

An Investigation of Jordanian English as a Foreign
Language (EFL) Teachers' Perceptions of and Practices in
Classroom Assessment

by

Malak Turki Ali Swaie

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

Master of Arts

in

Applied Linguistics and Discourse Studies

Carleton University

Ottawa, Ontario

© 2019, Malak Turki Ali Swaie

Abstract

This study investigated the purposes and methods of classroom-based assessment (CBA) used by Jordanian English as a Foreign Language (EFL) teachers in secondary schools and explored their perceptions about assessment for learning. Data were collected through an online questionnaire (n=54) and semi-structured interviews (n=3). The study found that 1) teachers used classroom assessment to achieve various purposes, primarily those related to students' learning; and 2) although teachers used a variety of methods to assess students in reading, writing, listening and speaking, tests were the most frequently used methods. The findings suggest that although teachers were favorably inclined towards assessment for learning, their perceptions might not be aligned with their reported assessment practices. It is hoped that the findings from this study will help raise awareness among Jordanian EFL teachers and educational policy makers about the importance of CBA and the need to develop teacher preparation and professional development programs.

This thesis is dictated to my parents, Aishah and Turki Swaie.

*Very special thanks and lifelong gratitude to you for your love, support and prayers
which have sustained me throughout my life.*

Acknowledgements

I would like to acknowledge and express my sincerest gratitude to all those who supported me in completing this master's thesis. Without each one of you, this endeavor would not have been possible.

First, I acknowledge with profound gratitude the support and guidance of my supervisor, Dr. Janna Fox throughout the development of the current thesis from the beginning. I would like to thank my thesis committee members, Dr. Trudy O'Brien and Dr. Angel Arias for their valuable comments and helpful suggestions.

I am also deeply grateful for my colleague, Christina Dore for her valuable comments and insights that have contributed greatly to the improvement of the thesis. I would like also to thank all the teachers in Jordan who participated in this study, without your input this study would not have happened.

Finally, I would like to acknowledge the constant love and support of my husband, Mu'ath Algazo and thank him for his encouragement and patience throughout this journey of completing my master's studies. To my lovely daughter, Salma for being the source of hope and motivation to work harder. My thanks are extended to my siblings, Ali, Noor, Farah and Areen for their love and support. Many thanks also go to my parents-in-law, Maryam and Abdullah Algazo, and my brother in-law, Dr. Mohammad and sister in-law, Mai for their encouragement and support.

Table of Contents

Abstract.....	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	vii
List of Figures.....	viii
Chapter 1 Introduction.....	1
Context of the Study.....	3
Purpose of the Study	6
Research Questions	6
Significance of the Study	6
Organization of the Thesis	7
Chapter 2 Literature Review	8
Defining Assessment and Testing.....	8
A Brief History of Second Language Testing and Assessment	9
Washback of Testing.....	14
Assessment and Second Language Learning Approaches	18
Traditional/behaviorist approach.....	19
Cognitivist approach.....	20
Constructivist approach	22
Assessment Shift	23
Classroom-Based Assessment.....	25
Teachers' Assessment Practices.....	29
Teacher Beliefs and Assessment Practices.....	33
Chapter 3 Methodology	38

Research Design.....	38
Data Collection.....	40
Recruitment	40
Participants	40
Instruments and procedures.....	43
Data Analysis	45
Chapter 4 Findings and Discussion.....	47
Purposes of Assessment (Research Question 1)	47
Results of Phase 1: Questionnaire	47
Results of Phase 2: Interviews.....	53
Discussion of Research Question 1	54
Methods of Assessment (Research Question 2).....	58
Results of Phase 1: Questionnaire.	58
Results of Phase 2: Interviews.....	72
Discussion of Research Question 2	74
Factors influencing teachers' choices of assessment.....	83
Teacher Perceptions about Assessment for Learning (Research Question 3).....	91
Chapter 5 Conclusion	107
Summary of Findings	107
Limitations of the Study	110
Recommendations for Further Research	111
Implications of the Study	112
References	114
Appendix A.....	134
Appendix B	135
Appendix C	136

List of Tables

Table 1. Questionnaire Respondents' Demographic Information	40
Table 2. Interview Participants' Demographic Information	41
Table 3. Questionnaire Results of Assessment Purposes.....	48
Table 4. Questionnaire Results for Methods of Reading Assessment	58
Table 5. Questionnaire Results for Methods of Writing Assessment	61
Table 6. Questionnaire Results for Methods of Listening and Speaking Assessment	66
Table 7. Questionnaire Results for Teachers' Perceptions About Assessment for Learning	89
Table 8. Frequency of Using Formative Purposes	93
Table 9. Frequency of Using Peer and Self-Assessment	94
Table 10. Form of Feedback Provided to Students During the Course	96

List of Figures

Figure 1. Frequency of Assessment Purposes	49
Figure 2. Reading Assessment Methods (teacher-made tests)	59
Figure 3. Reading Assessment Methods (student-conducted)	59
Figure 4. Writing Assessment Methods (teacher-made tests)	63
Figure 5. Writing Assessment Methods (student-conducted)	63
Figure 6. Listening and Speaking Assessment Methods (student-conducted)	67
Figure 7. Teachers' Perception about Assessment for Learning	91

Chapter 1 Introduction

The second half of the twentieth century showed an increased focus on teaching English as a second/foreign language (ESL/EFL) with the rapid spread of English in many parts of the world. This growth led to enormous development in the field of second language (L2) pedagogy, which had only recently moved away from behaviorist models of L2 acquisition (i.e., language learning as a process of habit formation through imitation, practice, and reinforcement) and towards theories of language learning based on interaction and negotiation with others. Thus, traditional pedagogical approaches to language learning—i.e., where language acquisition is perceived to occur by means of drilling, repetition, and memorization—were deemed inadequate to equip learners with the skills they need to be able to use the language efficiently in meaningful communication (Nunan, 1999). Instead, there was a shift to communicative approaches that enhance students' communicative skills through involvement in meaningful interactions using the target language. Such approaches consider language as a resource for exchanging meaning both in and out of the classroom rather than language as solely a classroom phenomenon.

These shifts in second language pedagogy impacted the field of language testing and assessment, which has likewise witnessed major changes in recent years. These changes are characterized by a shifted emphasis: from exclusively tests and testing to assessment in language classrooms in broader terms (Cheng & Fox, 2017). The process of monitoring and evaluating students' language development now incorporates measures of what students know and are able to do with language, rather than focusing on tests scores for isolated language elements, such as vocabulary and grammar.

In the classroom setting, assessment has been long used for evaluating students' learning and recording their achievement at the end of the learning process, referred to as *summative* assessment. Recently, there have been calls to use assessment throughout the learning process, referred as *formative* assessment, to inform teachers' decisions about what students have learned and how to move forward to achieve their learning goals (Black & Wiliam, 1998). More recently, emphasis has been placed on the use of formative assessment practices to support students' learning - students' self-awareness, autonomy, and goal setting (Fox, 2014). This coincides with the worldwide movement to adopt a balanced concept of assessment, which combines both types of assessment (Cheng, 2011) and has changed the teacher's role from being a judge of students' learning achievements to that of a facilitator of student learning (Rea-Dickens 2004; Teasdale & Leung, 2000). Thus, there has been growing interest in exploring teachers' practices regarding classroom-based assessment (CBA), which is the assessment of student learning in the classroom setting.

Nevertheless, the study of assessment practices used by ESL/EFL teachers is relatively limited compared to the much larger body of research on language testing practices (Cheng, Rogers, & Wang, 2008). A number of studies have explored teachers' assessment practices in different ESL/EFL contexts (e.g., Cheng et al., 2008; Cheng, Rogers & Hu, 2004; Cumming, 2001; Davison, 2004; Wang, 2017). Little research has been conducted on teachers' beliefs and perceptions regarding classroom assessment especially those related to formative assessment. Nevertheless, most studies have focused on exploring either aspect; practices or beliefs. Recently, researchers have begun to address ESL/EFL teachers' beliefs and perceptions about assessment and their relation to

teachers' assessment practices (e.g., Giraldo, 2018; Muñoz, Palacio, & Escobar, 2012; Yin, 2010).

However, this is a fairly recent development and has not extended to many ESL/EFL contexts. For instance, in the Jordanian context, there have only been a handful of studies about EFL teachers' classroom assessment practices. Some studies have investigated the washback (i.e., influence of testing on teaching and learning) of the National Examination on teachers' assessment practices (e.g., Al-Jamal & Ghadi, 2008), while others have examined the effectiveness of various types of classroom assessment—e.g., self-assessment (Baniabdelrahman, 2010) and portfolio-based assessment (Obeiah & Bataineh, 2016). No study has yet explored Jordanian EFL teachers' purposes and methods of CBA, or their perceptions about formative assessment in the EFL classroom. The present study therefore explores Jordanian EFL teachers' perceptions about assessment and CBA practices in the context of secondary EFL classes.

Context of the Study

Jordan is a small country situated in the heart of the Middle East. The number of primary and secondary schools reached 7,272 in 2017, with 1,992,481 students enrolled in private and public schools (Queen Rania Center, 2018). There were 126,262 qualified teachers (i.e., have at least a Bachelor's degree in an English language-related major); more than 20% of these also have advanced degrees, e.g., Diploma, Master's, or PhD (Al Sawalmeh, 2014).

The educational system in Jordan consists of two main stages: *primary* comprises Grades 1-10 and *secondary* comprises Grades 11 and 12. After completing ten years of compulsory primary education, students can enroll in an optional two-year study

program, either academic or vocational, which ends with a General Secondary Education Certificate Examination, or *Tawjihi* (MoE, 2010). The Tawjihi is a mandatory educational requirement for all students who plan to attend a Jordanian university and may determine what subjects they can study at university. The Tawjihi is also required for those who expect to work in public or private sectors in Jordan, as scores can be used in hiring processes. Due to its impact in the short term (graduation, entry into university) and long term (future career prospects), the Tawjihi is considered a *high-stakes examination*.

Regarding English language education, prior to 2000, English was taught starting in Grade 5 in all public schools in Jordan. However, following the 2000 educational reform, English became a primary subject in all public schools from Kindergarten to Grade 12 (Al-Khatib, 2008). To teach English, the Ministry of Education (MoE) requires teachers to have at least a Bachelor's degree in English or English language and literature (Algazo, 2018). Notably, graduates of English language and literature are often hired to teach English, even though these program lack courses about teaching methodologies (Alhabahba, Pandian, & Mahfoodh, 2016). The result of this hiring practice is a gap in teacher preparation. In response, in 2012, the MoE launched a mandatory training course to prepare all new teachers in public schools in all subject areas. This four-week course includes topics on teaching methodologies, classroom instruction planning, assessment and evaluation, as well as methodologies for teaching specific subjects, such as Science and Mathematics (MoE, 2012).

Despite this, the short-term course was not adequate to equip new teachers with the knowledge and skills they needed to perform their tasks effectively in the classroom.

Thus, in 2014 the MoE launched a nine-month program in partnership with Queen Rania Teacher Academy (QRTA) to better equip pre-service teachers with the theoretical and practical knowledge they need to teach professionally in public schools. The QRTA is a non-profit organization which provides teachers with training courses in order to enhance teaching and learning in Jordanian schools. The program, called the Teacher Education Professional Diploma (TEPD), offers two tracks—one for Grades 1-3 and one for Grades 4-12—for four subjects: Mathematics, Science, Arabic, and English (QRTA, 2015).

The program was originally designed to fulfill the needs of the MoE to recruit new qualified teachers in different school districts. The program is either free of cost or paid, depending on the applicant. A limited number of candidates are invited to enrol in the program and granted a full scholarship based on their high scores on the admission exams (written and oral). These candidates will be offered full-time jobs in public schools after graduation (QRTA, 2018). Other applicants who passed the exams but with lower scores can still enroll in the program but they have to pay the expensive tuition fee and have no job guarantees.

Overall, the TEPD program is likely an improvement to past programs since it covers a wide range of teaching-related topics over a longer duration and allows pre-service teachers to apply what they have learned during the program in a real classroom under the supervision of professional teachers. Nonetheless, the expensive tuition fees and the limited number of scholarships means that not all pre-service teachers can benefit from this program. Therefore, it is assumed that teachers in this study may lack the assessment knowledge required to conduct classroom assessment that follows the current trend and combines both summative and formative assessment.

Purpose of the Study

The purpose of the current study was two-fold: (1) to explore CBA purposes and methods reported by Jordanian EFL teachers in secondary classrooms and understand what influences their assessment choices regarding the methods used to assess students' competence in English language; (2) to examine teachers' perceptions and attitudes towards Assessment for Learning (AoL)/formative assessment in the language classroom and whether their perceptions are in line with their reported assessment practices.

Research Questions

Drawing on teachers' accounts of their assessment practices, the primary research questions driving this inquiry are:

- 1- For what purposes do EFL teachers in secondary schools in Jordan conduct CBA?
- 2- What methods do EFL teachers in secondary schools in Jordan use to assess students in reading, writing, listening, and speaking? And what influences their choices of assessment?
- 3- How do EFL teachers in secondary schools in Jordan account for their beliefs about Assessment for Learning (AfL)/formative assessment? Are their reported beliefs aligned with their assessment practices? What are the implications?

Significance of the Study

This research aims to contribute to further understanding the realities of CBA in the Jordanian EFL context. It may also provide evidence that can contribute to improving teacher preparation and professional development programs in Jordan to enhance teachers' expertise in teaching and assessment. This contribution may be especially valuable because in Jordan, like many countries where teachers are required to develop

their own assessments, initial teacher training programs may not include sufficient courses in assessment in general and/or test design, analysis, or evaluation (Cheng & Fox, 2017). This research may provide teachers who participated in the study with the opportunity to reflect on their teaching practices, particularly why they do what they do when assessing their students. This study may also help raise awareness of this issue among other teachers and EFL policy makers in Jordan and in the wider EFL context.

Organization of the Thesis

This thesis is organized into five chapters. Chapter 1 introduced the rationale, aims, and research questions guiding this study. Chapter 2 reviews previous research on teachers' assessment practices and the role of beliefs in shaping these practices. Chapter 3 describes the study design, including the participants, instruments, and procedures used for qualitative and quantitative data collection, as well as the types of data analysis that was undertaken. Chapter 4 presents and discusses the results of the study. Chapter 5 summarizes the main findings, outlines the limitations of the study and possible pedagogical implications, and proposes directions for future research.

Chapter 2 Literature Review

This chapter reviews the literature on CBA and teachers' assessment practices. The review begins with defining assessment and testing and providing an overview of the emergence and development of testing and assessment in the field of second language teaching. The next section sheds light on the influence of testing on teaching and learning (i.e., *washback*). The review then discusses the main approaches to second language learning and how these approaches led to major changes in language assessment. The rest of the review is devoted to the literature about focus of this study, CBA in ESL/EFL classrooms. This section overviews previous research on teachers' assessment practices and their perceptions about classroom assessment.

Defining Assessment and Testing

In the literature, the terms *testing* and *assessment* are sometimes used interchangeably; however, assessment is “an umbrella term encompassing measurement instruments administered on a “one-off” basis such as tests, as well as qualitative methods of monitoring and recording student learning such as observation, simulations, or project work” (Coombe, 2017, p. 2). A test is therefore a form of assessment and can be defined as “a method of measuring a person's ability, knowledge, or performance in a given domain” in order to determine what that person knows or has learned at particular point of time (Brown, 2004, p. 3). Assessment, meanwhile, is defined as “a systematic approach to collecting information and making inferences about the ability of a student or the quality or success of a teaching course on the basis of various sources of evidence” (Richards & Schmidt, 2010, p. 35-36).

