



University of Zawia

Post Graduate Studies & Training

Faculty of Arts

Department of English Language

Applied Linguistic MA Program

Speaking Assessment Strategies used by Libyan EFL University Teachers

*A Dissertation Submitted in Partial Fulfilment of the Requirements for the Degree
of Master of Art in Applied Linguistic*

Submitted by:

Amel Belqasem Mohammed Omar

Supervised by:

Dr. Abdussalam Saleh Tantani

Academic year - 2022

Abstract

Among the four key language skills, speaking is deemed to be the most important skill in learning a foreign or a second language which reaffirms that teaching this particular skill is an essential element in second language teaching, in addition to its assessment. Assessment helps teachers to understand whether or not their students have learned what have been taught. This study therefore, investigates EFL teachers speaking assessment strategies and the challenges they might encounter whenever they assess their students' oral skills in class. To achieve the goal of this study, both quantitative and qualitative methods were utilized. The quantitative data were represented through administering a questionnaire to investigate teachers' background knowledge about speaking assessment in general and to what extent they use the speaking assessment strategies in their classes in particular. The researcher selected a random sample that comprised of thirty-five teachers from different colleges in Zawia University. The data obtained from the questionnaire were analysed by SPSS software. On the other hand, in order to gain in-depth data about teachers' challenges, semi-structured interviews with five teachers were conducted as a qualitative data instrument. The qualitative data were analysed by employing Thematic Analysis Method. The findings obtained revealed that most of the teachers do consider accuracy, fluency, communicative skills, considered time, reliability, validity, vocabulary, pronunciation and the students' use of transitions as categories of speaking assessment as a theoretical knowledge, but they do not apply that in their real speaking classes when they assess their students. This implies that these teachers do not apply what they know about teaching and assessing students in English speaking classes. Moreover, the research findings showed that the majority of teachers do not employ formative assessment, set speaker assessment tasks for their students at an appropriate degree of difficulty, use of oral presentation task in speaking assessment, use discussion and decision tasks, retelling story or text from aural stimuli, use analytic scores in their tests to assess students. In contrast, the findings obtained showed that almost all of the teachers use summative assessment, role-play tasks and a holistic score for each category in speaking test in their classes. The research findings also reported that all of these failures were due to the internal and external challenges that teachers encounter. These challenges include difficulties in choosing appropriate assessment categories, difficulties in assessing different types of speaking, representative assessing tasks, plan and structure the oral testing carefully, ensure valid and reliable scoring, elicit a valid sample, choosing tasks for speaking assessment, classroom size, lack of resources and training sessions. Finally, the research in hand provided suitable solutions through some implications and recommendations.

Declaration

I declare that work in this dissertation “Speaking Assessment Strategies used by Libyan EFL University Teachers” has been carried out by me in English department at Zawia University. I further declare that this dissertation has not previously formed the basis for the award of any degree, diploma or other similar title of recognition.

Signature.....

Date:

Dedication

This dissertation is proudly dedicated to...

All my beloved family

My dear mother, my father, my brother, my second family especially my brother “Hassan”, my beloved husband “Mohammed” and my sweet children “Ayla” and “Diala” who have supported me throughout the process. I will always appreciate all they have done...

To the soul of my brothers “Sofian” and “Adel” ...

To my lovely sisters “Nafessa” and “Hajer” who have never left my side and are very supporting in my life.

To Dr. Abdussalam, without him I would be drifting like a ship without a sail

Acknowledgments

First and foremost, I thank Allah for granting me the ability to complete this work. It would not have been possible to write this dissertation without Allah's mercy, grace and conciliation.

My deep thanks and gratitude go to my supervisor, Dr. Abdussalam Salah Tantani for his moral support, excellent guidance and for his patience during my journey of master study. His inspiring comments and invaluable feedback have enabled me to gain a thorough understanding of my study. He has been my source of inspiration throughout my journey without his support I could not reach to this stage.

My sincere thanks go to the members of the discussion committee, for their approval to discuss this dissertation. Moreover, my thanks should be given to all the teachers from different colleges at Zawia University who participated in this study. They have been kindly co-operative and provided me with invaluable insights and perspectives.

My appreciation goes to my dearest friends Nafessa and Hajer for their support and assistance throughout my study, and with whom I have had stimulating discussions and gained useful feedback.

I owe my deepest gratitude to my beloved husband Mohammed Kashada and my sweet children, Ayla and Diala for their patience and constant moral support without which I could not have done this journey.

Finally, my warmest appreciation and gratitude go to my parents, my parents-in-law, my brother "Motaz", my husband`s brothers especially "Hassan" for their passion and love, and permanent support during my journey in master study.

Table of Contents

Title	Page
Abstract.....	I
Declaration.....	II
Dedication.....	III
Acknowledgments.....	IV
Table of Contents.....	V
List of Tables.....	IX
List of Figures.....	X
Abbreviations.....	XI
Chapter One	
Introduction	
1.0.Introduction.....	1
1.1.Background of the Study.....	1
1.2. Statement of the Problem.....	4
1.3. Aims of the Study.....	4
1.4. Research Questions.....	4
1.5. Significance of the Study.....	4
1.6. Methodology of the Study.....	5
1.7. Overview of the Study.....	5
Chapter Two	
Literature Review	
2.0. Introduction.....	6
2.1. Defining Speaking Tasks	6
2.2. Speaking and Assessment.....	7
2.3. Types of Speaking Assessments.....	9
2.3.1. Summative Assessment as a Formal Assessment	9
2.3.2. Formative Assessment as an Informal Assessment	10
2.4. Strategies of Speaking Assessment.....	11

2.4.1. Tests.....	12
2.4.2. Role Plays.....	12
2.4.3. Interviews.....	13
2.4.4. Dialogues.....	13
2.4.5. Storytelling.....	14
2.4.6. Observation.....	14
2.4.7. Discussion.....	15
2.4.8. Academic Presentations.....	15
2.5. Principles of Assessment	15
2.5.1. Validity.....	16
2.5.2. Reliability	17
2.6. Speaking Assessment Tasks.....	17
2.7. Types of Scoring and Rating Scales.....	20
2.8. Assessing Linguistic Constructs in Speaking.....	22
2.8.1. Vocabulary.....	22
2.8.2. Pronunciation.....	23
2.8.3. Accuracy	23
2.8.4. Fluency.....	24
2.9. Assessing Communicative Constructs.....	24
2.10. External Challenges in Speaking Assessment.....	25
2.10.1. Class Size Effects.....	25
2.10.2. Lack of Facilities	26
2.10.3. Lack of Training Sessions.....	27
2.11. Previous Studies in Speaking Assessment.....	27
2.10. Summary of the Chapter.....	30
Chapter Three	
Methodology	
3.0. Introduction.....	31
3.1. Research Design.....	31
3.2. Data Collection Instruments.....	32

3.2.1. Structured Questionnaires.....	32
3.2.2. Semi-structured Interviews.....	33
3.3. Pilot Study.....	34
3.4. Sampling and Population.....	35
3.5. Procedures of Data Collection.....	36
3.6. Ethical Issues.....	37
3.7. Procedures of Data Analysis.....	37
3.8. Summary of the Chapter.....	38
Chapter Four	
Data Analysis and Results	
4.0. Introduction.....	39
4.1. Quantitative Analysis.....	39
4.1.1. Teachers` Assessment Speaking Strategies to be Considered.....	40
4.1.2. Teachers` Assessment Speaking Strategies Used.....	45
4.2. Qualitative Analysis	52
4.2.1. Teachers` Challenges about Assessment of Spoken English	52
4.3. Summary of the Chapter.....	59
Chapter Five	
Discussion of Results	
5.0. Introduction.....	60
5.1. Research Questions.....	60
5.2. Teachers` Assessment Speaking Strategies Use and their Challenges.....	60
5.3. Summary of the Chapter.....	73
Chapter Six	
Conclusion	
6.0. Introduction.....	74
6.1. Conclusion of the Whole Study.....	74
6.2. Implications and Recommendations of the Study.....	77
6.3. Limitations of the Study.....	78
6.4. Suggestions for Further Research.....	79

6.5. Summary of the Chapter	78
References.....	80
Appendices	
Appendix A: Teachers' Questionnaire.....	
Appendix B: Teachers' Interviews Questions	
Appendix C: Letter Permission from Post Graduate Studies & Training at Zawia University for Collecting Data	

List of Tables

No	Subject	Page
4.1.	The use of summative assessment	45
4.2.	The use formative assessment	46
4.3.	To set speaker assessment tasks at an appropriate level of difficulty	46
4.4.	The use oral presentation tasks	47
4.5.	Use learner- learner joint discussion and decision taking tasks	47
4.6.	The use of role – play tasks	48
4.7.	The use of interview tasks	48
4.8.	The use of retelling story	49
4.9.	The use of a holistic score	49
4.10.	The use of an analytic score	50
4.11.	Developing my own assessment criteria	50
4.12.	Connecting the selection of assessment criteria to the aim of speaking assessment	51
4.13.	Encouraging students to read aloud in front of class	51

List of Figures

No	Subject	Page
4.1.	Identifying the purpose of assessment when teachers assess students' speaking language performance	40
4.2.	Using different speaking assessment tasks in a speaking test	40
4.3.	Explaining to the students how to accomplish speaking assessment tasks	41
4.4.	Using speaking assessment criteria by outside resources	41
4.5.	Considering accuracy as a category of speaking assessment	42
4.6.	Considering fluency as a category of speaking assessment	42
4.7.	Considering communicative skills as speaking assessment	43
4.8.	Considering range of vocabulary as a speaking assessment	43
4.9.	Considering pronunciation as category of speaking assessment	44
4.10.	Considering the students' use of transitions as a category of speaking assessment	44
4.11.	Considering the reliability and validity in speaking assessment	45

Abbreviations

EFL.....	English as a Foreign Language
SPSS	Statistical Package for the Social Sciences
ESL.....	English as a Second Language
FSI.....	Foreign Services Institute
ACTFL	The American Council on the Teaching of Foreign Languages
ASLPR	Australian Second Language Proficiencies Rating

Chapter One

Introduction

1.0. Introduction

This chapter is initiated with a discussion of the research background. Then, it goes further and discusses the problem statement, aims, research questions as well as the methodology adopted. An overview of the dissertation is also presented in this chapter.

1.1 . Background of the Study

Language means communication; human being uses it in transmitting information or messages and expressing their thoughts and emotions (Jondeya, 2011). Thus, teaching speaking is an important part in language teaching, as well as in assessing oral skills. Teachers who use communicative language teaching methods are responsible for providing assessment information that is “detailed, innovative, relevant, and diagnostic and addresses a variety of dimensions” (Shohamy, 1992, p. 515). Moreover, Canale (1983, p. 40) further argued that such “Language tests should ideally reflect the properties of communication: contextuality, productivity, process orientation, interactivity, and adaptively”. This leads to the argument that the assessment of EFL students is vital in the teaching/learning process. Wintle and Harrison (1999, p.10) concluded that assessment as the collection of data is the most “important ingredient that fuels much of our education system”. Assessment helps teachers investigate whether or not students have learned what has been taught to them; if they have, this means that the teaching process is effective and vice versa.

Furthermore, assessment is a way of judging students’ achievement throughout a course or a training programme. However, the notion of utterance fluency refers to the temporal values of speech or the “oral features of utterances that reflect the operation of underlying cognitive processes” (Segalowitz, 2010, p. 48). Teachers who conduct assessments are recommended to take care of selecting any strategy to be used for collecting information about students (Enerson, 1994). This is because each strategy can work best in some situations and not in others. This means that assessment tell teachers whether the methods, materials, or techniques used are suitable, applicable, and reliable. Not only does this help in finding the best ways to improve,

but it can, to the contrary, negatively affect the achievement that teachers want to assess. Trying to find the most appropriate methods and materials that might lead students to the desired improvement is as essential as employing a suitable assessment strategy. It could be very beneficial to use one assessment strategy rather than another; but it could also be discouraging, misleading, or damaging, if it is not carefully selected.

Assessment also can be linked to the following educational purposes: diagnosis, prediction, placement, evaluation, selection, grading, guidance or administration (Harmer, 2009). In all fields of education, assessment results are used to decide about students (i.e., student progression), to decide about teaching and learning (i.e., curriculum decisions) and increasingly assessments are linked with certification of competence and the validation of performance on job-related tasks (Taras, 2001). Assessment should provide multiple measures and opportunities for students to create and demonstrate what they can do with a language. To provide a comprehensive picture of a student's language ability, the teacher should strive to create a balance between formative assessment and summative assessment. Assessments should be authentic and include alternative and integrated performance-based assessments (ibid). The teacher and language learner need to identify goals, objectives, and expected results before beginning to plan a lesson or activity. In other words, it is essential to determine what the learner should know, understand, and be able to do.

In general, in carrying out any type of assessment with the aim of getting specific feedback for teachers or students, assessing agents such as teachers or educational authorities should use the appropriate strategy to obtain useful information. In the research conducted by Taras (2001) into classroom assessment techniques, she argued that teachers (who are responsible for most assessment) should take the utmost care when selecting a strategy to be used for collecting information about students, because each procedure will work in some situations but not in others. For example, tests are not useful when assessing learners' attitude; but another strategy, observation, can help to fulfil that task. To be more specific, a various number of speaking teaching strategies are used in the classrooms for many circumstances. Among others, the strategies of teaching speaking are; cooperative activities, role-play, creative tasks, and drilling. Cooperative activities can encourage negotiation of language item (Newton & Nation, 2009).

Role plays are activities where students are asked to pretend to be in various social contexts and various social roles (Harmer, 2001; Thornbury, 2005; Solcova, 2011). Creative tasks resemble real-life tasks as Solcova (2011) asserts that students develop their fluency best, if engaged in tasks where all their concentration focuses on producing something, rather than on the language itself. Drilling, as Thornbury (2005) argues, is a strategy to improve pronunciation by imitating and repeating words, phrases, and even whole utterances. It functions to make students pay attention to the new materials and emphasize words, phrases, or utterances on students' mind, move new items from working memory to long term memory, provide means of gaining articulatory control over language (ibid).

In addition, appropriate assessment is very crucial for the teachers in evaluating their students. It is one of the important aspects of teaching and learning process that influences the students. It concerns the quality of the teaching as well as the quality of the learning. Therefore, teachers should have an appropriate assessment that will not interfere with their students' language development. Harris and McCann (1994) pointed out that information about students' performance can be collected by teachers in ordinary classroom conditions without the setting of tests which happen when learners are assessed formally. It seems only logical that teachers should take the lead, because they are responsible for what occurs in their classrooms. Obviously, assessing speaking skills is a challenge in language teaching, especially when it relates to English learners. It is a quite difficult task since they have higher activity levels and get easy to be distracted by others. They also have a shorter attention span, wariness of strangers, and inconsistent performance in unfamiliar environments (Lee, 2007). Consequently, teachers in Zawia University need to use different strategies in assessing their students' oral skills. The use of oral assessment motivates students to practise and improve their English-speaking skills (Huang, 2012; Huxham et al., 2012; Lee, 2007). Based on that issue, this research is aimed at exploring EFL teachers' strategies in assessing speaking classes and the challenges they face. Thus, this study looked closely at the way teachers assess their students' speaking ability. It worth noting that the focus of the study was only restricted on the assessment of speaking skills.

1.2.Statement of the Problem

According to my experience as student and teacher assistance in English department, faculty of Arts at Zawia University, it has been noted that most of students fail in their speaking exams throughout the recent years. The reason behind students' failure to successfully pass their exams could be the misusing of teachers' assessment in most cases because most of teachers do not utilise appropriate strategies and techniques in a correct way. This means that the speaking assessment still encounters some obstacles and it needs to be investigated.

1.3. Aims of the Study

This study aims to identify the speaking assessment strategies that teachers use to assess undergraduate students' oral skills at Zawia University. It also aims to find out the challenges EFL teachers encounter when they assess their students in speaking classes.

1.4. Research Questions

The research questions to be answered are:

1. What speaking assessment strategies EFL teachers do currently use at the undergraduate level at Zawia University?
2. What challenges do EFL teachers encounter when they assess their students in speaking classes?

1.5. Significance of the Study

The main focus of this study is to assess the development of English-speaking skills among prospective teachers at University of Zawia in order to improve the use of their assessment strategies. The findings of this research could broaden teachers' knowledge about strategies in assessing speaking challenges in terms of the purpose and the type of assessment used and teachers' perceptions of speaking assessment. Moreover, the research findings could help students to overcome their speaking difficulties and improve their speaking proficiency. This research could be useful for teachers and designers in implementing amendments to the curriculum.

1.6. Methodology of the Study

For the purpose of the research, a mixed method research is designed and used. In other words, both quantitative and qualitative methodologies are utilised. Cohen and Manion (1994: 233) considered triangulation as “the use of two or more methods of data collection in the study of some aspect of human behaviour”. It has often been observed that no single research methodology is intrinsically better than any other methodology and that many authors such as Cohen et al., (2007) call for a combination of research methods in order to improve the quality of research. Therefore, this study employed a structured questionnaire and semi-structured interviews to answer the research questions. The two instruments were designed to be more appropriate for the study in terms of reliability and validity. The data obtained from the questionnaire were analysed by SPSS software program, whereas the data gained from the interviews were analysed by Thematic Analysis Method. Moreover, regarding to the sample which used in this study; all the participants were EFL Libyan teachers who teach speaking skills at different colleges at Zawia University where English is being taught as a foreign language. 35 participants were randomly chosen to fill the questionnaire, whereas 5 teachers were chosen in purpose according to their experience of teaching speaking to answer the semi-structured interview questions.

1.7. Overview of the Study

This study includes six chapters. Chapter one works as an introduction in which a general background about the topic being investigated. Chapter two is devoted to a wide review of the literature about the investigated topic. This includes some information that has been provided by previous studies and the findings these studies reached. Chapter three demonstrates the methods of collecting the data necessary to carry out the study. Chapter four shows the data analysis and the results reached by the researcher. Chapter five discusses the results obtained. Chapter six presents the conclusion, implications and recommendations.

Chapter Two

Literature Review

2.0. Introduction

The purpose of this chapter is “to get a sense of what we already know about a particular question or problem, to understand how it has been addressed methodologically, and to figure out where we need to go next with our research” (Norris & Ortega, 2006, p. 5). It reviews and addresses the theoretical concepts about speaking assessment in general and the strategies EFL teachers use in particular. It also addressed the challenges EFL encounter teachers when they assess their students in speaking classes. Therefore, this chapter reviews the following titles; definition of speaking, speaking and assessment, types of speaking assessments, strategies of speaking assessment, principles of assessment, speaking assessment tasks, types of scoring, rating scales, assessing linguistic and communicative constructs, challenges in speaking assessment, previous studies in speaking assessment. The summary of the chapter is also provided.

2.1. Defining Speaking Tasks

Speaking English is one of the important aspects amongst students’ purposes in learning the language. Since the beginning, they learn the language; they were ready and trained to talk the language. Richards (2008, p.19) considered speaking skills as a necessity for most EFL learners which means that the learners have to be compelled to be in a position to have sensible speaking because it influences on their ability in mastering the language. Speaking is considered as a vital for teachers to teach because it is initial skill that they have to learn and it is fundamental skill to develop alternative language skills (Linse, 2005). As a result, there are many studies conducted on the development of the learners’ level of speaking proficiency begin with the appearance of communicative language teaching (Nakamura, 1993). Moreover, Richards (2006) contends for the characteristic language use which happens when a speaker takes part in important cooperation and keeps up understandable and progressing correspondence in spite of limits in the incommunicative competence. Richards' clarification of speaking echoes the interactional part of the language that is utilized to fuel social intercourse, to advance enthusiastic knowledge and to satisfy certain objectives or to pass on specific thoughts, expectations, preferences or opinions.

Furthermore, the significant job of speaking is keeping up an oral correspondence. Seen from this viewpoint, speaking helps strengthen interpersonal ties particularly when it is utilized properly in social interaction (Zaremba, 2006). It is viewed as “the vehicle par excellence of social solidarity, of social ranking, of professional advancement and business” (Bygate, 1987, p.vii). Moreover, the complexity of speaking is identified with a bunch of capabilities which can be controlled by the capacity to occupy time with talk, the capacity to talk in coherent, contemplated and semantically dense sentences, the capacity to have suitable comments in a wide scope of contexts and the capacity to be inventive and creative in language use. (Richards, 1990).

In addition, since speaking is considered as the most basic means of human communication, Kingen (2000, p. 218) states that speaking helps fulfil the following twelve functions: (1) Personal – expressing personal feelings, opinions, beliefs and ideas; (2) Descriptive - describing someone or something, real or imagined; (3) Narrative – creating and telling stories or chronologically sequenced events; (4). Instructive – giving instructions or providing directions designed to produce an outcome; (5) Questioning – asking questions to obtain information; (6) Comparative – comparing two or more objects, people, ideas, or opinions to make judgments about them; (7) Imaginative – expressing mental images of people, places, events, and objects; (8) Predictive-predicting possible future events; (9) Interpretative – exploring meanings, creating hypothetical deductions, and considering inferences; (10) Persuasive – changing others’ opinions, attitudes, or points of view, or influencing the behaviour of others in some way; (11) Explanatory – explaining, clarifying, and supporting ideas and opinions; (12) Informative – sharing information with others.

2.2. Speaking and Assessment

The speaking skill is considered as one of the productive skills which is very important in teaching a foreign/second language. Therefore, the teachers should provide students with opportunities to practise the language orally for the sake of proving their students’ mastery of the language. After that the next important step for language teachers is to measure their students’ abilities in their speaking performance (Galata,2021). Assessing the speaking ability is a difficult task as it needs much time to administer to a large class. Teachers of English language are

intended to judge their learners' oral performance through observing students' interaction and participation in class activities (Ghermaoui, 2018). It was argued that the process of speaking assessment in language teaching is considered as extremely hostile and mind boggling (O'Sullivan, 2006). Thus, different scoring instruments are introduced for the assessment of students' oral performance, for instance, the use of checklists and the two main kinds of rubrics which are the logical and the holistic ones.

Speaking assessment tools in general are procedures used to measure the students' abilities, knowledge or performance with the use of different strategies. In other words, assessment involves development of materials, processes, activities, and criteria to be used as a tool for determine the progress of the students during their learning process (Haley & Austin, 2004, p.117). It was argued that one of these tools is the speaking test which is used to measure students' speaking proficiency and testing second language speaking is considered to be more difficult than testing other language skills (Fulcher, 2003). Assessing speaking focus much more on the oral production in which students faced talking tasks and have to speak while the teacher's role is to evaluate the students' level of speech and correct them. Brown (2007, p.352) stated that when teachers give a speaking test to the students, normally, there are going to measure the following criteria: pronunciation, fluency, vocabulary, grammar, discourse features, cohesion, sociolinguistic appropriateness, task accomplishing the objective of the task is also closely dependant on comprehension.

