QATAR UNIVERSITY

COLLEGE OF EDUCATION

ASSESSMENT OF EFL SPEAKING SKILLS IN QATARI PUBLIC SECONDARY

SCHOOLS: TEACHERS' PRACTICES AND CHALLENGES

BY

TAHA OSMAN

A Thesis Submitted to the Faculty of

the College of Education

in Partial Fulfillment

of the Requirements

for the Degree of

Masters of Arts in Curriculum and Instruction

January   2018

COMMITTEEPAGE

The members of the Committee approve the Thesis of Taha Osman
defended on 23/12/2017.

_____
Dr. Areej Isam Ibrahim Barham
Thesis Supervisor

_____
Dr. Michael Romanowski
Thesis Co-supervisor

_____
Dr. Yousef Mohammad Khaled Alshaboul
Committee Member

_____
Dr. Elsayed Elshabrawi Hassanein
Committee Member

Approved:

_____
Ahmed Al-Emadi, Dean, College of Education

ABSTRACT

OSMAN, TAHA, M., Masters:

 January: 2018, Masters of Arts in Curriculum and Instruction

Title: Assessment of EFL Speaking Skills in Qatari Public Secondary Schools: Teachers'

Practices and Challenges

Supervisor of Thesis: Dr. Areej Isam Barham.

This thesis aims to conduct a quantitative investigation into the practices and challenges of EFL teachers in assessing their students' speaking skills. To collect data for this study, all EFL teachers currently working for the Ministry of Education and Higher Education in Qatar were invited to participate in an online survey using Google Forms Software. A total of 120 teachers took part in the data collection process by completing the questionnaire. Using SPSS 23 Software, the data was analyzed under five sets of assessment practices and three categories of challenges. Descriptive statistics revealed that EFL teachers were committed to providing enough time for the assessment of students' EFL speaking skills. In addition, results proved that teachers were careful to differentiate speaking assessment tasks, use a rating scale in scoring students' performance and provide students with feedback. However, teachers' challenges in the assessment of EFL speaking skills were mainly related to practicality issues, the lack of relevant training and the students' low levels of motivation and English proficiency.

DEDICATION

I dedicate this thesis to my dear parents, my brothers and sisters, my beloved wife and my

children

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Background

English has become the most dominant language in the world and the preferred means of communication in business and politics (Crystal, 2012; Pennycook, 2017). Ammon (2001) argues that English is the principal means of communication among scholars and scientific researchers and this has encouraged educational systems worldwide to choose English as the medium of instruction for most subjects and learning resources (Nunan, 2003). Dearden (2014) explored the use of English as a medium of instruction and found that there was a prevailing trend towards the spread of using English as a medium in teaching academic disciplines in countries where the mother tongue of the majority of the population is not English. For example, in Hong Kong, India and the Philippines, English is not only the preferred medium of instruction, but is also recognized as an official language for these countries, even though regional languages are still dominant (Tollefson & Tsui, 2003). As a result, decision-makers in educational fields pay close attention to the teaching, learning and assessment of English as a foreign or second language in order to ensure an active role in the fields of education, science, politics and business at an international level (Phillipson, 2003; Sahlberg, 2011). Luoma (2004) advocates the most important skill for learners of English, as a foreign language is, the ability to speak fluently and accurately for communication and interaction. Hence, stakeholders in education often evaluate success in language learning, as well as the effectiveness of a language course, on the basis of how well they feel that learners have improved in their spoken-language proficiency (Richards & Rodgers,

2014). Therefore, when a foreign language is a school subject, the assessment procedures used to gain insight into learners' ability to speak is of high priority for schools, teachers and learners (Davison & Leung, 2009; East, 2016).

One of the principal goals of contemporary approaches to language teaching and learning is the need for authentic and reliable assessment of English-speaking skills (Bachman, 2000). However, assessing students' speaking skills is one of the basic challenges in teaching English as a foreign language (McNamara, 1996; Hughes & Reed, 2016). It requires the design of authentically valid tasks that urge students to respond spontaneously to genuine communicative situations in order to show their learning of different cognitive, linguistic and social skills (Luoma, 2004). These assessment tasks should be administered according to a pre-determined procedural framework, sometimes referred to as 'blueprint' or 'table of specifications', to ensure the reliability of the assessment results (Linn, 2008).

Embracing the perceived importance of the English language for the country's progress, Qatar has ventured, through its 2007 initiative called "Education for a New Era", to use English as a medium of instruction in public schools and state universities (Dearden, 2014). Although a decree in 2012 by the Supreme Education Council required that Arabic would be the language of instruction in Qatar University, the educational system persisted in promoting the instruction and assessment of English as a school subject (Asmi, 2013). In the Qatari Educational System, English as a foreign language (EFL) is an established school subject for all grades from Kindergarten through grade 12 (Brewer et al., 2007). Moreover, demonstrating proficiency in the English language is not only an admission requirement for higher education institutions in Qatar and most

countries in the region, but also the medium of instruction for most specializations (Weber, 2014; Dearden, 2014). These endeavors of the Qatari educational policymakers are meant to ensure that graduates from Qatari educational institutions are well-equipped for playing leading roles in different fields at international levels, and thus achieving the country's vision 2030 (Massialas & Jarrar, 2016).

To demonstrate the importance of the EFL speaking skills, Qatari curriculum standards allocated 30% of the total content teaching and assessment of the EFL course for speaking skills in all grades (Ministry of Education and Higher Education, n.d.1). In addition, the Qatari National Professional Standards for Teachers (NPST) highlighted the importance of assessment by necessitating teachers to assess students' learning and to manipulate the assessment results and data to enhance teaching and learning activities (Ministry of Education and Higher Education, n.d.2).

Since the Qatari government seeks to promote the teaching of EFL speaking skills, exploring the status quo of the EFL speaking assessments seems inevitable in order to ensure the validity, reliability and objectivity of the assessment results. Different teachers may provide students with varying conditions for the assessment of EFL speaking skills, which may result in unfair practices. These assessment practices might have a negative backwash effect on the teaching and learning of speaking skills in particular and the EFL courses in general (Brown, 2004).

Since assessment procedures should follow the Ministry of Education and Higher Education's requirement for standards-based assessment (Omran, 2005), empirical studies need to be conducted by researchers and practitioners. These studies will address the requirement for validating assessment tasks of these EFL skills and ensure the

reliability of the assessment results in order to achieve the standards-based assessment mandated by the Ministry of Education and Higher Education in Qatar (Brewer et al., 2007).

Investigating the present status of EFL speaking assessment in the Qatari educational context will reveal many issues that might need to be resolved. These issues are related to the assessment practices including preparing students for assessment, differentiating assessment tasks, allocating enough time for assessment, scoring the spoken responses and providing students with feedback. Therefore, it is essential to investigate the current practices in assessing EFL speaking skills in the Qatari context and to explore teachers' challenges as assessors of EFL speaking skills.

1.2 Research Questions

The current study seeks to investigate the EFL teachers' practices and challenges in assessing their students' speaking skills in Qatari public secondary schools. This aim can be achieved by addressing the following research questions:

RQ (1). What are the EFL teachers' current practices in assessing their students' speaking skills in Qatari public secondary schools?

1. Do teachers prepare students for assessment?

2. Do teachers differentiate the assessment tasks?

3. Do teachers allocate adequate time for students to complete the assessment?

4. What scoring techniques do teachers manipulate to score performances?

5. Do teachers provide students with feedback on their performance?

RQ (2) What are the challenges that EFL teachers face in assessing their students' speaking skills in Qatari public secondary schools?

1.3 Significance of the Study

The current study provides research-based information that might support the Qatari Ministry of Education and Higher Education in making informed decisions regarding the development of educational methods. Based on results from this research, relevant professional development workshops might be planned and conducted to equip EFL teachers in the secondary stage with the essential knowledge and skills required for performing the standardized assessment of their students' speaking skills. These professional development programs could help teachers in designing valid assessment tasks and in maintaining the objectivity and reliability throughout the different assessment stages. In addition, the study will offer scholars and researchers an opportunity to get empirical data about this area of language assessment that was seldom tackled in the field of teaching English as a foreign language in general and in the GCC region in particular.

1.4 Operational Definitions

There are certain words used in this study that may need operational definitions.

EFL Speaking Assessment: Is a performance-based assessment that is administered to measure non-native students' proficiency in using English for communication and interaction.

Assessment Practices: According to Wilson (2009), this term covers a wide range of activities starting from planning for assessment, designing assessment tasks, constructing relative rubrics and guiding material, implementing assessment to grading the output of the assessment process and using the assessment data. In this study, the researcher focused on practices for a specific performance-based assessment, i.e. the EFL speaking

skills assessment. Assessment practices include: preparing students for assessment, differentiating assessment tasks, providing ample time for assessment, using scoring techniques and providing students with feedback.

1.5 Organization of the Thesis

This chapter aimed at introducing the current study by providing the background of the study, the research questions, significance, the definitions of key terms and the thesis organization. The next chapter will display a review of the relevant literature tackling the assessment practices and challenges of EFL speaking skills. In the third chapter, the methodology used for completing the study is described, outlining the participants' data collection methods, data analysis procedures and the ethical considerations. The quantitative data collected to answer the research questions will be reported and analyzed in the fourth chapter. Finally, the fifth chapter will discuss the findings, present pedagogical implications and suggestions for further research.

CHAPTER 2: LITERATURE REVIEW

The previous chapter developed the significance of exploring EFL teachers' practices in assessing their students' speaking skills in Qatari public secondary schools. This chapter draws on research concerning the assessment of speaking skills, presenting the literature which addressed the different practices and challenges that EFL teachers encounter. The first section discusses the performance-based models of speaking assessment and how the researcher used them in defining the conceptual framework for his study. The second section highlights research on the different themes that affect the validity and reliability of EFL speaking assessment. The third section presents the research findings from previous studies that aimed at exploring teachers' practices and challenges in assessing their students' speaking skills. In the conclusion section, arguments are developed for the necessity of conducting investigative studies of EFL teacher's practices in assessing their students' speaking skills in the Qatari context.

2.1 Conceptual Framework for the Assessment of EFL Speaking Skills

O'Sullivan et al. (2002) points out that the assessment of EFL speaking skills is exclusively performance-based, and therefore it is necessary to elaborate the models used as frameworks for performance assessment tasks. The most common performance-based models for the assessment of speaking skills define the relationships between the speaking construct being measured, the assessment tasks used, the speaking performance elicited and the scoring of that performance which is used to make inferences about students' language ability (Milanovic & Saville, 1996; McNamara, 1996;, Skehan, 1998; Fulcher, 2003).