In this study, *assessment* is used to refer to “all those activities undertaken by teachers—and by their students in assessing themselves—that provide information to be used as feedback to modify teaching and learning activities” (Black & William, 1998, p. 2). Thus, assessment includes classroom tests and teachers’ daily assessment practices, and large-scale testing such as the Test of English as a Foreign Language (TOEFL; Cheng & Fox, 2017). This definition indicates that assessment is an interactive process that involves both teachers and students, where the ultimate goal of assessment is not only to make judgments about students’ performance, but also to support their learning.

A Brief History of Second Language Testing and Assessment

Language testing and assessment has emerged as a relatively new subfield within applied linguistics. The field has its roots in classical and modern test theory in education (e.g., Bachman, 1990; Linn, 2010) and the communicative theory of language teaching and learning (e.g., Canale & Swain, 1980; Savignon, 2003). The emergence of this subfield can be traced to 1961, when Robert Lado published his seminal book *Language Testing* (Tsayari & Banerjee, 2016). At that time, language testing was influenced by theories that viewed language ability as a set of skills (writing, reading, listening, and speaking) and components (e.g., grammar, vocabulary, and pronunciation) that were taught as discrete elements (Lado, 1961) using methods such as grammar-translation, the direct method, and audiolingualism (Bachman, 2000). Accordingly, most language proficiency tests, including large-scale tests, were structured and designed in the way that focused on testing learners’ knowledge of discrete elements of language in isolation (e.g., vocabulary and grammar tests). Tests made use of multiple choice, true-false, and other such question structures that focused on single independent items; for example, testing

the conjugation of the verb *to have* in one item, and the definition of *anger* in the next (Shohamy, 1996).

In the late 1970s, the above-described discrete-point approach to language testing (i.e., testing isolated and discrete language components) was criticized for failing to produce useful information about learners' performance because tests were designed to assess only one attribute of a particular skill (e.g., comprehension of reading material) (Gipps, 2002). The complaint was that items in such tests had little relation to authentic language use (Brindley, 2001; Cziko, 1981). For this reason, there was a shift towards integrative tests (e.g. cloze and dictation), which were based on the unitary trait hypothesis of language proficiency that views language competence as an indivisible and unified set of interacting abilities that cannot be tested separately (Carroll, 1972). These tests were claimed to be direct measures of overall language proficiency, as they required the learner to use the knowledge they have about various language components, such as knowledge of vocabulary, grammatical structure, discourse structure, and reading skills and strategies (Cziko, 1981; Oller, 1979). Thus, the tests included comprehension of whole texts rather than merely isolated language elements; for example, in cloze tests, learners fill in blanks missing words in paragraphs taken from longer texts (Shohamy, 1996). However, the integrative approach was criticized for its restricted view of how language works because it mainly measures linguistic competence in artificial tasks, which determines what the learner knows about the language rather than how to use it (Tzagari & Banerjee, 2016).

By the 1980s, proponents of the theory of *communicative competence* (e.g., Canale & Swain, 1981; Savignon, 1983; Widdowson, 1983) adopted a more

comprehensive approach to language testing inspired by Hymes' (1972) theory of language competence as knowledge and ability (Bachman, 2000). The communicative approach involved testing learners' performance on the three main components of the communicative competence model: grammatical competence (e.g., morphology, syntax, sentence-grammar semantics, and phonology), sociolinguistic competence (e.g., sociocultural appropriateness and rules of discourse), and strategic competence (e.g., verbal and non-verbal communication strategies) (Canale & Swain, 1980). Tests were designed to reflect actual demonstration of the three components of language knowledge in real second language situations and for authentic communication purposes (Canale & Swain, 1980).

The theory of communicative competence further evolved in the years that followed, including the development of several models of communicative competence (e.g., Bachman, 1990; Bachman & Palmer, 1996, 2010; Celce-Murcia, 2007; Celce-Murcia, Dornyei, & Thurrell 1995). Most notable in terms of its effect on testing was Bachman's (1990) multi-componential model of communicative language ability (CLA) and its upgraded version (Bachman & Palmer, 1996, 2010), which provided a comprehensive theoretical framework for second language testing research and practice (Bachman & Cohen, 1998; Purpura, 2004). According to this model, language ability was conceived as "consisting of both knowledge, or competence, and the capacity for implementing or executing that competence in appropriate, contextualised communicative language use" (Bachman, 1990, p. 84). This framework took into consideration how individual attributes interact with one another and with the characteristics of the target language use contexts (Liao, 2009). This and other models

communicative competence highlighted the importance of considering the sociolinguistic aspects of language and adjusting language testing to its social context through consideration of the domain to which it applies (Coombe, Davidson, O'Sullivan, & Stoyhoff, 2012).

The notion of *language proficiency* evolved to refer to “language knowledge needed for a specific future situation” (Shohamy, 1996, p. 145), which was different from the previous emphasis on *achievement*, i.e., a learner’s knowledge of language learned during a course. In communicative language testing, the tests were designed to simulate real-life situations where the learner performs in actual communicative settings using authentic language, such as being engaged in a dialogue to test oral language. However, since communicative testing required including all the principles of the communicative competence model in the constructs of language testing and the performance of test takers, there were challenges regarding the authenticity of tasks and the genuineness of texts (Brown, 2004).

Since then, *performance testing* has emerged as a new approach to language testing (Shohamy, 1996). A performance test is “a test in which the ability of candidates to perform particular tasks, usually associated with job or study requirements, is assessed” (Davies et al., 1999, p. 144). This approach is closely related to the communicative approach in the sense that they both focus on the authentic use of language. However, performance testing is concerned with measuring learners’ productive language skills while they are performing actual or simulated real-world tasks in which the language should be similar as possible to the language that would be used in non-testing situations (Bachman, 1990; Bailey, 1985). Test tasks are designed based on

what target language tasks learners would expect to perform in future situations, such as job-related tasks like filling in forms, answering phones, and writing letters. An example of a language performance test is the English for Special Purpose (ESP) test, where learners are required to use the target language in academic situations in order to cope with their study-related demands (Shohamy, 1996).

Since real-life language use is unpredictable and depends highly on context, testing that uses a specific task or procedure may not reveal all facets of a learner's language knowledge that enable them to use the language effectively (Wigglesworth & Frost, 2017). Thus, there is a need to include various types of measures that can help reveal more information about what learners know and can do using the target language in order to assess their proficiency. As Shohamy (1996) argued, "there are different types of language knowledge and mastering one type is no guarantee for mastering another... different instruments are capable of seeing different things" (p. 152). In response, there have been calls to focus on the achievement component of language knowledge because testing learners' proficiency using structured tasks that may not replicate real-life tasks cannot predict learners' ability to use the target language effectively in future situations (Lynch & Davidson, 1994; Shohamy, 1992).

Recently, *alternative assessments* have grown in popularity in the language testing and assessment domain. Alternative assessments refer to using multiple assessment procedures, such as portfolios, peer- and self-assessment, and observation to gather evidence about students' knowledge of the target language on various components and in different situations (Fox, 2017). Unlike standardized tests, alternative assessments employ various procedures that collect information about students' language knowledge,

where each procedure adds information that cannot be obtained using each procedure alone (Shohamy, 1996). Moreover, some researchers have argued that traditional standardized testing is a tool of power used to influence and manipulate educational systems in which students have to respond to external tests while having no say about the content and goals of these tests (Shohamy, 2001). Instead, alternative assessment has “the potential for the sharing of power and for the valuing of the individual, because it allow[s] for a more collaborative, dialogic interchange between the assessor and the assessed” (Fox, 2017, p. 140). This approach gives both teachers and students the opportunity to engage in the assessment process where students can discuss their learning goals and their progress towards achieving these goals.

Washback of Testing

In the field of education, it is widely accepted that there is a relationship between testing, teaching, and learning, where testing can shape teaching and the way in which knowledge is presented (Gipps, 2002). Large-scale and high-stakes tests can have a significant impact on what teachers do in the classroom since the results of these tests are seen by teachers, students, administrators, parents, and sometimes even the general public in order to make important decisions (Madaus, 1989). Consequently, teachers are more likely to “teach to the test” by focusing on, for example, the test’s items and format, and adjusting their instruction according to the form of the questions on the high-stakes test (Madaus, 1989, p. 95). This phenomenon is known as *measurement-driven instruction* (Popham, 1987), where high-stakes tests influence the instructional programme that prepares students for such tests. This influence is considered to narrow the curriculum to

include only the knowledge and skills assessed by the tests in a phenomenon known as *test-curriculum alignment* (Shepard, 1990).

In language testing, the term *washback* has been used to refer to how testing influences language teaching and learning at the classroom level (Tsagari & Cheng, 2017). Messick (1996) defines washback as “the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning” (p. 241). Studies on washback have investigated the effects of testing on different aspects of teaching, including the content of teaching, teaching methods, and assessment methods used by teachers (Cheng, 2008). Findings have revealed that while tests can have a direct effect on the content of teaching, the effect on teaching methodology may not be explained in relation to the test content (Alderson & Hamp-Lyons, 1996; Green, 2006). In other words, these studies suggest that tests can influence *what* teachers teach but they have little influence on *how* they teach because other factors—such as teacher’s beliefs, academic background, and previous experiences—play a more central role in shaping teacher’s practices (Watanabe, 1996).

Yet, other studies have found that the teaching approach used by teachers might also be influenced by external testing. For example, in the Jordanian context, Al-Jamal and Ghadi (2008) found that EFL teachers of Grade 12 tended to modify their instruction according to the content and the objectives of the National Exam (Tawjihi) so they can better prepare their students for the exam. Particularly, teachers were more likely to use traditional teaching approaches that focus on vocabulary and grammar rather than using communicative approaches that aim to improve students’ communicative abilities across the four skills (reading, writing, listening, and speaking). The reason for such practices is

that the Tawjihi exam only assesses students' competence in reading and writing and excludes listening and speaking skills.

The extent to which a test influences teaching can also be attributed to other factors, such as the nature of the test (low- or high-stakes) and the uses for the test scores (Shohamy, Donitsa-Schmidt, & Ferman, 1996). For example, the washback of a large-scale high-stakes test, the International English Language Testing System (IELTS), is evident in the instruction of IELTS preparation classes, where teaching focuses mainly on activities to practice the test and exercises that focus on the skills that will be tested (Mickan & Motteram, 2008). In contrast, in ESL/EFL classes not aimed at high-stakes test preparation, the teaching content and instruction may not be exclusive to the content of the course's tests, but rather focus on a wider range of general language skills (Alderson & Hamp-Lyons, 1996; Hayes & Read 2004).

Like the IELTS, high-stakes tests such as national examinations like the Tawjihi can have a similar influence on teaching. For example, studies on the washback of national EFL examinations for higher education admission (e.g., Azadi & Gholami, 2013; Manjarrés, 2005) have revealed that teachers tend to focus only on teaching and practicing the aspects of language assessed in the test (grammar and vocabulary) while ignoring other important aspects of language competence such as speaking and listening skills that students need in learning a second language. Similarly, Wall and Alderson's (1993) study on the washback of an English examination in Sri Lanka on language teaching revealed that teachers consistently skipped over listening lessons in their textbooks because listening was not part of the exam, yet other teachers taught listening

with no intention to improve students' listening skills but rather for the reason that listening exercises might be an item type that might appear in the reading test.

Teaching to the test may come as a result of the pressure on teachers to help their students to achieve their primary learning goal, which is passing the test (Harmer, 2010). Teachers' fear poor results—and the associated feeling of guilt, shame, or embarrassment—might lead them to wish for their students “to achieve high scores in whatever way seems possible” (Alderson & Wall, 1993, p. 118). Thus, teachers may select the teaching materials and design assessment tools that focus on the content of the test while ignoring the overall course learning outcomes (Shohamy et al., 1996; Wall, 2005).

Of course, students' learning is affected by this approach. As Shohamy (2006) states:

Centralized tests are capable of dictating to teachers what to teach and what test-takers will study, as teachers, and test-takers, comply with the demands of the tests by changing their behaviour so as to maximize the scores, given the detrimental effects of the tests. (p. 104)

The impact of testing on students may be seen in their attitudes to focus on practicing the tests while giving more attention to the knowledge and skills assessed in these tests. This might also have an influence on students' motivation to improve their overall proficiency especially in areas that are not assessed in the tests.

Is it worth mentioning that in the language testing literature, washback is a neutral term since it can be used to refer to positive or negative effects on teaching and learning. In

other words, any test can have either negative or positive washback “to the extent that it promotes or impedes the accomplishment of educational goals held by learners and/or program personnel” (Bailey, 1996, p. 268). Positive washback may occur when there is a match between what students are taught and the content of the test (Bachman, 1990). Qi (2005) adds that high-stakes tests can have a greater potential to achieve positive washback if learning goals are taken into consideration by policy-makers and test designers.

Therefore, when the testing process is clearly designed to achieve positive effects, tests can play an important role in enhancing learning. For example, Hirai and Koizumi (2009) explored the impact of using a practical speaking test as a part of classroom assessment in order to help Japanese university students improve their speaking skills. The study found that students’ perceptions about the test were positive in the sense that the test had a positive impact on their learning and motivation to enhance their speaking ability.

Assessment and Second Language Learning Approaches

It is widely accepted that testing and assessment influence teaching. Assessment and testing are also influenced by current learning theories. Different approaches to second language teaching, such as the grammar-translation method, the audio-lingual method, and communicative language teaching imply different theories of language and language learning, philosophies, learning skills, and methodologies, all of which impact how tests and assessments are designed (Spolsky, 1985). The following sections discuss shifts in the second language acquisition domain by focusing on three major approaches in second language teaching and learning.

Traditional/behaviorist approach. The *traditional approach* to language learning was developed during the early decades of the twentieth century after the emergence of the behaviorist model of language teaching. According to this model, learning and language development occur through habit formation and practice, where the “learning of complex competencies could be broken down into discrete skills learnt separately, through developing individual stimulus-response bonds” (Gipps, 2002, p. 18). From the behaviorist view, methods such as repetition, memorization of dialogues, and various forms of guided speaking and writing practice help form habits that can later be maintained by reinforcement and feedback (Richards, 2006). Moreover, learning is seen as a linear and sequential process, and complex understandings cannot occur unless simple structures and forms of the target language are first mastered (Mitchell & Myles, 2004).

Accordingly, tests were designed to assess language components and skills in isolation in the way that reflected students’ mastering of these skills. The assumption was that including “various recall, completion, matching, and multiple-choice test types, along with some essay questions, fit closely with what was deemed important to learn” (Shepard, 2000, p. 5). To be in line with the behaviorist model, language proficiency test items should cover all four skills and as many discrete linguistic points as possible so that the test reflects students’ competence in the target language based on their performance (Brown, 2004).

The traditional approach to language assessment was criticized, however, because tests that consist of identifying the correct answer in multiple-choice and true/false questions only assesses students’ ability to recognize the right answer, not their ability to

produce language (Popham, 2003). Therefore, relying on behaviorist theories of language learning in designing language tests was ultimately considered inadequate to collect information about students' ability to use the target language.

Cognitivist approach. The *cognitive approach* emerged in response to criticisms of behaviorist theories that saw language learning as a process of habit formation. The cognitive model highlighted the role of the brain in storing, processing, and analyzing information learned in the second language (Lightbown & Spada, 2014). Cognitivists rejected the behaviorist model, which fell short of explaining second language acquisition, and argued that learning factual knowledge about the target language and practicing this knowledge through drills activities cannot help students go forward and develop the ability to produce the language.

According to the cognitive model, language acquisition is a build-up of knowledge about language that can be stored, processed, and eventually retrieved automatically for speaking and understanding through processing (Mitchell & Myles, 2004). Shepard (2000) explains that learning is a complex process that cannot be explained through mere habit formation and knowledge accumulation but rather through using existing knowledge structures in higher levels of thinking and solving problems while developing self-monitoring and awareness about when and how to use skills. This view of learning draws attention to the importance of context in learning. Unlike the behaviorist view, which holds that knowledge can be learned in isolation and then used in any context, the cognitive approach indicates that there is a strong relationship between knowledge/skills and the context of use (Gipps, 2002). For example, having good

knowledge of grammar and vocabulary would be useless unless the learner knows how to employ this knowledge to produce the language in writing and speaking.

