

# Speaking Skill Assessment Instrument Validity: An Investigation into Instructors' Perceptions

Abdullah Alshakhi<sup>1</sup>

## Abstract

*To EFL learners, speaking the target language with satisfactory fluency is a highly significant aspect of learning the language. Assessment of learners' progress and achievement in speaking skills is equally significant to identify the course of further action, whether it is for feedback, placement, or grading. At the English Language Institute (ELI), King Abdulaziz University in Saudi Arabia, learners' proficiency in spoken English is assessed for end-of-the-term grading using Interview and Speaking Project as test instruments. The Interview questions are random, while the Speaking Project involves collection of materials on a chosen topic and report writing, followed by presentation and question-answer session. An empirical research study was conducted to examine ELI teachers' perceptions on the reliability and validity of the two assessment instruments to accurately measure learners' proficiency in speaking skill and meeting the requirements to reflect the learning outcomes designed for the speaking English course. The data were collected through a survey questionnaire and a structured interview. The surveyed instructors were also encouraged to express their opinions on the role of formative assessment in determining learners' final grades in English speaking skill.*

**Keywords:** *Speaking skill; Assessment techniques; Instrument validity; Formative and summative assessment*

## Introduction

Fluency development in a foreign language depends upon so many factors, proper assessment being one of them. Right kind of assessment is especially significant in a Foreign Language learning situation, such as English as a Foreign Language (EFL) situation, where learners have only minimal chances for direct exposure to native spoken language, and even the indirect exposure to the spoken form of the target language is restricted to the classroom hours where most often the learners are just passive listeners (Al-Seghayer, 2022). Leeming and Harris (2022) consider measurement of language learners' development in speaking proficiency very important for practicing language teachers. However, in the words of the researchers, effective and efficient measurement is challenging (Taylor, 2011). Under such circumstances, proper and timely assessment of learners' achievements in spoken English is very crucial to identify the gaps in learning, monitor the progress of learners, provide feedback on their linguistic development, and implement remedial teaching or other such measures to address the issue (Alharbi & Surur, 2019). English Language Institute (henceforth ELI) is a centre of education dedicated to the teaching of English to first year undergraduates. ELI represents a centralized education system governing the teaching of English and conducting examinations at all the campuses affiliated to the university. As regards the assessment in learners' spoken English, the ELI has developed a set of assessment techniques. For final grading, the assessment test consists of a face-to-face interview of the candidate, and a speaking project. In the interview technique, the examiner asks a set number of questions to the candidate. The questions are based on the units from the book prescribed for study in the class. For the Speaking Project, the learners select a topic for presentation, collect materials from various sources, prepare a project report, and finally present the topic before the examiner who puts up several questions related to the project. The ELI has a patterned rubric for the oral examination as well as the grading scale for the final grading of both types of speaking assessment.

In general, EFL teachers in Saudi Arabia lack a voice in the assessment process (Ali et al., 2019; Ariani, 2014). The ELI assessment process follows a set pattern; therefore, the teachers have very little say, either in determining the examination questions for the face-to-face interview or speaking project, or in marking and grading the responses of learners being assessed. However, an informal opinion-gathering from a few ELI teachers on the quality of assessment pattern revealed that teachers believe that commonly learners

---

<sup>1</sup> English Language Institute, King Abdulaziz University, Saudi Arabia; amalshakhi1@kau.edu.sa

display more spontaneity in their responses for the interview questions, whereas, for the speaking project their responses sound more scripted in nature since they are aware of the type of questions expected of them at the end of presentation. The informal opinion gathering proved to be the genesis of the present study as it led the researcher to believe that since teachers are also stakeholders in the success of their students, their perceptions on the speaking assessment pattern are very important and may throw more light on the reliability and validity of the assessment techniques in use. In addition, the significance of alternative forms of assessment, such as enhanced value of formative assessment for final grading, is also set to be explored further.

### *Research Background*

Speaking the target language with satisfactory fluency is a highly significant aspect of learning the language. Assessment of the learners' achievement in the speaking skill is equally significant to identify the course of further action, whether it is for feedback, placement, or grading. Identifying learning gaps, providing feedback and remedial teaching are the common actions taken after formative assessment, while candidate placement and grading are determined after summative assessment. Adult EFL learners enrolled at the English Language Institute (ELI) in Saudi Arabia are assessed on their progress in speaking skill formatively as continuous assessment in class, whereas for the final grading, the assessment is largely summative. The interview conducted for the final speaking skill assessment, although based on the units taught in class, is largely unstructured. On the other hand, the speaking project is a structured assessment pattern in which the learners are assigned a project in advance.

Both interview and speaking project are summative assessment patterns. Formative assessment marks play little or no role in determining the final grades of learners. The pertinent question to ask at this juncture is: in the opinion of ELI teachers, which pattern of assessment of the speaking skill is a more reliable and valid instrument of assessment of the skill in question? Summative assessment is not the sole form of speaking assessment at ELIs as teachers conduct formative assessment in their daily classes and provide the learners feedback on their progress. However, formative assessment is not marked for the final grading, although it may play an indirect role in the final assessment. A corollary question is: In the opinion of ELI teachers, what is the role of formative assessment in the final determination of grades in speaking skill in this scenario? And another related question as well as concern is whether formative assessment should bear more weightage than it has at present in the assessment of speaking skill of adult EFL learners in Saudi universities, particularly enrolled in ELI courses.