Milanovic and Saville's (1996) framework includes: specifications and construct, the test-taker, the examiner, the assessment criteria, the task, and the interaction between these factors (see Figure 1). This framework draws attention to the factors to be taken into consideration while designing a performance-based assessment task. For designing speaking assessment, this model defines three phases; the development phase, the administration phase and the marking phase. During the first phase, designers of the speaking test are held accountable for ensuring the reliability and validity of the assessment task in measuring the speaking construct. The assessment tasks designed in the first phase are used to elicit the candidates' speaking samples in the administration phase. Milanovic and Saville ascribed some factors affecting the speaking performance of candidates throughout this phase. These included such factors as their knowledge, ability, the examination conditions, the assessment tasks and the assessment criteria. In the third phase, examiners rate the candidates' performance in accordance with the assessment criteria provided under the assessment conditions. This framework is considered one of the earliest and most comprehensive models in elaborating variables involved in a performance test (O'Sullivan et al, 2002).
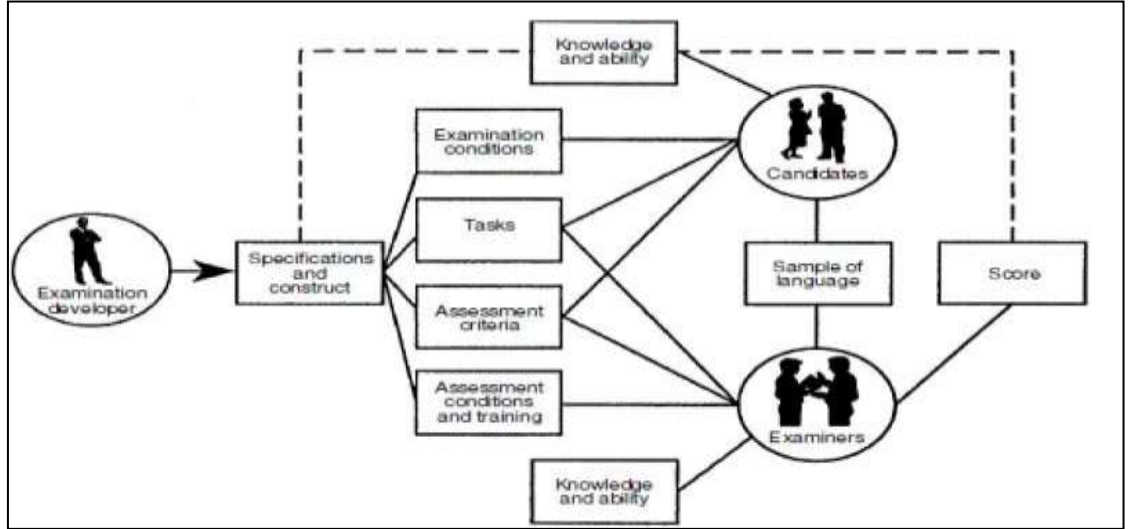
Figure 1. A Conceptual Framework for Performance Testing (Milanovic & Saville, 1996, P. 16)

McNamara (1996) put forward a model to illustrate the interactional nature of performance assessment with a focus on the rating process. This model describes how the interlocutor elicits the candidate's performance and how the rater rates that performance (see Figure 2). The model indicates that speaking performance is being influenced by several factors including the tasks which drive the performance and the raters who judge the performance using rating scales and criteria. Therefore, the final score can only be partly considered as a direct index of performance which is influenced by other contextual factors such as the test taking conditions. This model also includes two processes of speaking assessment. One is the candidate's test-taking process and the other is the rater's rating process. These two processes are of crucial importance to a speaking test in terms of reliability and validity.
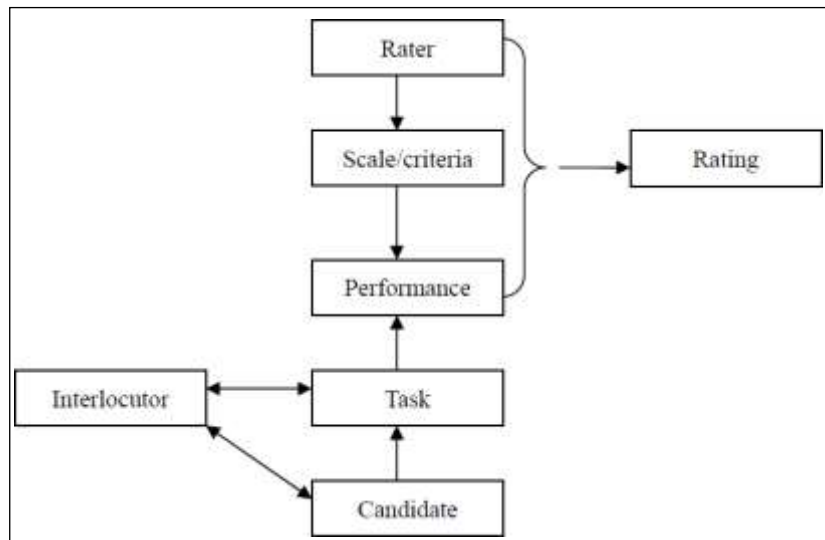
Figure 2. Proficiency and its Relationship to Performance (McNamara, 1996: 86)

Skehan's (1998) model for performance-based speaking assessment (see figure 3) strived to define additional factors that explain the intricacy of the speaking assessment. Skehan argues that assessment tasks need to be analyzed further to account for task characteristics and task implementation conditions. In addition, Skehan maintains that students' abilities require not simply an assessment of competences, but also an assessment of the ability to use the language. Fulcher (2003) comments that one distinctive feature of Skehan's model is that it depicts three factors mainly affecting test scores. These are "the interactive conditions of the performance, the abilities of the test taker, and the task used to elicit the speaking performance" (Fulcher, 2003, p. 113). The description of task qualities and task conditions in the model makes it much easier for language testers to develop and compare tasks.
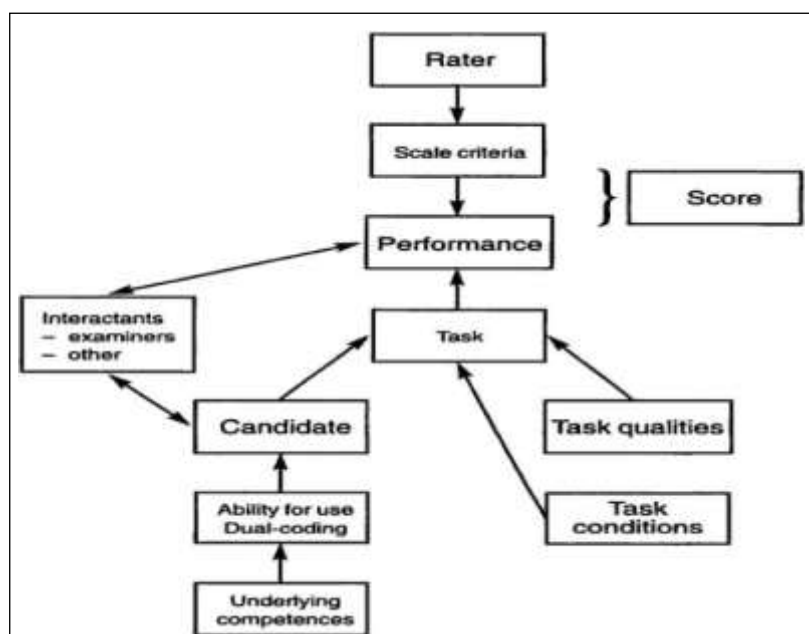
Figure 3. Model of Oral Test Performance (Skehan, 1998: 172)

Fulcher's (2003) model is more comprehensive and exhaustive than the other three models in defining the different variables affecting the candidate's performance and the test results in a speaking assessment task (Zhao, 2013). Zhao regards the relationships described in this performance-based model as especially expressive of the complexity of designing valid and reliable assessment tasks for assessing the EFL speaking skills. At the heart of his model, Fulcher places the construct definition together with the design of the rating-scale, the understanding of what constructs are being assessed, and the inferences that are drawn from scores. This model shows the effect of the nature of the rating scale and the scoring philosophy on test scores. Moreover, Fulcher acknowledges that rater training and characteristics play a role in the scoring process. He also indicates that there is an interaction between the rating scale and a test taker's performance which results in the score and any inferences that are made about the test taker.

In the same way as Skehan (1998), Zhao (2013) believes that Fulcher considers the variables that have an effect on the assessment task in his model. He believes that the orientation, the goals, and the topics of the assessment task together with the context-specific characteristics or conditions are among these variables. Moreover, Fulcher's model, as reflected by Zhao, demonstrates a number of factors that have an impact on the test taker. These include any individual differences between candidates (like personality), their ability for real-time processing and any task- specific knowledge or skills they might possess.
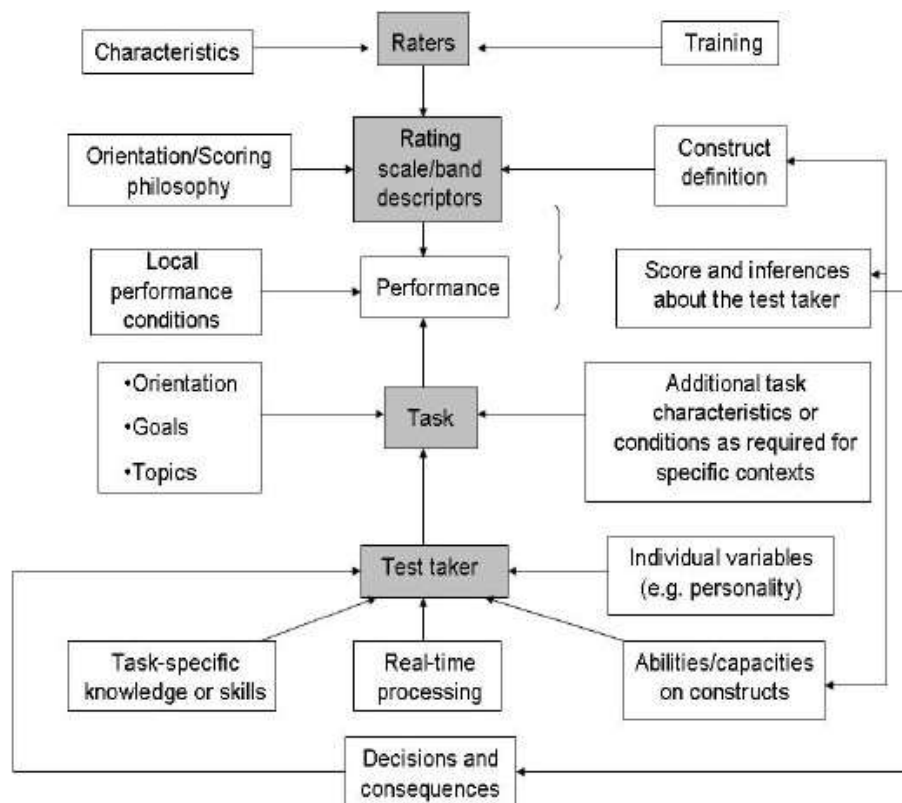
Figure 4. An expanded model of speaking test performance (Fulcher, 2003: 115)

The previous models for assessing speaking skills show the importance of each and every element in the assessment framework and how they are interconnected and influencing each other. Teachers' knowledge, skills and attitudes affect their assessment practices either in preparing students, selecting tasks, providing adequate conditions, scoring speaking performances or giving feedback. These practices have an effect on the results of speaking assessment and the interpretations of learners' performance. In order for these results and their accompanying interpretations to be valid and reliable, Brown (2004) highlights the need for a procedural framework that ensures that every component in the assessment process supports the elicitation of the intended performance. Luoma (2004) emphasizes the importance of examining all testing procedures and argued that all testing procedures should conform to the pre-determined blueprint.

2.2 Assessment Practices of EFL Speaking Skills

Results from speaking assessment as explained by Luoma (2004) could be affected by a myriad of factors that arise from the nature of the task type, inappropriate assessment procedures or the teachers' background. Without the relevant training and the availability of guiding materials, these factors may cause a number of challenges for English teachers ending in flawed assessment practices that might have a negative backwash effect on teaching and learning (Brown, 2004). For this reason, researchers have focused on investigating the assessment of speaking skills from different angles. For instance, research has been conducted to review the authenticity and validity of assessment tasks (Chinda, 2009; Khamkhien, 2010; Sinwongsuwat, 2012; Sook, 2003) while other studies explored the reliability of measurement related to the assessment

conditions (Dandonoli & Henning, 1990; Huang, 2016; Luoma, 2004; Nakatsuhara et al., 2017; Zhao, 2013). In addition, researchers have examined the rating practices in scoring speaking performances (Kim, 2011; Rodríguez, & Guiberson, 2011; Sawaki, 2007). preparation, feedback

2.2.1 Preparing Students for Assessment

Luoma (2004) required that language learners should be well informed about what and how they will be assessed for speaking skills. Prior to speaking assessment, there is a number of factors that should be considered. For Luoma, depending on teaching and learning resources, which do not consider the intended speaking types and conditions to be assessed, would challenge the validity and reliability of the assessment. In support, Dandonoli and Henning (1990) found that the use of assessment guidelines by language learners enhances the reliability and validity of the ACTFL Oral Proficiency Interview. Besides, Huang (2016) found that specific test-taking strategy use had a more direct effect on speaking performance than others. Hence, he recommended the inclusion of strategy instruction in the language classroom as a speaking test preparation practice.