The development of cognitive theories to explain language learning also influenced assessment methods and practices. Instead of assessing language components in isolation as per the traditional method, cognitivism-informed tests were designed to assess students' ability to use the language in tasks that were considered to be similar to real language use. This goal was achieved by providing contextual information about the language structures that were being tested, often in the form of *integrative tests*, i.e., tests that aim to measure overall proficiency by integrating various components of language knowledge and skills. Common examples of integrative tests include *dictation* and *cloze* tests. In dictation tests, learners write what they hear while listening to a passage read aloud by administrator or audiotaped. Dictation is an integrative test, because it involves listening skills, writing skills, and recognition of specific language items such as grammar and vocabulary. A cloze test is "a reading passage (perhaps 150 to 300 words) in which roughly every sixth-seventh word has been deleted; test taker is required to supply words that fit into those blanks" (Brown, 2004, p. 8). Such tests were believed to reveal useful information about learners' ability to use the language since using appropriate words in blanks requires a number of abilities including knowledge of vocabulary, grammatical structure, discourse structure, reading skills and strategies, and an internalized "expectancy" grammar (i.e., ability to predict the word that comes next in a sequence; Brown, 2004). Thus, learners had to know not only the language structures (e.g., grammar rules) but also how to use it in the given context.

The cognitive approach puts a heavy emphasis on the active role of the language learner in the assessment and learning process, unlike earlier behaviourism-informed approaches that viewed the learner as a passive absorber of knowledge provided by the teacher.

Constructivist approach. The basic premise behind the *constructivist approach* is that learning is an active process of knowledge construction within a shared social context and the learner plays an active role in this process through analyzing new knowledge and mapping it to their previous knowledge (Fox, 2001). In this sense, language is learned through building associations of language patterns in the brain and learners develop their language as a result of “the billions of associations which are made during language use, and which lead to regular patterns that might look rule-like, but in fact are merely associations.” (Mitchell & Myles, 2004, p. 98). In order to build these associations, learners need to construct meaning and build their own theories through taking part in their learning with the help of the teacher whose role it is to facilitate and monitor the learning process. An illustration of a constructivist-informed classroom teaching approach is communicative language teaching, where students construct and communicate meaning by engaging in interactive tasks.

The emergence of cognitive and constructivist theories marked a shift from traditional approaches that focused on memorization and repetition to more interactive approaches that take into consideration the importance of communicative interaction. Communicative language teaching evolved as a response to this shift and views language as a means of communication rather than as a system of rules. According to this theory, learning a language successfully requires the learner to know not only the forms and

structures of the language but also the functions and purposes of the language in different communicative settings (Lightbown & Spada, 2014). In particular, communicative competence involves: grammatical competence (i.e., knowledge of linguistic structures); sociolinguistic competence (i.e., knowledge of the rules of language use in different settings); and strategic competence (i.e., knowledge of strategies, such as gestures, that can maintain communication; Canale & Swain, 1980).

Assessment Shift

In the wider educational domain, the use of traditional *psychometric testing* has dominated the field for most of the twentieth century (Gipps, 2002). Psychometric testing makes inferences about learners' learning based on their performance on *standardized tests* (i.e., tests administered to large populations of students and scored in the way that allows for performance comparison). A psychometric model is rooted in psychological theories of learning, such as traditional instructional theory that “assumes that knowledge and skill can be analyzed into component parts that function in the same way no matter where they are used” (Resnick, 1989, p. 3). According to this model, curriculum is considered to be a body of knowledge that needs to be transmitted to the students, and the role of assessment is to check if this knowledge has been received.

The assumption about traditional testing is that “one can specify and measure all important learning objectives, and furthermore that mastery on the test items implies mastery of the intended skills and concepts” (Gipps, 2002, p. 20). However, with the emergence of cognitive and constructivist theories of learning, the concept of learning has evolved as a process of knowledge construction. Knowledge is now seen as a cohesive and holistic scaffolding means for the learner to use for later learning, where the

learner plays an active role in it through interaction (Gipps, 2002). Thus, teaching students isolated facts and discrete skills without connecting this knowledge to meaning and uses makes it difficult and even impossible for students to make use of this knowledge and move to the next level of understanding.

As a result of these developments, there has been a corresponding shift in the field of language assessment towards using multiple assessment measures to measure students' learning rather than relying only on external tests (Huerta-Macias, 1995). As a result, there has been a shift from a *testing culture* to an *assessment culture*. In a testing culture, learning is an accumulation of knowledge that can be measured using tests as the only tool to judge students' learning. In an assessment culture, learning is viewed as skills acquisition and needs multiple sources of evidence that can reflect students' ongoing growth and track their performance over time (Perkins, 2004; Wolf, Bixby, Glenn, & Gardner, 1991).

Birenbaum and Dochy (1996) describes the shift in culture:

A strong emphasis is now put on integrating assessment and instruction, on assessing process rather than just products and on evaluating individual progress relative to each student's starting point. The position of the student with regard to the evaluation process has also been changing from that of a passive, powerless, often oppressed subject who is mystified by the process, to that of an active participant who shares responsibility in the process, practices self-evaluation, reflection and collaboration and conducts a continuous dialogue with the teacher.

(p. xiii)

This change in the assessment view over the past two decades has led to a growing interest in examining teachers' assessment practices in the classroom setting in terms of their forms and functions. The following sections discuss classroom assessment and teachers' practices in the language classroom.

Classroom-Based Assessment

The role of assessment in the language classroom has attracted great attention in the language learning and teaching literature. Classroom-based assessment (CBA) is defined as “[a]ny reflection by teachers (and/or learners) on the qualities of a learner’s (or group of learners’) work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialisation purposes” (Hill & McNamara, 2012, p. 397). It is an ongoing process in which a set of procedures, techniques, and strategies are used by both teachers and students to collect, evaluate, and report student achievement (McMillan, 2015).

CBA involves both of two forms of assessment: summative and formative. *Summative assessment* refers to the process of collecting and interpreting evidence to make judgments about students' learning. This type of assessment takes place at the end of instruction (i.e., the end of instructional unit, programme, or school year) in the form of final exam or projects to evaluate students' learning and academic achievement. The results of assessment take the form of grades, scores, and ranks that are used to make statements about a student's learning status at a particular point in time (Brown, 2003; Popham, 2003). Since assessment is used in this sense for the purpose of grading and reporting students' achievement and progress, it is also known as *assessment of learning*.

On the other hand, *formative assessment*, also known as *assessment for learning* (Gipps, 1994; Stiggins, 2002, 2007), refers to the process of gathering evidence in a continuous and systematic way to be used by students and their teachers to determine the level of students' learning and to modify instruction in order to better meet students' needs to help them achieve their learning goals (Black & Wiliam, 1998; Popham, 2003). According to Black and Wiliam (2009), formative assessment involves eliciting and interpreting evidence about students' progress and used by teachers, learners, or their peers, to make decisions about future lesson plans.

As stated by Broadfoot et al. (1991), some of the principles that guide assessment for learning include:

- It is embedded in a view of teaching and learning of which it is an essential part.
- It involves sharing learning goals with pupils.
- It aims to help students to know and to recognize the standards they are aiming for.
- It involves students in self-assessment.
- It provides feedback, which leads to students recognizing their next steps and how to take them.
- It is underpinned by the confidence that every student can improve.
- It involves both teacher and students reviewing and reflecting on assessment data. (p. 7)

Thus, based on these principles, teachers should (1) understand and articulate in advance the achievement targets of the course, (2) inform students about those learning goals, (3) use assessment to inform and adjust their teaching practices to better support students' learning, (4) involve students in the assessment process through self-assessment, and (5) provide students with feedback to give them insight into how to improve (Stiggins, 2002).

In sum, assessment for learning employs a learner-centred approach in which learners play an active role in their learning. This type of assessment has the potential to enhance students' learning through providing feedback for both teachers and students that guide them toward achieving learning goals (Black & Wiliam, 1998; Rea-Dickins, 2001). CBA is therefore primarily concerned with monitoring students' needs and creating opportunities to adjust instruction to better meet students' needs to achieve their learning goals using both forms of assessment, summative and formative.

In CBA, assessment can be used to achieve various purposes. Cheng and Fox (2017) present three common purposes of assessment in the language classroom:

- *Instructional*: to gather information about students' understanding and skills so that teachers can plan and adjust their instruction to support students' learning. For example, asking students if they understood the point being presented in order to check their understanding and to adjust instruction accordingly.
- *Student-centred*: to achieve purposes related to identifying students' strengths and weaknesses in language knowledge or use (i.e., *diagnostic assessment*), where that information can be used to inform decisions about the focus of further

instruction. For example, using a formal written test or informal teacher questioning to obtain feedback that helps adapt course instruction based on the results/feedback (Black & Wiliam, 1998).

- *Administrative*: to assign values to students' output at the end of the instruction using a numerical (e.g., 88) or ordinal (e.g., A) scale to formally report students' achievements to parents or principals. For example, preparing a formal report of students' learning achievement at the end of the academic year. This purpose is more associated with summative assessment.

The above purposes play vital roles in the language classroom to enhance and support students' learning. For example, using diagnostic assessment can provide information for both teachers and students and help teachers to direct their instruction towards achieving students' needs which has a positive impact on students' language development (Fox, 2009). Moreover, this type of assessment can also improve teaching and support teacher change especially when it is used as on-going process during the curriculum renewal and innovations (Colby-Kelly & Turner, 2007).

With regards to assessment methods, CBA can be conducted using various types of methods to gather information about students' learning and achievements. In the literature, assessment methods are categorized in different ways. According to Cheng and Fox (2017), one way is to categorize methods according to the type of items that make up the assessment, such as open-ended questions or supply-type questions where students are asked to give detailed answers (e.g., short-answer questions, oral presentation). Another way is to categorize the methods as observations, conversations and product. Observations involve teachers observing students as they are performing tasks, such as

group discussions or rehearsals. Conversations include teacher-student interviews and conferences. Product methods include tests, presentations, and performances. A third way is to categorize the methods as teacher-made tests (i.e., designed and administered by the teacher), students-conducted (i.e., require students' involvement in the process), and standardized testing (Cheng et al., 2008). These methods can be used to assess students in reading, writing, listening and speaking through assessing each skill independently or in combination with another skill. For example, some argue that listening should be assessed as a separate skill (e.g., Vandergrift, 2006) while others (e.g., Brown, 2004) see that listening and speaking can be assessed together because listening is often implied as a component of speaking, as the learner needs to listen to be able to respond orally.

In closing, identifying and understanding the purposes of assessment is important for conducting assessment in the classroom, as the purpose influences teacher decision-making regarding how often they use assessment and what methods they need to use in order to achieve the purposes of assessment. The following section discusses teachers' assessment practices and sheds lights on the factors that influence their decisions about assessment in the language classroom.

Teachers' Assessment Practices

In assessment, teachers continuously make decisions throughout the process of gathering, interpreting, and evaluating evidence, and then taking actions based on the results. Rea-Dickens (2001) identifies four stages in the assessment process that require teacher decision-making: (a) planning (i.e., purpose, how to assess, preparation), (b) implementation (i.e., introducing the assessment, scaffolding, self- and peer-monitoring, immediate feedback), (c) monitoring (i.e., recording evidence of achievement,

interpreting evidence, revising, and delayed feedback), and (d) recording and disseminating formal results (i.e., preparing formal records of students' achievement).

Brookhart (2004) noted that assessment decision-making is complex due to the wide variety of purposes that assessment of student learning can fulfill. This complexity may be related to the various factors that influence teachers' decision making regarding assessment including the use of summative and formative assessment. One of the factors that was found to have a prominent influence on choosing assessment methods is the purpose of assessment. In a comparative interview study of three ESL/EFL university contexts in Canada, Hong Kong, and China, Cheng et al. (2008) found that purpose of assessment was a significant factor which influences how ESL/EFL instructors choose methods when assessing their students. For example, instructors in China reported that they used standardized tests as a main method of assessment because of the need to prepare their students for the College English Test they will need to take in the future. Similarly, other studies (e.g., Andrews, Fullilove, & Wong, 2002; Cheng, 2004; Qi, 2004) found that teachers in Hong Kong and China were more likely to follow the assessment formats used in external testing, especially toward the end of the term.

Teachers' knowledge of assessment and teaching experience were also found to play a role in how teachers use assessment in their classrooms. Studies (e.g., Cheng et al.; 2004; Cheng et al., 2008) found that ESL/EFL instructors who have more years of teaching experience and are better trained in assessment appear to be less reliant on published material when designing their assessment methods. On the other hand, teachers who do not have sufficient knowledge of assessment may rely more on traditional

methods of assessment (e.g., tests and quizzes) and disregard alternative methods (e.g., self- and peer assessment).

Tsagari and Vogt (2017) explored the assessment knowledge of EFL teachers in seven European countries and found that teachers tended to use traditional summative methods, mainly paper-and-pencil tests in their classrooms, more than alternative or formative methods. This practice was attributed to the lack of knowledge on assessment in the pre-service teacher preparation programs and professional training programs. Similarly, in the Australian context, Breen et al. (1997) explored how ESL teachers in different primary schools used assessment in their classrooms. The study found that teachers were using assessment frameworks (e.g., First Steps Developmental Continua in Western Australia, ESL Scales, and ESL Bandscales) that they had been familiarized with during their training and professional development.

Another factor that can play a role in how teachers assess their students is the context of the instruction. In a study on writing assessment, Cumming (2001) interviewed 48 highly experienced ESL/EFL teachers and found that teachers' conceptualizations of student assessment varied depending on whether the courses were for general English or English for specific purposes (ESP). Although, the focus of these courses was to teach students English writing skills, the teachers varied in their selection of assessment tasks and evaluation criteria. For example, teachers who taught ESP courses (e.g., ESL writing for engineering students) used a limited range of assessment tasks and had specific criteria for evaluation. On the other hand, in general English courses (e.g., immigrant settlement programs), teachers tended to use a variety of methods and broad-base evaluation criteria (Cumming, 2001).

Another aspect of the influence of the context is demonstrated in contexts where external testing is commonly used. In such contexts, teachers are more likely to adapt their assessment and design their tests using the same types of questions used in the external exams. Studies on ESL/EFL contexts in Hong Kong and China (e.g., Andrews et al., 2002; Cheng, 2004; Qi, 2004) found that external testing often influenced how ESL/EFL teachers assess their students. Teachers were more likely to follow the assessment formats used in external testing, especially toward the end of the term. These practices can be noticed more clearly with higher-stakes tests, where teachers used external testing formats more frequently (Cheng, 1999).

Another external contextual influence could be district policies that may require teachers to use certain procedures to report students' evaluation. For example, in Breen et al.'s (1997) study, teachers assessed their students using the mandated frameworks that their schools required. McMillan (2003) investigated classroom assessment and grading decision-making in order to understand the reasons behind teachers' choice of specific grading procedures in their assessments. The study found that teachers' assessment decisions were influenced by the external testing and district policies imposed on them. Moreover, the study found that teachers' beliefs and values (e.g., promoting understanding and motivating students to succeed by using different tools of assessment to accommodate individual) can also have an influence on teachers' assessment practices.

In closing, teachers' assessment practices are influenced by factors related to teachers, such as their beliefs, perceptions, experience, and knowledge; as well as by external factors, such as external testing and school policies. Since this study is concerned with teachers' perceptions about assessment particularly, assessment for

learning or formative assessment, the following section highlights the role of teachers' beliefs and perceptions in the language classroom and their relationship to teachers' assessment practices.