### *The Research Problem*

Undergraduate students enrolled at ELI for compulsory English requirements are required to take four English-speaking tests over two semesters, which are of two different types. Both are examples of summative assessment, though. The emphasis is, thus, on summative assessment. This type of assessment does not test the learners' progress in spoken skills achievement properly since the test primarily assesses learners' memorization and rote-learning which reflect their speaking skill only partially. The assessment tasks are pre-designed, and instructors have no say either in their design or grading process. The researcher's observation, and an informal opinion gathering with a few instructors, made him realize that it is important to know how the instructors' involved in the assessment process feel about the instruments of assessment, that is, in the opinion of instructors how reliable and valid are the instruments of speaking skill assessment techniques in use at ELI. The researcher also realized that there is no previous research literature available on this academic issue. For instance, there might be some incongruity between speaking assessment tasks and the learning outcomes. Or, in the opinion of teachers, the present assessment tasks may not be fully aligned with student learning outcomes. In other words, do the assessment tests used by ELI teachers in Saudi Arabia test what they are supposed to test? The corollary issue is: Is teacher-oriented assessment the only way to assess spoken English? Can peer assessment be also considered as a valid and reliable form of assessment? Moreover, can formative assessment be given more weightage in the final grading process? These are some of the preliminary issues related to present assessment techniques.

However, the researcher's observation is crystalized on the two major issues:

- At the present moment, there is no research on the validity and reliability of the speaking assessment instruments used in ELI at KAU. There is a lack of research literature on the issue.
- Both assessment instruments are summative in nature. One of them gives the learners an opportunity to be spontaneous and more natural in speech, while the other is more of a scripted exercise in nature and leaves little room for the learner to be more creative in responses. The difference between the two, in the eyes of the real users, hasn't been explored in a research study.

## Literature Review

### *Language Assessment*

Language assessment measures the level a learner has achieved in learning the target language as a result of continuous teaching. It monitors and checks whether the intended learning objectives have been achieved in the language. In Macalister and Nation's (2019) opinion, there are six major types of monitoring and assessment. They include placement assessment, observation of learning, short-term achievement assessment, diagnostic assessment, achievement assessment, and proficiency assessment. Some of these types fall in the formative assessment category, while the others can be categorized as summative assessment. To Brown (2019), summative and formative assessments are the most common types of assessment.

### *Formative Assessment*

Formative assessment is used to monitor student learning on a regular basis. The purpose is to provide feedback to students by identifying their strengths and weaknesses. Formative assessment has always been in practice, however, the assessment caught researchers' attention prominently after the publication of a review by Black and Wiliam (1998) where the researchers presented formative assessment as a powerful classroom intervention. In an attempt to push for educational reforms, Black and Wiliam (1998) critiqued formative assessment as a primary approach, especially concerning the low achievers. Their research study was supported by researchers and embraced by the teaching communities and policymakers worldwide. Formative assessment primarily uses assessment as a launching pad to devise subsequent instruction (Black & Wiliam, 2004). This also involves constructive feedback to learners and identifying gaps in their learning. However, teachers should handle formative assessment only after being trained in the foundational content-pedagogical knowledge. Otherwise, their assessment may lead to wrong decisions resulting in students' regress, instead of progress.

Alahmadi et al. (2019) observe that formative assessment has not been used to overcome the challenges faced by the Saudi students at Taibah University in the speaking test. This study concludes that formative assessment helps Saudi students to overcome the challenges they face in speaking test. It is also recommending constructive feedback to improve their speaking performance.

### *Summative Assessment*

Summative assessment is the end-of-the-term assessment, and its purpose is the final grading of learners. The focus of summative assessment is checking to what extent the learning outcomes set in the beginning of the course have been met, and where the individual learners stand when rated on a scale measuring the achievement of the learning objectives. Summative assessment has its own positive and negative aspects. For example, Shaaban (2000) believes that foreign language classes predominantly emphasize summative assessments designed to gauge learners' proficiency in language structures and linguistic precision. Shaaban (2000) believes that the purpose of language assessment must be to evaluate learners' ability to effectively communicate in the target language. Green (2013), on the other hand, considers the positive aspect in teachers' use of summative assessment. He says that even though summative speaking assessments often elicit resistance from students, educators persist in utilizing this method to ascertain whether learners have achieved language learning objectives or not.

### *Assessment of Speaking Skill*

If a language is learnt as a foreign language, assessment of learners' performance in speaking the language gets a heightened emphasis since the language in question is neither used as a second language in the learners' community nor as a language of communication. For example, in Saudi Arabia English is taught as a Foreign Language (EFL), and as a result, learners' exposure to spoken English outside formal instruction environment is either minimal or completely negligible. Assessment of adult EFL learners' spoken English in Saudi Arabia needs special attention as the spoken language bears multiple dimensions, and all those dimensions need to be assessed in a fair manner. In speaking skill assessment, several factors (such as, time allowed to speak, topic of speech, learners' range of vocabulary, learner anxiety, etc.) affect the learner performance, which get more accentuated in summative assessment as compared to formative assessment which is not so much affected by the factors cited above. It's for such factors that more and more researchers favor classroom formative assessment to yield substantial results in learning gains (Black & William, 1998).

### *Interviews as an Assessment Technique*

Face-to-face interview/interaction with learners is widely used as an assessment technique for foreign language speaking skill (O'Sullivan et al., 2002; O'Sullivan, 2002). However, some researchers (e.g. Ross, 2017) have cast doubts about the validity of the interview technique to assess learners' conversational ability. Ross (2017) cites several researchers who feel that interview is a form of institutional interaction rather than a fair assessment of speaking skill. Ross (2017) presents a few alternatives to the face-to-face interview, such as the simulated oral proficiency interview, the computer-delivered oral proficiency interview, speaking assessments that deliver narrative tasks via the Internet, and fully automated semi-indirect tests such as the Versant test as more viable assessment techniques of adult learners' speaking skill.