Furthermore, teaching and learning speaking requires to be mastered, of course it makes speaking to be one of the important parts in language teaching assessment. Teaching speaking is developing the students' communicative skills, however, almost many students, speaking need hard efforts to be mastered. Moreover, it can be argued that all learning processes need to be evaluated by assessing or measuring students' understanding or ability in mastering the lessons. Soliha (2019) stated that doing speaking requires establishing ideas and plan for somewhat to deliver to the listener, correct grammar, vocabulary, and pronunciation to use and give a clear understanding. A test or an assessment can be realized as the way to do some evaluation. Assessment can be used to control students' learning progress. That is more likely to be accomplished when assessment is authentic and tied to the instructional goals of the program

(ibid). As result of the all views, speaking is considered as a complex skill in learning and teaching. Therefore, the teacher should apply a good and appropriate method for the students to have better understanding.

2.3. Types of Speaking Assessments

It is widely shown that researchers such as Harris and McCann (1994) classify assessment types according to their formality and informality. That is assessment is seen as formal when it employs tests for assessing students, and informal when other tools are used such as observation. On the other hand, other researchers like James (1998, pp.85-86) does not prefer such classification when she argues that “such classification has dangers because it often carries unjustified value judgements”. According to Thornbury (2005, p.124) assessment of speaking can be done by using two forms: summative assessment (formal assessment) and formative assessment (informal assessment). However, in this respect, the first point of view is dealt with as being more familiar than the latter. The type of assessment teachers should use should match the intended purpose of the assessment. For example, if teachers want to assess the students’ academic achievement and compare it with other students, then they can use the formal assessment. If teachers want to use assessment to monitor students’ progress and help them maximize their own learning, or use assessment to improve instruction, then teachers can use the informal assessment (Williams, 2015). However, in this respect, the first classification of formal and informal assessment is dealt with as being more familiar.

2.3.1. Summative Assessment as a Formal Assessment

Summative assessment as a formal assessment is contrary to informal assessment as it is conducted by means of examinations or tests (Jones & Bray, 1986), formal assessments are based on systematic and planned techniques. It can be defined as a test that actually requires the learners to demonstrate their ability in the speaking skill performance and interact directly with the examiner, i.e., it requires face to face interaction (Ghermaoui, 2018). Hedge (2000) also confirmed the importance of using summative assessment in speaking classes. The main use of this type of assessment is to record or summaries the outcomes and the overall achievements of learners at a certain point in time such as the middle, end of term or end of year. This recording is usually done in a systematic way. However, when teachers choose direct assessment, they

assess what students are actually performing since direct assessment strongly facilitates the teachers' control of the learner performance (Luoma, 2004).

Moreover, summative assessment usually serves other parties rather than students and teachers. These parties could be schools, employment institutions, parents and others (Jones & Bray, 1986). It is referred to as formal because of the fact that it is "systematic, planned sampling technique constructed to give teacher and student an appraisal of student achievement" (Brown, 2004, p. 6). There are several types of tests such as progress tests, summative tests, placement tests, diagnostic tests and proficiency tests (Harris & McCann, 1994, p. 90). All these types are carried out by teachers for collecting information about the teaching and learning process except the last one which is conducted by other agencies aiming to measure which individuals are abler to perform in a foreign language than others (ibid). In the field of language testing the speaking skill, can be directly tested since it is considered as a productive skill that gives the examiner a direct output. Therefore, all the teachers should use summative assessment in speaking classes in order to know to what extent their students' proficiency level has been improved.

2.3.2. Formative Assessment as an Informal Assessment

Formative assessment is defined by Harris and McCann (1994, p. 5) as "a way of collecting information about our students' performance in normal classroom conditions". In other words, informal assessments are those spontaneous forms of assessment that can easily be incorporated in the day-to-day classroom activities and that measure the students' performance and progress. Informal assessments are content and performance driven (Weaver, 2013). Formative assessment is important because is meant to monitor and keep track of students' progress as well as the teaching process, it is always undertaken to fulfil different purposes. This type of assessment is not carried out at a certain point of time as is the case for some other types, but it is an ongoing process which could be conducted as Sutton (1991, p. 3) stated "every few minutes". He also added "without formative assessment teachers could not function effectively". Thus, teachers can informally assess their students without feeling that they are being assessed. For example, a teacher may assess how his or her students speak, write, or read. This type of assessment can be continuous during the teaching and learning process. Ghermaoui (2019) argued that informal means indirect assessment which is used to measures the ability or knowledge that underlies the

skill that is trying to be measured. Pronunciation tests in which students are required to read some words out loud are an example of direct testing, whereas asking students to check the word that sounds different in a group of words is an example of an indirect test of pronunciation.

Moreover, Harris and McCann (1994, p. 20) further argue that when carrying out informal assessment; three factors should be taken into consideration. These are: what to be assessed has to be worked out firstly. For assessing student, rough impression should not be relied on, but clear criteria must be established secondly. Linkage between this assessment and other types such as formal and self-assessment is significant lastly. Hedge (2000) argued that formative assessment by which teachers could make use of any information that has been gained about their students' progress as a means for any future procedure aimed to support students' learning. Teachers also use formative assessment as a means of judging their teaching in terms of its effectiveness. Thus, teachers should informally assess their students' performance without making them know that they are being assessed. This type of assessment can be continuous throughout the teaching and learning process.

2.4. Strategies of Speaking Assessment

Speaking assessment is equally important like the other language skills. The process of teaching speaking cannot go forward without employing appropriate assessment strategies. Teachers therefore, should recognize the effectiveness of their teaching strategies, they have to use these strategies for obtaining feedback by which they can have a complete picture of the whole process. Therefore, accurate testing is necessary in skills development because without testing, the process of achieving competence remains incomplete. More to the point, testing makes students sincere enough to learn things properly. In this regard teachers should accept that "if you want to encourage oral ability, then test oral ability" (Hughes ,1989). If students know they have to present something or give an oral test in English at the exam and their success in language use will depend on their performance, they will be serious enough to develop their speaking skills. Assessing speaking is integrated in teaching it, because students' progress is recognised by means of assessment conduct and the performance of students cannot be known without assessment. In this regard, Rowntree (1987, p. 162) points out that any assessment strategy to be used should "go with the conduct and style of the teaching and learning

experienced by our students”. Cooperative activities can encourage negotiation of language item (Newton & Nation, 2009). Thus, the strategies of teaching speaking are cooperative activities such as tests, role plays, interviews, dialogues, storytelling, observation, discussion and academic presentations. All of these strategies are reviewed below.

2.4.1. Tests

A test is another method of assessing students’ speaking abilities which can be used by classroom teachers. Broughton et al., (1980, p. 146) have defined a test as “a short, quick teacher-devised activity carried out in the classroom, and used by the teacher as the basis of ongoing assessment”. Therefore, according to this definition speaking teachers can test their students after each lesson to see how well they have understood. These tests usually serve teachers by providing them with information about learners’ progress in terms of their learning speaking process. Moreover, they can be beneficial in testing learners’ ability to recognise connection. On the other hand, James (1998) argues that in order to avoid any negative and possible side effects on learners that may be caused by testing, they should only be used when necessary.

2.4.2. Role plays

Teachers of oral skills may assess students through role play activities where students are asked to pretend to be in various social contexts and various social roles (Harmer, 2001; Thornbury, 2005). Creative tasks resemble real-life tasks asserts students to develop their fluency best, if engaged in tasks where all their concentration focuses on producing something, rather than on the language itself (Solcova, 2011). This means that role play is always used as a simple practice activity in information exchange when there is a limited investment that students want to make in it (Edge, 1992). Moreover, Muhd Syahir (2016) and Liu (2010) have found that the implementation of role-play as a speaking assessment helps students to be more confident as it arouses their motivation in speaking English than using other oral English tests. However, there is a problem which encountered by students when conducting role-play is in term of psychological aspect in which becomes the major problem to introverted students. Some students may be unwilling to participate and others may feel apprehensive (Kerr, et al., 2003, p.171). In this respect, Sano (1989) argued that students tend to feel nervous when making role play in

front of an audience. Therefore, using role play would help the teachers get to know them better, provide more innovative or authentic feedback and support, and better prepare for role play and other active learning strategies (Watkins, 2011).

2.4.3. Interviews

An interview is an assessment method that classroom teachers of speaking or any other assessing agent can use for gathering some information about their students. Such a method can be used when other methods such as testing do not work. For instance, teachers can conduct interviews to obtain a better idea about students such as how they think or their level of understanding, their ability to communicate or how to apply any concepts (Thornbury, 2005). In addition, interviews can also provide more information on exceptional students as is the case for those who suffer from some psychological problems which in turn may hinder their learning process. Dakowska (2005, p. 245) argued that interviews are difficult tasks in the sense that they require from the interviewer some preparation in the form of research, the selection of relevant questions and the prior analysis of native speaker interviews so as to properly evaluate both the questions and the elicited information. It is hard to assess a number of students within a short time because the teacher is put under pressure (*ibid*). Moreover, scoring is also difficult in testing communicative as the assessment includes pronunciation, grammar, fluency, content, organization, and vocabulary (Thornbury, 2007). These are perhaps the reasons for which EFL teachers remain less interested to assess students' speaking skills.

2.4.4. Dialogues

It is considered as conversations between two or more students regarding a topic being studied in class. In foreign language classrooms conversations are considered to be not the result of language learning but rather the context in which learning actually occurs (Thornbury, 2007). In order to assess students, they are asked to talk and introduce themselves to each other, and talk on different issues. Nolasco and Arthur (1987) argued that conversation activities may proceed from more controlled ones, in which the language is limited by instructions, through awareness activities which make use of audio-visual materials, fluency practice, to feedback sessions during which students analyse their own interactions.

2.4.5. Storytelling

One of the assessment strategies that EFL teachers could use in their speaking class is story telling. Wang and Lee (2007) assert that storytelling is an excellent verbal activity to help learners imagine and creates plots. It also encourages them in building up details, plot prediction, and drawing conclusion. Thornbury (2007) argued that this strategy is common ingredient of casual conversations and drama, role-play and simulation activities. These elements increase the range of registers and social roles that learners may encounter in speaking classroom. Therefore, it must be remembered that each speaking task needs to be productive, purposeful, interactive, challenging, safe and authentic (Thornbury, 2007, p. 90). The difficulty of speaking tasks, however, may be different in different students and they depend to some extent on the personality of the individual student.

2.4.6. Observation

Observation is another assessment method available for classroom teachers for assessing students speaking skills informally. Harris and McCann (1994) point out that information about students' performance can be collected by teachers in ordinary classroom conditions without setting any tests as happens when learners are assessed formally. Moreover, "Classroom observation is an important part of initial teacher training and feedback providing an invaluable opportunity for trainee teachers to improve their teaching, to get advice from, and exchange ideas with, experienced teachers. This practice should boost the trainee's confidence, enthusiasm and expertise and in this context, "the observation should be arranged in an optimistic climate of professional trust ... and should be practitioner-led and not observer-led" (Harvey, 2006, p.11). Observation as an assessing method is effective when other strategies are not. For example, it can be used when a teacher wants to know how his/her students behave when solving problems or how they work effectively with each other in completing any given speaking task. Thus, observation may also tell teachers what are the next steps to be taken; that is whether teachers need to go forward or they need to revise certain points first because the students still do not comprehend them.

2.4.7. Discussion

EFL teachers may assess their students' speaking abilities through discussion, which is based mainly on a particular chosen topic to be argued and discussed by students. Harmer (2001, p. 272) suggested a solution for teachers that allow them to avoid any difficulties which are the buzz group: where students have a chance for quick discussion in small groups before any of them is asked to speak in public. Dakowska (2005) argued that it should be stressed that the actual potential of this activity for developing an individual learner's speaking ability can be implemented only under any conditions. Therefore, it is important for teachers when they assess their students to make the arguments to be more logical and well-balanced and that they are presented in accordance with the social and cultural norms of the target language community.

2.4.8. Academic presentations

Using academic presentations by teachers are important to assess students speaking abilities in classrooms. It is useful for learners who study language for academic purposes and need practice in giving presentations or conference papers (Thornbury, 2007, p. 94). Thus, teachers should discuss the formal features of the genre and analysing the specific language patterns typical of each stage of an academic presentation. This can be conducted before students actually start preparing their own speeches, as well as discussing the effectiveness of individual presentations.

To conclude, designing activities for teaching and assessing speaking skills requires some principles to be considered. In this regard, Brown (2001) stated that speaking activities need to maximize the production of language to provide the best conditions for autonomous language use. Therefore, EFL teachers should bear in mind what student needs, from language-based focus on accuracy to message-based focus on interaction, meaning and fluency to encourage the use of authentic language in meaningful contexts (ibid).

2.5. Principles of Assessment

In order to measure what teacher intends to assess, there are a set of requirements that in every test should be followed. These requirements are called validity and reliability in order to design effective and appropriate assessment.

2.5.1. Validity

Validity is considered to be one of the most significant factors of any instrument. That is; it means that the instrument is capable of measuring what it is originally designed for. Hence, assessment as a means that is widely employed in educational processes for performing several purposes should be as valid as possible in order to fulfil its set goals. Language test validity is concerned with testing what it purports to test. Fulcher and Davidson (2007, p. 4) stated that, when designing a test there is a specific intention to measure something, that the ‘something’ is ‘real’, and that the principle of validity require finding out whether a test ‘actually does measure’ what is supposed to measure. According to Satterly (1989) there are several types of validity such as face validity and content validity. Face validity means that a test should seem to test what it aims to test. Black (1998, p. 42) states that “If for example, there is some agreed aim of assessing competence in the subject, then experts might look at a test and consider whether in their judgment the questions would call for those performances which characterised that competence”. In other words, if we want to test students in any language skill such as reading, the test to be carried out in this respect should not test any other skills rather than reading, for example, to introduce a question about writing skill. Therefore, face validity is regarded as an important feature in any test. Further account should also be taken for the other types of validity which Gipps (1990) and Murphy (1999) point to such as predictive validity, construct validity and concurrent validity. Predictive validity has correlation with future performance of individuals. For example, if we take any level, such as ‘A’ we can ask whether it can really predict the future performance of those students who may reach this level. Concurrent validity means whether a certain test about any skill such as reading, for example is substantially capable of giving the same results of any other test for the same skill. Another type of validity is called construct validity which according to Black (1998), means identifying the factors behind the variation between the scores gained by learners in some tests. In addition, it could also mean whether or not a test is sufficient to measure a skill. For example, if a test is set for measuring a student’s reading, it should be able not only to test if such a student knows how to read, but it should also test as Gipps (1990) stresses other aspects in the same skill such as comprehension, accuracy and enjoyment. In conclusion, validity is related to the content and construct of a test while reliability is related to the score (Li, 2011).

2.5.2. Reliability

Reliability means when a test is said to be reliable and if the results do not differ for no logical reason. In other words, reliability consists on getting similar scores for the same test giving to the same student. Test reliability refers to the consistency and stability of test results on different situations. Lado (1964, p. 330) argued that “Reliability has to do with the stability of scores for the same individuals. If the scores of students are stable the test is reliable; if the scores tend to fluctuate for no apparent reason, the test is unreliable”. Moreover, there are two components of reliability: the performance of students from occasion to occasion, and the reliability of the scoring. This means that in order to make an oral test reliable, testers should try out to achieve consistent performances from students and to achieve scoring reliability, however, James (1998, p.157) states that “reliable assessment is one which gives consistent results by eliminating errors arising from different sources”. Furthermore, score reliability is the other essential component to test reliability. Score reliability consideration of an oral test involves scoring and criterial levels. Criterial levels are specified to obtain valid and reliable scoring. Therefore, efforts should be made to ensure that all students have the opportunity to learn just what will be required of them. Then, uniform and non-distracting conditions of administration should be also provided.

To conclude, it can be argued that validity and reliability are interrelated with each other but they have different focuses. When validity focuses on test content such as test purpose or use while reliability, it focuses on the result or response or score. Reliability is a necessary condition for validity but it alone is not sufficient. In other words, validity is rather more comprehensive while reliability is much narrower. Thus, a balance should be made between validity and reliability in designing an oral test.

2.6. Speaking Assessment Tasks

Choosing or designing speaking assessment tasks is a big challenge for EFL teachers. While a definition of speaking is fundamental for speaking assessment, a further central segment is simply the test task itself. Kim (2009) argues that a test task could be considered as the vehicle of evaluation by which a sample can be evoked, hence, scored and, based on this idea-basification. She also states that the research in speaking assessment has the practical functioning of a wide range of speaking tasks in terms of their appropriateness and effectiveness for bringing-out

samples of language performance in different settings. Moreover, it was argued that made a qualification between two distinct yet related sorts of informational talk which are expository and evaluative. These include factual information, for instance, narration, comparison, description and instruction, while the last includes thinking, for example, clarification, avocation, expectation and decision (Bygate, 1987, p. 32).

Luoma (2004) supported Bygate's conceptualisation of the variety of speaking tasks, pointing attention to that “it is helpful to test the sorts of talk independently, as they may give diverse data about the learners' skills” (p.32). Bygate's idea of the qualification of classifications and their operationalisation in speaking tasks formed a reason for creating and explaining task classifications. However, this idea does not expressly perceive the essence of interactional capability which lies at the core of conversational discourse. It was argued that intuitiveness as perhaps the main parts that add to the test characteristics. From their point of view, intuitiveness can be characterised as far as the manners by which the test taker's areas of language information, metacognitive procedures, effective information, and powerful schemata are engaged by the test task (Bachman & Palmer, 1996).

Moreover, the task difficulty in speaking tests is considered as a significant issue in language testing (Galaczi, 2010). The issue has been fundamentally tended to regarding making deductions from the test-takers' oral language execution at various language levels. Performance conditions gave an order of the classifications regarding certain conditions that distinguish task difficulty (Muñoz & Álvarez,2010). The instance of these conditions is preparing under regular time requirements and level of correspondence. Therefore, it can be argued that these conditions reveal the difficulty in speaking tasks, particularly inside the level of interaction appointed in the task. Furthermore, the literature review also indicates another way to deal with task difficulty which is the classification of tasks as one-way and two-way tasks. Swain (2001) contended that multi-ways tasks can be led to make derivations from more complex constructs. These early examinations about task classifications and tasks difficulty essentially centre on psycholinguistic classes, for example, accuracy, fluency and the intricacy of information base of the language and its effect on the level of difficulty of the task (ibid).

Additionally, Muñoz and Álvarez (2010) stated that one of the challenges that should be considered is the conceptualisation of task difficulty in speaking assessment and classifications to incorporate test-takers' L1 cultural contexts as a classification. In such manner, Fulcher and Reiter (2003) contended that task difficulty in speaking tests cannot be only characterized regarding parameters like task condition, individual capacity and rather severity. All things being equal, they came to perceive that the idea of task difficulty should be connected to the cultural and social requests made by the task. Fulcher and Reiter recommended another approach to deal with task difficulty which involves the examination concerning pragmatic task features inside the cultural assumptions for speakers in communicative circumstances. The use of different tasks in speaking assessment has recently been the focal point of numerous investigations led in ESL/EFL contexts (Galaczi, 2010; Muñoz and; Poonpon, 2010). The use of numerous tasks is said to have multiple preferences; among these advantages is their potential contribution to the reliability and validity of speaking assessment. For instance, Muñoz and Álvarez (2010) thought about that the utilization of different tasks considers for more reliability in the assessment process and gives students opportunities to show a greater amount of their knowledge base.

Furthermore, Galaczi (2010) argued that a variety of tasks can possibly give a valid and reliable speaking assessment; however, consideration likewise should be paid to whether the tasks really inspire the language needed to assess test-takers' oral language performance. As Luoma (2004, p. 36) claimed, “the choice of tasks for the main part of the test (description, narrative, topic-based discussion, etc.) depends on the kind of information that is needed from the scores”. This concludes that the suitability of the task assumes a significant part in achieving validity in speaking assessment. O’Sullivan, et al., (2002) raised the issue of test task validity regarding the congruity between the test taker's real presentation on a given task and the test planner's assumptions dependent on his/her definition of the construct(s) that structure the task. They clearly contended for the significance of the match between the proposed content and the real matter, without considering the examination of the construct validity would be based on sand. They thought that the meaning of the construct is underpinned by establishing the nature of the actual performance elicited by test tasks, i.e., the true content of tasks.

Finally, what has already been discussed is also supported by Kim (2009), who featured the significance of validating a given speaking test prior to surmising examinees' language capacity from the test outcomes. That is, tests' developers and users need to "ensure what the test plans to quantify (e.g., different language segments, performance on tasks) and whether a test really measures what it expects to measure" (p. 2). Therefore, teachers should know that there are numerous examinations tended to speaking tasks regarding their use and appropriateness.

2.7. Types of Scoring and Rating Scales

Assessing the speaking skill is an extremely difficult and complex matter. There are two main types of rating scales holistic and analytic which identify the overall approach to assessment of a particular task. Two methods (holistic and analytic scoring) of testing oral performance are used in speaking assessment (Luoma, 2004). These scorers assign a single score based on the general quality of students' oral performance without relying on scoring specific ability (Brown, 2004, p.191). Holistic rating captures an overall impression of the speaker's performance, analytic scoring assesses the speaker's performance on a variety of categories, such as delivery, organisation, content and language (Chuang, 2009). The differences in the use and the objectives of holistic and analytic rating scales reflect the advantages and disadvantages of each type. Holistic rating scales have the main advantage of being very rapid (Hughes, 2003). In addition, they are more practical for decision-making since the assessors only have to produce one score (Chuang, 2009), and make scoring easier in that there is less to read and remember than in a complex grid with many criteria (Luoma, 2004). However, that may lead to a disadvantage of lacking reliability in the scores obtained from such a type of scale because it does not guide and help students to progress in their speaking ability as it gives no details about their weaknesses in speaking.

In contrast, the analytic rating requires the assessor to pay attention at individual criteria separately such as, pronunciation, fluency and accuracy and gives each of them its own score, then create a final score (Hughes, 2003, pp. 94 – 95). In other words, the score in this approach is the sum of several aspects of the oral performance rather than a single score. Therefore, employing analytic rubrics helps students to enhance their oral performance by recognize their area of weaknesses and which aspect of speaking they need to work on for further improvement.

However, it takes more time and energy for teachers to apply it. With regard to analytic scales, the time factor is considered a main disadvantage in that they take longer practice of scoring, yet they provide more reliable scoring (Chuang, 2009; Hughes, 2003). Another disadvantage of the analytic scale is that ratters may find it difficult and confusing to concentrate on different aspects in a sample of language to be assessed and to evaluate them simultaneously. Moreover, the focus on discrete elements of the performance may divert a rater's attention from its overall effect (Hughes, 2003; Davies et al., 1999; Kim, 2005). Therefore, language teachers need to examine "the suitability of the given rating scale to the quality of language samples actually elicited in each testing situation" (Nakatsuhara, 2007, p. 84).

The literature review also shows that a rating scale is a way of measuring the quality of the output that results from the spoken language tasks in a specific assessment context (Underhill, 1987; Weir, 1993). It is also considered a tool by which EFL teachers can gain a clear picture about students' performance in terms of different skills within a sample of language that needs to be assessed (McNamara, 1996). These definitions have the potential to contribute to EFL teachers' knowledge and perception of the practice of oral language assessment. According to Kim (2006) the choice or design of a rating scale depends on the purpose of the assessment and it has the potential to define the speaking ability to be assessed. Crocker and Algina (1986) argued that a scale represents the range of values that is associated with particular levels of performance, and scaling rules represent the relationship between the characteristic of interest and the value assigned. Therefore, it is essential to consider the scoring rubric which provides the test users with the actual aims of the assessment (ibid).