Teacher Beliefs and Assessment Practices

Ultimately, teachers have been found to develop their own theories about teaching and learning based largely on their assumptions, beliefs, and prior knowledge (Woods, 1996). These together shape what teachers do in the classroom, how they do it, the way they perceive their roles as teachers and the role of students, the role of other factors such as classroom settings, and the course content in the learning process (Fox, 2009). In the literature, teachers' beliefs have been used to refer to a variety of concepts such as perceptions, attitudes, perspectives, personal theories and, implicit theories, which all refer to teacher thinking (Chang, 2005; Pajares, 1992; Mansory, 2016).

In the language classroom, teachers' beliefs about how language should be learned and taught is found to be strongly related to what they do in their classrooms (Breen, 1991; Woods, 1996). Such beliefs stem from teachers' earlier experiences as learners and, later on, as teachers. For example, Johnson (1994) examined the beliefs of pre-service ESL teachers during the practicum teaching experience in American universities. The study found that teachers make their instructional decisions mainly based on the mental images of language teachers they had encountered as language learners rather than on what they had learned during the teacher education courses they attended. In other words, teachers may be more likely to be influenced by their experiences as learners in the way that lead them to somehow imitate their teachers' practices rather than developing their own ways of teaching.

Nevertheless, studies have found that teacher education can have an influence on what teachers know, believe, and do in the classroom. Freeman (1993) investigated how the practices of four foreign language teachers changed during an in-service Master's in Education degree program. Freeman found that the program had an impact on the way teachers think about what they experience in the classroom using new ideas and concepts they had learned during the program. This finding indicates the important role of teacher education in shaping teachers' knowledge and beliefs about teaching and learning, which they mirror in their classroom practices.

Beliefs can influence not only teaching practices and the decisions teachers make about classroom instruction, but also the way they evaluate students' performance and assess their learning. In his research on teachers' beliefs about assessment, Brown (2002, 2003) found that teachers hold one of four major conceptions of assessment: (a) it is useful because it can provide information for improving instruction; (b) it is necessary for making students accountable; (c) it is necessary for making institutions accountable; and (d) it is irrelevant to teaching and learning.

Such conceptions can lead to different assessment practices. Bliem and Davinroy (1997) reported a project tended to investigate teachers' beliefs about assessment and their relationship to instruction in literacy. They found that teachers' belief systems about assessment were separate from those about instruction in the way that makes assessment limited to documenting students' performance and instructional tasks cannot be used for both purpose; assessment and learning. Such beliefs may lead teachers to limit their assessment to merely summative while disregarding formative assessment, which is important to inform both teachers' instruction and students' learning.

With regards to using tests as an assessment tool particularly, teachers may hold misconceptions about designing language tests. For example, teachers may think that there is one “best” test for any given situation or put unreasonable expectations about what language tests can do and what they should be (Bachman & Palmer, 1996). These misconceptions may prevent teachers from designing and using tests effectively in their language courses which may result in producing inappropriate tests that do not meet students’ learning needs. Moreover, believing that there should be a “perfect” test may cause frustration for some teachers that they are not able to design their own tests and result in relying on mentors or colleagues (Bachman & Palmer, 1996). This belief may also encourage teachers to adopt specific testing procedures not because they are effective but because they are popular in the field.

Teachers’ beliefs also include their perceptions about students and their role in the teaching and learning process. Black and Wiliam (1998) argue that teachers’ beliefs about learning and their students’ potential to learn can impact the extent to which they can move toward adopting more formative assessment rather than relying on external testing and other summative assessment tools. The assumption is that if teachers believe that learning mainly occurs through receiving knowledge transmitted by teachers, they may be more likely to use summative tools to check if that knowledge has been received. Furthermore, if teachers believe that students have potential to improve their learning and understanding, they are more likely to engage students in formative assessment such as self- and peer assessment and provide them with feedback to enhance their learning.

Song and Koh (2010) explored the relationship between secondary EFL teachers’ beliefs and practices regarding formative assessment in Singapore. The data obtained

from a self-reported questionnaire revealed that teachers who believe that students should be active participants in the learning process are more likely to use formative assessment, such as written and oral feedback, and diagnosis of students' learning needs. Hence, teachers' beliefs can have prominent influence on their assessment practices especially with regards to the extent to which they tend to use formative assessment in their classroom and include alternative methods to assess students' language proficiency.

In conclusion, the above literature provided an overview of the history of language assessment and the shift that occurred over the last decades in foreign language assessment from a testing culture in which traditional methods (e.g., tests) are used to measure knowledge about language, to an assessment culture where other forms of assessment are used to reveal information about students' ability to use the language. This shift is characterized by integrating assessment with instruction in the classroom to better support students' learning through using assessment to serve both summative and formative purposes. Thus, classroom assessment is no longer limited to evaluating and recording students' achievement using traditional tests; it should focus on using other alternative methods in order to inform teachers' instruction about how to enhance students' learning.

Assessment knowledge can help teachers integrate assessment with instruction in a way that leads to enhance students' learning (Coombe, Al-Hamly, & Troudi, 2009; Inbar-Louri, 2008b; Malone, 2008; Stoyhoff & Chapelle, 2005; Taylor, 2009). However, a lack of assessment knowledge can influence teachers' ability to choose appropriate assessments for specific purposes and adjust instruction to better meet students' learning needs. Studies on assessment practices have found that pre- and in-service EFL teachers

often do not have sufficient knowledge to enable them to conduct assessment effectively in their classrooms (Brookhart, 2001; Campbell & Collins, 2007; Mertler, 2005).

One of the factors that may be contributing to this lack of knowledge is that teacher programs do not always equip teachers with sufficient knowledge regarding assessment. Some programs provide limited assessment training, for example; these programmes may provide curriculum courses for general education without a clear focus on second or foreign language assessment components (Coombe, Troudi, & Al-Hamly, 2012; Inbar-Lourie, 2008a). Or, in teacher programs that offer courses that cover the basic aspects of language testing, less attention may be given to educational and psychological measurement and student classroom practice (Jin, 2010).

Hence, exploring teachers' practices and their beliefs that may influence their practices can help improve teacher programs and professional development programs to better equip them with the knowledge and skills they need to improve their teaching practices in general and conduct classroom assessment effectively. Informed by the literature above, this study aims to explore EFL practices and beliefs about CBA in Jordanian secondary schools and sheds lights on teachers' assessment practices and perceptions regarding classroom assessment.

Chapter 3 Methodology

Having introduced the study and discussed the literature that informs it, in this Chapter, I provide an overview of the research methods that were used to address the research questions. As noted in Chapters 1 and 2, the purpose of this study was to investigate Jordanian EFL teachers' practices and perceptions about classroom assessment, and given my research questions, a mixed methods research design was chosen, combining both quantitative and qualitative methods (QUAN¹→qual) in an explanatory sequential mixed methods design (Creswell & Creswell, 2018).

Research Design

This study used mixed methods to investigate Jordanian EFL teachers' practices and perceptions about classroom-based assessment (CBA). In the field of assessment, researchers have recently started to integrate qualitative and quantitative data in a single study within a mixed methods framework (Cheng & Fox, 2013; Turner, 2013). The introduction of mixed methods as a new research paradigm in the field of language assessment in general and classroom-based assessment in particular "has provided more flexibility to investigate emerging issues that call for different approaches from the two pre-existing research methodologies [quantitative and qualitative]" (Tsushima, 2015, p. 109). Because mixed methods research "enables researchers to obtain knowledge that ranges from a general, macro picture to more detailed insights of specific cases" (Tsushima, 2015, p. 115), it is well suited to examine the complexity of classroom assessment practices where diverse types of data (e.g., surveys, interviews, observation)

¹ Although this signals dominance of quantitative, given the nature of the quantitative data, dominance was minimal.

can be used to develop a stronger understanding of micro and macro classroom phenomena.

Collecting and analyzing different types of quantitative and qualitative data to study the same phenomenon is known as *triangulation*, which is adding another source of data to validate research outcomes through the convergence and corroboration of the findings (Dörnyei, 2007). However, mixed methods design allows the researcher to conduct a multi-level analysis through incorporating different sources of data (quantitative and qualitative) at the level of design and across phases which result in convergence or merging for interpretation. This helps explore issues that requires explaining more than one dimension of the phenomenon in order to be better understood and provides the researcher with a more comprehensive understanding of the research problem (Creswell & Creswell, 2018) and more robust results.

The present study used a two-phase explanatory sequential mixed methods design (Creswell & Creswell, 2018), using questionnaires (quantitative) and interviews (qualitative) to address the research questions. In the first phase, a closed-ended questionnaire was used to survey teachers' assessment practices and beliefs. The results of this analysis were used to develop the interview guide for the second phase, qualitative interviews, which were intended to help clarify confusing, contradictory, or unusual questionnaire responses. In the second phase, semi-structured interviews were conducted to achieve a deeper understanding of the initial quantitative results (Creswell & Creswell, 2018).

Data Collection

Prior to beginning data collection, ethics clearance was obtained from the Carleton University Research Ethics Board-A (CUREB-A; see Appendix A for clearance certificate). An email request was sent to the Ministry of Education in Jordan to seek permission to conduct the study. Data collection started after the letter of permission was received (see Appendix B).

Recruitment. A recruitment poster was posted on two Facebook groups for Jordanian EFL teachers. Unfortunately, only a few teachers responded to the poster, and so a snowball sampling strategy was adopted to increase the number of participants. Interested teachers were contacted by email and sent an invitation message with the questionnaire link. At the end of the questionnaire teachers were invited to participate in a follow-up interview to be conducted via online phone chat. An informational letter and consent form were sent to interested participants prior to the interview. Interviews were conducted between December 2018 and January 2019 and audio-recorded for later transcription and analysis.

Participants. In total, 58 teachers from different secondary public schools in three governorates in Jordan participated in the questionnaire. The reason for choosing teachers from public schools only is that there might be less restrictions on teachers regarding their classroom practices; private schools may usually impose some restrictions on teachers in terms of teaching instruction and classroom assessment. Of these, 54 submitted completed questionnaires; the four incomplete questionnaires were excluded from analysis. Three respondents volunteered to participate in semi-structured interviews upon completion of the questionnaire. It should be noted that initially no teachers

approached the researcher to participate in the interviews; other methods of finding participants were carried out, namely, convenience sampling based on ease of contact.

The three teachers were from different schools and different governorate.

Detailed demographic information about the participants is presented in Table 1.

Table 1

Questionnaire Respondents' Demographic Information

	n=	percentage
<u>Gender</u>		
Male	30	56%
Female	24	44%
<u>Age</u>		
26-30 years	12	22%
31-35 years	16	30%
36-40	15	28%
41-45 years	6	11%
46-50 years	5	9%
<u>Educational Qualifications (Highest)</u>		
BA in English Language & Literature	36	67%
BA in English Language/Field Teacher	2	4%
Higher Diploma	5	9%
MA/Med	10	18%
Other	1	2%
<u>Years of Teaching Experience</u>		
Less than 5 years	10	18%
6-10 years	13	24%
11-15 years	17	32%
16-20 years	7	13%
More than 20 years	7	13%

Assessment Training

A course or workshop	25	46%
None	29	54%

All questionnaire respondents had at least a Bachelor's degree in English Language and Literature—the minimum qualification requirement to teach English in Jordanian schools. Roughly 28% (n=15) had further qualifications such as diplomas or Master's degrees. The majority of teachers had 11 or more years of teaching experience. The majority of them were teaching both Grades 11 and 12, while the remainder were teaching one or the other. The average number of classes per week was 14, with an average class size of 30 students.

Table 2

Interview Participants' Demographic Information

Name	Educational Qualifications	Years of Teaching Experience	Assessment Training
Rose	BA in English Language and Literature Current MA student in Applied Linguistics	5 years (2 in secondary education)	None
Seren	BA in English language and Literature MA in English Translation	6 years	Online course on formative assessment

Zeina	BA in English Language	3 years (1 in	None
	MA in Business	secondary education)	
	Management		

The teachers who participated in the interviews were from different public schools in different governorates in Jordan. Table 2 summarizes their demographic information. All names are pseudonyms.

Instruments and procedures. Data collection took about one month from December 2018 to January 2019. Two data collection tools were used to achieve the aims of the study: an online questionnaire and online interviews. The questionnaire helped me to collect information about the type and frequency of assessment methods that EFL teachers use in their classroom, as well as about teachers' beliefs regarding the role of assessment in the language classroom. Although questionnaires are ideal to collect data from a large group of participants, they are limited in terms of producing rich information to enable in-depth investigation of the research problem (Dörnyei, 2007). Therefore, interviews were used as an additional tool to gain deeper insight into teachers' use of classroom-based assessment and their perceptions. The questionnaire and the interviews are further detailed below.

Questionnaire. The online questionnaire was designed using Qualtrics, a web-based software used to design and analyze online surveys and prepare reports of the results. The use of a web-based questionnaire helped to reach EFL teachers in various cities in Jordan, which would not be easy using traditional types of administration, such as lettermail.

The questionnaire consisted of four main parts (see Appendix C). The first part asked about participants' background information: gender, age, educational qualifications, years of experience teaching English, the grade/s they teach, and their knowledge about assessment. The second part asked teachers about the purposes of assessment and how frequently they assess for each of these purposes, using a scale from 1 (Never 0%) to 5 (Always 100%). The third part of the questionnaire was to determine the type of assessment methods teachers use to assess the four language skills (reading, writing, listening, and speaking) by indicating how often they use each of the listed methods on a 1 to 5 scale (least to most frequent). The first three parts of the previously validated questionnaire adapted from a study by Cheng, Rogers, and Hu (2004) with some modifications, such as adding the 5-point Likert scale for frequency of use.

The fourth part asked teachers to indicate the extent to which they agree or disagree with 7 statements about assessment for learning using a scale from 1 (strongly disagree) to 5 (strongly agree). There was also one question asked participants to identify the methods they use to provide feedback during the course.

Interviews. Follow-up semi-structured interviews were conducted with three teachers using Facebook Messenger (an online calling app), which was the most convenient contact method for participants. All of the interviews were conducted in English. English was preferred since it would eliminate the need for time-consuming in translation of transcripts. Moreover, the participants were English language teachers who taught the highest-level grades and they were fluent in English so they did not mind using English in the interviews.

Interviews were semi-structured, which was intended to encourage participants to share their experiences, knowledge, and perceptions more freely than would have a standardized format, which can limit the flexibility and the richness of responses (Dörnyei, 2007). The interview questions were informed by the results of the questionnaire (see Appendix D). The questions sought additional explanations about their responses and further explored participants' assessment practices and the beliefs that drive these practices. Two interviews were one hour each, and the third was half an hour.

Data Analysis

Quantitative questionnaire data were analyzed first and those results were used to inform the follow-up interview guide. Quantitative data were analyzed using SPSS Statistics Premium (Version 25). The data were exported into SPSS directly from Qualtrics, which had organized the data into spreadsheets, thus reducing data entry errors. Incomplete questionnaires were eliminated from analysis. Descriptive statistics—specifically, the mean, median, and standard deviation—were used to summarize general data trends. Crosstab analysis was conducted to explore significant relationships between demographic groups (e.g., age, sex, academic qualifications). The reliability of the questionnaire was ascertained using Cronbach's Alpha coefficient. The computed values were purposes= .867, methods= .931, and perceptions= .713, all exceeded .7, which is the minimum value of validating the internal consistency (Dörnyei, 2003).

Audio interview data were transcribed by the researcher and underwent preliminary coding in order to gain an overall sense of themes. Then, the data were coded in NVivo, a qualitative data analysis software program. First, the data underwent initial coding in which the data were broken down into discrete parts and assigned initial codes.