### *Review of Related Research Literature*

Research on the reliability and validity aspects of assessment patterns, particularly in EFL learners' spoken English assessment techniques, is very scanty in Saudi Arabian context. There is plenty of research literature available on the difficulties of adult EFL learners in developing fluency in English and the issues impacting the spoken English performance of learners in Saudi Arabia, touching on the topic of assessment in general, not specifically the assessment of spoken English skills (see, for instance, Ahmed & Alamin, 2014; Al-Hassaani & Al-Saalmi, 2022; Ali et al., 2019; Almoosa & Alzahrani, 2022; Almuntasheri, 2016; Alrasheedi, 2020; Al-Seghayer, 2022; Alsubai'ai, 2021). Though a few scholars have dealt with speaking skill assessment as well (Huang et al., 2020; Jayaraman, 2017; Leeming & Harris, 2022; Luoma, 2004) but their studies are out of Saudi Arabian contexts. The research paper by Varela and Palacios (2013), for instance, examines some of the best-known proficiency tests in English, with particular focus on the oral component. The researchers contend that testing systems in English, especially concerning oral skills, need to be continuously evaluated and reformulated as is the case with teaching processes in general, since English holds a special nature in the TEFL contexts. In the assessment of speaking, Luoma's (2004) focus is on construct validity. Luoma (2004) emphasizes that teachers and researchers should first define what kind of speaking they want to test, and only then develop or adapt test tasks and rating criteria. Macalister and Nation (2019), on the other hand, emphasize that monitoring and assessment must take account of the environment in which the course will be used. The scholars look at important considerations in assessment, such as criterion-referenced and norm-referenced tests, tests of language knowledge and language use, performance objectives, and determining the reliability, validity, and practicality of assessment instruments.

In research studies, the qualitative value of speaking skill assessment patterns is hardly investigated. For instance, the study of Ali et al. (2019) is focused on the theoretical aspects of general testing in English language teaching and its significance in EFL contexts, while Almoosa and Alzahrani (2022) have explored the assessment practices used by teaching staff in Saudi universities, particularly how the practices have undergone a change during the coronavirus 2019 pandemic. Ahmed and Alamin's (2014) findings show a positive improvement in Saudi EFL learners' speaking abilities, though there is still room for development within the realm of possible future improvement.

Some attention has been paid to the evaluation of formative assessment too (Almuntasheri, 2016; Alrasheedi, 2020; Alsubaiai, 2021; Herman et al., 2015). Alsubaiai (2021), for instance, notices that formative assessment influences students' outcomes positively. The researcher opines that English teachers' perception is an important factor in determining the assessment strategies used in class and teachers' experience and education level influences their perception towards formative assessment. Investigation of factors influencing speaking performance of Saudi EFL learners has also been carried out at great length (Alsubaiai, 2021; Al-Seghayer, 2022). Al-Seghayer's (2022) study also considers the overarching features of existing assessment methods practiced in English classrooms in Saudi Arabia. The researcher opines that the assessment practices followed by Saudi English teachers have detrimental effects on the development of Saudi learners of English as a foreign language. According to Al-Seghayer (2022), the current practices in language assessment do not reflect cutting-edge or modern approaches to assessment. Herrera and Macías (2015) strongly believe that formative assessment deserves a due place. In their words, assessment continues to be regarded mainly as the summative evaluation which informs teachers of students' success or failure in their learning process based on a numeric scale. This is a very narrow approach and shows a lack of preparation and training in teacher education and professional development programs. To Herrera and Macías (2015), focusing on the accuracy of the instrument (e.g. test, exam) is just a small part of a larger picture in language assessment. The main focus should be on the acquisition of skills, knowledge, and competencies in assessment. Al-Wassia et al. (2015) investigate the challenges faced by faculty and students in implementing assessment for learning. They take into account the activities, capabilities, enablers, and indicators which could impact learner performance. Ali et al. (2019) note that in Saudi EFL context, EFL teachers lack a voice in the process of language assessment. The researchers suggest that teachers need training to develop their assessment literacy. On similar lines, Ariani (2014) says that only a handful of EFL teachers are aware of classroom assessment techniques, while all other make the same, reckless decisions assessing students' learning and use the same and unspectacular assessment techniques.

The other assessment research areas closely related to spoken English assessment which can boast of sufficient research are teachers' assessment literacy (Al-Wassia et al., 2015; Boyles, 2006; Engelsen & Smith, 2014; Herrera & Macías, 2015), lack of teachers' agency in spoken English assessment (Ali et al., 2019), teachers' lack of awareness of spoken English assessment techniques (Ariani, 2014), and teachers' challenges in assessing speaking skill. Lack of assessment literacy among English teachers has drawn scholars' attention in Saudi Arabia too. Rauf and McCallum (2020) note that in Saudi Arabia there are major gaps in EFL teachers' assessment literacy and incongruities between the tasks and learning outcomes. Their study is oriented towards investigating Saudi EFL teachers' development of assessment tasks and the extent that these tasks are aligned with student learning outcomes. Similarly, Umer et al. (2018) observe that in the context of Saudi higher education, EFL teacher assessment literacy is replete with loopholes and a serious incongruity between teachers' assessment tasks and course learning outcomes. For instance, higher order learning outcomes were not assessed at all.

A few studies that explore the challenges with speaking skill assessment patterns in Saudi Arabian EFL education contexts are discussed here. To Alahmadi et al. (2019), at Taibah University formative assessment has not been used to overcome the challenges faced by the Saudi students in the speaking test. The present summative test consists of three tasks and students are required to go through them all. The researcher concludes that formative assessment helps Saudi students to overcome the challenges in speaking test. Alharbi and Surur (2019), on the other hand expresses satisfaction with the variety of communicative oral assessment techniques teachers use in Saudi Arabia and that they are utilizing effective assessment procedures in assessing their students' speaking skills. Alharbi and Meccawy (2020) suggest the use of Online Student Response System (OSRS), Socrative, which is a web-based platform, as a tool for formative assessment in EFL classrooms. The tool provides several pedagogical benefits for language learners. In the words of Bailey (2017), assessment should not remain only a teacher-fronted exercise. Students can be involved in characterizing their own progress and in setting goals or objectives for their language learning. The practice can contribute to a personalized-learning approach to instruction. The study by Elshawa et al. (2017) is focused on English language instructors' beliefs on the purpose of assessment and their analysis shows that diagnosis of the strength and weaknesses of students' English, providing them information about their progress and providing them feedback were the main aims of assessment.