Torky (2006) highlighted the importance of preparing students for speaking skills assessment and suggested a number of procedures to be implemented. She recommended that students should be aware of the criteria according to which their speaking is evaluated in order to work hard to meet these criteria. To achieve this, she suggested that speaking instruction should be given more attention in EFL classes, and students should be offered enough opportunities to practice speaking for authentic purposes on a daily basis. In addition, she argues that EFL teachers should focus equally on the different speaking sub- skills, such as conversation management, pragmatic competence and

fluency beside their usual focus on grammar and vocabulary. She also recommended offering students the opportunities to self-evaluate their oral performance.

2.2.2 Assessment Tasks

A speaking assessment task could be seen as a tool by which a spoken sample can be elicited and scored (Kim, 2009). Early trials for developing speaking assessment tasks were concerned about the oral product that these tasks would elicit. For example, Brown and Yule (1983) necessitated the evaluation of four types of talk - description, storytelling, instruction and expressing and justifying opinions. Meanwhile, later scholars heightened the need for a speaking assessment to involve tasks that cover all the types of oral production to align with the nature of the assessed construct. For example, Luoma (2004) and Thornbury (2005) required speaking skills assessments to include interactive tasks and monologues. The trend seems to be moving from focusing on the oral product towards caring for the performance and the underlying context that formulates the whole discourse.

Brown (2004) distinguished six major assessment task types for examining speaking skills; imitative tasks to check learners' ability to repeat linguistic features; intensive tasks that entice the performance of specific language aspects; responsive tasks which expect peculiar replies to others' prompts; transactional tasks for exchanging definite information; interpersonal tasks testing the ability to complete a social goal and extensive tasks that elicit a prolonged purposeful speech.

Thornbury (2007) defined five oral test types, live monologues, recorded monologues or dialogues, interviews, role plays, collaborative tasks and discussions. Regarding the limitations in assessing language speaking skills, Thornbury states that live

monologues do not examine learners' ability to conduct a casual conversation while recorded monologues and dialogues would contribute to better reliability. Furthermore, he argued that during interviews and collaborative tasks and discussions, the examinee is affected by the interviewer or the interlocutor. These previous types of assessment tasks showed different levels of authenticity, validity and reliability.

Wigglesworth (2008) suggested that authenticating speaking assessment tasks is a challenge. For Wigglesworth, language assessment tasks should be designed in a way that elicits the desired language skills in the real world situations. Assessment tasks need to urge learners to use authentic language in performing a pragmatic goal (Brown, 2004). In this way, these tasks may gain the validity to measure speaking skills which is natural and spontaneous. EFL teachers have used many performance-based tasks in assessing their students' speaking skills such as playing roles, having an interview, giving a presentation, narrating a story, responding to an oral, visual or written prompts (Sook, 2003; Chinda, 2009; Khamkhien, 2010). However, different tasks may elicit different interactional features of spoken responses (Sinwongsuwat, 2012).

For assessing Thai students' EFL speaking skills, Sinwongsuwat (2012) advocated the use of non-scripted role-plays as an alternative assessment task. She investigated the capability of face-to-face interview and role-play tasks to measure students' proficiency in speaking skills and found that during the interview, teachers used to take the initiative to start the talk, ask questions, bring up another topic and end the talk, depriving students from displaying their abilities in performing these actions. Meanwhile, the naturally-developed, non-scripted role plays showed high efficiency in assessing learners' competence in speaking English for communication and interaction,

especially when an appropriate rating scale is used. These interactive tasks involve peers in a natural conversation where they may utilize their linguistic knowledge in natural communicative practices such as initiating a talk, taking turns, developing and shifting topics.

2.2.3 Assessment Conditions

Taking the principle of fair assessment into consideration, East (2016) believes it is essential to tell the students in advance about the time they will be assessed as this would give students the opportunity to do their best and get ready for the assessment.

Due to the time-consuming nature of face-to-face speaking tests, Öztekin (2011) declared that EFL teachers feel obliged to abandon the task of assessing their students' speaking skills or simply tend to ignore the need to assess. Although face-to-face mode for the assessment of speaking skills has dominated for a long time as Luoma (2004) reported, modern technology has allowed for a computer-mediated semi-direct speaking assessment to exist. Chapelle and Douglas (2006) point out that computer technology has been widely used in assessing language skills other than speaking such as listening, reading and writing. This new mode has an advantage in assessing speaking skills for large numbers of candidates simultaneously by responding to a set of prompts released by a computer (Galaczi, 2010).

Researchers were not in agreement concerning the validity of using technology in the assessment of EFL speaking skills. For example, Csépes (2010) and Xi (2010) tested the validity of computer-based tests and concluded that they just provide partial assessment of the ability to communicate and interact in English. Therefore, the

conventional face-to-face direct mode of speaking assessment is still recognized by teachers and practitioners as the most important and practical in assessing EFL learners' speaking proficiency. However, Nakatsuhara et al. (2017) compared two modes of conducting and scoring speaking assessment; internet-based video-conferencing and face-to-face interaction. The two modes were found to give similar results after analyzing the test-takers' spoken responses and the score received via the two modes.

2.2.4 Scoring Techniques

Hongwen (2012) listed many factors which are responsible for the variability in scoring spoken language performance including: the rater's gender, native language, educational and professional background and training. These factors affect the accuracy of the scores which students receive for their speaking performance.

In order to lessen the effect of variability, Luoma (2004) suggests that the expected type of speaking skills to be assessed should be identified before developing the rating scales and the assessment tasks. For her, it would seem void if the teacher, for example, tests pronunciation using the same rating scales and assessment tasks as those used for assessing their spoken grammar or the ability to have a meaningful interaction. In accordance with that was Sinwongsuwat's (2012) suggestion to modify the commonly used rubric which did not include the pragmatic skills that characterize the natural flow of a conversation since it was only concerned with linguistic features such as vocabulary, grammar, fluency, accuracy and listening comprehension.

For a peer-interaction assessment task, Luke and Pavlidou (2002) highlighted the need for creating a rubric which includes social and interactional features such as greetings, initiating and managing topics. In their rating scales to be used in assessing

speaking skills, Hayati and Askari (2008) discriminated between linguistic and interactive elements.

In order to facilitate the assessment of speaking skills for a large number of candidates, a number of studies focused on using automated scoring systems. For example, Zechner et al. (2014) created a model system for evaluating 21 speaking subskills in order to assess the speaking proficiency of nonnative EFL teachers.

2.2.5 Feedback

Although providing each student with specific, constructive feedback on their speaking performance requires substantial amount of time and effort from EFL teachers, it is essential for language learners so as to realize what they have absorbed and what they still need to study more (Brown 2004). In this regard, Brown stresses the weakness of using marks as the only feedback format because marks do not provide specific information about the learner. According to East (2016), interpreting students' performance by reference to some pre-established criteria is more effective, informative and constructive than giving a mark or a grade. That is why feedback is recommended to be preceded by explanation of detailed rubrics and guidelines. Hattie & Timperley (2007) consider feedback as a part of the dynamic assessment with which students can improve their future performances by being self-dependent learners who acts responsibly towards their learning endeavors.

2.3 Previous investigations of EFL speaking skills assessment

Investigating assessment practices for EFL speaking skills has been of interest for many researchers (Bengqing, 2009; Chuang, 2007; Grada, 2014; Kim, 2003; Lee, 2010; Xu & Liu, 2009). Bengqing (2009) surveyed 12 EFL teachers for their perceptions and

practices in respect to the assessment of speaking skills. Bengqing's study demonstrated that speaking skills were mostly assessed in Chinese middle schools using tasks that depend on memorization rather than tasks that give rise to communicative interaction.

In an attempt to explore novice EFL teachers' knowledge of speaking assessment and the effect of this knowledge on their practices in Libyan secondary schools, Grada (2014) found that teachers had prepared a set of questions to be answered and they may had assessed a written conversation by students instead of performance assessment. Grada found that, f or these teachers, summative assessment was more important than formative assessment, and linguistic content of responses was more important than the communicative strength. Such views of the speaking assessment were found to have a negative effect on their assessment practices.

Kim (2003) analyzed the speaking assessment tasks used by Korean secondary school teachers and how these teachers' perceptions affect their practice. Kim concluded that teachers' lack of care about the validity and reliability of the assessment tasks, the lack of a relevant speaking assessment background and other practical issues such as time constraint and large classes were responsible for the poorly developed assessment tasks of speaking skills and the lack of teachers' confidence in conducting these tasks.

Lee (2010) Surveyed 51 EFL teachers in South Korean secondary schools to explore their assessment practices concerning English speaking skills. She found that large class size and lack of time were the most challenging factors for EFL teachers in assessing their students' speaking skills.

These attempts to explore EFL teachers' practices for the assessment of speaking skills offer two advantages. On one side, they emphasize the need for developing

teachers' knowledge and skills in such a challenging area of language assessment. On the other side, they provide empirical data which assist in determining the strengths and weaknesses of the existent assessment and act as a foundation for further studies which may customize solutions or offer new assessment practices.

Based on the previous studies, and in light of the performance-based assessment models, the researcher found that there are three main phases in the assessment of speaking skills. During these phases, EFL teachers should pay attention to their practices. In table 2.1 the researcher proposed a procedural framework for teachers' practices which clearly defines the three phases in the timeline of the assessment process.

Table 2.1
Assessment practices of EFL speaking skills

Before Assessment	Preparing Students	use same language
		use same task format
		share the rubric with students
		self-assessment
		peer-assessment
During Assessment	Standardizing Conditions	provide time for students to get ready
		provide enough time for the assessment
	Differentiating Tasks	according to students' abilities
		According to students' views
		Use pair and group work
	Scoring Performance	Use rating scales
		Use checklists
		Score in panels
		Write narrative reports
		Score from memory
After Assessment	Providing Feedback	use students' recordings
		provide ample feedback

2.4 Conclusion

Performance-based assessment of EFL speaking skills is likely to be less reliable than other standardized language tests as a result of depending highly on teacher's judgment and using different assessment tasks (Birenbaum, 1996). Few studies were conducted all over the world to explore EFL teachers' practices and challenges in assessing their students in secondary schools; none in the Qatari context. Therefore, the current study aims at expanding the scope of this specific research field by investigating teachers' practices and challenges in assessing their students' EFL speaking skills in Qatari public secondary schools.

CHAPTER 3: RESEARCH METHODOLOGY

This chapter presents a detailed description of the methodology used in the current survey study. It begins by describing the study participants and then defines the set of data generation methods and data analysis procedures. Finally, the researcher addresses the ethical considerations for this study.

3.1 Participants

The objective of this study was to investigate the practices and challenges of EFL teachers in assessing their students' speaking skills. This objective is accomplished by surveying EFL teachers in selected schools. The selection of participants was based on a convenience sampling method. A convenience sampling is a fast, simple and low-cost way of selecting participants, in which the researcher chooses candidates from a well-defined population depending on the topic of the study (Etikan & Alkassim, 2016). This technique of sampling guaranteed the collection of relevant data as the researcher was investigating specific assessment practices and challenges. In this case, all EFL teachers in Qatari government schools.

One-hundred and twenty EFL teachers from Qatari governmental secondary schools participated in the study. Participants' profiles were recognized in terms of five background variables: participants' gender, the highest earned degree, years of experience as EFL teachers, grades they teach, and their familiarity level with the Qatari curriculum standards for speaking skills. The EFL teachers in this study were represented adequately by both male and female teachers with 52% and 48% respectively. The highest earned degree for the majority of participating teachers was the bachelor degree and only 25% completed post-graduate studies such as diploma and master degrees.