Furthermore, measurement of a speaking performance, however, requires a different kind of scale in order to facilitate description, standard performances. Such descriptors are typically associated with, but are not limited to, descriptions of the following components of a speaking performance at different levels of the scale: pronunciation; phonological control; grammar/accuracy; fluency; vocabulary; coherence; and organization (Kim, 2006). Moreover, a number of published rating scales are designed on the basis of native-speaker standards, such as those of the FSI and ACTFL, as well as the Australian Second Language Proficiency Ratings (ASLPR) (as cited in Fulcher, 2003a). This idea was challenged by Davies (1991) who argued

that the term ‘native speaker’ in rating scales was not sufficiently defined in order to make it useful in a testing situation. Davies (2003) also added another challenge when he said that the difficulty of defining the form of Standard English of the native speakers because of the variability of their language proficiency. This means that recognised rating scales on the standards of native speakers is not appropriate in the assessment of EFL speaking skills.

2.8. Assessing Linguistic Constructs in Speaking

The construct of speaking help tests developers and enables language teachers to be more skilful at manipulating speaking tasks and to identify the language ability to be assessed, which can best meet their methodological teaching of speaking skill (Hiebert & Kamil, 2007). It was argued, “no operational construct definition can ever capture the richness of what happens in a process as complex as human communication, even if the speaking test is mediated by tape and computer” (Fulcher, 2003, p. 19). This might be due to the diversity not only in the purposes of speaking tests, but also in the interpretations of the definition of the construct, who provides these interpretations, how they rationalise such interpretations, and in what assessment context. This means that speaking constructs can serve as a corner stone upon which speaking tasks and rating scale can be designed (ibid). Thus, an analysis of constructs of speaking should to be considered when designing rating criteria or selecting a speaking task. These constructs are discussed below.

2.8.1. Vocabulary

The assessment of vocabulary in oral language performance is essential (Iwashita et al., 2008). Vocabulary is defined as the words of a language and word is defined as a unit of language, consisting of one or more spoken sounds or their written representation, which function as a principal carrier of meaning (Ferreira, 2007). This implies the range of lexicon is another linguistic factor in the assessment of speaking skill. Researchers in speaking assessment have considered the relative weight of individual features of oral language performance, and vocabulary is one of the recurrent linguistic features that is examined in different levels of language proficiency (Iwashita et al., 2008). It was argued that descriptions of vocabulary are placed in most rating criteria for speaking as they reveal the speaker’s richness of lexicon (Luoma, 2004). Although rating scales vary in terms of the criteria by which vocabulary is assessed at different language levels, two main aspects are considered principal while assessing

spoken performance, namely the number of words produced and the range of words, which (Iwashita et al., 2008) referred to in their investigation of the nature of spoken proficiency in an ESL context as ‘token’ and ‘type’ respectively. Therefore, the assessment of vocabulary as speaking construct is essential regarding to the variation, amount and suitability of vocabulary used.

2.8.2. Pronunciation

Pronunciation means sounds, intonation, stress and voice quality. It is considered as speaking sub-skill; these elements should be taken into account when speaking assessment is carried out. Sakale (2012) argued that the sound of speech or pronunciation is one of the main characteristics of spoken language and is viewed as the departure point of any oral interaction since the effectiveness of any communication process is linked to mastery of the sound system. Schmitt (2002) defined pronunciation as “a term used to capture all aspects of how we employ speech sound for communication” (as cited in Roohani, 2013, p. 88). However, the sound of speech is a ‘thorny issue’ for language assessment since pronunciation is usually judged depending on a native speaker standard of pronunciation which in turn may result in inaccurate assessment on the part of the examiner (Luoma, 2004). Therefore, identifying and achieving a valid assessment need to draw clear criteria of pronunciation.

2.8.3. Accuracy

Accuracy is described as “the simplest and most internally coherent construct, referring to the degree of conformity to certain norms” (Pallotti, 2009, p. 592). This indicates that accuracy is a clearly-defined construct in that it essentially represents the linguistic aspects such as the appropriate use of grammatical structure and lexicon at different levels of oral proficiency. The increasing recognition of considering grammar of speech in the assessment of speaking seems to coincide with the re-conceptualisation of spoken grammar in English language teaching. The assessment of spoken grammar is one of the important issues addressed in the definition of accuracy as a construct in oral language performance (Goh, 2009). This leads to identify a number of grammatical aspects that characterise spoken grammar such as spoken idea units and grammar in planned and unplanned speech (Luoma, 2004).

2.8.4. Fluency

Fluency is a “multidimensional construct, in which sub-dimensions can be recognised, such as breakdown fluency, repair fluency, speed fluency. Once it is established which of these sub-dimensions is at issue, it is in principle relatively transparent what is being measured” (Pallotti, 2009, pp. 591-592). Fluency is viewed as a controversial issue in terms of its definition as a construct and the measure by which this construct is assessed. When evaluating accuracy, language teachers try to find out whether students are capable of constructing sentences of language that follow acceptable rules of usage (Fulcher, 2003). It was questioned the concept of fluency in terms of the lack of clear definition that defines what fluency actually means (Fulcher, 1996). This reflects the various perceptions of fluency as reflections of speech planning, thinking processes and a hesitation phenomenon (Freed, 1995). This highlights the issue of validity which mainly requires a clear definition of the construct that is intended to be assessed.

2.9. Assessing Communicative Constructs

Kramsch (1986) argued that existing proficiency tests at that time focused on the grammar and lexicon while the dynamic process of communication was relatively ignored (as cited in May, 2011). Therefore, researchers in the social aspect of speech highlighted the importance of developing the notion of interactional competence (Bygate, 1987; Fulcher, 2003; Galaczi, 2010; May, 2011). The interaction is described as “the skill of using knowledge and basic motor-perception skills to achieve communication”. It mainly involves “making decision about communication, such as: what to say, how to say it, and whether to develop it, in accordance with one’s intentions, while maintaining the desired relations with others” (Bygate, 1987, p. 6).

Moreover, it is important to highlight the features that characterise speaking as interaction in order to gain insights into the elements to be tested within an interactional framework in a certain speaking test context (Weir, 1993). More to the point, Weir designed a framework of spoken language interaction that helps test developers to construct features of spoken interaction in the design of their tests: A three-part framework is proposed, covering: the operations (activities/skills) that are involved in spoken interaction such as informational routines, e.g. telling a story, and improvisational skills that might be called into play when the performance of these routines breaks down, e.g. requesting clarification...; the conditions under which the tasks

are performed, e.g. time constraints, the number of people involved and their familiarity with each other(...); and the quality of output, the expected levels of performance in terms of various relevant criteria, e.g. accuracy, fluency or intelligibility (ibid, 1993, p. 30).

The complexity of testing interaction in speaking assessment in that talk is co-constructed by participants in a conversational discourse and, as such, responsibility for talk cannot be assigned to a single participant, but to the context-bound joint construction that is formed in communications. This means that this co-construction makes the assessment of interactional competence challenging for language testing (Fulcher & Davison 2007). The fundamental interdependence of the two participants in paired speaking tests which derives from the interaction they produce has led to the re-conceptualisation of awarding shared scores for interactional competence (Galaczi, 2010). It was argued in a similar vein that the implications of assessing interactional competence in speaking assessment are an area of concern (Fulcher, 2010; May, 2011). This is because the co-construction of the interaction by the participants, especially in paired speaking tests, has been found to have the potential to elicit a range of interactional competences such as management of interaction, turn taking, and initiating and ending topics.

2.10. External Challenges in Speaking Assessment

Speaking assessment is challenging because there are many external factors that might influence teachers' impression on how well someone can speak a language and these may be reflected in the assessing or scoring of learners' speaking (Madsen, 1983; Taylor, 2006; Chuang, 2007). The most important issues such class size, lack of facilities and lack of training sessions which are respectively discussed below.

2.10.1. Class Size Effects

A classroom size is an essential element for successful teaching and assessing English speaking. It is defined as an educational tool that can be used to describe the average number of students per class in a school (Adeyemi, 2008). Moreover, researchers such as (Achilles, 1999 & Bennett, 1996) concerned with classroom interactions between teacher and learners' interactions. They believe that bigger classes will decrease the amount of time because teachers will spend time on

instruction and when they dealing with individual students. It can be argued that although the teachers may suffer from lack of performance but the effects of class size in Libyan Universities are existed. Angrist and Lavy (1999) found that the significant effects of class-size in their schools are much larger when the average being 32 students. This is true because teachers in small size classes could apply their knowledge of teaching and assessing better than when they deal with large size classes. Thus, classroom size assumed has great significance effects on teachers' performances and students' achievements.

2.10.2. Lack of Facilities

Libyan Universities are suffering from lack of facilities which could help teachers in teaching and assessing speaking skills. These materials provide the language learners with valuable speaking opportunities. For example, videos are often more preferred by language students, as they are more attractive and motivating than audio recordings. Videos enable teachers to display a variety of challenges, such as: using pictures only, sound only, or pictures and sound together (Miller, 2003). These methods or activities aim at “generating new ideas and engaging in real communication” (Stempleski, 1990, p. 12).

Moreover, there is no language liberators and internet access in these universities which affect negatively on teachers and students. The internet is considered as a very important and rich source for authentic materials (Arianie, 2017). It is likewise important to note that a large amount of the materials mentioned are now easily available online. In this situation, students do not watch a film in a class represents, actually, a rich source of language input with verbal and visual stimulation that can develop speaking abilities. Films and TV shows represent a rich source of input for learners which can be exploited in various ways and on different levels to improve their communicative competence (Gilmore, 2007).

To sum up this discussion, for language teachers, the lack of materials in EFL classes causes many problems when they teach and assess their students in speaking classes. Therefore, the classrooms in universities should be provided by the necessary materials such as projectors, computers, photocopy machines and flash cards.

2.10.3. Lack of Training Sessions

Training sessions for teachers is essential for the successful teaching and assessing of speaking. EFL teachers need to know about new methods and techniques in order to manage their activities. Karavas-Doukas (1998) confirmed that innovation can lead to positive changes in curricula, the beliefs and behaviour of teachers. Therefore, it can be argued that training sessions are needed for teacher development but these should be focused and organized so as to produce positive outcomes. It is important for teacher trainers to encourage participants in teacher training programmes to discuss both ‘overt’ and ‘hidden’ barriers to the successful implementation of change in their own teaching/learning contexts (Shamim, 1996). As it has been acknowledged, the effective English language teaching mostly depends on training of EFL teachers. However, in many cases, EFL university teachers do not receive any effective, meaningful training in teaching and assessing speaking skills. Allen (2008, p.23) pointed out that, “if teachers are not equipped to deal with the implications of a new approach, they are likely to revert to the security of their previous behaviour and the desired change may not take place”. In brief, training courses are important to update teachers’ knowledge in order to be able to approach their teaching and assessing tasks more effectively.

2.11. Previous Studies in Speaking Assessment

The literature showed that there are many studies conducted on teaching and assessing speaking skills. Assessing speaking has recently been viewed as an important issue in language testing research. There are many researchers from different contexts, such as (Agasøster, 2015; Kim, 2003; Lee, 2010; Muñoz et al., 2003; Chang, 2006; Grada, 2014; Tran & Nguyen, 2017; Kormos and Denes, 2004; Abd El-Razek 2010) who explored the teachers English speaking assessment in class and their beliefs about the orientation and purpose of assessment practices. For example, Muñoz et al., (2003) investigated EFL teachers’ understanding of oral assessment in the Colombian context. One of the research questions related to the speaking assessment constructs that the participants considered important in oral language assessment. They found that the participants follow the traditional way of oral assessment in that they paid attention to accuracy. Moreover, a tendency to assess language from a communicative perspective was also existed.

Another study conducted by Kormos and Denes (2004) in which they examined native and non-native speaking teachers' perceptions in the assessment of fluency of second language learners' speaking performance. Their study was aimed to investigate the linguistic and temporal variables that have the potential to predict native and non-native speaking teachers' perceptions of fluency. They used tape-recorder speech samples which were collected from sixteen Hungarian EFL students. They divided the participants into two proficiency groups. The findings showed that there was a set of variables that were good predictors of fluency scores both for native and non-native speaking assessors: speech rate, the mean length of runs, phonation-time ratio and space.

Furthermore, Abd El-Razek (2010) carried out a study in order to explore the effectiveness of a program in order to develop instructional speaking skills of English department students in faculties of education. The participants in this study were twenty-two students who study in the third year. The researcher designed a checklist of the instructional speaking skills to be developed throughout the program and to be measured by the pre-post-test to evaluate the students' performance. The findings of his study revealed that there was a statistically significant difference between the mean scores of the students in the pre-post-test of instructional speaking skills, in favour of the post-test. This means that the effectiveness of the proposed program is in developing the participants' instructional speaking skills.

Moreover, Muñoz and Álvarez (2010) investigated the washback effects of an oral assessment system in a Colombian EFL context. This oral assessment was designed to promote positive washback in the teaching and learning of English language. The researchers referred to washback as the influence of testing on teaching and learning. Only 14 EFL teachers and 110 EFL college students participated in the research. Structured observations and external evaluations were used as methods to collect data from students and teachers. Their study was purely quantitative and the findings revealed positive washback in some of the areas investigated.

A study conducted in a Taiwanese context by Chou (2013) who investigated teachers' interpretation of students' speaking scores through the use of two different types of rating method. It was aimed to investigate how students could benefit from the feedback of the description of the two rating methods: level descriptors and checklist. The sample included 15

English language teachers and 300 college students. The researcher found that the rating type seemed to have an impact on teachers' interpretation of students' language performance.

There was also a study based on the theoretical framework that underpins authentic assessment. This study was conducted by Muñoz, et al., (2003) who investigated teachers' understanding of speaking assessment in order to establish speaking assessment criteria for the teachers and the institution. The researchers used a focus group technique with 30 teachers in order to examine the teachers' beliefs. The questions were related to teachers' understanding of assessment, aspects of speaking assessment, assessment techniques, the rationale behind conducting the assessment and frequency of assessment. The findings of the study reported that most of the teachers lack of knowledge of assessment. This was reflected in their misunderstanding of the difference between the functional aspects of both formative and summative assessment. They also found that their assessment practices were unsystematic and lacked planning. Therefore, they recommended establishing in-service training courses for language assessment as that would help teachers recognise that assessment is a powerful tool for developing language teaching and learning.

In conclusion, teacher's assessment strategies are usually investigated along with teaching language practise in general. It is obvious that the previously discussed studies provide insights into the importance of considering teachers' knowledge of speaking assessment and how they employ it in classroom. However, little attention has been given to teachers' assessment strategies in Libya particularly at universities. This research therefore, was aimed to fill this gap. Thus, the current study was designed to the shed light on the speaking assessment strategies that EFL Libyan teachers use at the undergraduate level at Zawia University and the challenges that they might encounter during assessing their students in speaking classes. This would take our understanding forward to see how teachers' current knowledge of speaking assessment is reflected in their assessment practices in the light of the context where they work. The following chapter discusses how the research methodology was designed.

2.12. Summary of the Chapter

This chapter reviewed the definition of speaking, speaking and assessment, types of speaking assessments, strategies of speaking assessment, principles of assessment, speaking assessment tasks, rating scales, types of scoring, assessing linguistic and communicative constructs, external challenges in speaking assessment and previous studies in speaking assessment. It was found that, while certain areas have been well researched and documented, gaps in the research still exist, namely speaking assessment strategies used by Libyan EFL University teachers. In the next chapter, the philosophical standpoint, methodology and processes of data analysis employed in the current study are presented.

Chapter Three

Methodology

3.0. Introduction

The purpose of this research is to investigate the speaking assessment strategies used at the undergraduate level in Zawia University, and to find out the challenges that EFL teachers might encounter when they assess their students in speaking classes. This chapter provides a description of how the research is designed to address the research questions. It demonstrates the design of research and describes the methods of data collection. It also explains the selection of participants, the pilot study and preparation data for analysis. Ethical issues are also provided.

3.1. Research Design

The research design is defined as the overall strategy that the researcher chooses to integrate the different components of the study in a coherence and logical way (Cohen et al., 2007). The process of developing a research design combines three broadly connected and interdependent components: the theoretical, methodological, and ethical considerations relevant to the specific project (Dornyei, 2003, p.12). The nature of this research forced the researcher to use two different types of data; quantitative and qualitative. It was conducted to investigate the speaking assessment strategies teachers used at the undergraduate level in Zawia University students, and to identify the challenges EFL teachers might encounter in their speaking classes. Creswell (2009) identifies the quantitative approach as a structured approach, where as he classifies qualitative approach as unstructured research. Therefore, a mixed research approach was used in this study.

Furthermore, using both quantitative and qualitative methods is a powerful technique that facilitates validation of data through cross verification from two or more sources (Dornyei, 2007). Consequently, a structured questionnaires and semi-structured interview were used in this study. The rationale behind using this framework in this study was based on the precept that a fuller understanding of the research focus can be obtained if it is investigated from different perspectives (Cohen et al., 2007). More importantly, the two methods were used because the findings from the different types of data collection and analysis can be crosschecked to identify

areas of similarity and, perhaps more productively, differences, anomalies and apparent contradictions in the findings (Dornyei, 2007).

3.2. Data Collection Instruments

This section outlines the data collection instruments used in the current study. Two methods are used for data collection. They include questionnaires and interviews. The structured questionnaire was used to identify the frequency of using the speaking assessment strategies by teachers at the undergraduate level in Zawia University. However, the semi-structured interviews tool was used to understand the challenges EFL teachers might encounter in their speaking classes. Both instruments are described below.

3.2.1. The Structured Questionnaire

In this study, a closed-ended questionnaire was used with teachers to identify the speaking assessment strategies they use in their classes. This type of questionnaires is considered as “questionnaires which are focused on getting participants responses to standardized items for the purpose of confirmatory research in which specific variables are measured and hypotheses are tested” (Johnson and Christensen, 2012, p. 170). It is a very good quantitative technique since it enables large-scale numerical data to be obtained in short period of time (Kothari, 2004). Most of the questionnaire items were adopted from Grada’s study (2014) entitled as (*An Investigation into Libyan EFL Novice Secondary School Teachers' Current Knowledge and Practice of Speaking Assessment: A Socio-cultural Perspective*). The questionnaire consisted of both questions and items. Eleven questions were designed in order to understand the teachers’ background knowledge about speaking assessment, while the thirteen items were designed to identify the frequency of using the speaking assessment strategies by teachers. Each statement was followed by a 5-points Likert Scale arranged from always, usually, sometimes, rarely and never (see Appendix A).

Furthermore, the reason behind using this type of questionnaires is that because it is effective mechanisms for efficient collection of certain kinds of information, particularly language learning strategies (Dornyei, 2003). In order to ensure the validity and reliability of the questionnaires, it was reviewed by five experienced teachers who have taught English speaking

at Zawia University. Denscombe (2007) pointed out that even small error in data entry can radically skew the overall direction towards which answers to questionnaires point. Therefore, a high degree of consistency between the component parts of the data was considered to be useful indicator in terms of validation.

3.2.2. Semi-structured Interviews

Generally speaking, the interview itself is considered as a conversation, usually between two people. But it is a conversation where one person, the interviewer is seeking responses for a particular purpose from the other person the interviewee (Cohen et al., 2007). It is known that “the interview method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses” (Kothari, 2004, p. 97). This type of interviews was used since the researcher was interested in teachers’ challenges about speaking assessment strategies. Moreover, semi-structured interviews were employed because the format of this type is open-ended and the interviewee can be encouraged to elaborate on the issues raised in an exploratory manner (Dorney, 2007). The interview questions were formulated (see appendix B) consisting of 10 questions that were carefully constructed by considering the theoretical framework of teacher’s challenges related to speaking assessment strategies. Furthermore, five teachers were interviewed. All interviews were recorded to ensure that what was written is accurate. According to Bell (2005, p. 164) “tape-recording is useful to check the wording of any statement you might wish to quote and to allow you to keep eye contact with your interviewee”.

To enhance validity and reliability of the interviews, firstly, teachers’ consistency over time was considered, as Punch (1998) stated, if the same interviews were to be carried out at another time but under the same circumstances, the same results are not guaranteed. This is because the teachers might change their responses as a result of experience they have gained since the last interview. Secondly, seeking corroboration from alternative sources of information was based on the same topic, is useful in supplying confidence in the data collected. A further way of assuring validity was considered in this research to estimate whether the interviewee is in a position to have expert knowledge in order to answer questions put with relative authority. This is because in some cases, respondents / informants may be tempted to respond quite persuasively into areas which are liable to be beyond their competence (Cohen et al., 2007). In order to increase the

validity and reliability of this study, interview questions were piloted by three experienced teachers who have taught English speaking at Zawia University.

3.3. Pilot Study

Pilot study can be considered as a small study to test the data collection tools. In this respect, Bell (1993, p. 84) reaffirmed that “all data-gathering instruments should be piloted to test how long it takes recipients to complete them, to check that all questions and instructions are clear and to enable you to remove any items which do not yield usable data”. The purpose of the pilot study is not only to acquire data but also to learn how to acquire data properly and accurately (Burns, 2000). Piloting research instruments are a significant for increasing the validity and reliability (Bryman, 2008). To evaluate the questionnaire in terms of its content and structure, it was checked by five experienced university teachers in order to examine the structure, clarity, logical flow of the questionnaire, length and the order of the statements. The pilot participants were asked to fill out the questionnaire to verify if the items are clear and easy to follow before used in the main study. Moreover, it provided some suggestions to develop and organise the questionnaire by deleting or modifying certain questions and items to be ready to use in the main study. These modifications are discussed below:

- Some items in part two which is related to teachers speaking strategies use were cancelled, because they did not serve the purpose of the questionnaire.
- One existing question was not clear according to the participants’ feedback in the pilot study. Therefore, “I connect the selection of assessment criteria” was modified to “I connect the selection of assessment criteria to the aim of speaking assessment”.
- Rearrange the numbers of questions and amend some grammatical mistakes.

Furthermore, three teachers were interviewed to shed light on the importance components of Libyan EFL teachers’ views and challenges faced in assessing speaking skills. Some interviews questions were modified to be ready for collecting the data in the actual study. Moreover, the pilot study enabled the researcher to improve interview schedules and specific questions. For example, the ordering of the interview questions was changed logically in terms of their difficulties. The researcher also became aware of the diversity in the construction and meaning of certain concepts. The data obtained from the pilot study was analysed to know if there any

difficulties when the researcher deals with the real data. One of the benefits gained from the pilot study was the interviews data collected should be transcribed immediately and organized in separate profiles. The participants of the pilot study of the interviews added two questions which are “Do you use different assessment categories for each task or you keep using the same ones for each task?” and “What factors do you think have an influence on the actual practice of speaking assessment?” to the questions of the interviews.