Contemporary research in assessment techniques also displays a major shift towards learner self-assessment and peer-assessment (Brown & Abeywickrama, 2010). Hua (2023), for instance, quotes Vuogan and Li (2023) and Yu and Lee (2016) reporting a comparable effectiveness of teacher assessment and learner assessment in promoting second-language writing skills. Hua's (2023) own study examined the effects, validity and reliability, and learners' perception of peer assessment and self-assessment in L2 English pronunciation and reported high validity and reliability for peer assessment but marginal validity and acceptable reliability for self-assessment. Kumar et al. (2023) have also examined the impact of self- and peer assessments on the development of EFL learners' self-regulation capabilities. Their study is founded on Pedrotti and Nistor's (2019) notions of self-regulation. Learners set their own goals, make plans to achieve those goals, evaluate their learning process, solve problems, and examine their own process in self-regulation practice (Pedrotti & Nistor, 2019). According to Simonsmeier et al. (2020), peer feedback bears the potential of improving academic achievement. In Simonsmeier et al.'s (2020) study, each student acted as an author and a reviewer. Their findings indicated significant improvements in academic achievements for the domain of academic writing over time. A new concept in language testing is the washback effect. Washback refers to the extent to which the introduction and use of an assessment influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning. Messick's (1996) research examines the concept of washback as an instance of the consequential aspect of construct validity. The researcher links positive washback to authentic and direct assessments.

To sum up, a review of major research studies focused on assessment of speaking skill of adult EFL learners reveals that not much attention has been paid to this important academic issue, particularly to the evaluation of predominant assessment techniques from the point of view of quality, reliability, and validity of the assessment techniques. This is specifically true with reference to Saudi Arabian academic contexts. Thus, there exists a research gap, and the present study has been designed to address the same issue.

## Objectives of the Present Research

In view of the research problem identified and the realization of a gap in the existing research literature after a preliminary review of literature in the research area, the primary objective of the present study was to gather the opinions of the teachers of English at the English Language Institute at King Abdulaziz University in Saudi Arabia concerning the quality, reliability, and validity of the speaking skill assessment instruments in use. The secondary objective of the study was to know the perceptions of the same teachers regarding the importance of formative assessment in the final grading process of learners in English speaking skill.

## Research Questions

To fulfill the stated objectives of the present research, the study was designed to answer the following research questions:

- RQ 1: What difference do ELI instructors perceive between the different instruments they use for the assessment of learners' achievement in English speaking skill for the final grading?
- RQ 2: In the opinion of instructors, which instrument is more reliable and valid to assess the English-speaking skill of adult EFL learners?
- RQ 3: In the opinion of instructors, what is the role of formative assessment in the final determination of learner achievement in English speaking skill?

## Research Design

The present study is perception-based research. The study involves data collection from ELI instructors, data analysis using simple statistical tools, such as tabulation, calculation of mean, etc., and deriving

conclusions of the study from data analysis. The study is designed to use both quantitative and qualitative methods. The major points of statistical analysis will be the following:

- the number of instructors expressing their opinions on different types of assessment instruments in use in ELI for the final grading, and the main differences between the instruments;
- the number of instructors favouring a particular assessment instrument for its higher reliability and validity to assess adult EFL learners' speaking skill; and
- the number of instructors favouring a more decisive role of formative assessment in determining learner's final grades in speaking skill.

Opinionated data will be collected from the practicing English teachers at ELI on the quality, reliability, and validity of the assessment technique in use at ELI and its affiliated centres. The collected data will be used to derive conclusions on what a majority of teachers perceive as the most reliable and valid instrument for the assessment of adult EFL learners' speaking skill. Teachers' opinions will also be collected on what they consider as the value of formative assessment of speaking skill.

## Research Methodology

As mentioned above, mixed-methods research methodology has been put to use to conduct the present study. The quantitative method will be used to analyze the percentage, means, and other numerical factors associated with the statistical analysis. Qualitative methodology will be used to interpret the numerical values to present the findings in a narrative fashion. At some point in the analysis, the two methodologies will be mixed as well.

### *Research Participants*

For the present research, opinionated data were collected from ELI instructors both male and female. The survey questionnaire data were collected from 100 instructors (76 male, 24 female), whereas twenty instructors (15 male, 5 female) were selected for the face-to-face interview. The participating ELI instructors' ages ranged variously from 34 to 55 years, length of service at ELI from 7 to 20 years, and total service experience ranging from 9 to 30 years.

The demographic data on research participants is presented in Table 1, given below. The teacher demographics are significant for data collection because the length of experience of teachers at ELI is very important concerning their opinions on the speaking assessment. Several ELI teachers have seen the changes taking place in assessment methods and curriculum development.

**Table 1.** Participants' demographic data

Data Collection Instrument	Participants	Number	Average Age	Mean of Length of service at ELI	Mean of Total experience
Questionnaire	Male	76	46	11.63	19.8
	Female	24	45	11	18
Interview	Male	15	46	12	20
	Female	5	45	11	18

### *Instruments of Data Collection*

Data were collected using the following tools:

- i. survey questionnaire, and
- ii. structured interviews with ELI instructors.

The survey questionnaire was comprised of a 5-point Likert scale. The scale contained 16 prompt statements followed by a series of five answer statements - Strongly disagree, Disagree, Neutral, Agree, and Strongly agree. The prompt statements were meant to elicit instructors' opinions on the quality, reliability, and validity aspects of the speaking skill assessment techniques in use at ELI, spontaneity in speech, and formative assessment. The structured interview contained 15 questions meant to elicit instructors' opinions on the reliability and validity of the assessment instruments to accurately measure learners' speaking skill, instructors' preference for a particular instrument, any challenges they might face in implementing the assessment techniques in class, and, opinions on the value of formatting assessment in the final grading of speaking skill, instructors' freedom of choice for the assessment type, and so on (see Appendix 1).