Around 80% of participants had over ten years' experience in teaching English as a foreign language, while only two participants were novice teachers with less than three years of experience. While almost 40% of the participants used to teach mixed grades in secondary schools, the rest gave classes to only one grade; 16.4% in grade 10; 18% in grade 11; and 27.9% in grade 12. The majority of participants expressed their moderate or extreme familiarity with the curriculum standards for speaking skills. (See Table 3.1)

Table 3.1
Background Data of Participants

Criteria	Categories	Numbers	Percentage
Gender	Male	67	55.8%
	Female	53	44.2 %
Certificate	Bachelor	91	75.8 %
	Post-graduate Diploma	15	12.5 %
	Master	14	11.7 %
	PhD	0	0 %
Experience of Teaching EFL	1-3 years	2	1.6 %
	4-9 years	25	20.8 %
	10-14 years	23	19.2 %
	15 or more years	70	58.4%
Classes Taught	Grade 10	18	15 %
	Grade 11	22	18.3 %
	Grade 12	34	28.3 %
	Mixed Grades	46	38.4 %
Familiarity with curriculum standards	Not at all Familiar	0	0 %
	Slightly Familiar	5	4.2 %
	Somewhat Familiar	12	10 %
	Moderately Familiar	36	30 %
	Extremely Familiar	67	55.8 %

3.2 Data Collection Methods

3.2.1 Choice of Methods

The current educational research sought to describe the EFL teachers' practices and challenges in assessing their students' speaking skills in Qatari public secondary schools. Cohen, Manion and Morrison (2011) argue that gathering data about a special group of people at a specific point in time with the aim of describing the existing conditions would be served best by conducting a survey study. MacDonald and Headlam (2008) suggest that large numbers of participants can practically be reached by means of a quantitative survey study. For these reasons, the researcher used a questionnaire to complete this quantitative survey study since it is the most popular, flexible and economic survey method of gaining data (De Leeuw & Dillman, 2008). Survey research is flexible because it allows the gathering of quantitative and qualitative data from participants who have the benefit of maintaining their anonymity (Oppenheim, 2000).

This descriptive research utilized Google Forms Software to administer an internet-based questionnaire. An internet-based questionnaire was a practical solution because it reduced the cost and time needed to complete the survey. In addition, it allowed the researcher to reach difficult population by accessing female teachers in secondary schools, who are difficult for a male researcher to reach in the Qatari context. Moreover, it permitted respondents to choose the suitable time and setting for them to complete the questionnaire, and the website software prompted them to complete missed items. One of the advantages of using an internet-based questionnaire, as argued by Cohen, Manion and Morrison (2011), is obtaining more authentic responses as volunteers come from a diverse population without being coerced to participate.

3.2.2 Instrument

The themes and items of the questionnaire were derived from the most common performance-based models for the assessment of speaking skills that were discussed in chapter two. These models defined the relationships between the speaking construct being measured, the assessment tasks used, the speaking performance elicited and the scoring of that performance that is used to make inferences about students' language ability (Milanovic & Saville, 1996; McNamara, 1996; Skehan, 1998; Fulcher, 2003). In composing the questionnaire, various references were valuable and beneficial such as Luoma's (2004) *Assessing Speaking;* Brown's (2004) *Language assessment: Principles and classroom practices;* and Pawlak & Waniek-Klimczak's (2014) *Issues in Teaching, Learning and Testing Speaking in a Second Language.* These sources provided assistance in outlining the construction of the questionnaire based on the discussed conceptual frameworks for language assessment in general and speaking skills in particular.

The first draft of the questionnaire consisted of 32 items. The assessment practices were targeted in three phases, before, while and after performance. Assessment challenges were explored by three open-ended items. To see the extent to which this first draft of the questionnaire measures what it was intended to measure, as Cohen et al (2007) necessitated, three experts in the field of language assessment examined the validity of the questionnaire.

The three experts expressed their concern regarding the length of the questionnaire and the possibility of causing respondents to become frustrated or bored. They recommended focusing on specific assessment practices that received more attention in the literature and the Qatari context. Also, the experts suggested organizing

the questionnaire items in groups rather than phases to facilitate the data analysis. To achieve this, some items were deleted because they did not fit under the main topics selected. For example, one item asked whether teachers follow a yearly plan for assessing their students' speaking skills. Another item investigated the extent to which speaking assessment tasks aligned with what was taught in class. These two items did not go under any of the five topics that were selected and presented later in this chapter.

In addition, the experts pointed to possible modifications of the wording and arrangement of some items that were thought to be vague or leading to specific answers. For example, the item, "I use various tasks for assessing students' speaking skills", was deleted because other items asked about specific methods differentiating assessment tasks. These suggestions assisted in producing well-written items that could generate the required data through a self-enumerated method of collecting data. After consulting the experts, the second draft of the questionnaire was more concise and focused and reduced from 40 to 26 items.

To ensure the reliability of this designed tool and the appropriateness of the survey operational procedures, a pilot study was conducted in five secondary schools in Qatar that were selected according to accessibility. Thirty responses were collected from teachers with similar background information to those participating in the study. These participants confirmed the importance and appropriateness of all items. Reliability was used here to check that the questions were efficiently well-expressed to enable a consistent understanding by participants. Using SPSS 23, Cronbach's Alpha test was conducted to test the internal consistency of the 5 Likert-scale items in the questionnaire

and it was calculated .846, a relatively high coefficient of inter-item correlation. The final version of the questionnaire was prepared in light of both experts and teachers' feedback (see Appendix A).

The computerized questionnaire opens with an introductory message, which presented the researcher and briefed the participants about the context and purpose of the research. It also assured participants about the confidentiality and anonymity of their responses as Shropshire et al. (2009) recommended, and that participation was voluntary. They were informed that they had the right to quit the study at any time with no harm or responsibility and to access the findings of the questionnaire if needed.

The first five items of the questionnaire were designed to collect the participants' background information. The remaining 21 items were divided into 18 items investigating EFL teachers' practices in assessing their students' speaking skills and three items exploring their challenges. The 18 items exploring the practices were composed of 17 items of 5-Likert-scale questions and one multiple- answers item. Meanwhile, teachers' challenges were investigated by one multiple answers item and two open-ended questions. For the first 17 items, participants selected one of the following frequency scales: (*never, rarely, occasionally, frequently or always*) in a 5-point Likert scale items. Item number 18 was a multiple answers question where participants were requested to select the assessment tasks they used in assessing their students' speaking skills out of ten suggested tasks. Item number 19 was another multiple answers question asking participants to select the challenges they face in speaking assessment. Items 20 and 21 asked participants to list the difficulties they find and their recommendations to improve the assessment of EFL speaking skills.

The questionnaire included both closed and open-ended items. As Johns (2010) suggests, Likert-scale (closed-ended) items facilitate the data gathering and analysis processes. Meanwhile, the open-ended items provide freedom of answers and opportunity to probe (Oppenheim, 2000). In addition, open-ended items avoid the bias that may stem from suggesting responses to participants in Likert-scale items (Reja et al., 2003).

3.2.3 Procedures

In January 2017, the researcher applied for and gained the approval from the Qatari Ministry of Education and Higher Education (see Appendix B) for conducting the current study since it involves surveying teachers in Qatari secondary schools. Qatar University's Institutional Review Board (QU-IRB) exempted the research proposal from the full ethics review and granted approval to conduct the study (see Appendix C). Upon getting these approvals, the researcher commenced the development of the internet-based version of the questionnaire using Google Forms Software.

In March 2017, an email addressing the target population of the study was composed. It contained a participation request, information about the purpose of the study and the rights of participants (see Appendix D). In that email, a link was provided for participants to click on, in order for the internet browser to open the questionnaire webpage. This email was formally forwarded to all public secondary schools in Qatar via the ministry's formal email addresses. By the end of March 2017, a total of 120 EFL teachers from Qatari public secondary schools responded to the internet-based questionnaire.

3.3 Data Analysis

In order to answer the research questions, the computer-assisted data gathered was summarized and interpreted under two main headings: assessment practices and assessment challenges. Assessment practices were discussed with regard to five topics: preparing students for assessment; differentiating assessment tasks; providing enough time for assessment; scoring students' performances; and providing students with feedback. However, assessment challenges were classified into three categories: challenges related to teachers; challenges related to students; and challenges related to the assessment conditions and available resources.

For analyzing the quantitative data gathered for exploring the assessment practices, the Statistical Package for Social Sciences (SPSS) version 23 was used to calculate frequencies, percentages, means and standard deviations. The raw data elicited by the 5-Likert scale items were expressed by one of five responses: *never, rarely, sometimes, frequently and always*. The researcher transformed these five scales into a machine-readable format by coding them into 1, 2, 3, 4 and 5 respectively. Meanwhile, item number 18 was a multiple-answers item that required the calculation of frequencies and percentages for each answer.

In order to investigate the challenges that EFL teachers face in assessing their students' EFL speaking skills, responses to three items were analyzed. In item number 19, a multiple answers question, responses were analyzed by calculating frequencies and percentages for each suggested challenge. Meanwhile, participants' responses to items 20 and 21, two open-ended questions, underwent an interpretative content analysis process.

To bring an interpretation to qualitative data, Cohen *et al*'s (2007) stresses the need for arranging the data and reducing it by gathering related ideas out of disarranged, conversational information. The data collected by the two open-ended items underwent three phases: coding, categorizing and calculating occurrences (Ezzy, 2002). In the first phase, the researcher went through all responses to identify emerging themes of common challenges (Fakis et al., 2014). The second phase witnessed the dissection of each response into its constituent codes and were classified under their relative categories. Finally, quantification of the data was applied by counting the codes to calculate the frequencies of different challenges (Neuendorf, 2016).

3.4 Ethical Consideration

Since the current study required the involvement of participants, it was the researcher's responsibility to pay due attention to the following ethical respects. First of all, the survey was designed so that it would take no more than ten minutes on average to be completed. It avoided sensitive or threatening questions or items that might invade the participants' privacy. Therefore, the researcher considered Cohen *et al*'s (2007) warning about using the questionnaire in a harmful way that may cause uneasiness for the participants.

The next ethical issue was obtaining the participants' consent to participate in the online questionnaire. A 'request for participation' email was sent which covered the participant's rights of voluntary participation, anonymity, confidentiality and the freedom to leave anytime without completing the questionnaire (James & Busher, 2007). In addition, the participants' rights were mentioned at the beginning of the questionnaire form. Participants were given the contact details of the researcher and his supervisor

in case they needed to ask for more information about the study.

The first section of the questionnaire included all the essential-to-know information about the study such as the research objectives and the data collection instrument. Participants were assured that their responses would be securely confidential and their identities would be kept anonymous. It was made clear that participation is voluntary, and participants had the absolute right to withdraw at any time (James & Busher, 2007). Moreover, the researcher notified participants that they have the right to be informed about the questionnaire findings in case they asked for that. Finally, for the sake of protecting data, the researcher stored anonymous copies of the questionnaire in a file with a password on his personal laptop.

CHAPTER 4: RESEARCH RESULTS

This chapter reports on the questionnaire results and presents the findings of the study in an attempt to answer the two main research questions.

RQ (1). What are the EFL teachers' current practices in assessing students' speaking skills in Qatari public secondary schools?

RQ1a. Do teachers prepare students for assessment?

RQ1b. Do teachers differentiate the assessment tasks?

RQ1c. Do teachers allocate adequate time for students to complete the assessment?

RQ1d. Do teachers manipulate different techniques to score students' performances? RQ1e. Do teachers provide students with feedback on their performance?

RQ (2). What are the challenges that EFL teachers face in assessing their students' speaking skills in Qatari public secondary schools?

4.1 Assessment Practices

This part of the results analysis seeks to answer the first research question of the study which aimed at describing the status of EFL teachers' practices in assessing their students' speaking skills in Qatari public secondary schools. These practices were explored by 18 items (from 1 through 18) on the questionnaire. Quantitative data collected was analysed under five headings: preparing students for assessment, differentiating assessment tasks; providing enough time for assessment; scoring students' performance; and providing students with feedback. These five headings correspond to the sub-questions for the first main research question.

4.1.1 Preparing Students for Assessment

To determine the extent to which EFL teachers are preparing their students for the assessment of speaking skills, items one through five elicited the regularity of using five essential preparation practices. Table 4.1 shows the frequencies that represent teachers' responses towards the items of the questionnaire in addition to the mean and standard deviation for utilizing these practices.