3.4. Sampling and Population

Sampling is considered as the process of selecting a few samples from a bigger group the sample population (Cohen et al., 2007). This means a sample is a subgroup of the whole population. Both probability and the purposeful sampling were used in this study. A probable sample was used because “all people within the research population have a specifiable chance of being selected. Only within random samples do participants have an equal chance of being selected” (Dawson, 2002, p.50). Being more specifically, the researcher used a cluster sample. Using this method, geographical ‘clusters’ are chosen. Thus, the researcher conducted her study only on the teachers who teach in English departments in colleges near to her residence. All the teachers in these departments were asked to fill the questionnaire. The researcher distributed 40 copies of the questionnaire in four departments at Zawia University but only 35 copies were returned. Moreover, the purposeful sampling was employed to select participants matching the main concerns of the research. Purposeful sampling is “based on the assumption that the investigator wants to discover, understand, and gain insight and therefore must select a sample from which the most can be learned” (Merriam,1998, p. 61). Therefore, only five Libyan EFL teachers were selected according to their experiences to be interviewed. The questionnaire background section guided the researcher to reach to these five teachers. All the participants were chosen from English departments distributed between Faculty of Art, Faculty of Education in Zawia, Faculty of Education in Abu-Essa and Faculty of languages in Zawia city, because these departments are located near to the researcher residence. These participants were chosen as they were teaching speaking skills or they taught it before at their English departments. The researcher used his judgment to select participants who can provide sufficient data. The researcher utilised different techniques of sampling, because using a mixture of sampling techniques within one project

which may help to overcome some of the disadvantages found within different procedures (Dawson, 2002).

3.5. Procedures of Data Collection

After the pilot study, the data was collected through a structured questionnaire and semi-structured interviews in two phases. In the first phase, the teachers' questionnaire was conducted. This instrument was administered in the middle of March and April 2021. After receiving approval from the University, the researcher visited only four English departments in Zawia University. The researcher introduced herself to the heads of the departments and explained the purpose of the survey. The questionnaire was handed to forty teachers in different English departments. Only thirty-five teachers returned the filled questionnaire. Bell (1999) argues that one limitation of questionnaires is the often-low response rate.

In the second phase, the participants were interviewed in order to explore the challenges they might encounter when they assess their students in speaking classes. Semi-structured interviews were used as a tool of data collection because such questions as Seidman (2013, p. 9) clarifies, "building and finding out how the participants see the phenomena of the study through their open responses". All the interviews were conducted in English and took place in May during the academic year 2020- 2021 in Zawia University. Furthermore, only five teachers (two teachers from Faculty of Art and one from each department) were interviewed in this study. The interview questions were set up in specific order, using clear language and simple words. The questions were carefully designed to cover aspects of teaching and assessing speaking skills challenges. In order to gain useful data and to make it more manageable, the researcher interviewed each teacher individually in their departments. The time of the interviews ranged between thirty-five and fifty minutes. The length of the interviews was dependent on the interaction between the interviewer and interviewee. The researcher stopped on only five teachers because the sixth and the seventh one did not add any new information which means a saturation point was reached. All the interviews were recorded and transcribed to be analysed by thematic analysis method.

3.6. Ethical Issues

Ethical issues require special consideration as they relate to the people with whom researchers conduct their research (Bryman, 2001). The consideration of the ethical issues was necessary for the purpose of ensuring the privacy as well as the safety of the participants. Creswell (2009) argues that ethics in educational research is viewed as one of the significant issues that researchers should take into account from the outset of the research because their studies entail contact with people as research participants. Researchers need to respect the participants by ensuring their anonymity and confidentiality, and to protect them from any potential harm while they are involved in the study (ibid). The current study carefully considered the ethical issues of data collection. A letter from the supervisor and university admissions confirming that the researcher wanted to collect data was issued (see appendix C). In the design of the questionnaire, the researcher took into consideration Cohen's et al., (2007) argument in which they said that "the questionnaire will always be an intrusion into the life of the respondent, be it in terms of time taken to complete the instrument, the level of threat or sensitivity of the questions, or the possible invasion of privacy" (p.317). Following the ethical principles of data collection, the researcher included in the cover sheet all the information that the participants needed to know about the study and the process of data collection. The researcher also provided them with information about the aim of the questionnaire and interviews. In addition, the participants were assured that their responses would remain confidential and their identity anonymous. Among the significant ethical issues that were considered in the research process include consent and confidentiality. In order to secure the consent of the selected participants, the researcher relayed all important details of the study, including its aim and purpose (Dörnyei, 2003). Furthermore, the researcher informed the participants of their rights to withdraw from taking part in this study at any time. The participants were also informed that they could access the dissertation after completion.

3.7. Procedures of Data Analysis

The researcher organized all the collected data in order to prepare it for analysis. The quantitative data was the first stage to be analysed by using descriptive statistical procedures. SPSS software was used to analysis quantitative data. Moreover, the data collected from semi-structured interviews were transcribed and kept in separate files for each teacher involved in this study. All

of the data were saved to the computer and this was the first stage of organizing themes in the data that were identified in order to use thematic analysis method processes. For more details about this point (see chapter four).

3.8. Summary of the Chapter

This chapter described the methodological part of study. Research design and the data collection instruments, namely a questionnaire and semi-structured interviews are presented. The pilot study and sampling to choose the participants also described. In addition to the methodological issues such as validity and reliability that have been discussed, this chapter also dealt with data collection procedures and ethical issues. The quantitative data were analysed by employing SPSS software while the qualitative data were analysed by using thematic analysis method.

Chapter Four

Data Analysis

4.0. Introduction

This chapter provides a detailed of the results obtained from the participants, and integrates the data gathered into the whole study. Thirty-five participants answered and returned the questionnaire, and five semi-structured interviews were conducted, and then analysed to answer the research questions (see sections 1.3). Glesne and Peshkin (1992, p.127) stated that "data analysis involves organizing what you have seen, heard, and read so that you can make sense of what you have learned". The data gained from the questionnaire were analysed by SPSS software while the data obtained from the interviews were analysed by employing thematic analysis method. Using mixed method research enables the reader to see the connection between the research questions and the data (Jang et al, 2008, p. 223). Each method is appropriate for analysing certain data (Cohen et al., 2007). Both types of data (quantitative and qualitative) are presented in two phases bellow.

4.1. Quantitative data Analysis

The structured questionnaire was administered to thirty-five EFL teachers who teach speaking classes at Zawia University. It presents the findings obtained from twenty-four open-ended questions and items that were administered to the participants in the questionnaire which has been analysed by SPSS software. The answers of the questionnaire reflect the speaking assessment strategies used at the undergraduate level at Zawia University. To illustrate the results quantitatively, the answers were counted and converted into percentage to facilitate comparison. It also represented rankings means values of the statement's responses from highest to smallest to indicate the level of importance (Cohen et al., 2007). The analysis of the quantitative data is presented in two main sections; teachers` speaking assessment strategies to be considered and teachers speaking assessment strategies use. For more details, see Appendix (A). /

4.1.1. Teachers` assessment speaking strategies to be considered

The main findings of the questionnaire related to section one are presented separately for each question, by means of a graph. A brief description is provided for each.

Q1. Do you firstly identify the purpose of assessment when you assess students` speaking language performance?

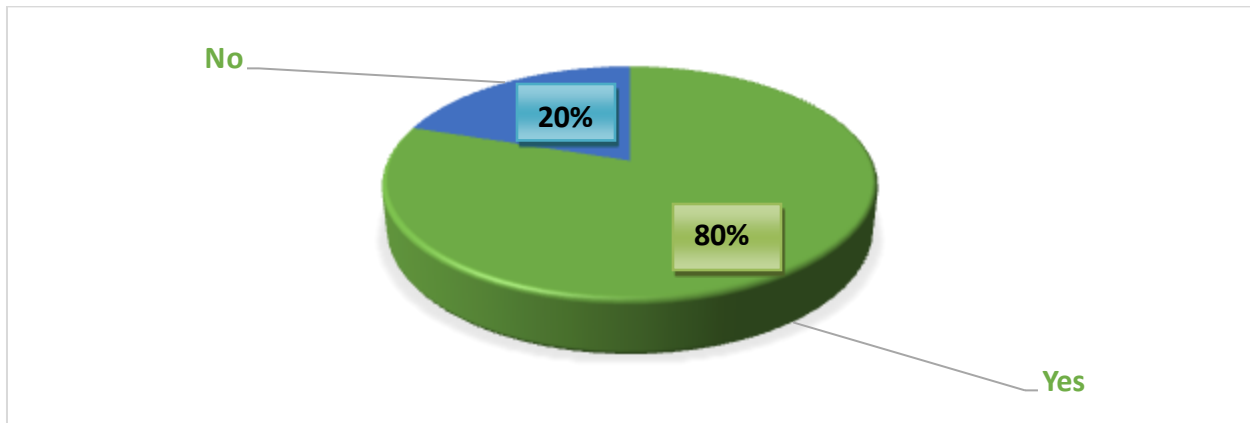


Figure (1) Identifying the purpose of assessment when teachers assess students` speaking language performance

The above figure displays that the teachers were asked if they firstly identify the purpose of assessment when they assess students` speaking language performance or not. The results obtained showed that the most of the teachers ($n = 28$, mean = 80%) answered 'yes' for this question, while ($n = 7$, mean = 20%) of them answered 'no'.

Q2. Do you use different speaking assessment tasks in a speaking test?

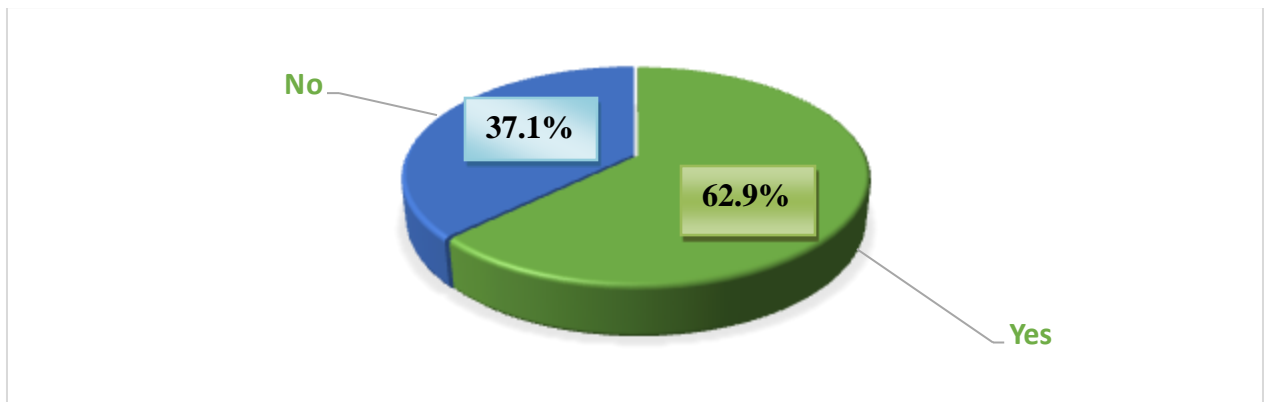


Figure (2) Using different speaking assessment tasks in a speaking test

The above figure presents that the teachers were asked in question (2) whether they use different speaking assessment tasks in speaking tests or not. However, the results showed that more than half of the teachers (n = 22, mean = 62.9%) did not use different tasks, whereas (n = 13, mean = 37.1%) use different tasks in their test to assess their students.

Q3. Do you clearly explain to the students how to accomplish speaking assessment tasks?

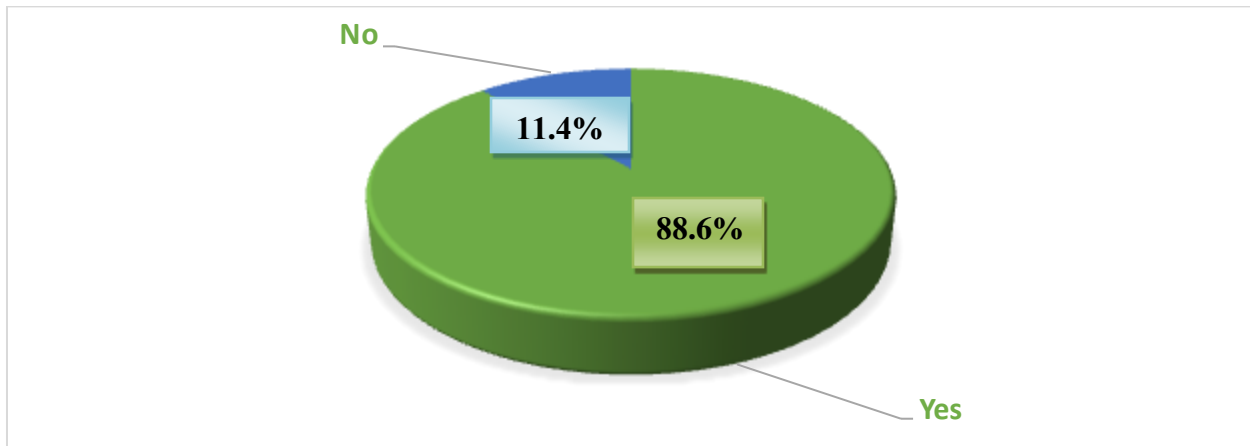


Figure (3) Explaining to the students how to accomplish speaking assessment tasks

Looking closely at figure (3) above reveals that the highest percentage of teachers were clearly explain to the students how to accomplish speaking assessment tasks (n = 31, mean = 88.6%), whereas the rest of them (n = 4, mean = 11.4%) did not do that.

Q4. Do you use speaking assessment criteria by outside resources?

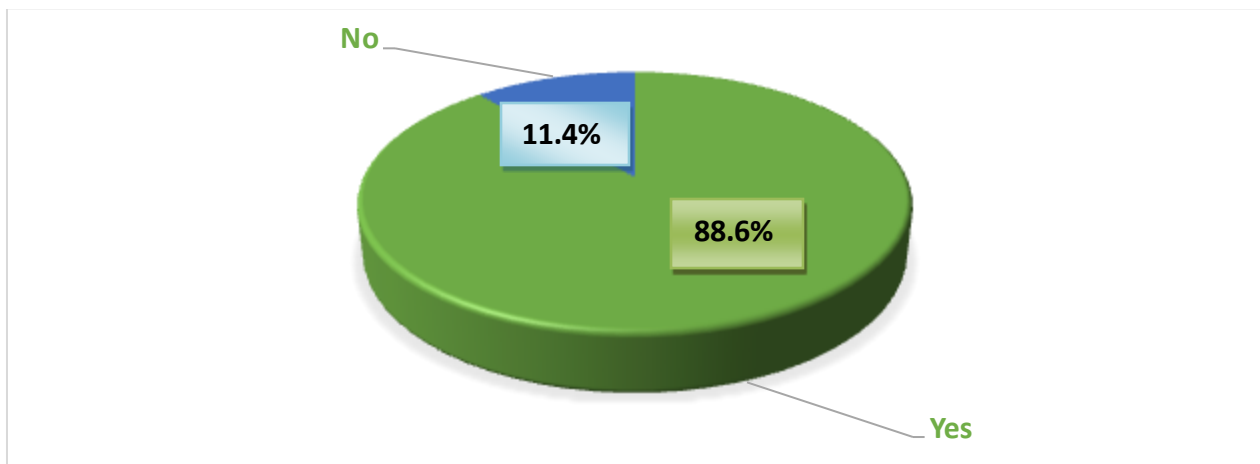


Figure (4) Using speaking assessment criteria by outside resources

The above figure illustrates the results of the question asking using speaking assessment criteria by outside resources or not: The answers were (n = 31, mean = 88.6%) said 'yes', and (n = 4, mean = 11.4%) of them said 'no'.

Q5. Do you consider accuracy as a category of speaking assessment?

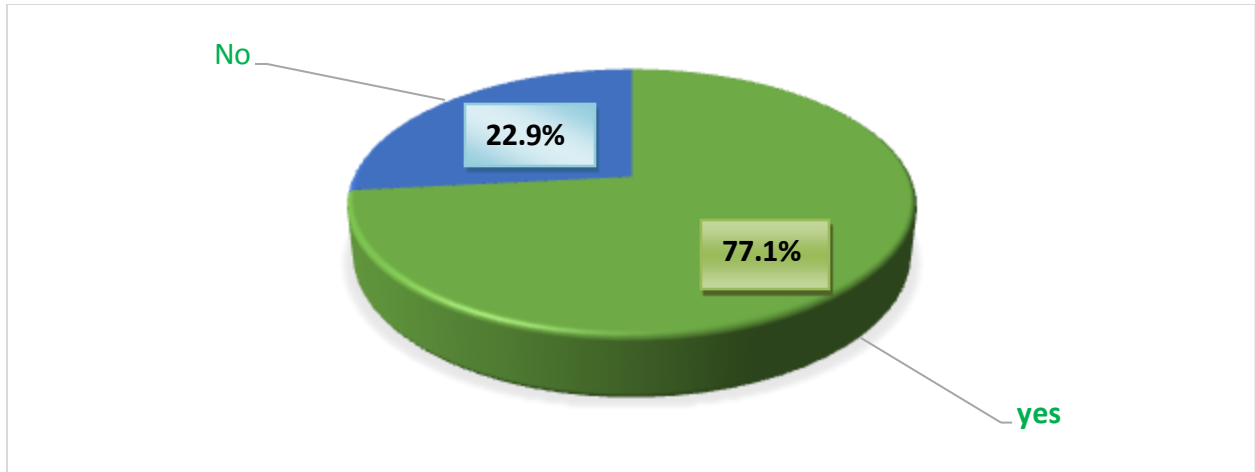


Figure (5) Considering accuracy as a category of speaking assessment

In figure (5) above the teachers were asked if they consider accuracy as a category of speaking assessment or not. Most of the teachers (n = 27, mean = 77.1%) answered 'yes' for this question, while (n = 8, mean = 22.9%) of them answered 'no'.

Q6. Do you consider fluency as a category of speaking assessment?

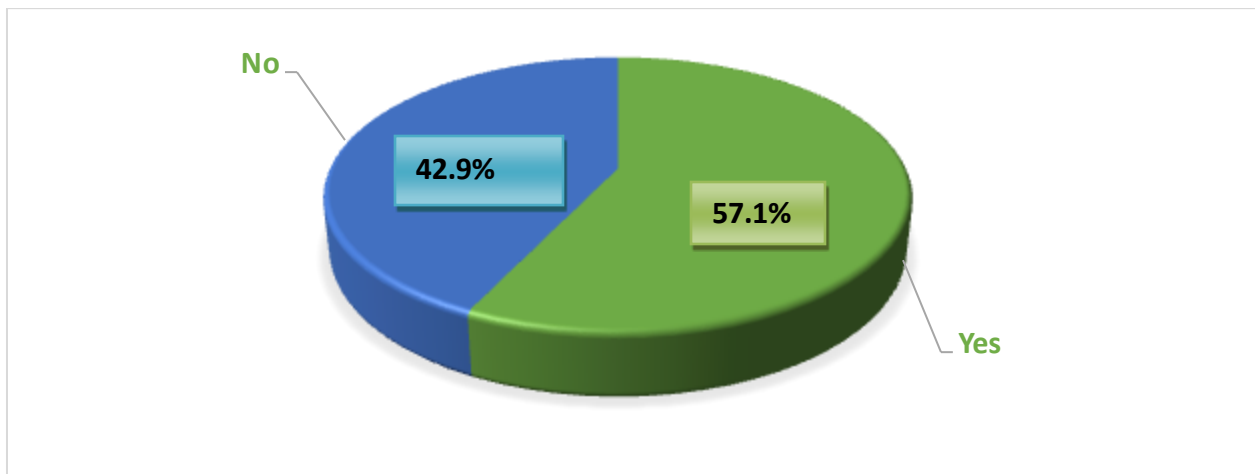


Figure (6) Considering fluency as a category of speaking assessment

In the figure above the data shows that the teachers were asked whether they consider fluency as a category of speaking assessment or not. The results revealed that (n = 20, mean = 57.1%) of the teachers consider fluency as a category of speaking assessment, whereas (n = 15, mean = 42.9%) of them they said that they did not consider it.

Q7. Do you consider communicative skills as speaking assessment?

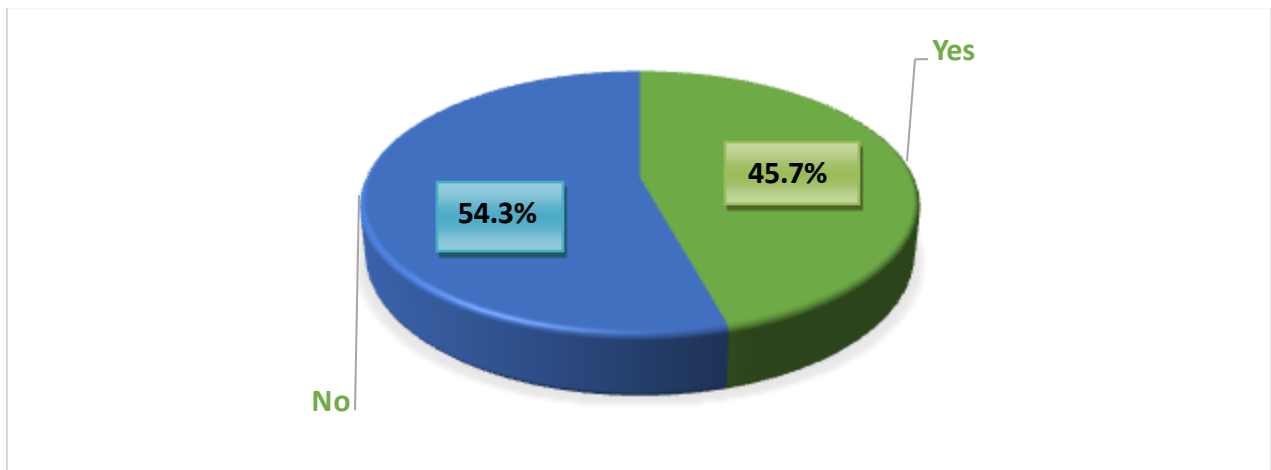


Figure (7) Considering communicative skills as speaking assessment

In figure (7), the teachers were asked whether they consider communicative skills as speaking assessment or not. However, more than half of the teachers (n = 19, mean = 54%) did not consider that, whereas (n = 16, mean = 45.7%) of the participants considered them as speaking assessment.

Q8. Do you consider range of vocabulary as a speaking assessment?

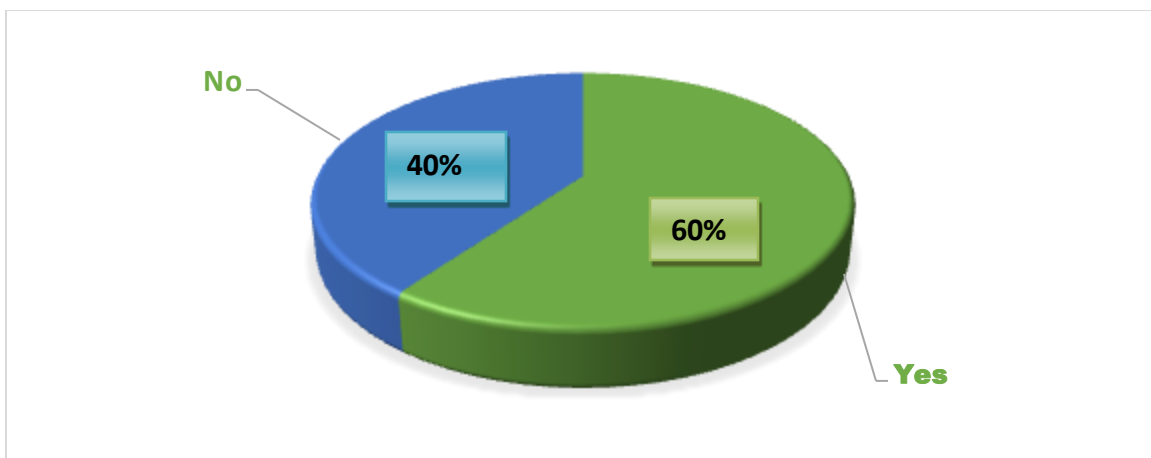


Figure (8) Considering range of vocabulary as a speaking assessment

The above figure shows that teachers were asked if they consider range of vocabulary as speaking assessment or not. More than half of the teachers (n = 21, mean = 60%) answered 'yes' for this question, while (n = 14, mean = 40%) of them answered 'no'.