Second, focused coding was conducted, where data were organized into *categories* by combining related initial codes; each category was then assigned a label (Creswell & Creswell, 2018). Third, related categories were grouped together to identify core patterns or *themes*. To improve the reliability of the analysis, a second researcher familiar with the issues considered in the thesis independently coded one of the transcripts in full. Codes and themes were then compared. The correspondence in our initial coding was approximately 80% agreement which is considered a satisfactory level of consistency according to Miles and Huberman (1994).

Chapter 4 Findings and Discussion

This chapter presents the findings of the study and discusses them in relation to the literature. The organization is according to the research questions. For each question, the findings of Phase 1 (questionnaire) and Phase 2 (interviews) are presented followed by a discussion of the results. Finally the results from both Phases were merged in response to the questions. Before presenting the findings, it may be useful to restate the research questions here.

- 1- For what purposes do EFL teachers in secondary schools in Jordan conduct CBA?
- 2- What methods do EFL teachers in secondary schools in Jordan use to assess students in reading, writing, listening, and speaking? And what influences their choices of assessment?
- 3- How do EFL teachers in secondary schools in Jordan account for their beliefs about Assessment for Learning (AfL)/formative assessment? Are their reported beliefs aligned with their assessment practices? What are the implications?

Purposes of Assessment (Research Question 1)

In this section, the findings of both phases regarding the purposes of assessment are presented followed by a discussion of the results using the data obtained from the interviews.

Results of Phase 1: Questionnaire. Not surprisingly, the findings from the questionnaire revealed that teachers use assessment in their classrooms to achieve different purposes. These purposes were classified into the three categories used in Cheng, Rogers, and Hu (2004): student-centred purposes, instruction-related purposes,

and administration-related purposes. Table 3 summarizes the results of assessment purposes and the frequency of each purpose as reported by the teachers.

Student-centered purposes. Among the various purposes that assessment can serve in the language classroom (listed in the questionnaire and Table 3) the following purposes related to students' learning and progress:

- *Obtain information on my students' progress.*
- *Diagnose strengths and weaknesses in my students.*
- *Determine final grades for my students.*
- *Formally document growth in learning.*
- *Motivate my students to learn.*
- *Make my students work harder.*
- *Provide feedback to my students as they progress through the course.*

As shown in Table 3 and Figure 1, the majority of teachers use assessment for purposes related to students more frequently than for other purposes. The most frequently used student-related purpose was *Motivate my students to learn* (87% responded "Always" or "Usually"). The second most frequent purpose was *Determine final grades for my students* (80% responded "Always" or "Usually"). The same frequency was reported for *Make my students work harder* and *Provide feedback to my students as they progress through the course* (79% responded "Always" or "Usually"). A less frequently reported student-related purpose was *Formally document growth in learning* (67% responded "Always" or "Usually"). Finally, while 74% of teachers reported that they

always or usually use assessment to *Obtain information on my students' progress*, 7% of teachers reported that they rarely use assessment for this purpose.

Although all purposes under this category are identified as related to student learning, some purposes could be used by teachers to support effective teaching (e.g., *Obtain information on my students' progress*, *Diagnose strengths and weaknesses in my students*, and *Formally document growth in learning*). On the other hand, purposes such as *Provide feedback to my students as they progress through the course*, *Make my students work harder*, and *Make my students work harder* can be used to support students' management of their learning. With regards to *Determine the final grades for my student*, although this is typically seen as summative assessment, it may be also seen to have formative potential for the teacher since it can inform teachers regarding the effectiveness of their teaching and help them see what worked for which students. The results in table 3 show that teachers reported using both types of formative assessment to support teachers' teaching and students' learning with varying frequencies. Figure 3 shows that teachers reported using assessment purposes to support students' management of their learning more frequently than those to support teaching.

Table 3

Questionnaire Results of Assessment Purposes

Purpose	Mean	Median	Mode	SD*	Always N(%)	Usually N(%)	Often N(%)	Rarely N(%)	Never N(%)
<u>Student-Centred</u>									
To obtain information on my students' progress	3.93	4.00	4	.866	14(25.9)	26(48.1)	10(18.5)	4(7.4)	0
To diagnose strengths and weaknesses in my students	4.13	4.00	5	.802	21(38.9)	19(35.2)	14(25.9)	0	0
To determine the final grades for my students	4.26	4.00	5	.732	23(42.6)	22(40.7)	9(16.6)	0	0
To formally document growth in learning	3.89	4.00	4	.793	13(24.1)	23(42.6)	17(31.5)	1(1.9)	0
To motivate my students to learn	4.46	5.00	5	.719	32(59.3)	15(27.8)	7(13)	0	0
To make my students work harder	4.31	5.00	5	.843	29(53.7)	14(25.9)	10(18.5)	1(1.9)	0

To provide feedback to my students as they progress through the course	4.24	4.00	5	.889	26(48.1)	17(31.5)	10(18.5)	1(1.9)	0
<u>Instructional</u>									
To plan my instruction	3.96	4.00	4	.971	17(31.5)	23(42.6)	11(20.4)	1(1.9)	2(3.7)
To diagnose strengths and weaknesses in my own teaching and instruction	3.91	4.00	4	.830	13(24.1)	26(48.1)	12(22.2)	3(5.6)	0
To group my students for instructional purposes in my class.	3.57	4.00	4	.924	8(14.8)	23(42.6)	15(27.8)	8(14.8)	0
<u>Administrative</u>									
To provide information to the school administration	3.85	4.00	4	1.123	18(33.3)	20(37)	8(14.8)	6(11.1)	2(3.7)

*SD = Standard Deviation

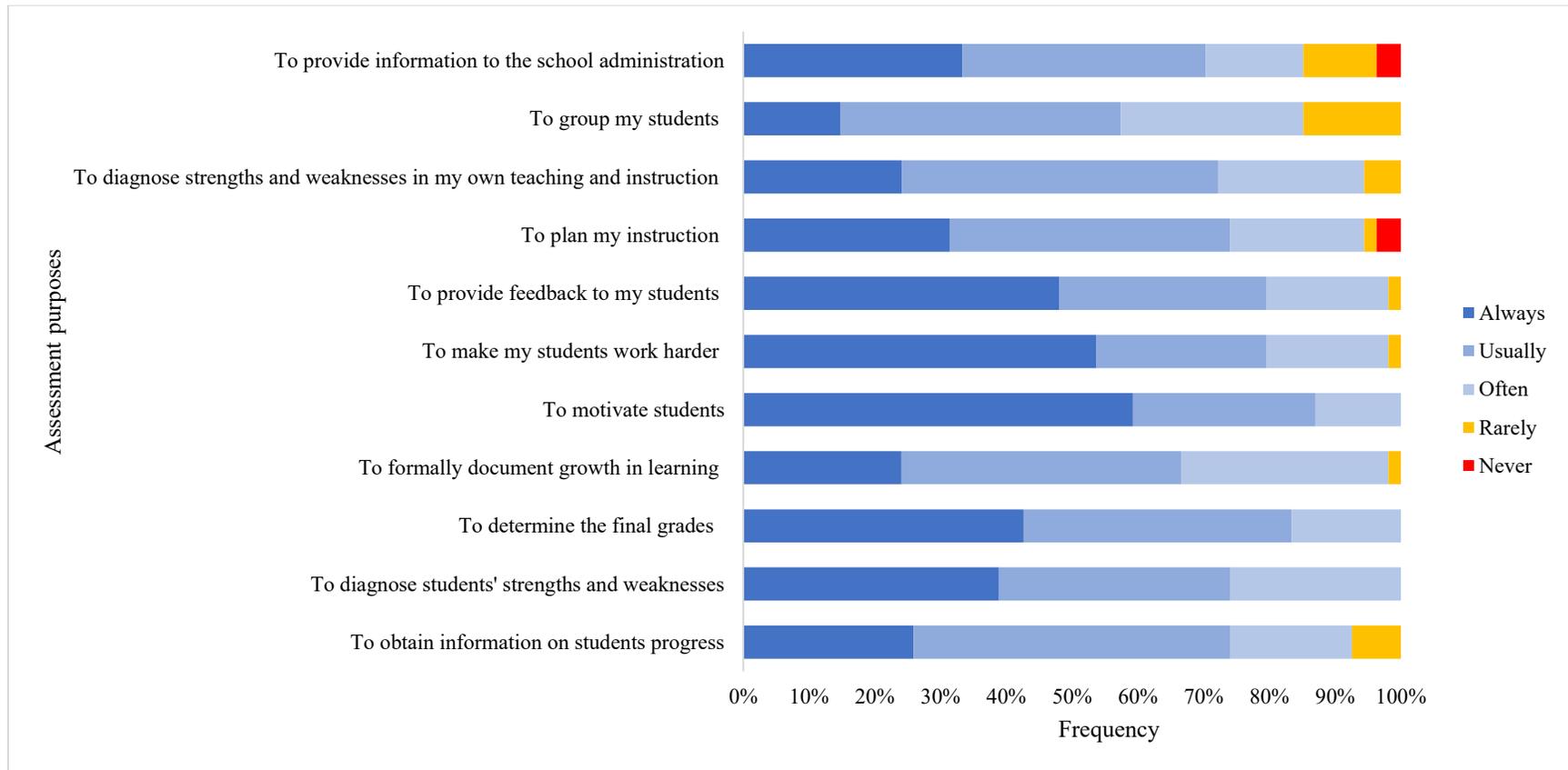


Figure 1. Frequency of assessment purposes.

Instructional purposes. The second category of assessment purposes, related to instruction, included:

- *Plan my instruction*
- *Group students for instructional purposes in my class*
- *Diagnose strengths and weaknesses in my teaching and instruction*

The most frequently reported purpose used by the majority of teachers was *Plan my instruction*, 74% answered “Always” or “Usually” followed by *Diagnose strengths and weaknesses in my teaching and instruction* (72% responded “Always” or “Usually”). The least frequently reported purpose was *Group students for instructional purposes in my class* (57% responded “Always” or “Usually”). However, the percentage of negative responses was 6% for the first two purposes and 14% for the last purpose.

Administrative purposes. The third category related to assessments used for an administrative purpose, i.e., *Providing information to the school administration*. Roughly 70% of teachers reported that they frequently use assessment for this purpose, yet 15% of teachers responded negatively (“Rarely” and “Never”). This is the highest percentage of negative responses among all purposes in the three categories.

Results of Phase 2: Interviews. The interviews reinforced that the main assessment purposes used by teachers were related to students’ learning and instruction. Teachers stated that the main goal of assessment is to evaluate students’ learning and obtain information about their progress. For example, Rose said, “The main goal of assessment is to assess the knowledge that students have got, at what level they are, and [to know] what their achievements in this course and if they achieved the goals that we

wanted them to achieve”. Seren added that assessment extends to motivating students to learn. She said, “It [assessment] is not just about keeping records of their [students] achievements but also helping them and motivating them to improve and go forward”.

Another purpose indicated by teachers was identifying students’ strengths and weaknesses. As Rose said, “I use diagnostic test at the beginning of the course to assess the strengths and weaknesses of students”. Rose also pointed to the use of assessment to prepare students for the Tawjihi. When she was asked about the main purposes of assessment from her perspective, she responded, “To see if they [students] are ready to pass the Tawjihi exam.”

Regarding using assessment for instructional-related purposes, the three teachers stated that assessment can be used to diagnose strengths and weaknesses of their teaching and plan their instruction. For example, Seren claimed, “Assessment can help teachers to improve their students’ learning as well as their own teaching methods.” Zeina said, “Also, to check if there is any need to use other teaching strategies to get the outcomes of the course.” Rose pointed to the use of assessment to plan instruction when she explained her use of diagnostic tests: “So I know where I should start with them [students]. After that, I should write a remedial plan for weak students and a developmental plan for good students.”

Discussion of Research Question 1. Overall, the findings of both phases revealed that the majority of teachers use assessment in their classrooms to achieve various purposes both summative (e.g., *Determine the final grade of students*) and formative (e.g., *Plan my instruction*). This finding may indicate that teachers are aware of the different purposes for using assessment in the language classroom. Their awareness

might be related to their teaching experience, as the majority of the teachers (more than 80%) have at least 6 years of teaching experience. Or, some teachers might have good assessment knowledge as a result of training courses or workshops on assessment, which half of the teachers reported attending. This finding is similar to previous studies that found that experienced teachers who had assessment training used more assessment purposes than teachers with less teaching experience and knowledge of assessment (e.g., Cheng et al., 2004).

Another finding indicated in both phases was that teachers used assessment more frequently for student-centred purposes, such as determining the final grade of students, obtaining information on students' progress, and motivating them to learn and work harder. Although some of purposes in this category, as noted earlier, can be used by teachers to support effective teaching (e.g., *Obtain information on my students' progress*), yet those related to supporting students' learning (e.g., *Motivate students to learn*) were reported to be more frequently used. This might suggest that a common assumption among teachers is that assessment should mostly relate to students and primarily be used to evaluate their learning. This assumption might lead teachers to focus their classroom assessment more on summative purposes and overlook the importance of using assessment to form and inform their instruction.

The interviews indicated that another assessment purpose reported by teachers was preparing students for the Tawjihi. This finding is not surprising since in most contexts where students need to take high-stakes examinations teachers tend to focus their instruction on preparing students for the exam. This practice is highlighted in the

assessment literature in general and language testing and assessment in particular (Cheng, 2008; Shohamy, 2006).

Regarding the use of diagnostic assessment (i.e., *Diagnose strengths and weaknesses in my students*), all teachers in the questionnaire and interviews responded positively, indicating this is a frequently used purpose for assessment. According to the teacher interviews, the reason for using this type of assessment is to help teacher plan their instruction based on the results of assessment. For example, Rose said:

I use diagnostic test at the beginning of the course to assess the strengths and weaknesses of students, to know their level in English so I know where I should start with them. After that, I should write a remedial plan for weak students and developmental plan for good students.

The use of diagnostic assessment may also be because of school administration requirements; as Zeina said, “It [diagnostic assessment] is required at the beginning of each year”. Thus, teachers may conduct diagnostic assessment only to observe the school policy. At the same time, interviews suggested that some teachers find it beneficial not only to assess students’ needs but also to evaluate their own teaching. For example, Zeina expressed her satisfaction with this type of assessment:

I find it very useful especially to see how much the student’s progress during the course. I feel so happy when I see that a weak student has become better at the end of the course. I feel that I have accomplished what I planned to do.

Zeina’s sentiment is reflected in the literature; studies (e.g., Fox, 2009) have suggested that using diagnostic assessment can provide information to both students and

teachers, and lead to instruction focused on students' needs. This purpose has the potential for supporting students' learning and effectively increasing their language development. Moreover, previous studies have suggested that the use of ongoing diagnostic assessment can also support teacher change (Colby-Kelly & Turner, 2007; Fox, 2009).

Another interesting finding was that teachers reported that they frequently use assessment to plan and improve instruction, which is considered the most essential purpose of using classroom assessment, as it has the potential to maximize students' learning (Black & Wiliam, 1998). For example, Rose stated:

When students do not do well in a test, I should rethink about my way of teaching. I share responsibility with my students for these bad results! So, I try to modify my teaching methods using different strategies to help students understand the material.

This finding may suggest that teachers in this study are aware of the vital role of assessment in improving both students' learning and teaching instruction. Although the questionnaire data showed a high percentage of positive responses to *Instructional-related purposes* which suggest a dominant trend in the teachers' responses, it was surprising to find "Rarely" and "Never" responses from a small sample. Their responses may suggest the absence of formative assessment in their classroom practice since these are primary formative purposes.

Methods of Assessment (Research Question 2)

In the following sections, the methods of assessment are categorized into *Teacher-made assessment methods*, *Student-conducted assessment methods*, and *Standardized testing* (Cheng et al., 2004). This classification structure is used to report the results from Phase 1 and Phase 2 about the assessment methods used for reading, writing, listening and speaking, followed by a discussion of the results.

