
- *You are in a gym. You see another student from your class.*
- *You are the only passenger in a bus. Another person gets on and sits next to you.*
- *You are in a friend's dorm room. A new person enters the room.*
- *You come to a party with a friend. The party is very crowded. You are ready to leave, but you cannot find your friend.*
- *You see a small child at the park. The child seems lost, and he starts to cry. You are the only adult around him.*
- *You just saw car accident. The person who caused the accident drove away, but you got a good look at that person.*
- *Your grandmother is visiting you from another city. She went out for a walk in the park and has not returned yet. She went at 4:00 p.m., and it is 9:00 p.m. now.*
- *You bought a coffee maker, but when you took it home and tried to use it, it did not work.*

- *Your friend asks you to return a CD you borrowed from her several weeks ago. You cannot remember what you did with it.*
- *You were at the mall. You bought an expensive watch. Then you entered another store and looked at some clothes. When you got home, you realized that you forgot the watch at the clothes store.*
- *You went to the mall. When you got home, you realized that you forgot your mobile there. It includes some of family, friends and your pictures.*
- *You invite your friend out for out for dinner to celebrate her graduation. When the waiter brings the bill, you realize that you forget purse/money at home.*
- *You are having people to your home for a tea party. Your mother asks to watch the cake. You forget about it, and the cake gets ruined (not good for eating).*
- *You are walking down the street. Suddenly, you see a friend in front of you. You run up to her, but when she turns around you, you discover she is a stranger.*
- *You sit down in a chair. Suddenly, you realize that you are sitting on your friend's glasses. The glasses are ruined.*
- *Someone comes up to you at a meeting. You do not remember that person, but the person knows your name and says that the two of you met before.*

TEST 3: Each student describes 1 picture or advertisement only.

- Look at the following picture and describe it. What do you like and do not like about it? Why?
- Look at the following advertisement. Is it a good or bad advertisement? Why?

APPENDIX B: SAMPLE TESTS FOR THE LANGUAGE LAB ASSESSMENT

TEST 1 (Speaking III): Each student answers all 10 questions.

- Read each word out loud and give its Arabic meaning:
 - *potable, desalination, degeneration, composer, tonsillitis, chancellor, cultural invasion, commodities, senate, brain drain, rehabilitation, reclamation, keynote speaker, archaeology, symposium, indigestion, consumption, favoritism, high-intensity sounds, delegates (10 marks)*
- What are the disadvantages of hunting animals like whales, elephants, tigers, snakes in large quantities? (10 marks)
- Give 5 characteristics of a professor that you admire. (10 marks)
- Give 5 causes of high blood pressure. (10 marks)
- Factors that make a small business successful. (10 marks)
- How can we improve the teaching of English in Saudi schools? (10 marks)
- Give 5 actions that would be taken to encourage unemployed college graduates to work as a cashier, taxi driver, at a gas station, and reduce foreign labor in the kingdom. (10 marks)
- How would you help students reduce wasted time and have more time for studying and recreation? (10 marks)
- How can we stop students from littering at the university? (10 marks)
- Give at least 5 advantages and disadvantages of social media? (10 marks)

UNDER PEER REVIEW