Table 4.1

Preparing students for the assessment of EFL speaking skills

Item	N	R	O	F	A	Mean	S.D	Mean	SD
1	3	11	31	57	18	3.64	.940		
2	1	9	38	54	18	3.66	.862		
3	6	13	22	46	33	3.73	1.125	3.33	1.087
4	15	29	45	26	5	2.79	1.024		
5	13	31	45	24	7	2.82	1.033		

The table shows that the five practices focusing on the preparation of students for the speaking skills assessment has the average mean value of 3.33, which indicates the regularity of using these practices. The mean value for the preparation practice of sharing the scoring rubrics with students was the highest in value (M=3.73). Data analysis reveal that using similar language and similar task format were the second and third most common practices that teachers utilize to prepare students for the speaking assessment.

On the other hand, the mean for preparing students via self-assessment and peer-assessment (M=2.79 and M=2.82 respectively) showed that these two practices

were utilized less frequently with greater disparity than other preparation activities. More than one third of participating teachers declared that they either rarely or never use these tools in getting students ready for the assessment.

4.1.2 Differentiating Assessment Tasks

To answer the first research question in terms of differentiating the assessment tasks, data from the three items (6, 8 & 10) and the multiple-answers item (18) were analyzed in tables 4.2 and 4.3 respectively.

Table 4.2
Differentiating the assessment tasks of EFL speaking skills

Item	N	R	O	F	A	Mean	SD	Mean	SD
6	11	36	35	30	8	2.90	1.088		
8	11	14	21	42	32	3.58	1.254	3.14	1.208
10	16	25	41	26	12	2.94	1.169		

*N = Never, R = Rarely, O = Occasionally, F = Frequently, A = Always, S.D. = Standard Deviation

Item 6 I consider students' views of how they want to be assessed
Item 8 I differentiate speaking assessment tasks according to students' abilities.
Item 10 Students are assessed by participating in paired and group discussions

The average mean value for item 8, (M=3.58), is relatively high which proves that differentiating speaking assessment tasks according to students' abilities is done more frequently than considering students' views or manipulating pair and group discussions in assessing the EFL speaking skills. However, the distribution of responses to items 6 and 10, marked by low mean values (M=2.90 and M=2.94 respectively), demonstrate the rarity of considering students' views or using paired or group discussions

in differentiating assessment tasks.

Item 18 sought to verify results concerning teachers' differentiation of assessment tasks. Table 4.3 illustrates the numbers and percentages of participants who announced their utilization of each assessment task.

Table 4.3
Speaking assessment tasks used by EFL teachers in secondary schools

Assessment Tasks	Number	Percentage
Memorize a model dialogue and demonstrate it	33	27 %
Answer a set of questions given before the assessment	54	44.3 %
A teacher - Student Interview	83	68 %
Reporting (giving an account of something seen, read, done or heard)	60	49.2 %
Describing (person, place, object, process, event, etc.)	88	72.1 %
Debating (arguing two sides of an issue)	64	52.5 %
Expressing an opinion or idea	89	73 %
Justifying something (defending a decision or an action)	50	41 %
Presentations and Speeches (improvised or rehearsed)	59	48.4 %
Role-playing (with no script)	28	23 %
Total	608	

Adding up the numbers of participants who selected each task amounts to 608 indicating that on average, every teacher had selected five assessment tasks out of the ten provided. This result confirms that speaking skills were being assessed by different assessment tasks. Three assessment tasks, namely expressing opinions, describing and

interviewing were commonly used by teachers as each task was chosen by more than two thirds of participants. Meanwhile, assessing students' speaking skills using role plays with no script was the least common assessment task.

4.1.3 Providing Enough Time for Assessment

Two items in the questionnaire were designed to provide insight into how EFL teachers in Qatari public secondary schools are caring for the assessment conditions by providing enough time for the assessment of speaking skills.

Table 4.4
Providing enough time for the assessment of speaking skills

Item	N	R	O	F	A	Mean	SD	Mean	SD
7	0	0	9	42	69	4.5	0.63		
9	0	0	18	34	68	4.4	0.74	4.45	0.712

* N = Never, R = Rarely, O = Occasionally, F = Frequently, A = Always, S.D. = Standard Deviation
Q 7   I tell my students when they will be assessed
Q 9   I give students enough time to complete the speaking assessment task

Table 4.4 highlights the high average mean values (M=4.5 and M=4.4) for items 7 and 8 which suggest that the majority of participating teachers were found to be providing enough time for the assessment of speaking skills, either through informing their students about the time they will be assessed or by giving ample time to complete their speaking assessment tasks.

4.1.4 Scoring Students' Performances

For investigating the techniques used for scoring the speaking performances, responses to item 11 through 15 were analysed in Table 4.5.

Table 4.5
Scoring Students' Performances

Item	N	R	O	F	A	Mean	SD	Mean	SD
11	34	20	37	26	5	2.57	1.228		
12	0	4	11	34	73	4.45	.787		
13	13	22	19	33	34	3.43	1.352	3.003	1.438
14	41	44	19	12	6	2.14	1.125		
15	35	38	23	15	11	2.42	1.275		

* N = Never, R = Rarely, O = Occasionally, F = Frequently, A = Always, S.D. = Standard Deviation
Item 11   I assess students' speaking skills in partnership with one or more colleagues
Item 12   I use a scoring rubric during the assessment tasks
Item 13   I use checklists to indicate whether or not certain elements are present.
Item 14   I write a narrative report of what is done.
Item 15   I observe and later I use the information from my memory to give a score

The relatively high average mean value for item 12 (M=4.45), is an indicator that participants were largely using a scoring rubric while assessing their students' speaking skills in secondary schools. Using checklists was the second common technique manipulated in scoring speaking performances as suggested by a mean value of 3.43. However, results for items 11 suggest that teachers were generally assessing their students' speaking skills as individual assessors rather than in panels. In addition, the distribution of responses to items 14 and 15 reveal that writing a narrative report and using memories of speaking performances were not regular activities in scoring speaking performances.

4.1.5 Providing Students with Feedback

The last investigation of EFL teachers' practices in assessing their students' speaking skills was the regularity of giving feedback to students. Table 4.6 shows descriptive statistics for items 16 and 17 in the questionnaire.

Table 4.6
Providing students with feedback

Item	N	R	O	F	A	Mean	SD	Mean	SD
16	51	31	19	14	5	2.09	1.195		
17	11	16	30	43	20	3.37	1.181	2.733	1.346

* N = Never, R = Rarely, O = Occasionally, F = Frequently, A = Always, S.D. = Standard Deviation

Item 16   I keep audio or video recordings of students' speaking responses

Item 17   I give students full-scale feedback on their performance in speaking assessment.

The low average mean value for item 16 (M=2.09) shows that the majority of participants are not keeping audio or video recordings of their students' speaking performances. However, descriptive statistics for item 17 (M=3.37) indicated that the majority of teachers give feedback to their students on their performance.

4.2 Assessment Challenges

This section analyzes responses to three items (19, 20 and 21) on the questionnaire in order to address the second research question that is concerned with the challenges that EFL teachers face in assessing their students' speaking skills in Qatari public secondary schools. First, participants selected the challenges they face from a list

in item 19. Then, they constructed their responses in items 20 and 21 to express their difficulties and recommendations for improving the assessment of speaking skills. Challenges were analysed under three categories: challenges related to teachers, challenges related to students and challenges related to the assessment conditions and available resources.

4.2.1 challenges related to teachers

The inadequacy of time and quality of speaking practice came on top of all recommendations given in item 21 in the questionnaire. More than 50% of participants magnified the importance of giving students more opportunities to practice speaking skills by means of authentic tasks. They suggested the involvement of students in '*real-life'* speaking activities in which they can express their ideas spontaneously. They illustrated that these speaking activities need to assimilate the assessment tasks.

Table 4.7
Assessment challenges for EFL teachers in assessing speaking skills

Challenge	Frequency					
	Item 19	**%**	**Item 20**	**%**	**Item 21**	**%**
A. challenges related to teachers	127		65		118	
Inadequate preparation of students	-		31	26%	63	53%
Lack of relevant training workshops	55	46%	2	2%	27	23%
Limited ability to design assessment tasks	38	32%	3	3%	2	2%
Absence of standardization sessions	34	28%	2	2%	7	6%
Absence of panel assessment	-		11	9%	19	16%

Nearly half the participants considered the lack of relevant training workshops on assessing speaking skills as a challenge to conducting the assessment properly. This

was confirmed by around one third of participants who stated that teachers lack the ability to design valid tasks for assessing speaking skills. Moreover, a number of participants specified the areas where teachers need professional development such as designing rating scales and using them in scoring students' spoken performances. In addition, around 28% of participants highlighted the need for conducting standardization sessions which help in adjusting their use of rating scales while scoring. Therefore, the limited ability to design assessment tasks, rating scales and implementing them poses a common challenging factor for EFL teachers. Hence, 27 participants suggested providing relevant professional development workshops as a step towards improving the assessment of speaking skills. For improving the accuracy of scoring students' speaking performance, 19 participants posed solutions in item 21 by suggesting that assessment be conducted by a panel of assessors.

4.2.2 Challenges Related to Students

As illustrated in table 4.8, more than half the participants indicated that the low levels of students' language competency, motivation and confidence were behind having difficulty in assessing their speaking skills. The first challenge related to students was their language competency levels. More than a quarter of participants argue that students' level of speaking English is not up to the curriculum standards for the secondary stage. This inability to speak fluently in English, as they explained, originated from the way they were taught and assessed in previous stages. In addition, teachers stated that students have the wrong concept about speaking assessment that it is the activity of memorizing a monlogue and demonstrating it in front of the teacher.

Table 4.8
Challenges related to students

Challenge	Frequency in item 19	%	Frequency In item 20	%	Frequency In item 21	%
Basic competence of English	-	-	35	29%	56	47%
Low motivation	-	-	16	13%	3	3%
Anxiety and low confidence levels	-	-	12	10%	5	4%

Sixteen participants were quoted saying that students are not adequately motivated to achieve well in this kind of classroom assessment. Students were seen as lacking the interest to achieve high scores, or even to speak. One participant explained that they might care about marks, but they do not do the due diligence for improving their language skills. The talk about student motivation in item 20 as a difficulty was developed and supported in item 21 by calling for increasing students' motivation to perform better in the speaking assessment. Some participants' recommendations were signified by posing solutions for increasing students' motivation such as tolerating students' mistakes or reducing the marks allotted for the mastery of speaking skills.

Twelve participants attributed their difficulties in assessing secondary school students' speaking skills to the anxiety level caused by speaking in English. For these participants, the lack of confidence may lead young learners to be reluctant or even refuse to talk in front of others. Two participants blamed students' shyness for that silent attitude.

4.2.3 Challenges Related to the Assessment Conditions and Resources

Results reveal that 'time restriction' and 'large class size' were the two most frequently reported challenges for EFL teachers in assessing speaking skills. In item 19, more than two thirds of participants selected these two practical constraints, making them

the two most common challenges. In addition, around 50% of participants in item 21 required more time for speaking assessment to be accomplished successfully, and 21 participants pointed out that time restriction causes them a difficulty in assessing their students' speaking skills.

Table 4.9

Challenges related to assessment conditions and resources

Challenge	Frequency in item 19	%	Frequency In item 20	%	Frequency In item 21	%
Time restriction	85	71%	21	18%	58	48%
Large class size	82	68%	13	11%	17	14%
Inaccuracy of the used rubric	37	31%	29	24%	13	11%

There is an evidence that the rating scale which is used for scoring students' speaking performances is a possible difficulty. Approximately one third of participants selected the 'inaccuracy of the used rubric' as one of their challenges in item 19 and around 25% showed their concern about the rubric in item 20. 13 participants in item 21 recommended using a simpler, more realistic rubric or even designing their own rubrics. Three participants recommended that the assessment of EFL speaking skills should be formative and to be carried out regularly. In support of this view, two participants advised to increase the frequency of these tests to be monthly or even weekly rather than being four times a year. Two participants explained that speaking skills assessment is not recognized with equal importance as other summative paper - and-pencil format tests. They suggested increasing the marks awarded for the speaking test.

4.3 Conclusion

In this chapter, the research questions were addressed by analyzing the collected data. The findings demonstrated that EFL teachers in Qatari public secondary schools attended to most practices while assessing their students' speaking skills such as providing enough time for assessment. However, the regularity of performing some assessment practices proved that teachers were not consistently keen on applying them especially those related to scoring speaking performances. In addition, findings indicated that the most common challenges for teachers in assessing their students' EFL speaking skills were related to the lack of time and students' low proficiency of English. In the next chapter, the study findings will be discussed and viewed in relation to the literature on the topic of EFL speaking skills assessment.