Results of Phase 1: Questionnaire.

Reading assessment. The data from the questionnaires revealed that teachers use a variety of assessment methods to assess students' performance in reading, including all three categories (teacher-made, student-conducted, and standardized test) with varied frequencies. Table 4 summarizes the methods that teachers use and how often they use each one.

Teacher-made assessment methods. As Figure 2 shows, the listed methods were used by the majority of teachers for assessing reading (more than 73% responded positively with "Always", "Usually", and "Often"). Among these methods, there were two methods used more frequently: short-answer items and multiple-choice items (78% responded "Always" and "Usually" to both methods), while only one teacher said that they never use multiple-choice items to assess reading. The least frequently used methods were matching-items and true/false items (more than 24% responded "Rarely" and "Never").

Student-conducted assessment methods. As shown in Figure 3, 63% of teachers responded positively to all of the listed methods under this category. *Read aloud* was the most frequently used method to assess students' competence in reading as about 80%

reported they Always or usually used it. *Student portfolio* was the least frequently used method (only 31% responded “Always” and “Usually” while 37% reported that they rarely or never use it). Half of teachers reported using *Peer assessment* always or usually. At least 15% of teachers reported that they rarely or never used *Student summaries*, *Student portfolio*, and *Self-assessment*. Overall, *Student-conducted methods* were used less frequently than *Teacher-made methods*, as shown in Figures 2 and 3.

Standardized testing. The majority of teachers (more than 74%) reported that they use standardized tests to assess reading, yet 26% said they rarely or never use it.

Table 4

Questionnaire Results for Methods of Reading Assessment

Methods	Mean	Median	Mode	SD*	Always N(%)	Usually N(%)	Often N(%)	Rarely N(%)	Never N(%)
<u>Teacher-Made</u>									
Cloze items	3.48	3.50	4	.818	5(9.3)	22(40.7)	21(38.9)	6(11.1)	0
True/false	3.61	4.00	4	1.28	19(29.6)	18(33.3)	7(13)	9(16.7)	4(7.4)
Sentence completion	4.00	4.00	4	.847	17(31.5)	22(40.7)	13(24.1)	2(3.7)	0
Matching items	3.33	3.00	4	1.19	10(18.5)	16(29.6)	14(25.9)	10(18.5)	4(7.4)
Multiple-choice	4.09	4.00	5	.976	22(40.7)	20(37)	8(14.8)	3(5.6)	1(1.9)
Short-answer items	4.19	4.00	5	.953	26(48.1)	16(29.6)	8(14.8)	4(7.4)	0
<u>Student-Conducted</u>									
Read aloud	4.13	4.00	5	.972	23(42.6)	20(37)	7(13)	3(5.6)	1(1.9)
Student summaries	3.41	4.00	4	1.22	11(20.4)	19(35.2)	8(14.8)	13(24.1)	3(5.6)
Peer assessment	3.54	3.50	3	.966	10(18.5)	17(31.5)	19(35.2)	8(14.8)	0

Student portfolio	2.98	3.00	3	1.17	7(13)	10(18.5)	17(31.5)	15(27.8)	5(9.3)
Self-assessment	3.54	4.00	4	1.07	9(16.7)	23(42.6)	13(24.1)	6(11.1)	3(5.6)
<u>Standardized Test</u>	3.44	4.00	4	1.39	15(27.8)	16(29.6)	9(16.7)	6(11.1)	8(14.8)

*SD = Standard Deviation

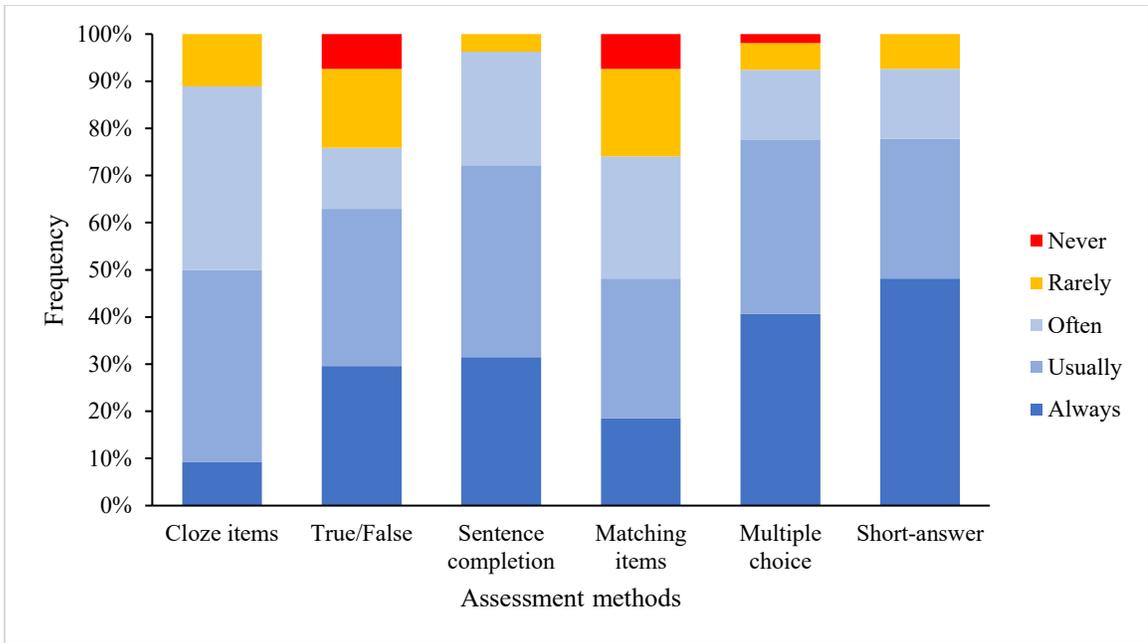


Figure 2. Reading assessment methods (teacher-made tests).

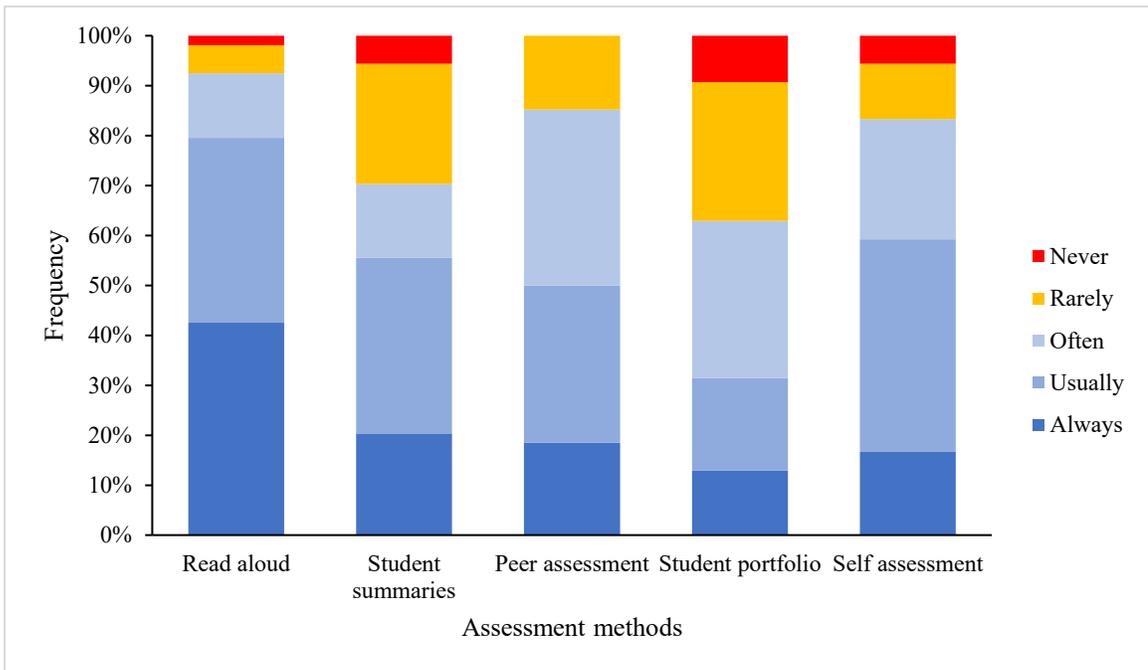


Figure 3. Reading assessment methods (student-conducted).

Writing assessment. The results of the questionnaire regarding the methods used by teachers to assess writing are summarized in Table 5.

Teacher-made assessment methods. Table 5 shows that teachers use all of the methods in this category with varying frequency. *Editing a sentence or a paragraph* was the most frequently used method to assess students' writing (87% responded "Always" and "Usually" none of the teachers answered "Never"). The second and third most frequently used methods were *Multiple-choice* (70%) and *Short essay* (63%). The least frequently used method was *Long essay*; half of teachers responded "Rarely" or "Never". Figure 4 illustrates the frequency of use for all methods in this category.

Student-conducted assessment methods. As Figure 5 shows, the most frequently used method in this category was *Peer assessment* (81% responded "Always", "Usually", or "Often"). The second most frequently used method was *Self-assessment* (about 78% responded "Always", "Usually" or "Often"). *Student portfolio* was the third most frequently used method (74% responded "Always", "Usually", or "Often), yet 23% of teachers reported rarely or never using this method. The least frequently used method was *Student journal*; more than 44% of teachers responded "Rarely" or "Never".

Standardized testing. The majority of teachers reported that they frequently use standardized tests to assess writing (more than 72% responded positively), while about 28% reported rarely or never using this method.

Table 5

Questionnaire Results for Methods of Writing Assessment

Methods	Mean	Median	Mode	SD*	Always N(%)	Usually N(%)	Often N(%)	Rarely N(%)	Never N(%)
<u>Teacher-Made</u>									
True/false	3.26	3.00	3	1.291	13(24.1)	9(16.7)	16(29.6)	11(20.4)	5(9.3)
Matching items	3.24	3.00	3	1.11	9(18.4)	13(24.1)	21(38.9)	8(14.8)	4(7.4)
Multiple-choice	3.89	4.00	5	1.16	20(37)	18(33.3)	9(16.7)	4(7.4)	3(5.6)
Editing a sentence/paragraph	4.26	4.00	4	.782	23(42.6)	24(44.4)	5(9.3)	2(3.7)	0
Short essay	3.78	4.00	4	.925	13(24.1)	21(38.9)	15(27.8)	5(9.3)	0
Long essay	2.89	2.50	2	1.284	8(14.8)	11(20.4)	8(14.8)	21(38.9)	6(11.1)
<u>Student-Conducted</u>									
Student journal	2.91	3.00	2	1.103	3(5.6)	17(31.5)	10(18.5)	20(37)	4(7.4)
Self-assessment	3.15	3.00	3	.998	5(9.3)	13(24.1)	24(44.4)	9(16.7)	3(5.6)
Peer assessment	3.35	3.50	4	1.06	6(11.1)	21(38.9)	17(31.5)	6(11.1)	4(7.4)

Student portfolio	3.07	3.00	3	1.079	5(9.3)	13(24.1)	22(40.7)	9(16.7)	5(9.3)
<u>Standardized Test</u>	3.28	3.00	3	1.280	11(20.4)	14(25.9)	14(25.9)	9(16.7)	6(11.1)

*SD = Standard Deviation

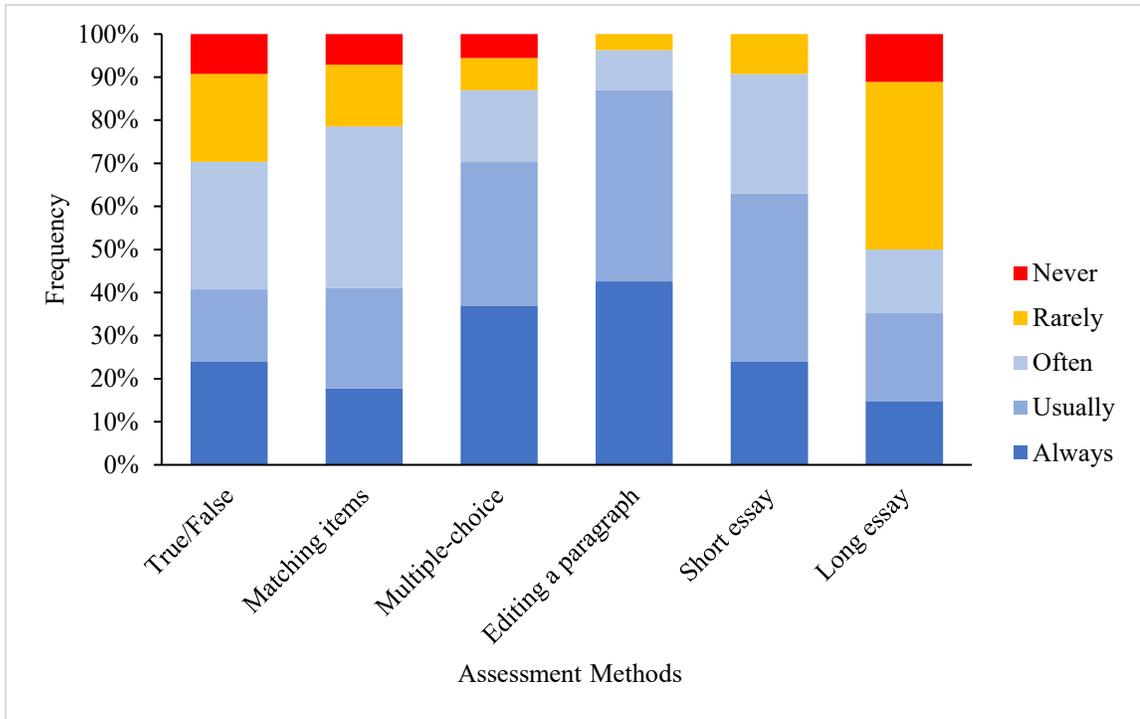


Figure 4. Writing assessment methods (teacher-made tests)

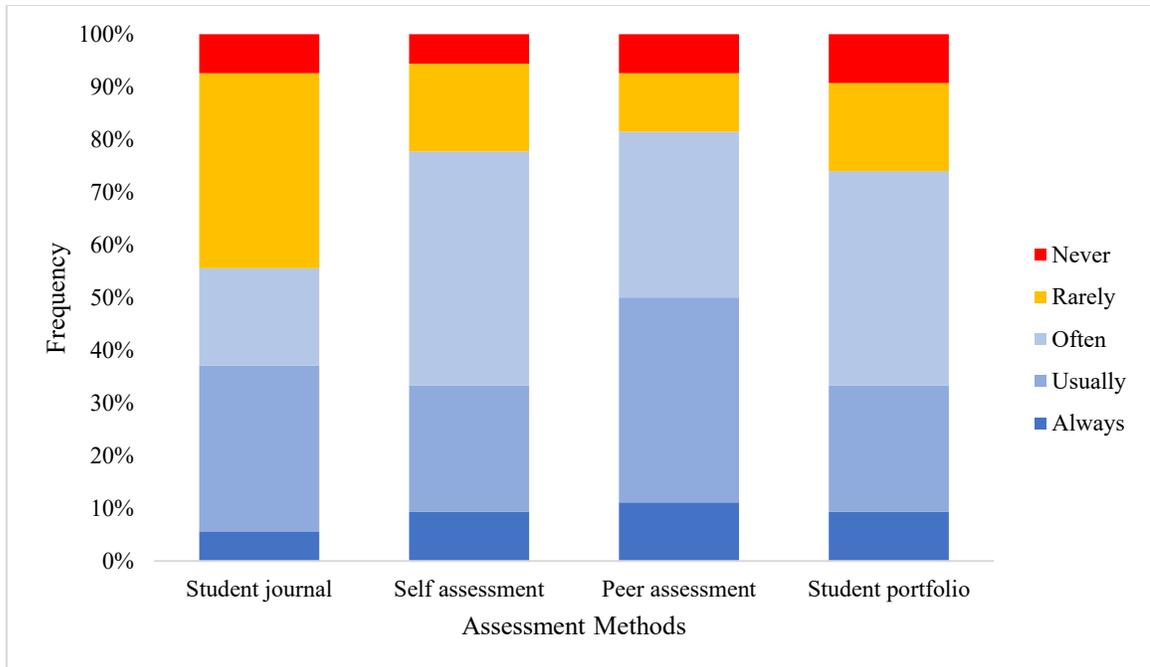


Figure 5. Writing assessment methods (student-conducted)

Listening and speaking assessment. Although researchers have argued that, listening comprehension is “a separate and important component of language learning,” (Vandergrift, 1999, p. 168), in this study data regarding listening and speaking assessment methods were collected and analyzed together, following Cheng et al. (2004). The questionnaire results about the methods that teachers use to assess listening and speaking are summarized in Table 6. The mode values show a tendency among teachers to usually use most of the listed methods. Figure 4 shows the most and least frequent methods.