The validity and internal consistency of the survey questionnaire was checked using Cronbach's Alpha calculations, whereas the reliability and validity aspects of the interview format were tested through piloting the interview and seeking suggestions for improvement from experienced teachers. Their suggestions were incorporated into the final interview format.

### *Theoretical Background*

Messick's (1996) idea of integrative view of construct validity and washback has been employed as the theoretical framework to conduct the present research. Kane's interpretation argument framework (Kane, 2013) has also been consulted. The theoretical focus was on two sources of validity evidence: internal structure and external relations to other variables (AERA [American Educational Research Association], APA [American Psychological Association], and NCME [National Council in Measurement in Education]). In this framework, there are two types of arguments specified during validation efforts: Interpretive/Use Argument (IUA): This argument outlines the claims supporting intended score interpretation and use. Validity Argument: It integrates various strands of evidence into a coherent account of how existing evidence and theory support the intended interpretation of test scores for specific uses.

Kane's approach emphasizes a systematic way to determine validity, moving away from various "types" of validity and focusing on the purpose of the test.

### *Research Procedure*

For data collection, the survey questionnaire was shared online with the selected participating ELI instructors with the request to return it filled in as soon as possible. Similarly, the interview format was sent to the selected participants online and the teachers were requested to send back their responses in typed format. The questionnaire responses were analyzed statistically. The numerical results obtained were interpreted and explained in a narrative format. The interview responses were organized thematically and interpreted accordingly. Both interpretations were finally merged to present a coherent qualitative analysis.

### **Data Analysis**

Validity and reliability of the questionnaire were checked through a pilot test conducted with 40 participants, followed by calculating Cronbach's Alpha with the scores obtained, as follows:

$$\alpha = (k/(k-1)) \times (1 - \Sigma \text{Var}/\text{Var})$$

$$\alpha = (40/(40-1)) \times (1 - 13.23/39.64)$$

$$\alpha = .68$$



The obtained value of Cronbach's Alpha is .68, which is acceptable since the minimum acceptable value of Alpha is .60. Correlational tests and confirmatory factor analysis also showed robust evidence for the test's internal structure. Correlation and multiple regression models also demonstrated strong positive relationships between the variables.

The raw scores obtained from the questionnaire were tabulated and used for further calculations, such as number and percentage of participants in agreement or disagreement with the questionnaire statements.

The interview responses received from instructors were organized into the following thematic categories: Validity and reliability of assessment components (interview and speaking project), instructors' preference for the speaking assessment technique, accuracy of assessment rating scale, accuracy of grading rubrics, formative assessment and its value in final grading, and examiners' challenges in assessment.

## Results

### *Results Obtained from the Questionnaire*

In Table 2, given below, are displayed the final percentage figures obtained after analysis of participants' agreement or disagreement with the questionnaire statements. The number of respondents who chose "Neutral" was discarded. The reason was that the choice neither referred to "agreement" nor to "disagreement" to the questionnaire statements. For a graphic representation of participants' agreement/disagreement with questionnaire statements, see Appendix 2.

**Table 2.** Percentage of Respondents in Agreement or Disagreement with the Questionnaire Statements

No.	Statement	Agreement (Strongly agree + Agree) %	Neutral %	Disagreement (Disagree + Strongly disagree) %
1.	In my opinion, both the assessment techniques, i.e. Interview and Speaking Project, are equally reliable and valid methods of assessment of speaking skill.	90.4	7	2.6
2.	Given a choice, I would prefer to use the Interview method to assess learners' speaking achievement.	63.3	20	16.7
3.	Given a choice, I would prefer to use the Speaking Project to assess learners' speaking achievement.	40.7	7	52.3
4.	To me, the interview method is a more reliable instrument of assessment to evaluate the speaking skill.	70	23.3	6.7
5.	To me, the speaking project method is a more reliable instrument of assessment to evaluate the speaking skill.	43.4	20	36.6
6.	To me, the interview method is a more valid instrument of assessment to evaluate the speaking skill.	70	23.3	6.7
7.	To me, the speaking project method is a more valid instrument of assessment to evaluate the speaking skill.	36.7	27	36.3
8.	In my opinion, the interview method is a more reliable and valid assessment technique since in an interview the learners display more spontaneity in speech.	63.4	0	36.6
9.	In my opinion, the interview method is not a reliable and valid assessment technique since in an interview	33.4	6.6	60

	the learners perform under pressure and may feel nervous.			
10.	In my opinion, the speaking project is a more reliable and valid assessment technique since the learners display more preparedness in speaking English in the presentation session.	43.3	20	36.7
11.	In my opinion, the speaking project is not a reliable and valid assessment technique for lack of spontaneity in speech on the part of learners and its scripted nature.	43.4	6.6	50
12.	Formative assessment in speaking skill plays a big role in determining the final grading of the learners.	86.7	13.3	0
13.	Marks obtained by learners in periodic formative assessment should bear more value in determining the final grades of learners.	36.7	7	56.3
14.	The writing part in the speaking project is important to assess the speaking skills of learners.	63.4	6.6	30
15.	The assessment techniques in use at ELI meet the requirements to evaluate the learning outcomes in speaking skill course.	94.5	0	5.5
16.	There are no challenges as such in conducting the speaking skill assessment through Interview or Speaking Project method.	90	0	10

#### *Results Obtained from the Interview Sessions*

The participating instructors' responses to the structured interview questions reveal that they are satisfied with the present speaking assessment techniques, that is, interview and speaking project. The instructors are not in favor of any major changes in the assessment techniques. The thematized interview responses are discussed below:

#### *Validity and reliability of assessment components*

A majority of instructors (16 out of 20) said that they find the speaking assessment components, i.e., interview and speaking project, valid and reliable and that the techniques meet the criteria and accurately assess the learning outcomes for the speaking skill course. To quote one of the participants:

*In fact, I am a bit confused about the learning outcomes for the course since there are no clearly defined outcomes for the course as such, yet the assessment components accurately measure students' speaking skill in English.*

#### *Instructors' preference for the speaking assessment technique*

None of the instructors specified any preference for just one component of the speaking assessment, though several of them agreed that the interview technique provides the learners opportunities to display more spontaneity in their speech, whereas, in the speaking project learners' responses look more scripted in nature. However, they said, since learners are required to use spoken English while collecting data for the speaking project, this method of assessment also achieves its objectives.