CHAPTER 5: DISCUSSION

5.1 Introduction

In the current study, the researcher surveyed the assessment practices and challenges of EFL teachers in assessing their students' speaking skills in Qatari public secondary schools. This study used information collected via an internet-based questionnaire targeting the EFL teachers in Qatari public secondary schools. 120 teachers participated in the study by responding to the questionnaire. The data collected through the questionnaire was analyzed in the previous chapter. This chapter discusses the findings of the major themes of the study in light of the research questions and the related literature. After discussing the findings, the researcher presents the pedagogical implications of the research findings and offers suggestions for further research before concluding the chapter.

5.2 Discussion of the study findings

This section presents a discussion of the findings with the aim of understanding the status of EFL teachers' practices and challenges while assessing their students' speaking skills in Qatari public secondary schools. Each research questions will be discussed in light of the relevant literature related to language assessment.

RQ1. (What are the EFL teachers' current practices in assessing students' speaking skills in Qatari public secondary schools?)

It is definitely unmistakable, when it comes to the assessment of speaking skills, that teachers should stick to standardized practices in order to enhance the reliability of assessment results (Brown, 2004). However, it is not always easy to maintain the

status quo to the expected standards. Hence, this study intended to examine the extent to which teachers' assessment practices are in congruence with what the literature has recommended. These assessment practices are discussed under five main categories that mark out the different stages of speaking assessment.

RQ1a. (Do teachers prepare students for assessment?)

Results demonstrated that teachers were frequently preparing students for the assessment of speaking skills. One important assessment practice that teachers are supposed to adhere to in preparing their students for speaking assessments is to share the rating scale (rubric) with students (East, 2016). Descriptive statistics in this study showed that sharing the rubric with students was relatively the most repeated preparation practice among participants.

Preparing students by using language and format similar to that on the assessment tasks was reported as a relatively more frequent practice.

Meanwhile, essential preparation practices such as self-assessment and peer-assessment received the lowest average mean values. Using self-assessment increases students' motivation and goal orientation (Todd, 2002), while using peer-assessment incites higher order thinking skills and motivates learning (Angelo & Cross, 1993; Freeman, 1994). Therefore, preparation practices seem to be far from achieving the primary objectives of motivating learning, promoting speaking skills, alleviating learners' anxiety and familiarizing students with assessment tasks.

RQ1b. (Do teachers differentiate the assessment tasks?)

EFL teachers are supposed to differentiate the speaking assessment tasks in order

to cater for students' needs and language abilities. In this study, teachers were found to be differentiating assessment tasks for measuring speaking skills. Accommodating students' language abilities seemed to be the principal basis for differentiating speaking tasks as this practice showed a mean value of (M= 3.58) which means that it is a frequent practice for the majority of teachers. This result is in congruence with Simin's (2014) suggestion that EFL teachers should consider varying the difficulty of speaking assessment tasks according to different proficiency levels of students.

The study results highlighted that specific speaking assessment tasks were more common such as a teacher-student interview in which students express an opinion, describe a person, place, object, process or event. This provides evidence suggesting that EFL teachers in Qatari public secondary schools view linguistic content of responses as more important than the communicative strength, a conclusion previously reached by Grada (2014). Grada found that such views could negatively affect teachers' assessment practices. Students should be assessed for EFL speaking skills by taking part in genuine interactions with their peers instead of rehearsing an unreal conversation that lacks purpose (Brown, 2004). To gain authenticity, speaking assessment tasks should simulate the real interaction where communicative roles such as negotiating meaning or collaborating are elicited spontaneously (Van den Branden, 2006)

In this study, involving students in pair and group work for speaking assessment was found to be a rare practice. In addition, role playing with no script was the least used assessment task. This may be explained by participants' opinion of students' low proficiency of English, which makes it difficult for teachers to involve them in pair or group discussions. However, Sinwongsuwat's (2012) found that these non-scripted role

plays elicit natural and spontaneous interactional features of spoken responses and consequently motivate students to use authentic language in performing a pragmatic goal.

RQ1c. (Do teachers allocate adequate time for students to complete the assessment?)

East (2016) highlighted the necessity of informing students about the time they will be assessed in order for them to do their best and be well-prepared for the assessment. When asked about the regularity of providing students with enough time, results showed that it is common for EFL teachers to tell students about the time they will be assessed and grant them ample time to complete assessment tasks. Although these practices proved to be the most common and consistent among participants, the majority of participants gave recommendations for increasing the time for practice and for assessment. Time restriction was the most common challenge among teachers in conducting speaking assessment, a challenge previously discovered by Lee (2010).

RQ1d. What scoring techniques do teachers manipulate to score performances?

In this study, EFL teachers were found to be largely using rubrics as rating scales while assessing their students' speaking skills in secondary schools. Using checklists came second as a commonly used rating scale for scoring speaking performances. However, results showed that teachers were rarely applying assessment practices such as writing narrative reports and scoring speaking performances using memories of performances.

Research studies suggest that reliability could be promoted by assessing speaking performances by a panel of two assessors or more. Rater variability has been

repeatedly discussed and investigated in the literature of psychological and educational measurement (Engelhard, 1994; McNamara, 1996; Myford & Wolfe, 2004). However, teachers in this study were generally assessing their students' speaking skills as individual assessors rather than in panels. When the teacher is the only assessor, many factors could affect students' scores such as the assessor's degree of leniency and the extent of consistency with his own ratings or with other raters (Mullen, 1980).

RQ1e. (Do teachers provide students with feedback on their performance?)

Giving students descriptive and evaluative comments support them to be more accurate in light of an established criteria (Booth-Butterfield, 1989; Johnson & Johnson, 1993; Smith & King, 2004). Results suggest that most EFL teachers were not keeping audio or video recordings of their students' speaking performances, a practice which is beneficial in decreasing the discrepancies between the actual speaking performance and what students might perceive. Asserting the necessity of recording students during speaking assessment, Yoshida (2001) highlighted the importance of giving students access to their recordings in order to improve the accuracy of their self-assessment practices.

RQ2. (What are the challenges that EFL teachers face in assessing their students' speaking skills?)

RQ2a. Practicality Challenges

Results showed that the majority of EFL teachers in secondary schools are suffering from the insufficiency of the time they were allowed to use for assessing their students' EFL speaking skills, and this becomes worse with the large class sizes. In a

similar vein, Lee (2010) identified that the most frequent cause for teachers' concern was lack of time allotted for the speaking practice and assessment. Classes with large numbers of students increase the difficulty of providing individual students with adequate time to practice EFL speaking skills. Hence, the majority of participants in this study expressed their need for solving this issue in order to provide more speaking practice for their students and in order to better assess the speaking skills, a demand previously suggested by Lee (2010).

RQ2b. Lack of relevant training

A considerable number of participants expressed their limited ability in developing valid assessment tasks to measure speaking skills. This finding can be supported by another finding which is the lack of relevant training workshops on speaking assessment, a common challenge for nearly half the participants. Such a challenge could negatively affect the validity of speaking assessment since assessment tasks need to be well-constructed so as to measure what they are supposed to measure.

This key challenge means that EFL teachers did not receive adequate professional development which focuses on developing teachers' skills in a demanding area like the assessment of speaking skills. Fulcher (2003) recommends the devotion of considerable time and effort to assessor training to improve the degree of assessment validity and to improve the inter-rater reliability of EFL speaking skills assessment. Testing Fulcher's recommendation, Xi and Mollaun (2009) found that raters from India scored spoken performances of both Indian and non-Indian examinees more accurately and consistently after receiving quality training package.

There is an evidence that teachers suffer from the lack or absence of these

calibration sessions in which teachers reach an acceptable degree of agreement on the use of rating scales to score speaking performances. Cognitive and behavioral characteristics of assessors can have a considerable effect on score variability as different assessors may focus on different aspects of speaking performance (Su, 2014).

RQ2c. Students' low levels of motivation, confidence and language proficiency

Teachers revealed their concern about students' lack of motivation as a barrier to successful speaking assessment. Students' motivation to learn and score high marks affects the degree of interactiveness during an assessment task (Fulcher & Davidson, 2007). Students' motivation could be enhanced by using preparation practices as self-assessment and peer-assessment, which, as discussed above, build up learners' self-awareness and supports them in identifying the areas which need improvement.

Teachers considered that students' anxiety is one of the difficulties they face in assessing speaking skills. This lack of students' confidence during the assessment of speaking skills could be interpreted as a consequence of the inadequacy of teachers' preparation practices. Overcoming this challenge would have been more likely if teachers were using video-recordings of students' performances as a feedback tool (Parr & Cartwright-Hatton, 2009). They found that using video recordings of students' speaking assessment decreases anxiety by showing learners the gap between their real performance and the expected one.

The most common challenge for teachers in assessing speaking skills was the unexpectedly low levels of English proficiency showed by their students in secondary schools. Teachers believe that students have substantial difficulty in producing grammatically and phonologically accurate utterances. This language incompetence of

students' speaking skills, as some teachers perceive, makes it difficult for teachers to elicit speaking performance and rate it. Hence, teachers recommended giving more classroom practice to these unprepared students in order to enhance students' EFL speaking skills.

5.3 Pedagogical Implications

The results attained in this study would propose a number of implications that would be valuable for EFL teachers and decision makers in the ministry of education and higher education in Qatar. Few examples of assessment practices were found to be inadequate. This is not just de-motivating for students, but it is also affecting the reliability of the assessment results and it may result in a negative washback effect on teaching and learning. Guskey (2003) stated that teachers who assess efficiently, provide constructive feedback, and help students demonstrate their achievements are able to offer better teaching and promote learning.

What is needed to improve the quality of speaking skills assessment in Qatari public secondary schools? The ministry is recommended to consider the implications of research on language assessment in general and the assessment of speaking skills in particular to provide teachers with adequate guidelines and relative training workshops, and to handle teachers' challenges all over the stages of assessment. In short, the ministry needs to create a procedural framework for the assessment of EFL speaking skills which gives an accurate picture of standardized assessment activities, well-constructed and reliable assessment tasks, implementation timeline and a follow up plan. Standardizing practices seeks to ensure that every student has more or less the same experience because variations in the administration may affect their speaking

performance or gained scores.

5.4 Suggestions for Further Study

The present study reveals the importance of continuously examining the existent practices for language assessment in general and the EFL speaking skills in particular. The reason is that new trends and reform acts in education are constantly changing in alignment with modern philosophies and EFL teaching and learning approaches. Since assessment has significant washback effects on teaching and learning, researchers should focus on ensuring the validity and reliability of speaking skills assessment.

This study gives an overview of what EFL teachers in Qatari public secondary schools are doing and the challenges they meet while assessing their students' speaking skills. Results outlined the range of assessment elements and factors which need special attention and research in order to see the extent to which the EFL speaking assessment is deemed valid. Further research should be undertaken to probe for the washback effect of the current assessment practices on teaching and learning EFL speaking skills as this will trigger more empirical studies into key assessment areas such as teacher competences, teaching material, assessment system and methodology. In addition, research studies can be carried out to design or recommend the implementation of specific assessment tasks and procedural frameworks for the assessment of speaking skills.

From a methodological point of view, it seems to be a practical approach to use mixed methods in another survey study investigating speaking assessment practices. This design of investigation enables the researcher to better describe these practices.

The participants of the present study were EFL teachers in secondary schools; in order to have a more comprehensive view of assessment practices, it would be desirable

to widen the scope by involving EFL teachers in all grades with other stakeholders such as students, parents, supervisors and administrators in one study to explore their perceptions of, and practices in assessing EFL speaking skills.

Results of the questionnaire indicated that a noticeable proportion of participants were not adequately familiar with the curriculum standards of speaking skills. This point should not be overlooked as it is fundamental for all teaching and assessment activities. Research studies need to find better ways of familiarizing teachers with curriculum standards and the effect of low familiarity with curriculum standards on teaching and assessment.