Teacher-made assessment methods. As shown in Table 6, the most frequently used methods were *Take notes* and *Multiple-choice items* (65% responded “Always” or “Usually” for both), followed by *Giving summaries after listening to a spoken passage* (50%).

Student-conducted assessment methods. This category included 10 methods that required students to produce language orally with almost no written production. As shown in Figure 6, the most frequently reported methods were *Oral reading/dictation*, *Giving oral directions*, and *Oral interviews* (76% at least responded “Always” and “Usually” to all). The least frequent method was *Peer assessment*; only 44% reported they always or usually use it while 22% of teachers said they rarely or never use it in their classrooms. Finally, 76% of teachers said they favour *Oral interviews*, compared to 61% who favour *Oral presentations*, as indicated by “Always” and “Usually” responses.

Standardized testing. The majority of teachers (79%) reported that they use standardized speaking tests more often than standardized listening tests. Moreover 20%

said they never use standardized tests to assess listening, while only 6% responded “Never” to speaking standardized tests.

The crosstab analysis did not show any significant relationships between demographic groups (e.g., age, sex, academic qualifications) and their reported assessment practices. For example, teachers who have high qualifications such as diploma or Master’s did not seem to use more formative assessment methods than those who had only BA in English language or English language and literature.

Table 6

Questionnaire Results for Methods of Listening and Speaking Assessment

Methods	Mean	Median	Mode	SD*	Always N(%)	Usually N(%)	Often N(%)	Rarely N(%)	Never N(%)
<u>Teacher-Made</u>									
Take notes	3.78	4.00	4	1.003	14(25.9)	21(38.9)	13(24.1)	5(9.3)	1(1.9)
Give summaries for what is heard	3.46	3.50	3	1.041	9(16.7)	18(33.3)	18(33.3)	7(13)	2(3.7)
Multiple-choice	3.76	4.00	4	1.045	14(25.9)	21(38.9)	13(24.2)	4(7.4)	2(3.7)
<u>Student-Conducted</u>									
Oral presentation	3.72	4.00	4	1.123	16(29.6)	17(31.5)	13(24.1)	6(11.1)	2(3.7)
Oral interviews	3.93	4.00	4	1.025	17(31.5)	24(44.4)	6(11.1)	6(11.1)	1(1.9)
Oral discussion with each student	3.81	4.00	5	1.083	18(33.3)	17(31.5)	10(18.5)	9(16.7)	0
Retell a story after listening	3.43	3.50	4	1.045	7(13)	20(37)	17(31.5)	9(16.7)	1(1.9)
Provide an oral description for an event	3.85	4.00	4	.920	13(24.1)	25(46.3)	12(22.2)	3(5.6)	1(1.9)

Give oral directions	4.04	4.00	4	.868	17(31.5)	25(46.3)	10(18.5)	1(1.9)	1(1.9)
Follow oral directions	3.80	4.00	4	.919	12(22.2)	24(44.1)	14(25.9)	3(5.6)	1(1.9)
Peer assessment	3.26	3.00	4	.975	4(7.4)	20(37)	18(33.3)	10(18.5)	2(3.7)
Self-assessment	3.43	4.00	4	1.075	6(11.1)	25(46.3)	13(24.1)	6(11.1)	4(7.4)
Oral reading/dictation	4.19	4.00	4	.870	22(40.7)	23(42.6)	7(13)	1(1.9)	1(1.9)
<u>Standardized Test (Speaking)</u>	3.46	4.00	4	1.111	9(16.7)	21(38.9)	13(24.1)	8(14.8)	3(5.6)
<u>Standardized Test (Listening)</u>	3.06	3.00	3	1.393	10(18.5)	12(22.2)	14(25.9)	7(13)	11(20.4)

*SD = Standard Deviation

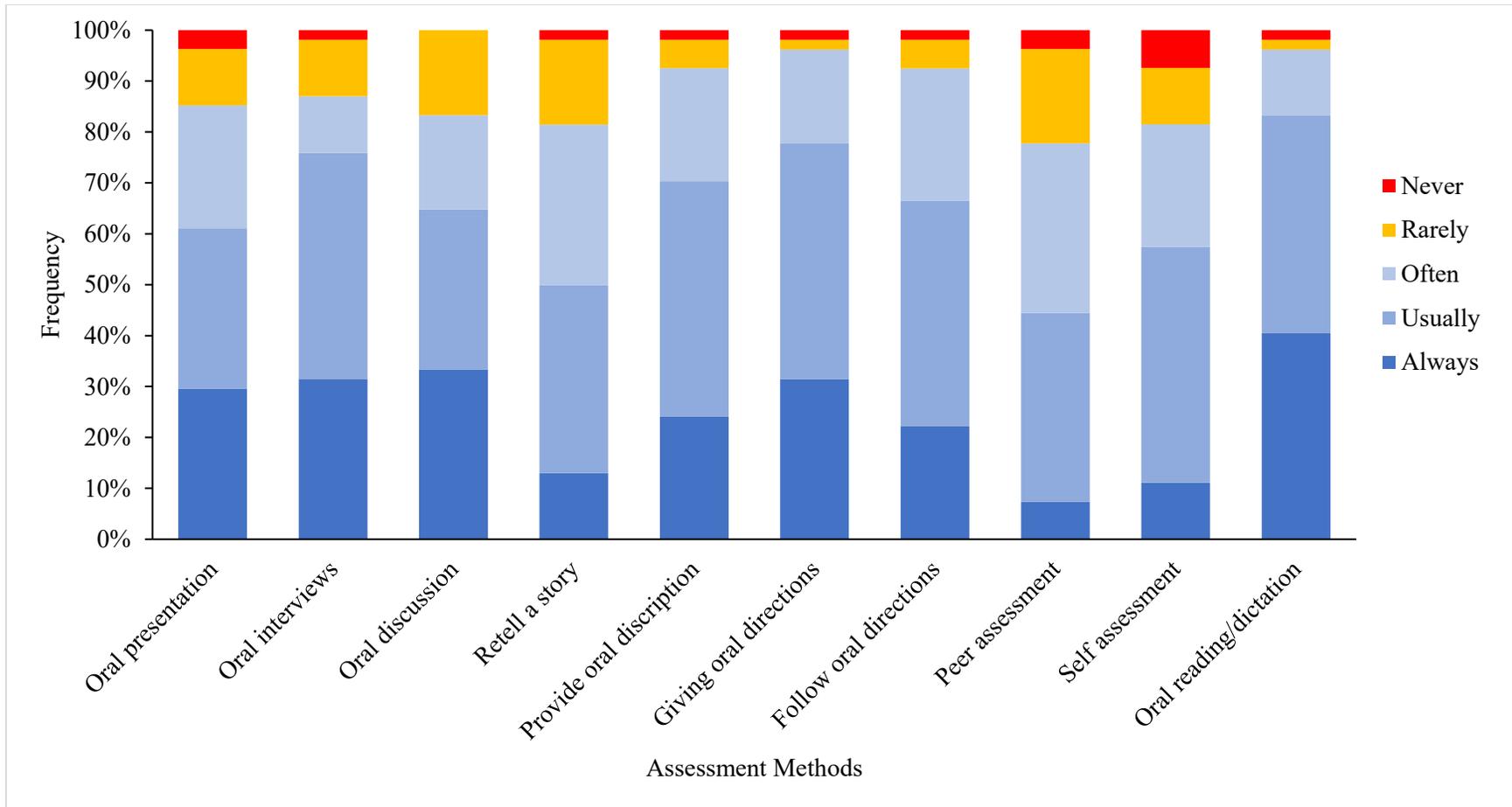


Figure 6. Listening and speaking assessment methods (student-conducted).

Results of Phase 2: Interviews. In Phase 2, the interviews findings were considered in light of the results from the questionnaire analysis in Phase 1. This section summarizes the methods of assessment used to assess each of the four skills as reported by the interviewed teachers.

Reading assessment. The results revealed that teachers use a number of methods to assess students' competence in reading, including both teacher-made tests and student-conducted methods. With regards to teacher-made tests, teachers stated that they usually use multiple-choice items and comprehension questions. For example, when Rose was asked if she uses multiple-choice items, she said, "Yeah, multiple choice items are very common in Tawjihi." Zeina also stated that she uses comprehension questions in assessing reading while avoiding true/false items: "Most of the time, I use comprehension questions.. I try to be away from yes/no questions."

Reading aloud and giving summaries were the main student-conducted reading assessment methods. For example, Zeina said, "I let them read aloud to check their pronunciation and understanding." Similarly, Seren said, "At the end, I check their [students'] understanding by asking them to summarize what they have read."

Writing assessment. All three teachers reported that most of the teacher-made tests they use to assess writing include editing a paragraph or writing an essay. For example, when Seren was asked if she includes editing a paragraph in her writing tests, she said, "Yeah, I always include this question in the test for Grade 11 and 12 because it is a common question in the National Exam." Rose also said, "I think for writing, the only method I use... maybe not only me but also most English teachers, is giving the students a topic and ask them to write an essay about it." Two teachers also reported that

they use some student-conducted methods for writing assessment, such as peer assessment. For example, Seren said, “I sometimes use peer assessment in writing classes, I ask students to exchange their essays with their partners and try to grade it and give feedback.”

Speaking and listening assessment. Teachers revealed that they use teacher-made tests that include taking notes, multiple-choice items, and fill in the blank items to assess listening. For example, Seren said, “I ask them to write some notes, main points or anything they feel that it is important such as numbers, dates, and famous names.” Zeina also said, “For Grade 11 & 12 listening is more about fill in the blank.”

With regards to speaking assessment, teachers reported that they sometimes use student-conducted methods such as oral presentation, while the others said they do not include speaking skills in their assessments. For example, Zeina said, “Actually, I ask them to prepare at home for the presentation and to write some sentences, but they do not feel comfortable doing that.” Rose said, “I do not focus on speaking and listening in my teaching in general or in assessment.”

All teachers confirmed that they never use portfolios to assess any of the language skills. For example, Zeina said, “No, we do not use it in Jordan.” Regarding standardized tests, all teachers stated that they use previous Tawjihi tests for their reading and writing assessments. For example Seren said, “Actually, for reading and writing tests, I usually use questions from the National Exam of previous years.” Rose also stated, “Teachers usually use questions from the National Exam for previous years, so the students will be familiar with the type of questions they would expect in their National Exam.”

Discussion of Research Question 2.

Reading assessment methods. The results from the questionnaire showed that teachers reported using different methods to assess their students' competence in reading. The most frequently used methods were short-answer items, multiple choice items, and reading aloud. The interviews allowed for a deeper and more meaningful interpretation of the questionnaire responses in Phase 1. For example, when Rose was asked if she use reading aloud to assess her students, she said:

At the end of each month, I have criteria to assess students for each of the four skills and reading is one of these skills. One of the points I focus on is pronunciation. It is very important to see how they [students] pronounce the English words.

Zeina also stated that she always uses read aloud in her instruction and assessment:

After reading, I usually check their understanding of the text. I always have a checklist and a rating scale for that. I check students' expression when they are reading because the way they read a question should differ from the way they read a sentence, pronunciation of each word, and the way they use punctuation in their reading, they for example, they need to stop at the full stop mark. I do this at the end of the lesson but first I read aloud for them or I use a CD, then we do some activities and lastly, I let them read aloud to check their pronunciation and understanding.

It seems from both excerpts that teachers use the read aloud method because of their belief that pronunciation is an important aspect of reading competence and reading aloud

as an assessment tool can help improve students' pronunciation. This matches the findings of previous studies about the effectiveness of using this assessment method in improving English pronunciation for ESL adult learner (e.g., Nurani & Rosyada, 2015).

The focus on pronunciation in teaching and assessing reading skills for students might be attributed to the fact that, in many EFL contexts, English proficiency may be more related to the ability to speak English intelligibly. Moreover, teachers may consider the classroom as the only place where students have the opportunity to learn how to pronounce English words properly with the help of their teachers. This was indicated by Rose as she said:

In my teaching I use it [read aloud] although my supervisor has said to me once not to read aloud for students. Actually, I find it a very beneficial way especially for second language learners. You know.. For students, the teacher is considered as a native speaker!.. Not exactly but the teacher could be the best to help students pronounce the words correctly. So, as a teacher I think I should read aloud for my students and encourage them to read aloud themselves to practice the good reading.

It is clear that Rose's focusing on pronunciation seems to stem from her beliefs about the effectiveness of using the read aloud method in improving students' pronunciation which can ultimately improve their reading skills including learning vocabulary. This may conform with the theories of second language acquisition that suggest that deeper knowledge of a word meaning is evidenced by spelling, and pronunciation (Nation, 2013). Rose's beliefs about language teaching and learning seem to influence her instruction and assessment accordingly. This finding is compatible with most studies on

language teachers' beliefs which have shown that teachers' beliefs about how language should be taught and learned can highly influence what they do in the classroom including teaching instruction and assessment (Breen, 1991; Woods, 1996).

The second most frequently used method was tests with short answer and multiple-choice items. The reason for using short answer items more frequently than other questions, such as true/false, was explained by Zeina: "Most of the time I use comprehension questions. I try to be away from yes/no questions because I want them to read and answer in details." Using this type of question in reading tests might be because students of Grades 11 and 12 are supposed to have an advanced level of English proficiency and be able to read and comprehend what they read and answer related questions. Moreover, using comprehension questions may reveal more information about students' reading skills as it requires the student to read, comprehend, skim and scan, make inferences, and use discourse markers (Brown, 2004).

Regarding multiple-choice items, teachers' frequent use of this type of question was because it is very common in the Tawjihi exam. For example, Rose said:

Multiple choice items are very common in the National Exam. Now, the Ministry of Education is talking about some modifications to the National Exam in 2019 that involve computer-based exam with multiple choice items only. So, multiple choice questions will be the common type of questions in the National Exam for the coming years and we should focus on this type of questions.

She also explained that she uses cloze items less frequently due to the influence of the Tawjihi:

I usually use that with younger students but not Grade 12 because we usually follow the same format of the National Exam (Tawjihi) and they do not use this type a lot. We try to make the students familiar with the types of questions in the National Exam so they can achieve good marks.

In the literature, cloze items are very common in reading assessment because they can reveal useful information about students' abilities and knowledge, such as knowledge of vocabulary, grammatical structure, discourse structure, and reading skills (Brown, 2004). Nevertheless, teachers use this type less frequently and focus more on the format of the Tawjihi to help students become familiar with the exam. Teachers focusing on the formats of large-scale high-stakes examinations was also found to be the case in many EFL contexts (e.g., Andrews et al., 2002; Cheng, 2004; Qi, 2004).