#### *Accuracy of assessment rating scale*

The participating instructors opined that the present assessment rating scale is fine. However, there is always some scope for improvement. For instance, one participant noted that,

*The distribution of scores across different categories in the interview part might be reconsidered. The focus may shift towards better pronunciation in connected speech and overall fluency in the language since, commonly, our students are slow speakers which affects their pronunciation and intelligibility negatively.*

#### *Accuracy of grading rubrics*

The participating instructors observed that the grading rubrics for both interview and speaking project are accurate. A few participants (7 out of 20) said that the rubrics may be rewritten making them simpler than what they presently are, with equal distribution across categories in interview as well as speaking project, leading to easier grading and better standardization.

#### *Formative assessment and its value in the final grading*

Almost all the participants said that formative assessment is helpful in monitoring student progress, determining the strengths of candidates and preparing them better for the final exams. Participants added that formative assessment motivates students, warns them of potential issues, and helps them adjust their strategies to prepare more effectively for the final exam. However, they said that formative assessment should not be formally marked, and the obtained marks should not contribute to students' final grades. Their reasoning was that the practice may lead to inflation of marks by instructors affecting the assessment process negatively.

#### *Examiners' challenges in assessment*

The general opinion of participants was that they don't face any challenges in conducting speaking assessment using interview technique or evaluating the speaking project. The instructors observed that the assessment rubrics guide them in a way that there are not many challenges in conducting the exams. There might, of course, arise several challenges in evaluating students' speaking skills if no rubrics are provided.

### **Discussion**

A cursory glance at Table 2, above, shows that the ELI instructors display satisfaction with the present speaking assessment pattern that uses both interview and speaking project as assessment techniques. 90.4 percent of teachers say that both assessment techniques, i.e. Interview and Speaking Project, are equally reliable and valid methods of assessment of speaking skill. Although 63.3 percent of instructors expressed the opinion that given a choice, they would prefer to use the interview method to assess learners' speaking achievement and only 40.7 percent of teachers say they would prefer to use speaking project as an assessment technique, yet they didn't assert their preferences in the interview sessions. Similarly, 70 percent of the participants expressed the opinion that interview is a more reliable and valid method of speaking assessment while only 43.4 percent of teachers find speaking project as valid and reliable, but their non-assertion of the choice in the interview leaves the issue unsettled. A higher percentage of participating instructors seem to prefer the interview method (63.4%) over the speaking project method (43.4%), but those who disagree with the statements (more than 36% in each case) are also considerable. Moreover, since a very high percentage of teachers (94.5%) agree with the statement that the present assessment techniques in use at ELI meet the requirements to evaluate the learning outcomes in speaking skill course accurately, the higher percentage of teachers expressing their preference for interview technique has no meaning.

The findings from the present research cannot be assessed from a comparative perspective since there is a lack of previous research studies on the topic. However, the findings from the present research corroborate research findings on the significance of formative assessment in student motivation and better preparation for the final exam, such as Black and Wiliam (1998), Shaaban (2000) and Alahamadi et al. (2019). The findings from the study by Elshawa et al. (2017) focused on English language instructors' beliefs on the purpose of assessment are also supported by the findings from the present study.

## Conclusions

To sum up, the results obtained from data analysis provide a mixed response from the participating instructors. The research findings show that the instructors are not in favor of any major changes in the speaking assessment pattern at ELL. Considering the research findings, the research questions can be answered as follows:

To the first research question regarding the differences perceived by ELL instructors between the interview and speaking project used for the assessment of speaking skill, the findings show that the teachers perceive only a slight difference between the two, that is, students are more spontaneous in speech in the interview method than in the speaking project. The answer to the second research question concerning instructors' opinions on the reliability and validity of assessment instruments is that a higher percentage of teachers find the interview method more reliable and valid. To the third research question on the role of formative assessment in the final determination of learner achievement in English speaking skill, the answer is that although the instructors acknowledge a valid role of formative assessment in preparing the students to face the final exams, they are not in favor of assigning any official value to formative assessment marks of students (if any) in determining students' final grades in speaking.

## Limitations of the Study

For lack of time and sufficient resources, a few points could not be taken into account while conducting the present study which may be considered as its limitations. The first limitation is that the interview sessions were limited to only twenty participants. That was because the researcher received interview responses from only twenty instructors. Interview sessions with a larger number of participants might have affected the results. The second limitation of the study was that the data collection was limited to only two sources. This factor has reduced the potential to use triangulation method for a more robust research outcome.

## Recommendations for Further Study

Future researchers can work on the issues identified as limitations of the present study. That is, they can involve more participants in the interview sessions in order to document a wider range of opinions. They can also collect data using more data collection sources so that triangulation method can be employed for a better understanding of the research issues. Triangulation method can also affect the obtained results.

## Author ethical declarations

I confirm that the work has not been published elsewhere in any form or language.