Q9. Do you consider pronunciation as category of speaking assessment?

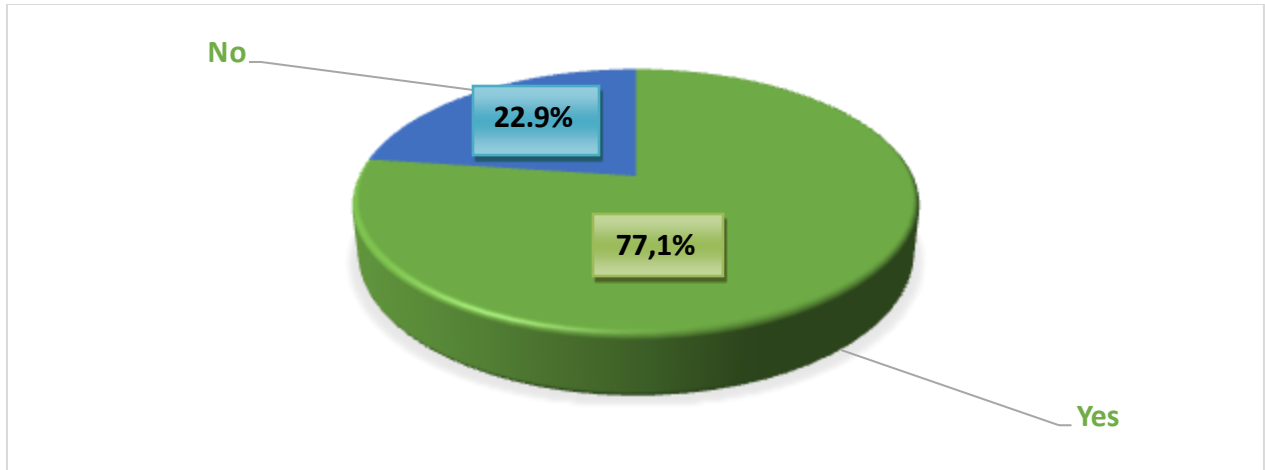


Figure (9) Considering pronunciation as category of speaking assessment

Figure (9) indicates that the majority of participants (n = 27, mean = 77.1%) who strongly agreed that they consider pronunciation as a category of speaking assessment. Besides only (n = 8, mean = 22.9%) of them said that they did not consider that.

Q10. Do you consider the students' use of transitions as a category of speaking assessment?

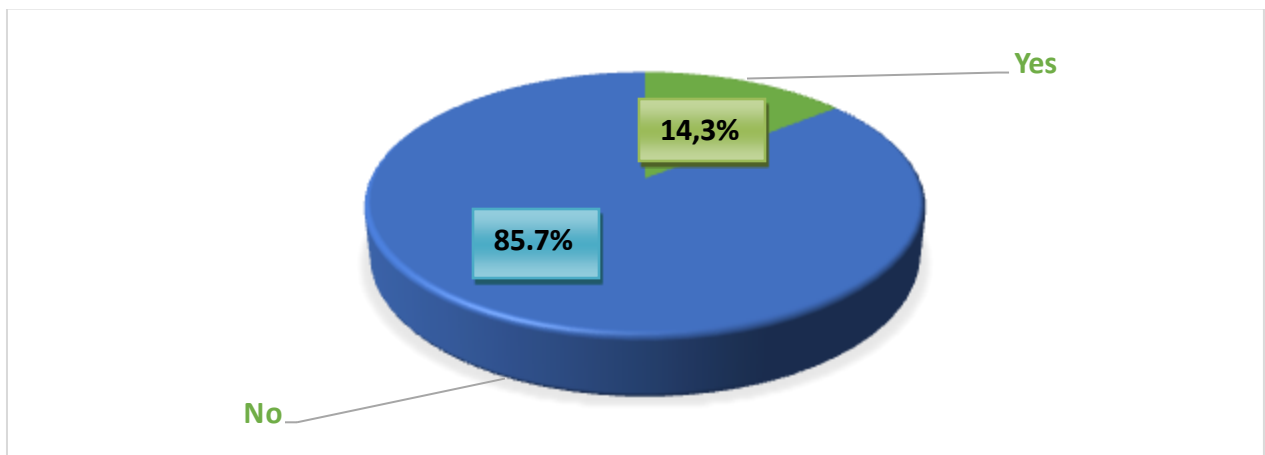


Figure (10) Considering the students' use of transitions as a category of speaking assessment

Figure (10) above clearly indicates that the highest percentage of teachers (n = 30, mean = 85.7%) consider the students` use of transitions as a category of speaking assessment. However, (n = 5, mean = 31.4%) of them did not consider that.

Q11. Do you consider the reliability and validity in speaking assessment?

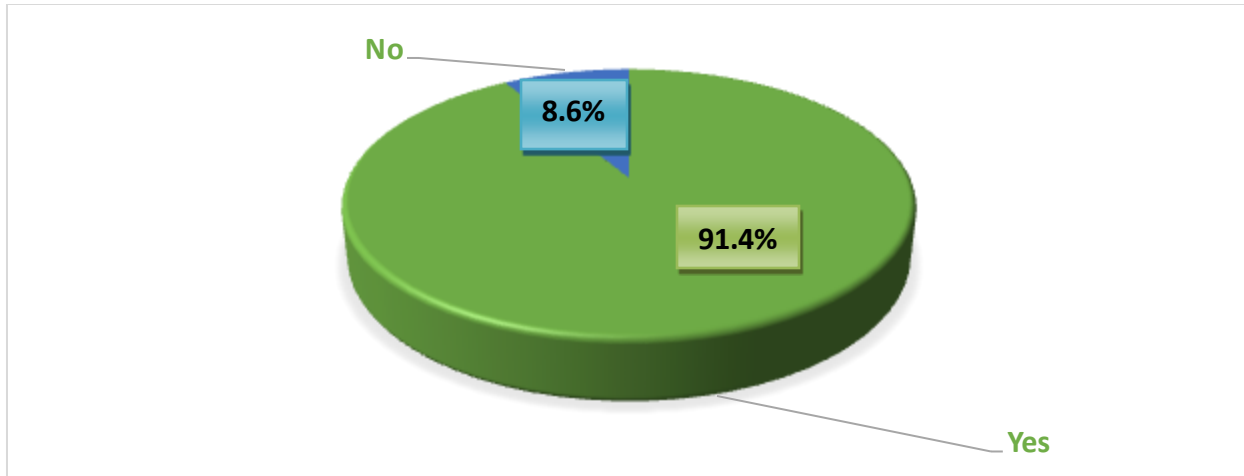


Figure (11) Considering the reliability and validity in speaking assessment

In figure (11), the findings show the majority of the participants (n = 32, mean = 91.4%) consider reliability and validity in speaking assessment, while only (n = 3, mean = 8.6%) did not consider them in speaking assessment.

4.1.2. Teachers’ Assessment Speaking Strategies Frequency Use

The main findings of the questionnaire related to section two are presented separately for each item. A brief description is also provided for each below.

1. I use summative assessment in assessing speaking ability

Table: 4.1. The use of summative assessment

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	30	85.7	85.7	85.7
	USUALLY	3	8.6	8.6	94.3
	SOMETIMES	2	5.7	5.7	100.0
	RARELY	0	0	0	
	NEVER	0	0	0	
Total		35	100.0	100.0	

Looking carefully at table (1) it shows that the majority of participants with a percentage of (85.7%) were *'always'* using summative assessment in assessing speaking ability in their speaking classes. Beside (8.6%) of them *'usually'* did that while a few numbers of participants with a percentage of (5.7%) *'sometimes'* used it. The data also shows that none of teachers select *'rarely'* and *'never'*.

2. I use formative assessment in assessing speaking ability

Table 4.2. The use formative assessment

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	6	17.1	17.1	17.1
	USUALLY	15	42.9	42.9	60.0
	SOMETIMES	12	34.3	34.3	94.3
	RARELY	2	5.7	5.7	100.0
	NEVER	0	0	0	
Total		35	100.0	100.0	

The above table illustrates that the teachers were asked in the question (16) whether they use formative assessment in assessing their students. The analysis shows that (42.9%) of the teachers were *'usually'* used it, whereas (34.3%) said that they *'sometimes'* use formative assessment in their assessment. Besides (17.1%) said they *'always'* use it, and only (5.7%) respond they *'rarely'* use that.

3. I set speaker assessment tasks at an appropriate level of difficulty

Table 4.3. To set speaker assessment tasks at an appropriate level of difficulty

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	8	22.9	22.9	22.9
	USUALLY	12	34.3	34.3	57.1
	SOMETIMES	9	25.7	25.7	82.9
	RARELY	4	11.4	11.4	94.3
	NEVER	2	5.7	5.7	100.0
	Total		35	100.0	100.0

Table (3) indicated that (34.3%) of the teachers are *'usually'* set speaker assessment tasks at an appropriate level of difficulty for their students. Besides (25.7%) of them *'sometimes'* did that

for their tasks. (22.9%) of teachers are ‘*always*’ did it in their tasks while (11.4%) said they ‘*rarely*’ did this in their tasks, and only (5.7%) of teachers ‘*never*’ did it in their tasks.

4. I use oral presentation tasks in speaking assessment

Table 4.4. The use oral presentation tasks

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	3	8.6	8.6	8.6
	USUALLY	5	14.3	14.3	22.9
	SOMETIMES	10	28.6	28.6	51.4
	RARELY	7	20.0	20.0	71.4
	NEVER	10	28.6	28.6	100.0
	Total	35	100.0	100.0	

As table (4) shows, (28.6%) of the teachers select ‘*sometimes*’ and ‘*never*’ use oral presentation tasks in speaking assessment. Besides (20%) of them said that they ‘*rarely*’ use that, while (14.3%) of them said they ‘*usually*’ do this type of tasks, and only (8.6%) of them said they ‘*always*’ use it.

5. I use learner- learner joint discussion and decision taking tasks

Table 4.5. Use learner- learner joint discussion and decision taking tasks

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	0	0	0	0
	USUALLY	2	5.7	5.7	5.7
	SOMETIMES	3	8.6	8.6	14.3
	RARELY	11	31.4	31.4	45.7
	NEVER	19	54.3	54.3	100.0
	Total	35	100.0	100.0	

The above table shows that the teachers were asked in the question (16) whether they use learner-learner joint discussion and decision tasks in assessing their students. The data gained show that none of the teachers selected the rating scale ‘*always*’. However, most of them (54.3%) selected ‘*never*’ use it, whereas (31.4%) said that they ‘*rarely*’ use it in their assessment. Besides (8.6%) said they ‘*sometimes*’ use it and only (5.7%) said they ‘*usually*’ use that.

6. I use role – play tasks in speaking assessment

Table 4.6. The use of role – play tasks

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	16	45.7	45.7	45.7
	USUALLY	10	28.6	28.6	74.3
	SOMETIMES	1	2.9	2.9	77.1
	RARELY	3	8.6	8.6	85.7
	NEVER	5	14.3	14.3	100.0
	Total	35	100.0	100.0	

According to table (6) above, the majority of the participants with a percentage of (45.7%) reported that they ‘*always*’ using role-play tasks in assessing speaking. Beside (28.6%) of them ‘*usually*’ did that while (14.3%) of the participants selected ‘*never*’ use it. However, (8.6%) of teachers ‘*rarely*’ use this type of tasks, and only (2.9%) of them ‘*sometimes*’ use that.

7. I use interview tasks in speaking assessment

Table 4.7. The use of interview tasks

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	15	42.9	42.9	42.9
	USUALLY	8	22.9	22.9	65.7
	SOMETIMES	9	25.7	25.7	91.4
	RARELY	3	8.6	8.6	100.0
	NEVER	0	0	0	
Total		35	100.0	100.0	

Table (7) indicates that (42.9%) of the teachers reported that they ‘*always*’ use interview tasks in speaking assessment. Besides (25.7%) of them said that they ‘*sometimes*’ use that, while (22.9%) of them said they ‘*usually*’ do this type of tasks, and only (8.6%) of them said they ‘*rarely*’ use it.

8. I use retelling story from written stimulus as speaking task when students are being orally assessed

Table 4.8. The use of retelling story

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	5	14.3	14.3	14.3
	USUALLY	4	11.4	11.4	25.7
	SOMETIMES	6	17.1	17.1	42.9
	RARELY	7	20.0	20.0	62.9
	NEVER	13	37.1	37.1	100.0
	Total	35	100.0	100.0	

Table (8) reveals that the majority of teachers with a percentage of (37.1%) reported that they ‘*never*’ use retelling story from written stimulus as speaking task when students are being orally assessed. Beside (20%) of them ‘*rarely*’ did that while (17.1%) of them said ‘*sometimes*’ use it. However, (14.3%) of teachers reported ‘*always*’ use this type of tasks and only (11.4%) of them selected ‘*usually*’ use that.

9.I use a holistic score for each category in speaking test

Table.4.10. The use of a holistic score

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	28	80.0	80.0	80.0
	USUALLY	7	20.0	20.0	100.0
	SOMETIMES	0	0	0	
	RARELY	0	0	0	
	NEVER	0	0	0	
	Total	35	100.0	100.0	

Looking closely at table (9) indicates that the majority of teachers with a percentage of (80%) reported that they ‘*always*’ using a holistic score for each category in speaking test. beside (20%) of the participants reported that they ‘*usually*’ did that. However, the data also shows that none of them select the rating scales ‘*sometimes*’, ‘*rarely*’ and ‘*never*’.

10. I use an analytic score for each category of speaking assessment test

Table 4.10. The use of an analytic score

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	1	2.9	2.9	2.9
	USUALLY	3	8.6	8.6	11.4
	SOMETIMES	2	5.7	5.7	17.1
	RARELY	9	25.7	25.7	42.9
	NEVER	20	57.1	57.1	100.0
	Total	35	100.0	100.0	

As table (10) shows, the participants were asked in the question (21) whether they use an analytic score for each category of speaking assessment test. The results show that more than half of the teachers with a percentage of (57.1%) said they ‘*never*’ use analytic score in their test, whereas (25.7%) said they ‘*rarely*’ use it. Moreover, (8.6%) of the teachers ‘*usually*’ did that. Besides (5.7%) of the participants ‘*sometimes*’ use this kind of score, and only (2.9%) of them ‘*always*’ use it.

11. I develop my own assessment criteria when I design speaking tasks

Table 4.11. Developing my own assessment criteria

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	8	22.9	22.9	22.9
	USUALLY	8	22.9	22.9	45.7
	SOMETIMES	11	31.4	31.4	77.1
	RARELY	3	8.6	8.6	85.7
	NEVER	5	14.3	14.3	100.0
	Total	35	100.0	100.0	

Table (11) indicates that (31.4%) of the teachers reported that they ‘*never*’ develop their own assessment criteria in their tasks when they design speaking tasks. Besides (22.9%) of them selected ‘*always*’ and ‘*usually*’ they develop their own assessment criteria in their tasks. While (14.3%) of the teachers said that they ‘*never*’ did it in their tasks while (8.6%) stated they ‘*rarely*’ did this in their tasks.

12. I connect the selection of assessment criteria to the aim of speaking assessment

Table 4.12. Connecting the selection of assessment criteria to the aim of speaking assessment

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	14	40.0	40.0	40.0
	USUALLY	15	42.9	42.9	82.9
	SOMETIMES	3	8.6	8.6	91.4
	RARELY	3	8.6	8.6	100.0
	NEVER	0	0	0	
Total		35	100.0	100.0	

According to table (12), the teachers were asked in the question (23) whether they connect the selection of assessment criteria to the aim of speaking assessment. The results show that (42.9%) of the teachers said they '*usually*' connect the selection to their aim, whereas (40%) said they '*always*' did it. Moreover, (8.6%) of teachers said that they '*sometimes*' and '*rarely*' use it in their speaking classes. The data also shows that none of the teachers stated that they '*never*' use it.

13. I encourage students to read aloud in front of class and assess them

Table 4.13. Encouraging students to read aloud in front of class

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ALWAYS	26	74.3	74.3	74.3
	USUALLY	8	22.9	22.9	97.1
	SOMETIMES	1	2.9	2.9	100.0
	RARELY	0	0	0	
	NEVER	0	0	0	
Total		35	100.0	100.0	

Table (13) indicates that the highest percentage of teachers (74.3%) reported that they '*always*' encourage their students to read aloud in front of class and assess them. Besides (22.9%) of them said that they '*usually*' did that. While (2.9%) of the teachers stated '*sometimes*' use it to assess their students in speaking classes. The data also shows that none of the teachers choose '*rarely*' and '*never*'.

4.2. Qualitative Analysis

The data gained from the semi-structured interviews were analysed qualitatively. Assessment is an activity that engages both students and teachers in judgment about the quality of student achievement or performance, and inferences about the learning that has taken place (Bryman, (2004). The current research used theoretical analysis because the research analysis is driven by research questions. This method is called thematic analysis method. It is a method of analysing reporting themes within qualitative research data (Braun and Clarke, 2006). The main goal of thematic analysis is to divide important and interesting data into themes and patterns (Maguire and Delahunt, 2017). In order to analyse the interviews data, the researcher followed the six-step framework about coding suggested by Braun and Clark (2006, p. 30). At the first stage, the researcher read the transcripts of interviews several times and created a series of analytic notes and familiarizing herself with the data. At the second step, data were organized and reduced in small meaningful chunks. Each segment that is related to research questions was coded. After that, for the third stage, these codes were compared manually using a highlighter and preliminary themes were identified. The codes were put together into broader themes. At the fourth stage, the researcher read that data of each theme and made sure that data fitted and supported its theme. At the fourth stage, the researcher reviewed each theme. The main goal of this stage is to identify the most important part of each theme of data. Moreover, at the fifth stage, the researcher created a series of analytic notes while reading the collected data such as key ideas, words, and phrases of each participant. In order to compare similarities and differences between the data of each participant, the researcher employed a coding system using various colours to mark responses that differ from each other and defining and naming each theme. Finally, at the sixth stage, the researcher produced the report of analysing the data. The researcher needs to transform her analysis into an interpretable piece of writing by using vivid and compelling extract examples that relate to the themes, research question and literature.

4.2.1. Teachers' Challenges about Assessment of Spoken English

The findings obtained showed many challenges that teachers face in the assessment of oral skills in a foreign language. These themes are: lack of how choose assessment categories to test students' oral ability, difficulty to understand assessing different types of speaking, difficulty with representative assessing tasks, difficulty to plan and structure the oral testing carefully,

ensure valid and reliable scoring, difficulty to elicit a valid sample of oral ability, choosing tasks for speaking assessment, classroom size, lack of resources and lack of training sessions. These challenges are analysed and presented below.

Lack of how choose assessment categories to test students' oral ability

The findings gained from the interviews revealed that all of the teachers ('A', 'B', 'C' and 'D') agreed that the main objective of teaching spoken language is the development of the ability to successfully interact in that language. However, teachers ('A', 'B', 'C' and 'D') face problems of how test oral students' ability in their classes. For example, teacher 'A' said "*testing speaking ability is the most challenging in terms of test preparation, administration, and scoring (grading)*". Teacher 'B' stated that "*one of the problems that I face is that designing the tasks which elicit behaviour truly represents the students' ability*". Teacher 'C' added that "*it is not easy for teachers to know the tasks that form a representative sample of the population of oral tasks that students are expected to perform*". Moreover, teachers 'D' and 'E' find difficulty to distinguish between the samples behaviour that can and will be scored validly and reliably. This indicates that these teachers do not have sufficient background knowledge about testing students' oral skills to overcome the challenges they might face when they teach speaking skills. In contrast, the data also showed that teacher 'E' was different when she argued, "*I have no problem to test students' oral abilities when I teach speaking skills*". This implies that this teacher has more experience about teaching and assessing speaking skills.

Difficulty to understand assessing different types of speaking

The qualitative findings confirmed that all the interviewees have difficulties to assess different types of speaking in some occasions. Speaking can be imitative, intensive, responsive, interactive or extensive. For instance, teachers 'A' and 'C' agreed that it is not easy to test students who are asked to imitate a word or phrase or possibly a sentence. This implies that the focus is placed on pronunciation rather than the ability to understand or convey meaning or to participate in an interactive conversation. In this respect, teacher 'B' stated that:

I face difficulty when I ask students to produce short stretches of oral language designed to show competence in a band of grammatical, phrasal, lexical, or phonological relationships such as intonation, stress, rhythm. Tasks include directed response tasks,

reading aloud, sentence and dialogue completion, limited picture cued tasks and translation up to the simple sentence level.

The extract above shows that the teacher has difficulties in assessing students' intensive tasks in speaking class. Similarly, teachers 'D' and 'E' argued that students' interactive is difficult to be assessed in some cases. This can be attributed to the length and complexity of the interaction which might be a transactional mean to exchange specific information or interpersonal to maintain social relationships. The data also showed that teacher 'B' encounters difficulty in assessing students' extensive knowledge which includes speeches, oral presentations and storytelling. The reason stands behind that might be the language style is deliberative (planned) and mainly formal.

Difficulty with representative assessing tasks

The qualitative findings revealed that the teachers have difficulties with representative assessing tasks. They face challenges in specifying the possible content of speaking tasks to be assessed. For example, teacher 'A' said, *"it is not easy to follow some international criteria such as the specified content of the Cambridge about Test of Oral interaction"*. Moreover, teachers 'B', 'C' 'D' and 'E' agreed that they find difficulties in some cases to operate tasks such as expressing, directing, describing, eliciting, narration and reporting to assess students in speaking classes. This can be interpreted as a lack of teaching experience.

Difficulty to plan and structure the oral testing carefully

The findings gained show that all the teachers interviewed ('A', 'B', 'C', 'D' and 'E') experience similar and different difficulties whenever they plan and structure oral tests. For example, teacher 'A' stated that *"it is not easy to do oral test unless you have sufficient knowledge about drawing the plan and the structure of testing speaking skill"*. Being the same track, teacher 'B' said, *"making the oral test as long as feasible is difficult because it is unlikely to obtain reliable information in less than 15 minutes"*. Teachers 'C' and 'E' admitted that they cannot use more than one format or more than one tester to encourage their students to make a second attempt if they face difficulty to express themselves in different words. More to the point, teacher 'D' mentioned that *"setting only tasks and topics that would be unexpected lead to cause students'*

difficulty in their own language". He also added that *"teachers have difficulty to find what level of English is the student at, the tester has to begin by guessing what this level is on the basis of an early response"*. This can be interpreted as inability to design and structure oral tests.