5.5 Conclusion

In this study, the researcher surveyed the EFL teachers' practices and challenges in assessing their students' speaking skills in Qatari public secondary schools. T h e research evidence on assessment practices such as preparing students for assessment, differentiating assessment tasks, and providing students with feedback, proved to adhere to the standardization levels. However, many challenges have been reported w h i c h teachers face in the assessment of these language skills.

REFERENCES

Ammon, U. (Ed.). (2001). *The dominance of English as a language of science: Effects on other languages and language communities*. Berlin / Germany: Walter de Gruyter.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers*, San Francisco, CA: Jossey-Bass

Asmi, R. (2013). *Language in the Mirror: Language Ideologies, Schooling and Islam in Qatar*. Columbia University.

Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language testing*, *17*(1), 1-42.

Bingqing, F. (2009). Middle school English teachers perceptions of oracy and their practice of speaking assessment, *Teaching English in China-CELEA Journal, 32*(6), 27-39.

Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2003). *Formative and summative assessment: Can they serve learning together?* Paper presented at the Annual meeting of the American Educational Research Association, Chicago, IL*AERA*.

Booth-Butterfield, M. (1989). The interpretation of classroom performance feedback: An attributional approach. *Communication Education, 38*(2), 119–131.

Brewer, D. J., Augustine, C. H., Zellman, G. L., Ryan, G. W., Goldman, C. A., & Ryan, G. (2007). *Education for a New Era: Design and Implementation of Kâ€"12 Education Reform in Qatar*. Rand Corporation.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Boston,

MA: Allyn & Bacon.

Brown, D. H. & Abeywickrama, P. (2010). Language Assessment: Principles and
Classroom Practices (2nd Edition). London: Pearson Education ESL

Brown, G., & Yule, G. (1983). *Teaching the spoken language* (Vol. 2). Cambridge UK:
Cambridge University Press.

Bygate, M. (1999). Task as context for the framing, reframing and unframing of
language. *System, 27*(1), 33-48.

Cai, H. (2015). Weight-based classification of raters and rater cognition in an EFL
speaking test. *Language Assessment Quarterly, 12*(3), 262–282.

Chang, C-W. (2005). *Oral language assessment: Teachers' practices and beliefs in
Taiwan collegiate EFL classrooms with special reference to Nightingale.*
(Unpublished Dissertation Doctoral). University of Exeter, UK.

Chang, Y. J., Wu, C. T., & Ku, H. Y. (2005). The introduction of electronic portfolios to
teach and assess English as a foreign language in Taiwan. *TechTrends, 49*(1), 30-
35.

Chinda, B. (2009). Professional development in language testing and assessment: A case
study of supporting change in assessment practice in in-service EFL teachers in
Thailand (Unpublished Dissertation Doctoral). University of Nottingham, UK.

Chuang, Y. Y. (2007). *Speaking Assessment: A Study of how Language Teachers Test
Students' Foreign Language Oral Proficiency and Their Attitudes Toward
Teaching Speaking in the EFL Classroom* (Unpublished Dissertation Doctoral).
Texas A & M University-Kingsville, Kingsville, TX.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research Methods in Education, 7th*

edition. London: Routledge.

Crystal, D. (2012). *English as a global language*. Cambridge, UK: Cambridge University

Press.

Csépes, I. (2009). *Measuring oral proficiency through paired-task performance* (Vol.

14). NY: Peter Lang.

Dandonoli, P., & Henning, G. (1990). *An investigation of the construct validity of the*

ACTFL proficiency guidelines and oral interview procedure. Foreign Language

Annals, 23(1), 11-22.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based

assessment. *Tesol Quarterly, 43*(3), 393-415.

Dearden, J. (2014). *English as a medium of instruction–a growing global phenomenon*.

UK: British Council. Retrieved from:

https://www.teachingenglish.org.uk/sites/teacheng/files/E484%20EMI%20-

%20Cover%20option_3%20FINAL_Web.pdf

de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). The cornerstones of survey

research. In E. D. de Leeuw, J. J., Hox, & D. A. Dillman (Eds.), *The international*

handbook of survey methodology, (1-16). UK: London: .

East, M. (2016). *Assessing foreign language students' spoken proficiency: Stakeholder*

perspectives on assessment innovation. NY: Springer.

Engelhard, G. (1994). Examining rater errors in the assessment of written composition

with a many-faceted Rasch model. *Journal of Educational Measurement,*

31(2), 93-112

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling

and purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1-4.

Ezzy, D. (2002). *Qualitative analysis: Practice and innovation*. London: Routledge.

Fakis, A., Hilliam, R., Stoneley, H., & Townend, M. (2014). Quantitative analysis of qualitative information from interviews: A systematic literature review. *Journal of Mixed Methods Research*, *8*(2), 139-161.

Fulcher, G. (2003). *Testing second language speaking*. UK: Pearson Education.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London, NY: Routledge.

Galaczi, E. D. (2010). Face-to-face and computer-based assessment of speaking: Challenges and opportunities. In L. Araújo (Ed.), *Computer-based assessment of foreign language speaking skills*. (29-51). Luxembourg: Publications Office of the European Union

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. (Unpublished doctoral dissertation). University of Exeter, UK.

Green, A. (2013). *Exploring language assessment and testing: Language in action*. Routledge.

Guskey, T. R. (2003) How classroom assessments improve learning. *Educational Leadership. 60*(5), 6-11.

Hayati, A. M., & Askari, E. (2008). Testing oral language proficiency of university EFL students. *Asian EFL Journal, 29*, 26-37.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77* (1), 81–112.

Hongwen Cai (2012) Weighting Patterns and Rater Variability in an English as a foreign language speaking test. (Unpublished doctoral dissertation). University of California, Los Angeles.

Huang, H. T. D. (2016). Exploring strategy use in L2 speaking assessment. *System*, *63*, 13-27.

Hughes, R., & Reed, B. S. (2016). *Teaching and researching speaking*. UK: Taylor & Francis.

James, N., & Busher, H. (2007). Ethical issues in online educational research: protecting privacy, establishing authenticity in email interviewing. *International Journal of Research & Method in Education*, *30*(1), 101-113.

Johns, R. (2012). *Survey question bank: method fact sheet 1. Likert items and scales*. Retrieved from: http://surveynet.ac.uk/sqb/datacollection/likertfactsheet.pdf

Johnson, D. W., & Johnson, R. T. (1993). Cooperative learning and feedback in technology-based instruction. *Educational Technology Research and Development, 40*(3), 21-38.

Khamkhien, A. (2010). Teaching English speaking and English speaking tests in the Thai context: A reflection from Thai perspective. *English Language Teaching, 3*(1), 184-190.

Kim, H. J. (2011). *Investigating raters' development of rating ability on a second*

language speaking assessment. (Unpublished doctoral dissertation). Teachers

College, Columbia University, NY.

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), 1-30.

Lee, S. (2010). *Current practice of classroom speaking assessment in secondary schools*

in South Korea. (Unpublished MA thesis). The University of Queensland,

Australia.

Linn, R. L. (2008). *Measurement and assessment in teaching*. London, UK: Pearson

Education.

Louma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

Luke, K. K., & Pavlidou, T. (Eds.). (2002). *Telephone calls: Unity and diversity in*

conversational structure across languages and cultures. Amsterdam, The

Netherlands: John Benjamins Publishing.

Massialas, B. G., & Jarrar, S. A. (2016). *Arab education in transition: A source book*

(Vol. 2). London: Routledge.

MacDonald, S., & Headlam, N. (2008). *Research Methods Handbook: Introductory guide*

to research methods for social research. Manchester, UK: Centre for Local

Economic Strategies.

McNamara, T. F. (1996). *Measuring second language performance*. Boston, MA:

Addison Wesley Longman.

Milanovic, M. & Saville, N. (1996). Performance testing, cognition and

assessment. *Selected papers from the 15th language testing research colloquium*

Cambridge and Amhem, Cambridge: Cambridge University Press.

Myford, C.M. and Wolfe, E.W. (2004). Detecting and measuring rater effects using many facet Rasch measurement: Part 2. *Journal of Applied Measurement,* 5(2), 189–227.

Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2017). Exploring the Use of Video-Conferencing Technology in the Assessment of Spoken Language: A Mixed-Methods Study. *Language Assessment Quarterly*, *14*(1), 1-18.

Neuendorf, K. A. (2016). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Nunan, D. (2003). The impact of English as a global language on educational policies and practices in the Asia Pacific region. *TESOL quarterly*, *37*(4), 589-613.

Omran, S. A. R. A. (2005). English language examination system in general secondary education certificate in Qatar: An evaluative study. (Unpublished doctoral dissertation). Sudan University of Science and Technology, Khartoum, Sudan.

Oppenheim, A. N. (2000). *Questionnaire design, interviewing and attitude measurement.* UK: Bloomsbury Publishing.

O'Sullivan, B., Weir, C. & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing, 19(1), 33-56.*

Öztekin, E. (2011). *A comparison of computer assisted and face-to-face speaking assessment: Performance, perceptions, anxiety, and computer attitudes* (Unpublished doctoral dissertation). Bilkent University: Ankara, Turkey.

Parr, C. J., & Cartwright-Hatton, S. (2009). Social anxiety in adolescents: the effect of video feedback on anxiety and the self-evaluation of performance. *Clinical psychology & psychotherapy*, *16*(1), 46-54.

Pawlak, M., & Waniek-Klimczak, E. (Eds.). (2014). *Issues in teaching, learning and testing speaking in a second language*. NY: Springer.

Pennycook, A. (2017). *The cultural politics of English as an international language*. UK: Taylor & Francis.

Phillipson, R. (2003). English-only Europe?: Challenging language policy. London: Routledge.

Qatari Ministry of Education and Higher Education (n.d.1). *Curriculum standards*. Retrieved from:

http://www.edu.gov.qa/Ar/SECInstitutes/EducationInstitute/CS/English/Pages/CS.aspx

Qatari Ministry of Education and Higher Education. Professional Development Office. (n.d.2)

http://www.edu.gov.qa/En/SECInstitutes/EducationInstitute/Offices/Pages/ProfessionalDevelopmentOffice.aspx

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Developments in applied statistics*, *19*(1), 160-117.

Rahman, M. F., Babu, R. & Ashrafuzzaman, M. (2011). Assessment and Feedback Practices in the English Language Classroom, *Journal of NELTA, 16*(1-2), 97-106.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching.* Cambridge: Cambridge University Press.

Rodríguez, B. L., & Guiberson, M. (2011). Using a teacher rating scale of language and

literacy skills with preschool children of English-speaking, Spanish-speaking, and bilingual backgrounds. *Early Childhood Education Journal*, *39*(5), 303.

Sahlberg, P. (2011). *Finnish lessons*. NY: Teachers College Press.

Sawaki, Y. (2007). Construct validation of analytic rating scales in a speaking assessment: Reporting a score profile and a composite. *Language Testing*, *24*(3), 355-390.

Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing, 11*(2), 99–123.

Shropshire, K. O., Hawdon, J. E., & Witte, J. C. (2009). Web survey design: Balancing measurement, response, and topical interest. *Sociological Methods & Research*, *37*(3), 344-370.

Simin, S., & Tavakoli, M. (2014). Assessing speaking ability in academic context: Focusing on a mixed methods approach. *International Journal of Research Studies in Language Learning, 4*(1), 71-82.

Sinwongsuwat, K. (2012). Rethinking assessment of Thai EFL learners' speaking skills. *Language Testing in Asia*, *2*(4), 75.

Skehan, P. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.

Smith, C. D., & King, P. E. (2004). Student feedback sensitivity and the efficacy of feedback interventions in public speaking performance improvement. *Communication Education*, *53*(3), 203-216.

Sook, K. H. (2003). The types of speaking assessment tasks used by Korean Junior

    Secondary school English teachers. *Asian EFL Journal*. Retrieved from

    http://www.asian-efl-journal.com/dec_03_gl.pdf.

Su, Y. W. (2014). *The impact of rater characteristics on oral assessments of second*

    language proficiency. (Unpublished dissertation). Austin, Texas: University of

    Texas at Austin.