**Funding information:** No funding was received for conducting this study.

**Conflict of interest:** The author states no conflict of interest.

## Declaration of interests

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Ahmed, S., & Alamin, A. (2014). Assessing speaking ability in academic context for fourth year Taif University students. *International Journal of English Linguistics*, 4(6), 97-103. <https://doi.org/10.5539/ijel.v4n6p97>
- Alahmadi, N., Alrahaili, M., & Alshraideh, D. (2019). The impact of the formative assessment in speaking test on Saudi students' performance. *Arab World English Journal (AWEJ)*, 10(1), 259-270. <https://dx.doi.org/10.24093/awej/vol10no1.22>
- Alharbi, A. F., & Surur, R. S. (2019). The effectiveness of oral assessment techniques used in EFL classrooms in Saudi Arabia from students and teachers point of view. *English Language Teaching*, 12(5), 1-19. <https://doi.org/10.5539/elt.v12n5p1>

- Alharbi, A. S., & Meccawy, Z. (2020). Introducing Socratic as a tool for formative assessment in Saudi EFL classrooms. *Arab World English Journal (AWEJ)*, 11(3), 372-384. <https://doi.org/10.24093/awej/vol11no3.23>
- Al-Hassaani, A. M. A., & Al-Saalmi, A. F. M. Q. (2022). Saudi EFL learners' speaking skills: Status, challenges, and solutions. *Arab World English Journal (AWEJ)*, 13(2), 328-337. <https://dx.doi.org/10.24093/awej/vol13no2.22>
- Ali, S., Ahmad, H., & Khan, A. (2019). Testing in English language teaching and its significance in EFL contexts: A theoretical perspective. *Global Regional Review*, 4(2), 254-262. [https://doi.org/10.31703/grr.2019\(IV-II\).27](https://doi.org/10.31703/grr.2019(IV-II).27)
- Almosa, S. Y., & Alzahrani, S. M. (2022). Assessment approaches of English language teachers in the Saudi higher education context. *Language Testing in Asia*, 12, Article 10. <https://doi.org/10.1186/s40468-022-00160-x>
- Almuntasheri, S. (2016). Saudi teachers' practices of formative assessment: A qualitative study. *Problems of Education in the 21st Century*, 74, 6-15. <https://doi.org/10.33225/pec/16.74.06>
- Alrasheedi, S. (2020). Investigation of factors influencing speaking performance of Saudi EFL learners. *Arab World English Journal (AWEJ)*, 11(4) 66-77. <https://dx.doi.org/10.24093/awej/vol11no4.5>
- Al-Seghayer, K. M. (2022). Continued concerns with language assessment practices in Saudi Arabian English education. *Education, Language and Sociology Research*, 3(3), 55-64. <http://dx.doi.org/10.22158/elsr.v3n3p55>
- Alsubaia, H. S. (2021). Teachers' perception towards formative assessment in Saudi universities' context: A review of literature. *English Language Teaching*, 14(7), 107-116. <https://doi.org/10.5539/elt.v14n7p107>
- Ariani, M. G. (2014). Iranian EFL teachers' techniques to assess student learning during class. *Procedia-Social and Behavioral Sciences*, 158, 23-26. <https://doi.org/10.1016/j.sbspro.2014.12.027>
- Al-Wassia, R., Hamed, O., Al-Wassia, H., Alafari, R., & Jamjoom, R. (2015). Cultural challenges to implementation of formative assessment in Saudi Arabia: An exploratory study. *Medical Teacher*, 37(sup1), S9-S19. <https://doi.org/10.3109/0142159X.2015.1006601>
- Bailey, A. L. (2017). Theoretical and developmental issues to consider in the assessment of young students' English language proficiency. In M. K. Wolf & Y. G. Butler (Eds.), *English language proficiency assessments for young learners* (pp. 25-40). Routledge. <https://doi.org/10.4324/9781315674391>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2004). Classroom assessment is not (necessarily) formative assessment (and vice-versa). *Teachers College Record*, 106(14), 183-188. <https://doi.org/10.1177/016146810410601408>
- Boyles, P. (2006). Assessment literacy. In M. H. Rosenbusch (Ed.), *New visions in action: National assessment summit papers* (pp. 18-23). Iowa State University. <https://files.eric.ed.gov/fulltext/ED527580.pdf#page=18>
- Brown, G. T. L. (2019). Is assessment for learning really assessment? *Frontiers in Education*, 4, 64. <https://doi.org/10.3389/feduc.2019.00064>
- Brown, H. D., & Abeywickrama, P. (2010) *Language assessment: Principles and classroom practices* (3rd ed.). Pearson.
- Elshawa, N., Abdullah, A. N., & Rashid, S. M. (2017). Malaysian instructor's assessment beliefs in tertiary ESL classrooms. *International Journal of Education & Literacy Studies*, 5(2), 29-46. <http://dx.doi.org/10.7575/aiac.ijels.v.5n.2p.29>
- Engelsen, K. S., & Smith, K. (2014). Assessment literacy. In C. Wyatt-Smith, V. Klenowski & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 91-107). Springer. [https://doi.org/10.1007/978-94-007-5902-2\\_6](https://doi.org/10.1007/978-94-007-5902-2_6)
- Green, A. (2013). Exploring language assessment and testing: Language in action. Routledge. <https://doi.org/10.4324/9781315889627>
- Herman, J., Osmundson, E., Yunyun, D., Ringstaff, C., & Timms, M. (2015). Investigating the dynamics of formative assessment: Relationship between teacher knowledge, assessment practice and learning. *Assessment in Education: Principles, Policy & Practice*, 22(3), 344-367. <https://doi.org/10.1080/0969594X.2015.1006521>
- Herrera, L., & Macias, D. F. (2015) A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302-312. <https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09>
- Hua, C. (2023). Pronunciation assessment of learners, by learners, and for learners: Effects, validity and reliability, and learners' perception. *Language Teaching Research*, Online first. <https://doi.org/10.1177/13621688231195143>
- Huang, B. H., Bailey, A. L., Sass, D. A., & Chang, Y-h. S. (2020). An investigation of the validity of a speaking assessment for adolescent English language learners. *Language Testing*, 38(3), 401-428. <https://doi.org/10.1177/0265532220925731>
- Jayaraman, S. (2017). EFL assessment: Assessment of speaking and listening. In R. Al-Mahrooqi, C. Coombe, F. Al-Maamari & V. Thakur (Eds.), *Revisiting EFL Assessment* (pp. 133-150). Springer. [https://doi.org/10.1007/978-3-319-32601-6\\_9](https://doi.org/10.1007/978-3-319-32601-6_9)
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73. <https://doi.org/10.1111/jedm.12000>
- Kumar, T., Soozandehfar, S. M. A., Hashemifardina, A., & Mombeini, R. (2023). Self vs. peer assessment activities in EFL-speaking classes: Impacts on students' self-regulated learning, critical thinking, and problem-solving skills. *Language Testing in Asia*, 13, Article 36. <https://doi.org/10.1186/s40468-023-00251-3>
- Leeming, P., & Harris, J. (2022). Measuring speaking proficiency growth in the language classroom: An investigation of practical approaches for teachers. *Language Teaching Research*, online first. <https://doi.org/10.1177/13621688221130856>
- Luoma, S. (2004). *Assessing Speaking*. Cambridge University Press. <https://www.doi.org/10.1017/CBO9780511733017>
- Macalister, J., & Nation, I. P. (2019). *Language curriculum design*. Routledge. <https://doi.org/10.4324/9780429203763>
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256. <https://doi.org/10.1177/026553229601300302>
- O'Sullivan, B., Weir, C. J., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing*, 19(1), 33-56. <https://doi.org/10.1191/0265532202lt219oa>