Ensure valid and reliable scoring

The data gained from the interviews revealed that one of the most significant difficulties that teachers face is how to make their scoring valid and reliable. All the teachers interviewed admitted similar problems in this regard. For example, teacher 'A' stated that

It is not easy to create appropriate scales for scoring because scoring oral ability test can be holistic or analytic or both. This can be occurred when teachers make interview procedure which requires assigning students to a level holistically and to rate them on a six-point scale for each of the following: accent, grammar, vocabulary, fluency and comprehension.

Moreover, the research findings revealed that teachers ('B' and 'D') agreed that it is not easy to design ratings then weighted and totalled to obtain the holistic described levels. Thus, teachers should have a clearer idea of how this scale works. Teacher 'C' added that *"I have problem with the following acceptable scoring procedure"*. Teacher 'E' said that:

I sometimes face problems with designing scoring techniques. However, I prefer to use an analytic scoring technique. I assess each category separately. I give each a category a mark, and then I give the biggest mark on how students could convey their ideas, messages, and her principles to the audience. After having considered that in detail, I give the overall mark.

This means that almost all of the teachers in this study need more attention to design valid and reliable scoring to assess students speaking skills in the right way. Therefore, a great care must be taken to ignore personal qualities of the students that are irrelevant to an assessment of their language ability.

Difficulty to elicit a valid sample of oral ability

The data obtained from the interviews show that only teachers 'B' and 'D' who stated that they have problems in choosing appropriate techniques to elicit a valid sample of oral ability. They stated these problems include three general format such as interview, interaction with fellow candidates and responses to audio-or video-recorded stimuli. For instance, teacher 'B' said that *"I sometimes find difficulty to run interview as a technique to assess students speaking skills. He went on saying, "a traditional interview has a serious drawback because the student feels as if he/she is talking to a superior and is unwilling to take the initiative"*. Teacher 'D' stated that *"It is not easy to use audio-or video-recorded stimuli and interpreting task in certain cases, however, simple interpreting tasks can test both production and comprehension in a controlled way"*. This means these teachers do not know how to use some techniques in an appropriate way, such as questions and requests for information, i. e. requests for elaborations, appearing not to understand, invitation to ask questions, interruption, and abrupt change of topic, or/and using pictures, role play, interpreting and reading aloud.

Choosing tasks for speaking assessment

The qualitative data show that all of the teachers agreed that choosing tasks that make students communicate in class is important; however, that is always challenging. For instance, teacher 'A' said that *"It is not easy to find tasks that include pair and group work very useful in that students are given a chance to discuss their ideas, identities, and above all to practise the language in a communicative way and then assess them"*. Teacher 'B' presented the most challenges that teachers face whenever they assign tasks for speaking assessment by saying,

I often ask myself questions like: what do I want to find out about my students' speaking skills? ... What aspects of my students' oral skills do I want to assess (... of course this is closely related to the purpose of conducting tests, which can be derived from the objectives of language learning)?... What do students already know (i.e., their current level of proficiency)? ... What do they need to learn to develop their oral English language skills? ... What task(s) will give me this sort of information? ... What sort of task is relevant to the students' level (language level, age/year level and stage of learning)? ... actually, my anticipated answers to these questions usually help me decide what task(s) I should choose to test my students' speaking skills.

Moreover, teacher 'C' said *"it is difficult to choose tasks for speaking assessment. Therefore, I choose tasks from the aspects of my students' oral skills and I choose tasks that are relevant to the students' level (language level, age/year level and stage of learning"*. This means that this teacher knows that tasks should be at an appropriate difficulty level for students. Thus, it is better to begin with a task that is too simple, and then to move on to more difficult tasks in order to know how to assess them in the right way. Teachers 'D' and 'E' agreed that choosing assessment tasks for speaking is big challenge in terms of using assessment categories. This was obvious when teacher 'E' said, *"it is difficult to find appropriate tasks contains certain categories to be focused on during the assessment operation such as the pronunciation, the use of vocabulary, grammar, confidence, communication, and showing a sense of authenticity"*. This indicates that these teachers suffer from insufficient knowledge to help them to find the appropriate speaking assessment tasks.

Classroom size

The research findings confirmed that class size as one of the most significant contextual challenges in the present study. According to the feedback obtained, it was difficult for teachers of speaking to manage all their various different activities and applying different speaking assessment strategies in classes with more than 25 students. According to the interviewees' responses, these issues need to be urgently considered. For example, teacher 'A' stated *"low or uneven participation: often caused by the tendency of some learners to dominate in the group and the large number of the students does not help in assessing them in the right way because they need much more time to assess them individually"*. Teacher 'B' said that:

Class size causes lack of target language exposure which is a main problem affecting speaking skills. Students may avoid engaging in conversation due to their incapability of social interaction in the target language. Not giving the chance to students to participate in conversation may be a reason for the difficulty in speaking assessment.

Moreover, the data collected also showed that teachers 'C', 'D' and 'E' agreed that some external challenges hinder speaking teachers to assess their students' achievements. This is because in large classes speaking assessment takes a lot of time as it tests one student at a time.

In this case, class size can be attributed to the mismatch between their aspirations and what they were actually able to do in the classroom. Therefore, the teachers might find themselves under pressure over what to do, or how to manage the speaking lessons and things that may result in negative consequences for the students' learning.

Lack of resources

The findings gained confirmed that the teachers in this study suffered from a lack of various different resources: For example, certain teachers lacked laboratories, internet, textbooks, tapes and assessment books. The absence of all these important facilities in class might negatively affect teaching and assessing students' speaking production. For example, teacher 'A' mentioned that:

Because of lack of teaching speaking materials, I sometimes run written tests in speaking classes; however, students do not show any interest in them. They feel bored since they know that they are going to memorise definitions or to do some matching. Therefore, I make great efforts while assessing students orally as a result of lack teaching materials. As such students always ask for feedback when they are orally assessed. They keen to know about their mistakes in order to improve their oral language skills.

Teachers 'B', 'D' and 'E' stated that one of the significant facilities they lack is the internet. They mentioned that there is no internet access in their collages. It is their favourite resource because there are hundreds of links and websites that provide insights into the communicative methods of language teaching and assessment. Thus, they need to frequently search links about teaching speaking and how can teachers make their students practise speaking in classroom. Teacher 'C' said that:

As a result of limited resources for teaching speaking in my collage, I do not use a variety of assessment tasks. The problem is that I know how to use tasks that assess individuals, in pairs or groups. I also find difficulty to use tasks contains role play, solving a problem, making up a dialogue about a certain situation, and describing a picture task in the assessment of pair or group work.

The above extract confirms that this teacher knows how to use and choose the speaking assessment tasks but she complained from lack of facilities in her collage. Nevertheless, this does not justify submission to the current situation; language teachers should be creative in their efforts to find different resources on which they can rely.

Lack of training sessions

The findings gained revealed that none of the teachers are involved in any training sessions related to assessing language skills in general and speaking skill in particular. For example, teacher 'A' said "*I have never had such a training session! Actually ... lack of language labs and technological aids in speaking classes make it even worse when it comes to designing speaking tests*". Teachers 'B' and 'C' stated that "*our University did not do anything related to training sessions at all. Unfortunately, up to the present time, some teachers still follow the traditional methods in teaching and assessing English language*". Moreover, teachers 'D' and 'E' mentioned that they had a training session on how to teach English but not how to assess. They involved in programme about three years ago about how to teach English language curriculum. It was a 6-week training programme. They also stated that it was really beneficial and informative, in that they provided them with a variety of communicative tasks that can be used to teach the skills in the curriculum. This indicates that the committees who are running training sessions do not consider the English language assessment in particular to help teachers for professional development.

4. 4. Summary of the Chapter

This chapter discussed the assessment strategies used in speaking classes, in addition to the challenges encountered. The results were obtained from analysing the quantitative and qualitative data obtained from the structured questionnaires and semi-structured interviews. SPSS software and thematic analysis method were used to analyse the data collected (see section 4.1). The following chapter discusses the most interesting findings in the light of the existing literature.

Chapter Five

Discussion

5.0. Introduction

This study was conducted to find out what speaking assessment strategies EFL teachers currently use at the undergraduate level at Zawia University. It was also aimed to investigate the challenges teachers encounter when they assess their students' performance in speaking classes. This chapter discusses the most interesting findings gained from the teachers' questionnaire and interviews. The main focus of interpreting the findings is to relate them to the original research questions and to the existing literature. Therefore, the findings are divided into two main sections according to the aims of the study.

5.1. Research Questions

The research questions to be answered are:

- What speaking assessment strategies do EFL teachers currently use at the undergraduate level at Zawia University?
- What challenges do EFL teachers encounter when they assess their students in speaking classes?

As it was presented in the previous chapter, the data collected were analysed and interpreted separately under frequencies entitled points. More to the point, the teachers' challenges and the reasons that might stand behind them were also discussed as justifications of why these problems occurred.

5.2. Teachers' Assessment Speaking Strategies use and their Challenges

The findings obtained from the questionnaire which concerned with teachers' background knowledge about speaking assessment revealed that the majority of the teachers stated that they firstly identify the purpose of assessment to students when they assess their speaking language performance. They also stated that they clearly explained to students how to accomplish speaking assessment tasks and add that they use speaking assessment criteria by outside resources. Moreover, most of these teachers stated that they consider accuracy, fluency,

communicative skills, considered time, reliability and validity, vocabulary pronunciation and the students' use of transitions as categories of speaking assessment. This means that these teachers have sufficient knowledge about assessing English speaking skills however, the results gained from the second section of the questionnaire concerning the use of assessment strategies in speaking classes were completely different. Most of the teachers failed to use appropriate assessment speaking strategies although they were aware of them. In this case, incongruence between what teachers knew and what actually did in their real speaking classes. This indicates that these teachers are suffering from certain challenges that prevent them to apply what they know about speaking assessment. As an example, teacher 'A' said, "*testing speaking ability is the most challenging in terms of test preparation, administration, and scoring (grading)*". Being on the same track, teacher 'B' argued "*one of the problems that I face is that designing the tasks which elicit behaviour truly represents the students' ability*". He also added "*I face difficulty when I ask students tasks include directed response tasks, reading aloud, sentence and dialogue completion, limited picture cued tasks and translation up to the simple sentence level*". These challenges negatively affect students' achievements. Moreover, there are also many other external challenges that influence teachers' impression on how well someone can speak a language and these may be reflected in assessing or scoring of learners' speaking (Taylor, 2006; Chuang, 2007). This finding disagrees with Brown (2007, p. 352) who argued that when teachers give a speaking test to the students, normally, there are going to measure the following criteria: pronunciation, fluency, vocabulary, grammar, discourse features (cohesion, sociolinguistic appropriateness, task accomplishing the objective of the task is also closely dependant on comprehension. Thus, the authority of education and the universities should support EFL teachers in order to be able to teach and assess students' speaking skills in appropriate way.

Furthermore, it is interesting to find that only few teachers who use various strategies in their test to assess their students. This indicates that these teachers have problem with speaking assessment. This conclusion falls in disagreement with Solcova (2011), who found creative tasks resemble real-life tasks asserts students to develop their fluency best, if engaged in tasks where all their concentration focuses on producing something, rather than on the language itself. It is also against Kim's (2009) research findings in which she argued that a test task could be considered as the vehicle of evaluation by which a sample can be evoked and, hence, scored and,

based on this idea-basification, research in speaking assessment has the practical functioning of a wide range of speaking tasks in terms of their appropriateness and effectiveness for bringing-out samples of language performance in different settings. Furthermore, Galaczi (2010) stated that a variety of tasks can possibly give a valid and reliable speaking assessment; however, consideration likewise should be paid to whether the tasks really inspire the language needed to assess test-takers' oral language performance. Luoma (2004, p. 36) also claimed, "the choice of tasks for the main part of the test (description, narrative, topic-based discussion, etc.,) depends on the kind of information that is needed from the scores". Therefore, teachers of speaking skills should use different tasks in their tests to assess their students' performance.

The findings also revealed that the majority of the teachers with a percentage of 85.7% reported 'always' use summative assessment in assessing speaking ability and only some of them reported 'usually' and 'sometimes' use this test. This means that not all of the teachers were aware of the importance of using summative assessment in speaking classes, however, they were supposed to report 'always' because it is necessary for evaluating and assessing students at end of any stage or semester. This finding falls in disagreement with Luoma's (2004) conclusions in which she stated that the teachers choose direct assessment, they assess what the students are actually performing since direct assessment strongly facilitates teachers' control over learners' performance. Hedge (2000) also highlighted the importance of using summative assessment in speaking classes. The main use of this type of assessment is to record or summarise the outcomes and the overall achievements of learners at a certain point in time such as the middle, end of term or end of year. This recording is usually done in a systematic way. Moreover, summative assessment usually serves other parties rather than students and teachers. These parties could be schools, employment institutions, parents and others (Jones & Bray, 1986). Therefore, all the teachers should use summative assessment in speaking classes in order to know to what extent their students' proficiency level has been improved.

Moreover, with a percentage of 42.9% the participants reported 'usually' use formative assessment in assessing their students, whereas 34.3% reported 'sometimes' use it. Besides 17.1% said they 'always' use it, and only 5.7% stated they 'rarely' use it. This result indicates that the majority of teachers were not regularly use formative assessment in speaking classes.

These teachers should use this type of tests because informal assessments are those spontaneous forms of assessment that can easily be incorporated in the day-to-day classroom activities and that measure students' performance and progress. Informal assessments are content and performance driven (Weaver, 2013). The qualitative findings justified this issue, for example teacher 'B' said that "*class size causes lack of target language exposure which is a main problem affecting speaking skills*". Being the same track, teacher 'A' argued, "... *large number of the students does not help in assessing them in the right way because they need much more time to assess them individually*". In this respect, Hedge (2000) confirmed that formative assessment by which teachers could make use of any information that has been gained about their students' progress as a means for any future procedure aimed to support students' learning. Teachers also use formative assessment as a means of judging their teaching in terms of its effectiveness. Formative assessment is important because is meant to monitor and keep track of students' progress as well as the teaching process, it is always undertaken to fulfill different purposes. This type of assessment is not carried out at a certain point of time as is the case for some other types, but it is an ongoing process which could be conducted as Sutton (1991, p. 3) stated "every few minutes". He also added "without formative assessment teachers could not function effectively". Thus, teachers should informally assess their students' performance without making them know that they are being assessed. This type of assessment can be continuous throughout the teaching and learning process.

The research findings also showed that 34.3% of the participants reported 'usually' set speaker assessment tasks at an appropriate level of difficulty for their students, whereas 25.7% of them chose 'sometimes' do that with their tasks. More to the point, 22.9% of participants reported 'always' do it in their tasks, and only 11.4% said they 'rarely' do this in their tasks and only 5.7% of teachers stated 'never' do it in their tasks. According to the frequency of using this assessment speaking strategy by teachers above, it can be concluded that more than half of these teachers did not set speaker assessment tasks at an appropriate level of difficulty for their students regularly in order to keep them in the safe side. However, there are certain teachers confirmed during the interview sessions that they consider the assessment tasks level of difficulty for their students. For example, teacher 'C' said "*it is difficult to choose tasks for speaking assessment. Therefore, I choose tasks from the aspects of my students' oral skills and I*

choose tasks that are relevant to the students' level (language level, age/year level and stage of learning". The discrepancy between the quantitative findings and the qualitative findings regarding this issue implies that not all teachers were aware that tasks should be chosen at an appropriate difficulty level for students. Interestingly, this conclusion does not go in line with Galaczi's (2010) research findings in which he argued that the task difficulty in speaking tests is considered as a significant issue in language testing. Moreover, Fulcher and Reiter (2003) reaffirmed that task difficulty in speaking tests cannot be only characterized regarding parameters like task condition, individual capacity and rather severity. All things being equal, they came to perceive that the idea of task difficulty should be connected to the cultural and social requests made by the task. Muñoz and Álvarez (2010) thought about that the utilization of different tasks considers for more reliability in the assessment process and gives students opportunities to show a greater amount of their knowledge base. Thus, teachers should use of numerous tasks to have multiple preferences; among these advantages in their potential contribution to the reliability and validity of speaking assessment.

Furthermore, the findings obtained indicated that 28.6% of participants reported 'sometimes' and 'never' use oral presentation tasks in speaking assessment. Besides 20% of them confirmed 'rarely' use that, while 14.3% of them chose 'usually' do this type of tasks, and only 8.6% of them choose 'always'. This result gives an undesirable indication about the teachers' use of oral presentation tasks in speaking assessment because it shows the majority of them do not use this strategy whenever they assess their students in speaking classes. The qualitative findings provided a number of interesting justifications and reasons of why teachers do apply certain assessment strategies in speaking classes. For example, teacher 'A' stated "*it is not easy to do oral test unless you have sufficient knowledge about drawing the plan and the structure of testing speaking skill*". By the same token, teacher 'B' argued "*making the oral test as long as feasible is difficult because it is unlikely to obtain reliable information in less than 15 minutes*". This finding falls in a mismatch with Thornbury's (2007) argument, which says that it is useful for learners who study language for academic purposes and need practice in giving presentations or conference papers. Galaczi (2010) added that a variety of tasks can possibly give a valid and reliable speaking assessment; however, consideration likewise should be paid to whether the tasks really inspire the language needed to assess test-takers' oral language performance. In

contrast, this finding is in line with Kerr Troth and Pickering (2003) who argued that some students may be unwilling to participate and others may feel apprehensive. Sano (1989) also stated that the students tend to feel nervous when making oral presentation in front of an audience. However, there are different arguments about using oral presentation tasks in speaking assessment but teachers should consider the suitability of the task assumes a significant part in achieving validity in speaking assessment.

It is interesting to discuss whether the teachers use learner-learner joint discussion and decision tasks in assessing their students. The findings gained showed that most of the teachers with a percentage of 54.3% reported 'never' use it; whereas 31.4% said they 'rarely' use it in their assessment. Besides 8.6% said they 'sometimes' use it and only 5.7% chose 'usually' use that and none of them reported 'always'. This result shows the highest percentages of teachers' responses were 'never' and 'rarely' which means that these teachers do not use discussion and decision tasks in assessing their students. The reason behind this finding was detected during the interview sessions when the interviewees admitted that they usually encounter significant difficulties regarding testing students' oral skills and designing appropriate assessments tasks for their students. For example, teacher 'A' said "*testing speaking ability is the most challenging in terms of test preparation, administration, and scoring (grading)*". Teacher 'B' stated that "*one of the problems that I face is that designing the tasks which elicit behaviour truly represents the students' ability*". Teacher 'C' added "*it is not easy for teachers to know the tasks that form a representative sample of the population of oral tasks that students are expected to perform*". This finding does go in harmony with Harmer's (2001: 272) suggestion that highlights a solution for teachers to easily assess their students' oral skills. Harmer recommends teachers to avoid such difficulties which are the buzz group: where students have a chance for quick discussion in small groups before any of them is asked to speak in public. Dakowska (2005) argued that it should be stressed that the actual potential of this activity for developing an individual learner's speaking ability can be implemented only under certain conditions. Therefore, it is important for teachers when they assess their students' oral skills to make the arguments to be more logical and well-balanced and that they are presented in accordance with the social and cultural norms of the target language community, which is based mainly on a particular chosen topic to be argued and discussed by students.

Analysing the quantitative data showed that nearly half of participants with a percentage of 45.7% reported 'always' use role-play tasks in assessing speaking, while 28.6% of them chose 'usually' use it. However, only 14.3% of them said 'never' use it. It was also noticed that 8.6% of the participants reported 'rarely' use this type of tasks and only 2.9% of them stated 'sometimes' use it in their speaking classes. It can be concluded therefore that most of the teachers reported 'always' and 'usually' use role-play tasks when they assess their students in speaking classes. This means that most of these teachers have background knowledge about this assessment strategy. The teachers' abilities of assessing students in speaking classes are also confirmed by certain teachers in during the interview's sessions. For instance, teacher 'E' said *"I have no problem to test students' oral abilities when I teach speaking skills"*. The literature also clarified the importance of using role-play tasks. Harmer, (2001) and Thornbury, (2005) argued that teachers of speaking skills may assess students through role play activities where students are asked to pretend to be in various social contexts and various social roles. Creative tasks resemble real-life tasks asserts students to develop their fluency best, if engaged in tasks where all their concentration focuses on producing something, rather than on the language itself (Solcova, 2011). Moreover, this finding is also in line with Edge (1992) who stated that role play is always be used as a simple practice activity in information exchange when there is a limited investment that students want to make in it. Furthermore, Muhd Syahir (2016) and Liu (2010) supported using the role-play as a speaking assessment to help students be more confident as it arouses their motivation in speaking English than using other oral English tests. Thus, using role play would help teachers get to know their students better, provide more innovative or authentic feedback and support, and better preparation for role play and other active learning strategies (Watkins, 2011).

According to the findings gained, 42.9% of the participants reported 'always' use interview tasks in speaking assessment. Beside 25.7% of them said they 'sometimes' 'use that, while 22.9% of them said they 'usually' do this type of tasks. This means that these teachers possess sufficient background knowledge about the assessment strategies. This finding agrees with Thornbury's (2005) argument in which he encourages teachers to use this particular assessment strategy. Thornbury confirmed that an interview is an assessment method that classroom teachers of

speaking or any other assessing agent can use for gathering some information about their students. Teachers can conduct interviews to obtain a better idea about students such as how they think or their level of understanding, their ability to communicate or how to apply any concepts. However, Dakowska (2005) argued that interviews are difficult tasks in the sense that they require from the interviewer some preparation in the form of research, the selection of relevant questions and the prior analysis of native speaker interviews so as to properly evaluate both the questions and the elicited information. It is hard to assess a number of students within a short time because the teacher is put under pressure. Thornbury (2007) added that scoring is also difficult in testing interviews such as communicative as the assessment includes pronunciation, grammar, fluency, content, organization, and vocabulary. These are perhaps the reasons for which EFL teachers remain less interested to assess students' speaking skills. Therefore, teachers should consider "the choice of tasks for the main part of the test (description, narrative, topic-based discussion, etc...) depends on the kind of information that is needed from the scores" (Luoma, 2004, p. 36). This implies that the suitability of the task assumes a significant part in achieving validity in speaking assessment.

Moreover, the quantitative findings revealed that the majority of the teachers with a percentage of 37.1% chose 'never' use retelling story or text from aural stimuli as speaking task in speaking assessment and 20% of them *reported* 'rarely' do that. This indicates that these teachers have limited knowledge about assessing speaking strategies. It can be also attributed to the challenges that teachers usually face. For example, teacher 'D' stated that "*It is not easy to use audio-or video-recorded stimuli and interpreting task in certain cases, however, simple interpreting tasks can test both production and comprehension in a controlled way*". Teacher 'A' confirmed that "*it is not easy to do oral test unless you have sufficient knowledge about drawing the plan and the structure of testing speaking skill*". Teacher 'B' also said, "*making the oral test as long as feasible is difficult because it is unlikely to obtain reliable information in less than 15 minutes*". Teacher B also admitted that he has difficulty to assess students' extensive knowledge which includes speeches, oral presentations and story-telling. The reason might be the language style is deliberative (planned) and mainly formal. Teachers 'C' and 'E' argued that they sometimes cannot use more than one format or have more than one tester and to encourage students to make a second attempt if they face difficulties to express themselves in different words. These findings

are not in line with Thornbury's (2007) argument in recommending EFL teachers to utilise story telling as one of the assessment strategies. This strategy is common ingredient of casual conversations and drama, role-play and simulation activities. These elements increase the range of registers and social roles that learners may encounter in speaking classroom. Thornbury added teachers must be remembered that each speaking task needs to be productive, purposeful, interactive, challenging, safe and authentic. The difficulty of speaking tasks, however, might be not the same for students depending to some extent on their personalities.