Thornbury, S., & Watkins, P. (2007). *The CELTA course: Trainee book*. Cambridge, UK:

    Cambridge University Press

Todd, R. W. (2002). Using self-assessment for evaluation. *English Teaching Forum,*

    40(1), 16-19.

Tollefson, J. W., & Tsui, A. B. (Eds.). (2003). *Medium of instruction policies: Which*

    agenda? Whose agenda? London: Routledge.

Torky, S. (2006). The effectiveness of a taskbased instruction program in developing the

    English language speaking skills of secondary stage students. (Unpublished PhD.

    Dissertation) Women's College, Ain Shams University, Cario, Egypt

Van den Branden, K. (Ed.). (2006). *Task-based language education: From theory to*

    practice. Ernst Klett Sprachen.

Weber, A. S. (2015). Linking education to creating a knowledge society: Qatar's

    investment in the education sector. In *Information Resources Management*

    Association (IRMA) (Eds.), STEM education: Concepts, methodologies, tools,

    and applications (pp. 818−839). Hershey, Pennsylvania: Business Science

    Reference, IGI Global. Emerald Group Publishing Limited.

Wigglesworth, G. (2008). Task and performance based assessment. In *Encyclopedia of*

language and education (pp. 2251-2262). NY: Springer.

Weir, C. J. (2005). Limitations of the common European framework for developing

    comparable examinations and tests, *Language Testing, 22*(3), 281-300.

Wiles, R., Crow, G., Heath, S., & Charles, V. (2008). The management of confidentiality

    and anonymity in social research. *International Journal of Social Research*

    Methodology, *11*(5), 417-428.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning

    progression. *Journal of Research in Science Teaching*, *46*(6), 716-730.

Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we

    heading? *Language Testing, 27*, 291–300.

Xi, X., Mollaun, P. (2009). How do raters from India perform in scoring the TOEFL

    iBT™ speaking section and what kind of training helps? (TOEFL iBT Research

    Report No. 11, RR-09-31). Princeton, NJ: Educational Testing Service.

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative

    inquiry of a Chinese college EFL teacher's experience. *Tesol Quarterly*, *43*(3),

    492-513.

Yoshida, Y. (2001). Authentic Progress Assessment of Oral Language: Oral Portfolios.

Zechner, K., Evanini, K., Yoon, S. Y., Davis, L., Wang, X., Chen, L. & Leong, C. W.

    (2014). *Automated scoring of speaking items in an assessment for teachers of*

    English as a foreign language. Proceedings of the ninth workshop on Innovative

    Use of NLP for Building Educational Applications (pp. 134–142). Baltimore,

    MD: Association for Computational Linguistics.

Zhao, Z. (2013). An overview of models of speaking performance and its implications for

the development of procedural framework for diagnostic speaking tests.

International Education Studies, *6*(3), 66-75.

APPENDIX A: THE DATA COLLECTION INSTRUMENT (QUESTIONNAIRE)

Assessment of Speaking Skills: Teachers' Practices and Challenges

About the Questionnaire

- It asks about practices and challenges in assessing students' speaking skills in secondary school
- It should take about 10 – 15 minutes to complete.
- Participation in this project is voluntary.
- You may withdraw from the study at any time without penalty or harm of any type.
- Your confidentiality and anonymity as a participant in this study will remain secure.
- It has been reviewed and approved by the Institutional Review Board in Qatar University.
- You may request a copy of any publications arising from the work.
- When in doubt about any aspect of the questionnaire, or if you would like more information about it or the study, you may contact the researcher by phone at the following numbers: [+974 6685 9939]

Thank you for your cooperation

* Required

1.
I agree to participate in the above mentioned research project conducted by Taha Osman under the supervision of Dr. Areej Barham from Qatar University.
Mark only one oval.

( ) Yes

( ) No

Background Information

2.
Gender *
Mark only one oval.

( ) Male

( ) Female

3.

Highest level of formal education *

Mark only one oval.

◯ Bachelor

◯ Post-graduate Diploma

◯ Master Degree

◯ PhD Degree

◯ Other: _____

4.

How long have you been working as a teacher? *

Mark only one oval.

◯ 1 - 3 years

◯ 4 - 9 years

◯ 10 - 14 years

◯ 15 0r more years

5.

What grade level(s) do you teach? *

Mark only one oval.

◯ Grade 10

◯ Grade 11

◯ Grade 12

◯ Mixed Grades

6.

To what extent are you familiar with the Qatari curriculum standards for speaking?
*

Mark only one oval.

◯ Not At All Familiar

◯ Slightly Familiar

◯ Somewhat Familiar

◯ Moderately Familiar

◯ Extremely Familiar

Teachers' Practices & Challenges

Please indicate how often do you do the following for speaking test preparation.

7.
1. I have students practice using the same format as the speaking assessment tasks. *

Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

8.
2. I have students practice using language similar to that on the assessment tasks. *

Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

9.
3. I share the scoring rubrics with my students. *

Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

10.
4. I have students practice using self-assessment. *

Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

11.
5. I have students practice using peer-assessment. *
Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

12.
6. I consider students' views of how they want to be assessed. *
Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

13.
7. I tell my students when they will be assessed. *
Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

14.
8. I differentiate speaking assessment tasks according to students' abilities. *
Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

15.

9. I give students enough time to complete the speaking assessment tasks. *

Mark only one oval.

◯ Never

◯ Rarely

◯ Occasionally

◯ Frequently

◯ Always

16.

10. Students are assessed by participating in paired and group discussions. *

Mark only one oval.

◯ Never

◯ Rarely

◯ Occasionally

◯ Frequently

◯ Always

17.

11. I assess students' speaking skills in partnership with one or more colleagues. *

Mark only one oval.

◯ Never

◯ Rarely

◯ Occasionally

◯ Frequently

◯ Always

18.

12. I use a scoring rubric during the assessment tasks. *

Mark only one oval.

◯ Never

◯ Rarely

◯ Occasionally

◯ Frequently

◯ Always

72

19.

13. I use checklists to indicate whether or not certain elements are present. *

Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

20.

14. I write a narrative report of what is done. *

Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

21.

15. I observe and later I use the information from my memory to give a score. *

Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

22.

16. I keep audio or video recordings of students' speaking responses. *

Mark only one oval.

- ( ) Never
- ( ) Rarely
- ( ) Occasionally
- ( ) Frequently
- ( ) Always

23.

17. I give students full-scale feedback on their performance in speaking assessment. *

Mark only one oval.

( ) Never

( ) Rarely

( ) Occasionally

( ) Frequently

( ) Always

24.

18. Which of the following tasks have you used While assessing students' speaking skills? Check all that apply *

Check all that apply.

[ ] Memorise a model dialogue and demonstrate it

[ ] Answer a set of questions given before the assessment

[ ] A teacher - Student Interview

[ ] Reporting (giving an account of something seen, read, done or heard)

[ ] Describing (person, place, object, process, event, etc.)

[ ] Debating (arguing two sides of an issue)

[ ] Expressing an opinion or idea

[ ] Justifying something (defending a decision or an action)

[ ] Presentations and Speeches (improvised or rehearsed)

[ ] Role-playing (with no script)

[ ] Other: _____

25.

19. What challenges do you face in assessing students' speaking skills? Check all that apply *

Check all that apply.

[ ] Time restriction

[ ] Lack of required resources

[ ] Large class size

[ ] Inaccuracy of the used rubric

[ ] Limited ability to design valid and reliable assessment tasks

[ ] Lack of relevant training workshops on assessing speaking skills

[ ] Absence of standardization sesions to adjust the use of the rubric

[ ] Other: _____

Final Comments

Kindly mention at least one point for the following items.

26.

20. Please list and explain the difficulties you usually find in conducting speaking assessments. *

_____

_____

_____

_____

_____

27.

21. What are your recommendations for improving the speaking skills assessment
*

_____

_____

_____

_____

_____

Powered by

Google Forms

التاريخ : 2017/1/16 م

تسهيل مهمة القائم بالبحث الميداني في المدارس

السيد : مدير إدارة التوجيه التربوي   المحترم

السلام عليكم ورحمة الله وبركاته

نود إحاطتكم علما بأن الباحث / الباحثون المذكورة أسماءهم أدناه ، بصدد إجراء دراسة ميدانية

في مدرستكم وبياناتهم كالتالي :

- **اسم الباحث :** طه عثمان مديولي عثمان عثمان
- **جهة البحث :** كلية التربية . جامعة قطر
- **عنوان البحث :** ممارسات و تحديات المعلمين في تقييم مهارات التحدث باللغة الانجليزية لطلاب المرحلة الثانوية
- **هدف البحث :** تحديد الممارسات التي يقوم بها معلمي اللغة الإنجليزية بالمرحلة الثانوية في دولة قطر عند تقييم مهارات التحدث للطلاب. و التعرف على الطرق المستخدمة لتقييم مهارات التحدث و أشكال التغذية الراجعة التي يقدمها المعلمين لطلابهم بعد تقييم مهارات التحدث لديهم. و تحديد التحديات التي يجدونها وكيفية قيام معلمي اللغة الإنجليزية بالمرحلة الثانوية بتحسين ممارساتهم في تقييم مهارات التحدث لدى الطلاب.
- **عينة البحث :** معلمي اللغة الإنجليزية بالمرحلة الثانوية
- **التاريخ :** 16 / 1 / 2017

عليه ، يرجى التكرم بتسهيل مهمة الباحث ، علما بأن البيانات ستكون سرية ولأغراض البحث العلمي..

مع شكرنا لحسن تعاونكم معنا ،،،

د. عزيزة أحمد السعدي

مدير إدارة السياسات والأبحاث التربوية

التاريخ : 16 / 1 / 2017 م

Qatar University Institutional Review Board
QU-IRB

جامعة قطر
QATAR UNIVERSITY

January 23, 2017

Mr. Taha Osman M. Osman
Graduate Student Project
College of Education
Qatar University
Tel.: 66859939
Email: 201505937@student.qu.edu.qa, tahatom@hotmail.com

Dear Mr. Taha Osman,

Sub.:   Research Ethics Review Exemption / CEDU Graduate Student Project
Ref.:   Project titled, "Assessment of English Speaking Skills in Qatari Secondary Schools: Teachers' Practices & Challenges"

We would like to inform you that your application along with the supporting documents provided for the above proposal, is reviewed and having met all the requirements, has been exempted from the full ethics review.

Please note that any changes/modification or additions to the original submitted protocol should be reported to the committee to seek approval prior to continuation.

Your Research Ethics Approval No. is: **QU-IRB 728-E/17**

Kindly refer to this number in all your future correspondence pertaining to this project.

Best wishes,

Dr. Khalid Al-Ali
Chairperson, QU-IRB

Institutional Review Board
(IRB)
Office Of Academic Research

Qatar University-Institutional Review Board (QU-IRB), P.O. Box 2713 Doha, Qatar
Tel +974 4403-5307 (GMT +3hrs) email: QU-IRB@qu.edu.qa

APPENDIX D: REQUEST E-MAIL FOR PARTICIPATION IN THE STUDY

Dear English teachers in Secondary Schools,

You are kindly requested to respond to a questionnaire

for the sake of completing a master degree thesis entitled,

<u>"EFL Speaking Skills Assessment in Qatari Public Secondary Schools:</u>

<u>Teachers' Practices and Challenges"</u>

You can reach the questionnaire through the following link:

https://goo.gl/forms/hkFwagZSsUITfKUJ2

Please be informed that

- This questionnaire asks about practices and challenges in assessing students' speaking skills in secondary school

- It should take about 10 - 15 minutes to complete.

- Participation in this project is voluntary.

- You may withdraw from the study at any time without penalty or harm of any type.

- Your confidentiality and anonymity as a participant in this study will remain secure.

- It has been reviewed and approved by the Institutional Review Board in Qatar University.

- You may request a copy of any publications arising from the work.

- When in doubt about any aspect of the questionnaire, or if you would like more information about it or the study,

you may contact the researcher by phone at the following numbers: [+974 6685 9939]

Thank you for your cooperation

Kind Regards

Taha Osman

Cell Phone: 00974 6685 9939

E-mail: tahatom@hotmail.com