- O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing*, 19(3), 277-295. <https://doi.org/10.1191/0265532202lt2050a>
- Pedrotti, M., & Nistor, N. (2019). How students fail to self-regulate their online learning experience. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, J. Schneider (Eds.). *Transforming learning with meaningful technologies* (pp. 377–385). Springer. [https://doi.org/10.1007/978-3-030-29736-7\\_28](https://doi.org/10.1007/978-3-030-29736-7_28)
- Rauf, M., & McCallum, L. (2020). Language assessment literacy: Task analysis in Saudi universities. In L. McCallum & C. Coombe (Eds.), *The assessment of L2 written English across the MENA region* (pp. 13–41). Palgrave Macmillan, Cham. [https://doi.org/10.1007/978-3-030-53254-3\\_2](https://doi.org/10.1007/978-3-030-53254-3_2)
- Ross S. J. (2017). *Interviewing for language proficiency: Interaction and interpretation*. Palgrave MacMillan. <https://doi.org/10.1007/978-3-319-60528-9>
- Shaaban, K. (2000). Assessment of young learners' achievement in ESL classes in the Lebanon. *Language, Culture and Curriculum*, 13(3), 306–317. <https://doi.org/10.1080/07908310008666606>
- Simonsmeier, B. A., Peifer, H., Flaig, M., & Schneider, M. (2020). Peer feedback improves students' academic self-concept in higher education. *Research in Higher Education*, 61(6), 706–724. <https://doi.org/10.1007/s11162-020-09591-y>
- Taylor, L. (2011). *Examining speaking. Research and practice in assessing second language speaking*. University of Cambridge Press.
- Umer, M., Zakaria, M. H., & Alshara, M. A. (2018). Investigating Saudi University EFL teachers' assessment literacy: Theory and practice. *International Journal of English Linguistics*, 8(3), 345–356. <http://doi.org/10.5539/ijel.v8n3p345>
- Varela, M. L., & Palacios, I. M. (2013). How are spoken skills assessed in proficiency tests of general English as a foreign language? A preliminary survey. *International Journal of English Studies*, 13(2), 53–68. <https://doi.org/10.6018/ijes.13.2.185901>
- Vuogan, A., & Li, S. (2023). Examining the effectiveness of peer feedback in second language writing: A meta-analysis. *TESOL Quarterly*, 57(4), 1115–1138. <https://doi.org/10.1002/tesq.3178>
- Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, 49(4), 461–493. <https://doi.org/10.1017/S0261444816000161>

## Appendices

### *Appendix 1: Interview Questions*

I would appreciate it if you could express your opinions on the following points concerning the reliability and validity of the ELI speaking skill assessment techniques, i.e. Interview and Speaking Project. Please feel free to elaborate as much as possible.

- Please elaborate on whether the Interview and Speaking Project meet the criteria to accurately assess the learning outcomes set for the speaking skill course.
- Given a choice, which assessment technique would you prefer to use to evaluate learners' speaking achievement, Interview or Speaking Project?
- Why do you prefer to use that particular assessment technique? Please cite as many criteria as possible for your preference.
- In your opinion, does the writing part in the Speaking Project affect learners' speaking skill development [positively or negatively]?
- If you prefer Speaking Project as a speaking skill assessment technique, what elements of the project make it a better speaking assessment technique?
- If you prefer the Interview method as a speaking skill assessment technique, what elements of the technique make it a better speaking assessment method?
- Do you conduct formative assessment in speaking skill and assign marks to learners?
- In your opinion, do formative assessment marks of learners also play a role in the summative assessment of speaking skill in the final grading?
- In your opinion, should formative assessment be given more weightage in evaluating the speaking skill of learners for final grading?
- What is your perception of Interview being a valid assessment technique to evaluate learner achievement in speaking skill?

- What is your perception of Speaking Project being a valid assessment technique to evaluate learner achievement in speaking skill?
- In your opinion, is Interview method a reliable assessment technique to evaluate learner achievement in speaking skill?
- In your opinion, is Speaking Project a reliable assessment technique to evaluate learner achievement in speaking skill?
- Are there any challenges you face in conducting the speaking skill assessment through Interview or Speaking Project method?
- Please elaborate your opinion on the speaking skill grading rubrics, the set of interview prompts, and learner peer interaction and note-taking for the Speaking Project, etc.

*Appendix 2: Graphic Representation of Participants' agreement/ disagreement with questionnaire statements.*