In addition, it is interesting to find the majority of teachers with a percentage of 80% reported 'always' using a holistic score for each category in speaking test and the rest of them responded with 'usually' do that in their classes. This implies that these teachers understand and know what to do with this strategy in order to assess students in speaking classes. Although the teachers apply this strategy but they reported that they have similar problems in this regard. For example, teacher 'A' stated that *"It is not easy to create appropriate scales for scoring because scoring oral ability test can be holistic or analytic or both. This can be occurred when teachers make interview procedure which requires assigning students to a level holistically"*. This finding agrees with Chuang's (2009) conclusions in which he confirmed that holistic rating captures an overall impression of the speaker's performance. Being the same track, Hughes (2003) wrote holistic rating scales have the main advantage of being very rapid. More to the same point, Luoma (2004, p. 36) argued, "the choice of tasks for the main part of the test (description, narrative, topic-based discussion, etc.,) depends on the kind of information that is needed from the scores". This discussion leads to the conclusion that the suitability of the task assumes a significant part in achieving validity in speaking assessment. In contrast, holistic rating scales may lead to a disadvantage of lacking reliability in the scores obtained from such a type of scale because it does not guide and help students to progress in their speaking ability as it gives no details about their weaknesses in speaking.

Analysing the quantitative data showed that the majority of the teachers 57.1% responded with never use analytic score in their test, while 25.7% said they rarely use it. It also reported that only a few of them reported 'always', 'usually' and 'sometimes' use it. Unexpectedly, during the interview's sessions, these teachers confirmed that they use this strategy in speaking classes. For

example, Teacher 'E' said *"I sometimes face problems with designing scoring techniques. However, I prefer to use an analytic scoring technique"*. She added that *"I assess each category separately. I give each a category a mark, and then I give the biggest mark on how students could convey their ideas, messages, and her principles to the audience. After having considered that in detail, I give the overall mark"*. This implies that these teachers face significant challenges that hinder them to use this strategy whenever they assess their students in speaking classes. Hughes, (2003) argued that the analytic rating requires the assessor to pay attention at individual criteria separately such as, pronunciation, fluency and accuracy and gives each of them its own score, then create a final score. In other words, the score in this approach is the sum of several aspects of the oral performance rather than a single score. Therefore, employing analytic rubrics helps students to enhance their oral performance by recognize their area of weaknesses and which aspect of speaking they need to work on for further improvement (ibid). In contrast, this finding agrees with Chuang (2009) and Hughes, (2003) who said that analytic test takes more time and energy for teachers to apply. With regard to analytic scales, the time factor is considered a main disadvantage in that they take longer practice of scoring, yet they provide more reliable scoring. Another disadvantage of the analytic scale is that raters might find it difficult and confusing to concentrate on different aspects in a sample of language to be assessed and to evaluate them simultaneously (Hughes, 2003). Thus, language teachers need to examine *"the suitability of the given rating scale to the quality of language samples actually elicited in each testing situation"* (Nakatasuhara, 2007, p. 84).

Furthermore, the questionnaire findings revealed that most of the teachers with a percentage of 31.4% reported 'never' develop their own assessment criteria when they design speaking tasks. This means that these teachers lack of the teaching experience that help them to develop and create different assessment criteria in their tasks. This can be attributed to the challenges that these teachers face when they assess their students in speaking classes. For example, teacher 'B' stated that *"I face difficulty when I ask students to produce short stretches of oral language designed to show competence in a band of grammatical, phrasal, lexical, or phonological relationships such as intonation, stress, rhythm"*. It can be concluded that this finding disagrees with Weir's (1993) argument in which he said that it is important to highlight the features that characterise speaking as interaction in order to gain insights into the elements to be tested within

an interactional framework in a certain speaking test context. Weir designed a framework of spoken language interaction that helps test developers to construct features of spoken interaction in the design of their tests: A three-part framework is proposed, covering: the operations (activities/skills) that are involved in spoken interaction such as informational routines, e.g. telling a story, and improvisational skills that might be called into play when the performance of these routines breaks down, e.g. requesting clarification...; the conditions under which the tasks are performed, e.g. time constraints, the number of people involved and their familiarity with each other(...); and the quality of output, the expected levels of performance in terms of various relevant criteria, e.g. accuracy, fluency or intelligibility (ibid, 1993, p. 30). Therefore, teachers of speaking skills should develop their own assessment criteria when they design speaking tasks.

The data gained from the teachers' questionnaires also revealed that most of the teachers, with a percentage of 42.9% responded with 'usually' connect the selection of assessment criteria to the aim of speaking assessment, whereas 40% reported 'always' do it. This means that these teachers understood how to draw their tests design appropriately; however, they stated during the interviews sessions that they face different challenges in this regard. For instance, teacher 'A' said that *"It is not easy to find tasks that include pair and group work very useful in that students are given a chance to discuss their ideas, identities, and above all to practise the language in a communicative way and then assess them"*. This is also confirmed by teacher 'C' who said *"it is difficult to choose tasks for speaking assessment. Therefore, I choose tasks from the aspects of my students' oral skills and I choose tasks that are relevant to the students' level (language level, age/year level and stage of learning"*. Teachers 'D' and 'E' added that choosing assessment tasks for speaking is a big challenge in terms of using assessment categories. Rowntree (1987, p. 162) argued that any assessment task to be used should "go with the conduct and style of the teaching and learning experienced by our students". Newton and Nation, (2009) added that using cooperative activities can encourage negotiation of language item. Therefore, the Universities should support teachers to be ready to select appropriate assessment criteria which lead to speaking assessment.

The questionnaire data showed that almost all of the teachers with a percentage of 74.3% reported 'always' encourage their students to read aloud in front of class and assess them, beside

22.9% of them chose 'usually' do that. This indicates that these teachers aimed to signify students' errors in order to be avoided by the rest of the students. However, the data gained from the interviews confirmed that these teachers face challenges that obstacle them to assess students in appropriate ways. For example, teachers 'A' and 'C' agreed that it is not easy to test students who are asked to imitate a word or phrase or possibly a sentence. This indicates that the focus is on pronunciation and not on the ability to understand or convey meaning or to participate in an interactive conversation. Teacher 'B' added that "*I face difficulty when I ask students to produce short stretches of oral language designed to show competence in a band of grammatical, phrasal, lexical, or phonological relationships such as intonation, stress, rhythm*". This result is in line with Dakowska (2005) who argued that it should be stressed that the actual potential of this activity for developing an individual learner's speaking ability can be implemented only under certain conditions. In contrast, this finding is not corresponded with Luoma (2004) who found that the sound of speech is a 'thorny issue' for language assessment since pronunciation is usually judged depending on a native speaker standard of pronunciation which in turn may result in inaccurate assessment on the part of the examiner. It is difficult because it reflects the various perceptions of fluency as reflections of speech planning and thinking processes and a hesitation phenomenon (Freed, 1995). Thus, it is significant for teachers to encourage their students to read aloud in front of class and assess them in order to be more logical and well-balanced and that they are presented in accordance with the social and cultural norms of the target language community.

Furthermore, one of the interesting issues which was raised by the teachers during their interview sessions was classroom size. The data obtained showed that it is difficult for teachers of speaking to manage all their various different activities and apply different speaking assessment strategies in classes with more than 25 students. For example, teacher 'A' stated "*low or uneven participation: often caused by the tendency of some learners to dominate in the group and the large number of the students does not help in assessing them in the right way because they need much more time to assess them individually*". In this regard, teachers 'C', 'D' and 'E' added that some external challenges hinder speaking teachers to assess their students' achievements. This is because in large classes speaking assessment take a lot of time, especially students should be tested individually. This finding is in line with Achilles, 1999 and Bennett, 1996) who believe

that bigger classes will decrease the amount of time because teachers usually spend more time on general instruction and on dealing with students individually. It can be argued that although the teachers might suffer from their limited experience but the effect of class size in Libyan Universities is notably exist. This conclusion also agrees with Angrist and Levy's (1999) findings in which they say that the significant effects of class-size in their schools are much larger when the average being 32 students. This means that classroom size assumed has a great significance effect on teachers' performance and students' achievements. Therefore, classroom size should be considered by universities in order to manage the teachers' speaking lessons.

The qualitative findings confirmed that all the interviewees suffered from a lack of various different resources such as, lack of laboratories, internet, textbooks, tapes and assessment books. All these are important facilities that might negatively influence teaching and assessing students' speaking skills. For example, teacher 'A' mentioned that "*Because of lack of teaching speaking materials, I sometimes run written tests in speaking classes; however, students do not show any interest in them. They feel bored since they know that they are going to memorise definitions or to do some matching*". This result disagrees Stempleski's (1990) research findings in which he argued that facilities such as video is often more preferred by language students, as it is more attractive and motivating than audio recordings. Video enables teachers to display a variety of challenges, such as: using pictures only, sound only, or pictures and sound together (Milliar, 2003). Moreover, the internet is considered as a very important and rich source for authentic materials (Arianie, 2017). It is likewise important to note that a large amount of the materials mentioned are now easily available online. In this situation, students do not watch a film in a class represents, actually, a rich source of language input with verbal and visual stimulation that can develop speaking abilities. Gilmore, (2007) added that films and TV shows represent a rich source of input for learners which can be exploited in various ways and on different levels to improve their communicative competence. All of these renounces, materials and facilities are required for both teachers and students because they aim at generating new ideas and engaging in real communication. Therefore, classrooms in universities should be equipped with varieties of materials and resources that make the teaching and learning process more interesting and more memorable.

In addition, the qualitative findings revealed that that none of the teachers are involved in any training sessions related to assessing language skills in general and speaking skill in particular. For example, teacher 'A' said "*I have never had such a training session! Actually ... lack of language labs and technological aids in speaking classes make it even worse when it comes to designing speaking tests*". Teachers 'B' and 'C' stated that our university did not do anything training sessions related to assessment at all. This means these teachers suffer lack of teaching and assessing speaking skills. This finding agrees with Allen's (2008) argument in which he pointed out that "The majority of teachers have an insufficient command and training of English to be able to teach it effectively and teachers with insufficient subject knowledge have very little if any confidence" (p. 2). In contrast, this finding disagrees with Karavas-Doukas's (1998) findings which say that innovation can lead to positive changes in curricula and the beliefs and behaviour of teachers. In this respect, Shamim (1996) added that it is important for teacher trainers to encourage participants in teacher training programmes to discuss both overt and 'hidden' barriers to the successful implementation of change in their own teaching/learning contexts. Thus, the authority of education and the universities should run training sessions because this kind of knowledge is essential for teacher development but these should be focused and organized so as to produce positive outcomes.

5.5. Summary of the Chapter

This chapter discussed the findings obtained from the teachers' questionnaire and interviews. It focused on what speaking assessment strategies EFL teachers do currently use at the undergraduate level in Zawia University and the challenges they encounter whenever they assess their students in speaking classes. The findings obtained are interpreted and discussed in order to make the picture very clear in each discussed issue. A brief summary of the most interesting conclusions is provided in the next chapter.

Chapter Six

Conclusion

6.0. Introduction

This chapter presents the whole summary of the current study. It includes some pedagogical implications of the research. In this chapter, the findings are viewed in the light of the implications, suggestions for further research and research limitations. A summary of the chapter is also provided.

6.1. Conclusion of the Whole Study

This study provides a deeper understanding of the obstacles that EFL Libyan University teachers encounter while assessing speaking skills. In this research, both quantitative and qualitative data were used, i. e. a questionnaire and semi-structured interviews were used as research tools. The quantitative data were analysed by using SPSS software, while the qualitative data were transcribed and analysed by using the thematic analysis method. The quantitative findings showed that the teachers used effective strategies weakly and in a very limited way. A significant finding obtained is that the teachers believe in the importance of using speaking assessment strategies in their classes, but surprisingly they do not apply what they know about teaching and assessing students in English speaking classes. Moreover, it was clear from the data analysis and discussion (see chapters four and five) that the teachers face several obstacles in assessing their students in speaking class. The study presents a number of significant conclusions which are summarised as follows.

- ❖ The findings of this study revealed that most of the teachers have a good understanding of how to assess English speaking skills. These teachers considered accuracy, fluency, communicative skills, considered time, reliability and validity, vocabulary pronunciation and the students` use of transitions as categories of speaking assessment. This indicates that these teachers have adequate knowledge about assessing English speaking skills; however, the results obtained from the second section in the questionnaires concerned with the use of the assessment strategies in speaking classes were different. It showed that the most of the teachers did not use the appropriate assessment speaking strategies

although they knew them. This indicates the existence of obstacles that prevent these teachers from applying what they know about speaking assessment. Some of these challenges were the test preparation, administration, scoring (grading), designing and the tasks which elicit behaviour truly represents the students' ability.

- ❖ The study showed that only few of teachers who employ different tasks in their tests to assess their students. This implies that these teachers did creative tasks which resemble real-life tasks asserts students to develop their fluency best, if engaged in tasks where all their concentration focuses on producing something, rather than on the language itself (Solcova, 2011).
- ❖ The findings obtained confirmed that almost all of the teachers reported 'always' use summative assessment in assessing their students' speaking ability in class. This means that all of the teachers were aware of the importance of using summative assessment in speaking classes.
- ❖ The quantitative findings gained also revealed that the majority of the teachers do not use formative assessment in their speaking classes on a regular basis. These teachers attributed that to the big class size. It is difficult to assess a big number of students appropriately because it takes considerably more time to assess them individually. This case is strange because informal assessments are those spontaneous forms of assessment that can easily be incorporated in the day-to-day classroom activities and that measure the students' performance and progress. Informal assessments are content and performance driven (Weaver, 2013).
- ❖ The research findings indicated a mismatch between the quantitative and the qualitative findings. Being more explicit and concise, more than half of the research participants did not set speaker assessment tasks for their students at an appropriate degree of difficulty on a frequent basis in order to keep them safe although they stated they know how to deal with this issue. This means not all teachers knew that tasks should be chosen at an appropriate difficulty level for students.
- ❖ The findings gained showed that the majority of the research participants did not make use of oral presentation task in speaking assessment in their classes. The reason was reported by certain teachers who stated that they did not have adequate knowledge of

using this strategy of assessment, and it is difficult to obtain reliable information in less than 15 minutes.

- ❖ This study also revealed that most of the teachers did not utilise discussion in assessing their students. This was sensed from teachers' contributions when they said that it is difficult to know the tasks that form a representative sample of the population of oral tasks that students are expected to perform. This can be interpreted as a difficulty in deciding how to assess students' performance in class.
- ❖ Moreover, the quantitative findings showed that most of the teachers reported 'always' and 'usually' use role-play tasks when they assess their students in speaking classes. This means that most of these teachers possessed good background knowledge about this assessment strategy.
- ❖ The results of the study revealed that most of the teachers did not utilise retelling story or text from aural stimuli as speaking task in speaking assessment. These teachers justified that they cannot use more than one format or more than one tester to make a second attempt if they face a difficulty to express themselves in different words.
- ❖ According to the research findings, almost all the participants reported using a holistic score for each category in speaking test in their classes. This means that these teachers understand and know what to do with this strategy in order to assess students in speaking classes. Holistic rating captures an overall impression of the speaker's performance (Chuang, 2009).
- ❖ The findings obtained confirmed that most of the teachers do not use analytic scores in their tests. This indicates that these teachers encounter challenges that prevent them from using this method regularly while evaluating their students in speaking classes. These teachers stated during their interview sessions that they usually face problems with designing scoring techniques.
- ❖ Furthermore, the qualitative findings showed that the classroom size was one of the serious challenges that teachers encounter. They stated that it is difficult for teachers of speaking to manage all their various different activities and apply different speaking assessment strategies in classes with more than twenty-five students. This is because in large classes speaking assessment take a lot of time, especially as it tests one student at a time.

- ❖ Additionally, this study confirmed the scarcity resources such as laboratories, internet, textbooks, tapes and assessment books. All these are important facilities and might negatively affect teaching and assessing students' speaking skills.
- ❖ Finally, the findings gained from the interviews revealed that that none of the teachers are involved in any training sessions related to assessing language skills, however, the innovation can lead to positive changes in curricula as well as in the beliefs and behaviour of teachers (Karavas-Doukas, 1998).

6.2. Implications and Recommendations of the Study

This study has several implications for EFL teachers in speaking classes.

- ❖ To measure their students' speaking abilities, teachers of speaking skills should employ a variety of tasks in their assessments.
- ❖ Explain the criteria on which the assessment is based on to students in order to facilitate and increase their awareness of the speaking activities.
- ❖ Teachers should use summative assessment in speaking classes since it is important for evaluating and assessing students at the end of any level or semester.
- ❖ Teachers should assess their students informally so that they do not feel judged. This form of assessment may be performed on a regular basis throughout the teaching and learning process.
- ❖ Teachers should use various tasks to have multiple preferences; among these advantages is their potential contribution to the reliability and validity of speaking assessment.
- ❖ Teachers of speaking should assess their students' speaking abilities through discussion in order to make the arguments more logical and balanced, and to ensure that they are presented in accordance with the social and cultural norms of the target language community, which is primarily based on a specific chosen topic to be argued and discussed by students.
- ❖ Using analytic assessment rubrics help students to improve their oral performance by identifying their areas of weakness and which aspects of speaking they need more concentration.
- ❖ Teachers should consider the task's appropriateness since it plays a vital role in obtaining validity in speaking assessment.

- ❖ It is essential for teachers to encourage their students to read aloud in front of the class and assess them so that their assessment would be more logical and balanced.
- ❖ When designing speaking assignments, teachers of speaking skills should create their own assessment criteria.
- ❖ Education authorities and institutions should provide EFL teachers different teaching materials in order to be able to teach and assess students' speaking skills in appropriate ways.
- ❖ It is important for teachers to remember that each speaking task must be productive, meaningful, participatory, challenging, safe, and real. The difficulty of speaking tasks, on the other hand, varied from one student to another and might be influenced by individual students' personality.
- ❖ Another important implication is that classroom size should be considered by universities in order to manage the teachers' speaking lessons. Moreover, the classrooms should be equipped with the necessary materials and resources.
- ❖ Libyan universities should run regular in-service training sessions about teaching and assessing speaking skills because this kind of knowledge is essential for EFL teacher development.
- ❖ To overcome the teachers' challenges and difficulties in terms of teaching and assessing speaking skills, the authority of education in Libya and EFL researchers, assessment professionals, curriculum developers, teacher educators, professional learning providers, and teachers should work collaboratively to strengthen teachers' abilities of teaching speaking and improve their assessment strategies in their classrooms.

6.4. Limitations of the Study

There were multiple limitations throughout this study that might have accounted to the lack of significant findings. The study only employed a questionnaire and semi-structured interviews. The researcher planned to do classroom observations during speaking lectures, but the Corona crisis and university closures hindered the researcher from doing so. Furthermore, the sample size was limited in this study however the researcher distributed the questionnaire to 50 participants at the beginning but only 35 copies were collected and only 5 interviews were conducted with teachers. Moreover, the conducted study is limited in terms of time, place and

population. This study was restricted only to Faculty of Art in Zawia University, Training Collage in Zawia, Faculty of Education in Abu-Issa and Faculty of languages.

6.3. Suggestions for Further Research

According to the research findings, an important trend in language assessment research is a better knowledge of speaking assessment strategies and the challenges that teachers encounter during teaching speaking. This study provided a clear understanding about how speaking assessment strategies take a big part in speaking assessment. Further studies can be carried out to investigate how EFL teachers construct their knowledge of speaking assessment and how the institutional contexts contribute to the improvement of teachers' knowledge of language assessment in the classroom. In addition, more studies can be conducted on teachers' role in the institutional context with regard to the design of large-scale speaking tests and its influence on their speaking assessment practices, and the relationship between the teachers and students can affect their understanding in the speaking classes. These issues can be further investigated to explore teachers' construction of knowledge of speaking assessment through their interaction with these contexts.

6.5. Summary of the Chapter

This chapter included a summary of the main conclusions regarding the assessment strategies teachers' use in their speaking classes in addition to the challenges that might be encountered. The implications, recommendations, suggestions for further research and limitation of the research were also acknowledged in this chapter.

References

- Abd El-Razek, S. (2010). *The Effectiveness of a Program Based on Reciprocal Teaching in Developing Instructional Speaking Skills for English Department Students in Faculties of Education*. thesis (M. Sc.). Faculty of Education. Port Said University.
- Achilles, C.A. (1999) *Let's Put Kids First, finally: Getting Class Size Right*. Thousand Oaks, California: Corwin Press; Cambridge: Cambridge University Press.
- Adeyemi, D. (2008). Bilingual education: meeting the challenges of diversity in Botswana, *Nordic Journal of African Studies*, 17, 1, pp. 20–33.
- Agasøster, S. (2015). *A study of assessment practice of oral English at lower secondary schools in Norway*. Unpublished MA thesis. University of Bergen, Norway.
- Akale, S. (2012). Rethinking speaking skills in EFL (English as a Foreign Language) settings. *Sino-US English Teaching*, 9(4), pp. 1100-1111.
- Allen, K. (2008). Teaching vocabulary with visual aids. *Journal of kao Ying Industrial & Commercial Vocational High School*, 1(9), pp.1-5.
- Angrist, J. & Levy, V. (1999). Using Maimonides' Rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114, pp. 533– 575.
- Arianie, M. (2017). Authentic Materials and interactive activities in EFL classroom. *English Franca*,1(2), pp.115-134.
- Arthur, L. Nolasco, R. (1987). *Conversation*, Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bell, J. (1993). *Doing Your Research Project*. Philadelphia: Open University Press.
- Bell, J. (1999). *Doing your Research Project: a guide for First-time researchers in education and social science*. Third edition. Buckingham: Open University Press.
- Bell, J. (2005). *Doing your research project. A guide for first researchers in education, health and social science.4th.ed. Berkshire*: Open University Press.
- Bennett, N. (1996). Class size in primary schools: perceptions of head teachers, chairs of governors, teachers and parents. *British Educational Research Journal*, 22, 1, pp. 33-55.
- Black, P. (1998). *Testing, Friend or Foe? Theory and Practice of Assessment and Testing*. London: Falmer Press.

- Braun, V & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, pp.77-101.
- Broughton, G., Brumfit, C., Flavell, R., Hill, P. & Pincas, A. (1980). *Teaching English as a Foreign Language*, 2nd ed. London: Routledge & Kegan Paul.
- Brown, A., N. Iwashita, and T. McNamara. (2005). *An Examination of Rater Orientations and Test Taker Performance on English for Academic Purposes Speaking Tasks*. (Monograph Series MS- 29). Princeton, NJ: Educational Testing Service.
- Brown, A., N. Iwashita, T. McNamara, & S. O'Hagan. (2002). 'Getting the balance right: Criteria in integrated speaking tasks.' Paper presented at the 24th Language Testing Research Colloquium, Hong Kong, December 12–15.
- Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy* (2nd ed.). New York: Addison Wesley Longman, Inc
- Brown, H. D. (2007). *Teaching by principles: An Interactive approach to language pedagogy* (3rd ed.). White Plains, NY: Pearson Education
- Brown, H. D. (2004). *Language assessment principles and classroom practices*. New York: Longman
- Bryman, A. (2001). *Social Research Methods*. Oxford: Oxford University Press.
- Bryman, A. (2004). Qualitative Review on leadership: A critical but appreciative review. *Leadership. Quarterly*, 15, pp. 729-769
- Bryman, A., (2008). *Social research methods*. 3rd ed. Oxford: Oxford University Press
- Burns, R. (2000). *Introduction to Research Methods*. London: Sage Publications
- Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J.C. Richards & R.W. Schmidt (Eds.), *Language and communication*, pp. 1-27. London: Longman.
- Chang, C-W. (2006). *Teachers' Beliefs towards Oral Language Assessment in Taiwan Collegiate EFL Classrooms, international conference on English instruction and assessment*, Fooyin University.
- Chou, M. (2013). Teacher interpretation of test scores and feedback to students in EFL classrooms: A comparison of two rating methods. *Higher Education Studies*, 3(2), pp. 86-95.

- Chuang, Y. (2009). Foreign Language Speaking Assessment: Taiwanese College English Teachers' Scoring Performance in the Holistic and Analytic Rating Methods. *The Asian EFL Journal*, 11 (1), pp.150-173.
- Chuang, Y. Y. (2007). A study of college English teachers' preferences and perceptions of using the types of tasks for assessing Taiwanese EFL students' foreign language oral proficiency. *Journal of Applied Foreign Languages*, 8, pp. 74-105.
- Cohen, L. & Manion, L. C. (1994) *Research Methods in Education*. London: Routledge.
- Cohen, L., Manion, L. & Morrison, K. (2007). *Research Methods in Education*. 6th ed. London: Routledge.
- Creswell, J. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage
- Creswell, J. W. Plano Clark, V. L., Gutmann, M., & Hanson, W. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook on mixed methods in the behavioral and social sciences* (pp. 209–240). Thousand Oaks, CA: Sage Publications.
- Creswell, W.J. (2009). *Research design: Qualitative, quantitative and mixed method approaches*. Thousand Oaks, CA: Sage.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth Group/Thomson Learning.
- Dakowska, M. (2005). *Teaching English as a Foreign Language. A Guide for Professionals*. Warszawa: Wydawnictwo Naukowe PWN.
- Davies, A. (2003). *The native speaker: Myth and reality* (2nd ed.), Buffalo, New York: Multilingual Matters.
- Davies, A. (1991). *The native speaker in applied linguistics*. Edinburgh: Edinburgh University Press.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Dawson, C. (2002). *Practical Research Methods. A User-friendly Guide to Mastering Research*. Oxford: How to books Ltd.
- Denscombe, M. (2007). *The Good Research Guide for Small-scale Social Research Projects. 3rd ed. projects*. Berkshire: Open University Press.

- Dörnyei, F. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press
- Dörnyei, Z. (2003). *Questionnaires in Second Language Research: Construction, Administration, and Processing*. Lawrence Erlbaum Associates.
- Edge, J. (1992). *Cooperative Development*. Harlow, UK: Longman
- Enerson, D., Plank, M. & Johnson, N. (1994) *An Introduction to Classroom Techniques*. Available at: <http://www.Psu.edu/celt/CATs.html/>.
- Ferreira, L. (2007). *How to teach vocabulary effectively*. Published dissertation, Cape Verde.
- Freed, B. (1995). What makes us think that students who study abroad become fluent? In B. Freed (Ed.) *Second language acquisition in a study abroad context*, Amsterdam, (pp. 123-148).
- Fulcher, G. & Reiter, R.M. (2003). Task difficulty in speaking tests. *Language Testing*, 20(3), pp. 321- 344.
- Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing*, 13(1), pp. 23-51
- Fulcher, G. (2003a). Interface design in computer-based. *language testing*. *Language Testing*, 20(4), pp. 384-408.
- Fulcher, G. (2010). *Practical language testing*. London: Hodder Education
- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment*. London: Routledge.
- Galaczi, E. D. (2010). Paired speaking tests: An approach grounded in theory and practice. In J. Mader & Z. Urkun (Eds.), *Recent approaches to teaching and assessing speaking (IATEFL TEA SIG Famagusta Conference Proceedings)*. Canterbury, UK: IATEFL Publications
- Galata, F. (2021). The Practice of English Language Productive Skills' Continuous Assessment in Preparatory School: *The Case of West Wollega Zone, Ethiopia*. *International Journal of Innovative Science and Research Technology ISSN No: -2456-2165*.
- Ghermaoui, I. (2018). *Assessing the Speaking Skill Among EFL Learners the Case of Third- Year Licence Students at the University of Tlemcen*. University of Tlemcen.
- Gilmore, A. (2007). Authentic materials and authenticity in the foreign Language Teaching, *ELT Journal*, 40/2, pp. 97-118.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24, pp. 355-392

- Glesne, C., & Peshkin, A. (1992). *Becoming Qualitative Researchers: An Introduction*. White Plains, NY: Longman.
- Goh, C. (2009). Perspectives on spoken grammar. *ELT Journal*, 63(4), pp. 303-312.
- Grada, T. K. A. (2014). *An investigation into Libyan EFL Novice Secondary School Teachers' current knowledge and practice of speaking assessment: A socio-cultural perspective* (Doctoral dissertation). University of Exeter, United Kingdom.
- Haley, M.H., & Austin, T.Y. (2004). *Content-Based Second Language Teaching and Learning an Interactive Approach*. New York: Allyn and Bacon.
- Harmer, J. (2009) *The Practice of English Language Teaching*. Pearson Longman, UK.
- Harmer, J. (2001). *How to Teach English*, Seventh Impression. Malaysia: Longman.
- Harris, M. & McCann, P. (1994). *Assessment*. Oxford: Macmillan Heinemann.
- Harvey, P. (2006). *Improving Teaching Observation Practice in the Learning and Skills Sector*. The Research Centre, City College Norwich, Ipswich Road, Norwich.
- Hedge, T. (2000) *Teaching and Learning in the Language Classroom*. Oxford: Oxford University Press.
- Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, 42(2), pp. 282–296. <https://doi.org/10.1598/RRQ.42.2.4>
- Huang, S. (2012). Pushing Learners to Work Through Tests & Marks: Motivating or Demotivating? A Case in a Taiwanese University. *Language Assessment Quarterly*, 9, pp. 60–77.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press. Print.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.), Cambridge: Cambridge University Press.
- Hughes, Arthur. (2000). *Testing for Language Teachers*. Beijing: Foreign Language Teaching and Researching Press
- Huxham M., Campbell, F. & Westwood, J. (2012). Oral versus written assessments: a test of student performance & attitudes. *Assessment & Evaluation in Higher Education*, 37(1), pp.125–136.

- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29(1), pp. 24–49.
- James, M. (1998) *Using Assessment for School Improvement*. Oxford: Heinemann Educational.
- Jang, E., McDougall, D., Pollon., Herbert, M. & Russell, P. (2008). Integrative Mixed Methods Data Analytic Strategies in Research on School Success in Challenging Circumstances. *Journal of Mixed Methods Research*. 2. 221-247. 10.1177/1558689808315323.
- Johnson, B. & Christensen, L. (2012). *Educational Research, Qualitative, Quantitative and Mixed Approach*. (4th ed). California: SAGE Publication.
- Jondeya, Rania S. (2011). *The Effectiveness of Using Information Gap on Developing Speaking Skills for the Eighth Graders in Gaza Governorate Schools-A Thesis*. Al- Azhar University-Gaza
- Jones, R.L & Bray, E. (1986), *Assessment; Form principles to action*, Guides to assessment in assessment. London MacMillan.
- Karavas-Doukas, K. (1998). Evaluating the implementation of educational innovations: lessons from the past. In P. Rea-Dickins & K. P. Germaine (eds.), *Managing evaluation and innovation in language teaching*, London: Longman.
- Kerr, D., Troth, A., & Pickering, A. (2003) The use of role-playing to help students understand information systems case studies, *Journal of Information Systems Education*, 14(2), pp.167-167.
- Kim, H. J. (2005). *World English and language testing: The influence of rater variability in the assessment of process of English language oral proficiency*, Published doctoral dissertation in the University of Iowa.
- Kim, H.S. (2003). The types of speaking assessment tasks used by Korean junior secondary school English teachers. *L2 Language Assessment in the Korean Classroom*, 5(4), pp.1-30.
- Kim, Y. H. (2006). Issues of rating scales in speaking performance assessment. *TESOL & Applied Linguistics*, 6, pp. 1-3.
- Kim, Y. H. (2009). An investigation into native and non- native teachers' judgement of oral English performance: A mixed methods approach, *Language Testing*, 26 (2), pp.187-217.
- Kingen, S. (2000). *Teaching Language Arts in Middle Schools.: Connecting and Communicating*. New Jersey: Lawrence Erlbaum Associates, Publishers.

- Kormos, J. & Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), pp.145-164.
- Kothari, R. (2004). *Research methodology: Methods and techniques*, (2nd ed.), New Delhi: New age international limited.
- Kramersch, C. (1986). From Language Proficiency to Interactional Competence. *The Modern Language Journal*, 70, 366-372. <https://doi.org/10.1111/j.1540-4781.1986.tb05291.x>
- Lado, R. (1964). *Language testing: The construction and use of foreign language tests*. 3 ed. London: Lowe and Brydone (Printers) LTD.
- Lee, S. (2010). *Current practice of classroom speaking assessment in secondary schools in South Korea*. Published MA. University of Queensland, Australia.
- Lee, Y. (2007) The Multimedia Assisted Test of English Speaking: The SOPI Approach. *Language Assessment Quarterly*, 4(4), pp.352–366.
- Li, W. (2011). Validity Considerations in Designing an Oral Test. *Journal of Language Teaching and Research*, Vol. 2, No. 1, pp. 267-269, January 2011 © 2011 ACADEMY PUBLISHER Manufactured in Finland.
- Linse, C. T. (2005). *Practical English Language Teaching: Young Learners*, in Nunan, David (ed.). Ney York: McGraw-Hill Companies.
- Liu, X. (2010). Arousing the College Students' Motivation in Speaking English Through Role-Play. *International Education Studies*, p. 3.
- Luoma, S. (2004). *Developing speaking tasks. assessing speaking*. Cambridge. Cambridge University Press.
- Madsen, H. S. (1983). *Techniques in testing*. Oxford: Oxford University Press.
- Maguire, M. & Delahunt, B. (2017). Doing a thematic analysis: Step by step guide for learning and teaching scholars, *AISHE*, 8(3), pp. 3350-33514.
- May, L. (2011). Interactional competence in a paired speaking test: features salient to raters. *Lang. Assess. Q.* 8, 127–145. doi: 10.1080/15434303.2011.565845
- McNamara, T. (1996). *Measuring second language performance*. London; New York: Longman
- Merriam, S.B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass
- Miller, L. (2003). Developing listening skills with authentic materials. *ESL Magazine*, 6(2), pp.16-18.

- Muhd Syahir, Z. I. (2016). The Implementation of Role-Play as A Means to Overcome Inhibition in English Speaking Assessment. *Simposium Pendidikan Bahasa Kebangsaan* . Ipoh: IPG.
- Muñoz, A., Dary, L., Crespo, F., Gaviria, S., Lopera, L. & Palacio, M. (2003). Assessing spoken language in EFL: Beliefs and practices, *Medellin: Universidad EAFIT*, 129, pp. 63-73.
- Muñoz, P. & Álvarez, M. (2010). Washback of an oral system in the EFL classroom. *Language Testing*, 27(1), pp.1-17.
- Murphy, P. (1999). *Learners, Learning & Assessment*. Paul Chapman Publishing Ltd. London.
- Nakamura, Y. (1993). *Measurement of Japanese college students' English-Speaking ability in a classroom setting*. Unpublished doctoral dissertation, International Christian University, Tokyo.
- Nakatasuhara, F. (2007). Developing a rating scale to assess English speaking skills of Japanese upper-secondary students. *Essex Graduate Student Papers in Language and Linguistics*, 9, pp. 83-103.
- Nation, I.S.P., and Newton, J. (2009). *Teaching ESL/EFL Listening and Speaking*. New York: Routledge.
- Norris, J & Ortega, L. (2006). *Synthesizing Research on Language Learning and Teaching*. John Benjamins Publishing.
- O'Sullivan, B. (2006). *Modelling Performance in Oral Language Tests: Language Testing and Evaluation*. Frankfurt: Peter Lang.
- O'Sullivan, B., Weir, C., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing*, 19(1), pp. 33-56.
- Pallotti, G. (2009). Defining, redefining and differentiating constructs. *Applied Linguistics*, 30(4), pp. 590-601.
- Poonpon, K. (2010). Integrating English speaking tests in an EFL classroom: What teacher can learn from. *The 36th International Association for Educational Assessment (IAEA) Annual Conference 2010*, pp.1-7.
- Punch, K. F. (1998). *Introduction to Social Research: Quantitative and Qualitative Approaches*. Thousand Oaks, CA: Sage
- Richards, J. (1990). *The language teaching matrix*. Cambridge: Cambridge University Press.

- Richards, J. C. (2006). *Communicative Language Teaching Today*. New York: Cambridge University Press.
- Richards, J. C. (2008). *Teaching listening and speaking: From theory to practice*. Cambridge: Cambridge University Press.
- Roohani, A. (2013). A comparative study of intuitive-imitative and analytic-linguistic approaches to teaching pronunciation: Does age play a role? *The Asian EFL Journal Quarterly*, 15(1), pp. 87-127.
- Rowntree, D. (1987). *Assessing Students*. London.1st Edition.
- Sakale, S. (2012). Rethinking speaking skills in EFL (English as a Foreign Language) settings. *Sino-US English Teaching*, 9(4), pp. 1100-1111.
- Sano, M. (1989). Speaking Shidou to Dramatization. *Speaking Instruction and Dramatization*. *Eigo-Kyouiku, The English Teachers' Magazine*. October, pp. 24–25.
- Satterly, D. (1989) *Assessment in schools*. Second edition. Basil Blackwell Inc. New York. USA.
- Schmitt, N. (2002). *An introduction to applied linguistics*. London: Oxford University Press.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
- Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. Thousand Oaks, Sage.
- Shamim, F. (1996). Learner resistance to innovation in classroom methodology. In H. Coleman (ed.), *Society and the Language Classroom*, pp. 105-121. Cambridge: Cambridge University Press.
- Shohamy, E. (1992). Beyond proficiency testing: a diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal* 76, pp. 513-21.
- Solcova, P. (2011). *Teaching Speaking Skills*. Thesis of Faculty of Arts of Masaryk University: Unpublished.
- Soliha, I. (2019). *Speaking Assessment Based on the 2013 Curriculum from English Teachers and Their Perception*. The University of Muhammadiyah Surabaya.
- Stempleski, S. (1990). Teaching communication skills with authentic video. *Video in Second Language Teaching: Using, Selecting and Producing Video for the Classroom*. Washington: Teachers of English to Speakers of Other Languages, pp. 7-24.
- Sutton, R. (1992) *Assessment: A framework for Teachers*. Routledge. U.K.

- Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn from tests scores. *Language Testing*, 18(3), pp. 275-302.
- Tantani, A. (2005). *Assessment of English as a foreign Language in Libyan Universities*. Dissertation of Faculty of Arts of Sunderland University: Unpublished.
- Taras, M. (2001). The use of Tutor Feedback and Student Self-assessment in Summative Assessment Tasks: towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, 26(6), pp. 606-614.
- Taylor, L. (2006). The changing landscape of English: implications for language assessment. *ELT Journal*, 60(1), pp. 51-60.
- Thornbury, S. (2005). *How to Teach Speaking*. England: Pearson Educational Limited.
- Thornbury, S. (2007). *How to teach Speaking*. Essex: Pearson Education Limited.
- Tran, T. T. N., & Nguyen, H. H. T. (2017). An investigation into the difficulties in implementing in-class English speaking and the required resources for its effective practice. *3rd National Conference on Interdisciplinary Research on Language and Language Teaching*, Hue University of Foreign Languages.
- Underhill, N. (1987). *Testing spoken language. A handbook of oral testing techniques*. Cambridge: Cambridge University Press.
- Wang, F., & Lee, S. (2007). Storytelling is the Bridge. Retrieved on January 12, 2015, from http://www.o.ntust.edu.tw/~syying.lee/publications/Story_telling_is_the_bridge_IJETL.pdf
- Watkins, C. (2011). *Learning: a sense maker's guide*. London, England. Association of Teachers and Lecturers. Available at: <http://www.atl.org.uk/Images/Learning%20a%20sense%20makers%20guide%20-%202011.pdf>
- Weaver, B. (2013) Formal vs. Informal Assessments. Scholastic, Inc. Available at: <https://beta.scholastic.com/teachers/articles/teaching-content/formal-vs-informal-assessments/>
- Weir, C. (1993). *Understanding and developing language tests*. Hemel Hempstead: Prentice Hall.
- Williams, D. (2015) *Speakout. Intermediate teacher's book*. (2d Ed.). Pearson Education Limited.

Wintle, M. & Harrison, M. (1999). *Coordinating assessment practice across the primary school*.
Routledge Falmer.

Zaremba, A. J. (2006). *Speaking professionally*. Canada: Thompson South-Western.

Appendices

Appendix A

Teachers' Questionnaire

The following questionnaire is part of MA research, which is carried out in English language departments at Zawia University. The aim of this questionnaire is to identify the speaking assessment strategies used by teachers in their classes. The provided information will be used for research purposes only and will not be revealed to any party. Thus, it would be a great help if you answer this questionnaire. The questionnaire will take about 15 minutes. Please put a tick in the appropriate box or choice whenever necessary. Your participation will be of a great value and will help me conduct the research.

Thank you for your help in advance.

Name

Background Information

- **Gender**

Male ()

Female ()

- **Age**

Under 25 ()

25-29 ()

30 or over ()

- **Qualification**

MA ()

PHD ()

- **Experience in teaching speaking**

Over year or less ()

1-2 years ()

more than 3 years ()

Part One: Teachers' speaking strategies to be considered

Language assessment criteria: they are standards whereby assessment categories are evaluated. For example, if vocabulary is selected as a category to assess students' oral language performance, a set of criteria is used to indicate how well a student can use vocabulary.

- 1- Do you firstly identify the purpose of assessment when you assess students` speaking language performance?
yes () no ()
- 2- Do you use different speaking assessment tasks in a speaking test?
yes () no ()
- 3- Do you clearly explain to the students how to accomplish speaking assessment tasks?
yes () no ()
- 4- Do you use speaking assessment criteria by outside resources?
yes () no ()
- 5- Do you consider accuracy as a category of speaking assessment?
yes () no ()
- 6- Do you consider fluency as a category of speaking assessment?
yes () no ()
- 7- Do you consider communicative skills as speaking assessment?
yes () no ()
- 8- Do you consider range of vocabulary as a speaking assessment?
yes () no ()
- 9- Do you consider pronunciation as category of speaking assessment?
yes () no ()
- 10- Do you consider the students` use of transitions as a category of speaking assessment?
yes () no ()
- 11- Do you consider the reliability and validity in speaking assessment?
yes () no ()

Language assessment tasks: Activities or techniques used to assess students` language performance.

Formative assessment: this type of assessment is used by teachers during classroom activity. It helps teachers to identify their students` understanding and the extent to which they and their students have met the goals prescribed in the curriculum.

Summative assessment: assessment given at the end of a prescribed period of teaching. Such an assessment takes place at the end of a semester or an academic year.

An analytic score: a mark given to each category in assessment criteria. For example, the teacher gives a separate mark to grammar, fluency, pronunciation, etc... while s/he assessing speaking skill.

A holistic score: a total mark given to the students` overall language performance.

Part Two: Teachers Speaking Strategies Use

For each statement, please tick the box which best represents your frequency of practicing these processes in speaking assessment.

Key: *A = always (100%); U = usually (75%); S = sometimes (50%); R = rarely (25%); N = never (0%).*

	Teachers` speaking strategies use	A	U	S	R	N
12	I use summative assessment in assessing speaking ability					
13	I use formative assessment in assessing speaking ability					
14	I set speaker assessment tasks at an appropriate level of difficulty					
15	I use oral presentation tasks in speaking assessment					
16	I use learner- learner joint discussion and decision taking tasks					
17	I use role – play tasks in speaking assessment					
18	I use interview tasks in speaking assessment					
19	I use retelling story or text from aural stimuli as speaking task in speaking assessment					
20	I use a holistic score for each category in speaking test					
21	I use an analytic score for each category of speaking assessment test					
22	I develop my own assessment criteria when I design speaking tasks					
23	I connect the selection of assessment criteria to the aim of speaking assessment					
24	I encourage students to read aloud in front of class and assess them					

Thank you for your time

Appendix B

Teachers' Interviews Questions

1. What do you know about speaking assessment in general?
2. Do you find any problem to use different assessing types of speaking?
3. To what extent do you understand how to test students' oral ability?
4. Do you have any difficulty with representative assessing tasks?
5. What do you know about how to plan and structure an oral test?
6. What does it mean to you valid and reliable scoring?
7. Do you have any problems with eliciting a valid sample of oral ability?
8. How do you choose tasks for speaking assessment?
9. Do you use any different assessment categories for each task or you keep using the same ones for each task?
10. What factors do you think have an influence on the actual practice of speaking assessment?

Thank you for your cooperation

Appendix C: Letter permission from post graduate studies & training at Zawia University
for collecting data

Ministry Of Higher Education
And Scientific Research
University Of Zawia
General Administration

وزارة التعليم العالي
والبحوث العلمي
جامعة الزاوية
الإدارة العامة

التاريخ: / / 20 م الموافق: / / 14 هـ الرقم الإشاري:

السادة/ بكلية التربية الزاوية
كلية الاداب الزاوية
كلية التربية ابو عيسى
كلية اللغات الزاوية

تحية طيبة وبعد

تحيةكم ... وتفيدكم إدارة الدراسات العليا والتدريب بجامعة الزاوية بأن الطالبة "امل
بلقاسم محمد عمر" إحدى طالبات الدراسات العليا لنيل درجة الاجازة العالية "الماجستير"
بقسم: (اللغة الانجليزية) ، وهي بمرحلة إعداد رسالتها .

نأمل ... مساعدة الطالبة المعنية للاطلاع و الحصول على البيانات والمعلومات ذات
العلاقة بموضوع الرسالة ..

شاكرين لكم حسن التعاون
والسلام عليكم ورحمة الله وبركاته

د/ يوسف عيسى المبروك
مدير إدارة الدراسات العليا والتدريب

مسورة إلى

الصفحة: 1 من 1
التاريخ: 2021/12/zainab