Writing assessment methods. The findings of Phase 1 revealed that teachers assess their students' writing skills through different methods—most frequently, editing a paragraph and short essay. These were also reported by interview teachers as the most common methods used in their classrooms. For example, Zeina said, “I ask them [students] to edit a paragraph by underlining any grammar, punctuation or word choice mistakes then correct them.” Seren also said, “I always include this question in the test for Grades 11 and 12.” The reason for using this type of assessment more frequently was that it is a common question in the Tawjihi. For example, when Rose was asked about using editing a paragraph in her writing tests, she responded, “I use it a lot because there is always a question like that in the National Exam every year.” This finding clearly shows that teachers tend to focus their assessment on preparing students to the Tawjihi by including the types of questions that students would expect in the exam. This again

demonstrates the influence of high-stakes examinations on teachers' classroom assessment.

Another frequent method teachers used to assess writing was asking students to write a short essay in a given topic. Using this method is also found to be related to the influence of the Tawjihi, as indicated by teachers in the interviews. For example, Rose said:

I think for writing, the only method I use.. maybe not only me but also most English teachers, is giving the students a topic and ask them to write an essay about it. I usually give them lessons to model the genre of writing and after that I ask them to write an essay following that genre. All the topics [of the essays] are from the textbook because this is what they are going to be asked about in the National Exam.

Zeina explained her reason for using this method: "There are usually 3 parts [in the writing section of the National Exam]: guided writing, short essay and correct the mistakes."

Using short essays more frequently than long essays to assess writing might be also related to the proficiency level of the students. Rose, for example, talked about the recent training courses offered by the MoE for in-service teachers to improve students' proficiency level in writing. She said, "One of these workshops is for training English teachers to teach English writing skills because they have found that students in public schools have low level of writing skills in English."

Students' low level in English writing might be related to the fact that writing is a challenging cognitive activity (Nunan, 1989) especially for ESL/EFL learners as factors such as their first language transfer and the lack of effective writing instruction might impede developing their writing skills (Al-Khairi,2013).

Listening and speaking assessment methods. As shown in the results of the questionnaire, the most frequently used methods to assess listening and speaking skills were all student-conducted, including providing an oral description for an event and giving/following oral directions. The use of these types of assessment tasks might be because such tasks can reveal more information about students' performance in listening and speaking. Moreover, it could be that they are commonly used in the textbook, which all teachers are required to follow, especially for Grade 12 since the Tawjihi is designed based on the material of the textbook.

Nevertheless, the interview results revealed that teachers tend to focus less on listening and speaking exercises in the textbook because these two skills are not assessed in the Tawjihi. For example, when Rose was asked if she include listening in her classroom assessment, she said:

Actually, I have to include it because I should assess the four skills, but it is usually on paper only! Students in Grade 12 do not care a lot about speaking and listening as they are not parts of the National Exam.

It is obvious that the dominant influence of Tawjihi and the need to prepare students for this high-stakes exam may lead teachers to focus mainly on the skills that are assessed in the exam.

This practice of “teaching to the test” has been observed in previous studies on the washback of large-scale and high-stakes examination in ESL/EFL contexts, where teachers tend to narrow their instruction to teach the skills and content of the exams (Azadi & Gholami, 2013; Manjarrés, 2005). Another teacher, Zeina, argued that the reason that listening and speaking skills may receive less attention is that teachers may not have sufficient training and experience on how to teach these skills, she stated “I think that teachers now are focusing more on listening. We have some supervisors who come and give workshops on how to teach listening.” This might lead teachers to pay less attention to listening and speaking skills in their instruction and assessment. Moreover, this might also have a negative impact on students’ motivation to learn and improve their proficiency in these skills.

With regards to teacher-made tests, teachers reported that they usually use multiple-choice items following a listening passage in listening tests. The interviews findings revealed that using this type is related to students’ low proficiency level. For example, Zeina said:

I usually ask them to fill in the blank..Or to choose from multiple answers the correct spelling. I sometimes write on the board 2 or 3 simple questions and they have to answer them in short while they are listening. Really it is just that simple because students are weak and not used to listening tests.

Choosing assessment methods that suit students’ low level may indicate that the proficiency level of students can influence what methods teacher use to assess their students. This was also noted by previous studies on assessment practices in the EFL

context, for example, teachers were found to use translation tasks more often because of the low level of their students (Cheng et al., 2008).

Overall, the findings from both phases showed that teachers reported using *Teacher-made assessment methods* more frequently than *Student-conducted methods* and standardized tests, especially in reading (e.g., short answer and multiple-choice items) and writing (e.g., editing a paragraph). This may be attributed to the influence of the Tawjihi exam and the need to prepare students for this exam as noted earlier.

Some *Student-conducted methods*, such as peer- and self-assessment and portfolios, were less frequently used by all teachers for assessing the four skills. One of the interview teachers attributed the less tendency to use self-assessment as an assessment tool to students unfamiliarity with such methods of assessment as she said:

I do not use it with my students because I think they might not be aware of it because what they usually care about is getting a good mark rather than being honest about their strengths and weaknesses! So, I do not consider it as a tool for assessment.

Another teacher said that she never uses peer assessment because some students are not willing to participate when she asks them to work in pairs or in groups. She said, “Most of the time, I do not use that because from my experience, I would have three or four students who are good and I can depend on but the rest of the class they refuse to participate in pair work.”

Although students’ willingness might influence teachers’ use of peer and self-assessment, the lack of assessment knowledge on the teachers’ part might also contribute to such

practice since fewer than half of the teachers reported having assessment training. The lack of assessment knowledge and training may impact teachers' use of this type of assessment, which can influence its effectiveness.

Portfolio assessment was also reported to be used less frequently. Although some of the questionnaire respondents reported they sometime use it, none of the interviewed teachers said they ever did. For example, Zeina, claimed that portfolio assessment is not used in Jordanian public schools, while Rose said she only heard about this method in the short course she attended for new teachers. She stated:

We do not have this type of assessment yet but I have heard that the Ministry of Education is going to apply this tool in public schools. I know that most private schools use portfolio as a good way of assessment, but it is not yet applied in public schools. Maybe we will see that in the coming years!

The reason for not using this method of assessment in public school might be related to the expensive cost of implementing this assessment tool, whereas private schools might find it worthy to use this method to attract students and parents who are interested in non-traditional teaching and assessment methods. Questionnaire respondents may have exaggerated their use of portfolios in the classroom out of a desire to show that they are familiar with non-traditional methods. This phenomenon, known as *social desirability* or *prestige bias*, refers to survey respondents' tendency to provide responses that seem good/desirable even if they are not true (Dörnyei & Taguchi, 2009). It may also be that they misunderstood what "portfolio" means. One of the interviewed teachers asked to explain what this tool is and how it is used, while another teacher said she had

some knowledge about this method because she used to live in the USA where her children used portfolios in schools.

Teachers' reliance on using methods that are designed and administered by them (i.e., tests) and do not directly involve students in assessment may influence their learning effectively. As noted by Cheng and Fox (2017), teachers' frequent use of assessment practices that directly involve students has the potential to increase students' awareness of their learning progress and as a result promotes and supports effective learning.

Factors influencing teachers' choices of assessment. As shown above in the discussion of the methods used by teachers, there could be some reasons behind why teachers chose some assessment methods more or less frequently than other methods. In this section, the findings of phase 2 (interviews) were used to identify some factors that might play a role in teachers' decision-making about what assessment methods to use in their classrooms. In this thesis, *factor* refers to considerations that may influence teachers' choice of assessment, as opposed to results from any type of questionnaire analysis (e.g., factor or correlation analysis). These factors were: the influence of the National Exam (Tawjihi), students' proficiency level, the curriculum and textbook, and teachers' knowledge of assessment. These factors are discussed in turn below.

Influence of the National Exam (Tawjihi). As the findings show, the Tawjihi may have a marked influence on teaching instruction in general and assessment methods in particular. This is demonstrated in teachers' tendency to focus on designing most of the tests to be similar to what students of Grade 12 will see on the Tawjihi. All interviewed teachers stated that they often use the same assessment formats as on the Tawjihi to familiarize their students with the exam. For example, Rose said, "We usually

follow the same format of the Tawjihi. We try to make the students familiar with the types of questions in the Tawjihi so they can achieve good marks”. Seren mentioned that she always include editing a paragraph, which is a very common format on the Tawjihi, in her writing tests not only for Grade 12 but also for 11 as well. This shows the prominent influence of this high-stakes exam on teachers’ assessment practices. This finding is compatible with previous findings on the washback of high-stakes tests, which revealed that teachers are more likely to follow the assessment formats used in external testing when designing their tests (Andrews et al., 2002; Cheng, 2004; Qi, 2004).

Nevertheless, this influence seems to affect not only teachers’ assessment practices but also their teaching instruction. All interview teachers said that they usually do not focus on teaching Grade 12 listening and speaking because these two skills are not assessed in Tawjihi, and they usually skip the listening and speaking exercises in the textbook. For example, Rose said “for Grade 12, because these skills [listening and speaking] are not included in the National Exam, they [students] do not care about them, so I do not focus on speaking and listening in my teaching in general or in assessment.” Rose attributed her ignorance of teaching and assessing listening and speaking to students who are less concerned about material not required in the Tawjihi. This was also indicated by Zeina when she compared teaching Grade 12 with Grade 5:

When I was teaching grade 5, I used to ask them to do some role-play and dialogues, but I do not use this with students at the secondary stage. Students think it would be a waste of time if I ask them to do the listening and speaking exercise in the textbook. They just want to focus on what they need for the National Exam. They always say “teacher, let’s just skip these exercises and focus

on what is important! Give us some grammar!.. I feel that they do not aim to develop their skills in English to be able to use it as a language in communication, they just want to pass the National Exam and I do not actually blame them for that!

The Tawjihi creates pressure on students which leads them to limit their learning goals to what is usually included in the exam instead of aiming to develop their skills in English language. Such attitude might also indicate the influence of testing culture on students since they become more concerned about achieving good grades in the language course rather than learning the language itself (Abdo & Breen, 2010).

Tawjihi clearly influences both students and teachers; students become more concerned about passing the exam rather than improving their general proficiency in English, and teachers narrow their instruction to teach students what they need to pass the exam. However, teachers' practice may not stem from their own beliefs about how language should be taught and learn. For example, Rose claimed that although she does not focus on listening and speaking because of the influence of Tawjihi, she believes that teachers should teach these skills and include them in assessment. Zeina also stated, "If I were teaching a different grade, I would focus on all four skills." These responses might show that although high-stakes exams influence teachers' decisions about instruction, it might be minimally, as was found in previous studies on the influence of high-stakes tests in teachers' literacy-related instructional beliefs and practices (e.g., White, Sturtevant, & Dunlap, 2003).

Students' level of proficiency. Students' proficiency level in English and their learning needs may play an important role in the way teachers design assessment tools in

the language classroom. As the results of the interviews showed, teachers considered the proficiency level of their students when they design assessment tasks or tests. For example, Zeina reported that she usually adjusts her instruction and chooses the assessment tools that suit the proficiency level of her students:

In writing classes, I even do not follow the textbook sometimes. I just try to help my students by making it easier for them. I ask them to prepare for the class at home by writing a few sentences about their future plans. I do this because my students are weak especially in writing.

The low level of proficiency of students in Zeina's class led her to choose what she thinks maybe "easier" for them, although Grade 12 teachers are required to follow the textbook since it is the only required material for the Tawjihi.

Rose argued that students' low level of proficiency might impede her use of assessment approaches that aim to assess students' ability to use the target language rather than their knowledge about it:

I think it is unfair to evaluate students according to what they do with the language because in our context where English is taught as a foreign language, the classroom is the only place where students hear and use English...there are some students who are high level and have very good skills in speaking and writing so it is unfair for the rest of the students. What we use to assess students in public schools is the material in the textbook, so it is mainly what they know about the language.

Zeina also expressed that designing tests that suit students' proficiency level is a challenging task:

Sometimes I find it challenging to write questions that suit my students' level. I have to include questions with different degrees of difficulty. I try to be fair with all students. I know sometimes it would not be fair for excellent students to not have questions that require higher level of thinking, but I have to think about the majority who are usually not good enough to be able to answer these questions. Most of the students are weak and they find learning English challenging.

It seems that Zeina's concerns about designing assessment tasks that satisfy the majority of students (who have a low proficiency level) might stem from an assumption about using assessment in a way that avoids creating negative feelings for students. This assumption was also reported in Chang (2005), which revealed that teachers often believe that assessment should be conducted in ways that motivate and support students' learning without generating negative emotions.

In the literature, language proficiency level of students was also found to have an influence on teachers' assessment practices in previous research. For example, Cheng et al., (2008) found that EFL teachers in China used translation tasks in assessment because of students' limited proficiency in English.

Curriculum and textbook. The use of the textbook may have an impact on teachers' decision-making regarding the methods of assessment they use in their classrooms. In this study, the interviewed teachers reported that the textbook constrains how they plan their instruction and design assessment tools. For example, when Seren

was asked if she sets course objectives based on the textbook or on students' needs, she said:

Actually, I try to rely on both, but you know the teacher's book limits what teachers can plan for their students. My supervisor always asks me to follow the textbook because as he says, "the people who write textbooks and teacher's books know better what students need at this level and what teachers should do"

The constraint that the textbook imposes on teachers' practices might be because students in Grade 12 are assessed based on the material in the textbook. Thus, teachers may feel pressured to cover all the material in the textbook before the end of the year (the time of Tawjihi). Seren elaborated:

As a teacher, you have to follow the curriculum and the textbook, and you have to finish the material by the end of the semester, you cannot skip anything. We [teachers] feel that we are restricted to the curriculum and the textbook and we cannot go beyond that because we will find many challenges from the school, parents and students themselves.

However, for other grades, teachers might be less restricted to use the textbook so they might choose to teach students the way they think and believe best works to enhance their learning. This was indicated by Zeina who used to teach younger students. She said:

When I was teaching Grade 5 and 6, I used to tell the school principle that I do not want to follow the curriculum and cover all the material in the textbook, I just want to focus on what I think is important for my students. I want them to be

comfortable with using English in their life. So, I used to ask them to do lots of presentations and other activities like ‘spelling bee’ and they loved that.

It can be said the pressure on teachers to teach the content of the textbook seems to be more related to the influence of the Tawjihi, as the textbook includes all the required material for the exam, rather than the influence of curriculum or school policy. However, this might not be the case for all schools that may be more restricted in terms of following the textbook as stated earlier by Seren.

Knowledge of assessment. The fourth and last factor that may be identified to have an influence on teachers’ assessment practices is teachers’ knowledge of assessment. Although this factor might not be directly related to teachers’ choices of methods, the lack of knowledge and training on classroom assessment may impact how teachers design their assessment methods and how they use them. The interviewed teachers claimed that although they had attended a training course for new teachers, they still do not have sufficient knowledge about how to conduct classroom assessment because that short course did not provide enough information about assessment. For example, Zeina said:

I have never attended any course in assessment. It is just the short course I attended at the beginning of my teaching in public schools. There were topics on teaching English in general, how to manage the classroom and a little information about assessment.

Rose also attributed her lack of knowledge about some assessment methods such as portfolios to the inadequacy of assessment training:

I have heard about it [portfolio assessment] once during the pre-service teachers' workshop when I first started teaching at public schools. In that workshop we had brief information about assessment tools and types, just few information nothing in deep. I do not actually have a clear idea about the concept of portfolio. I am really not aware of this concept.

The lack of sufficient knowledge of assessment may lead teachers to use methods that are commonly used by teachers or to seek help from their colleagues who are more experienced. For example, Seren stated:

I usually discuss with other English teachers how to design the tests and tasks especially for the final exam. And we can share questions because you know it is the same textbook. Also, for Grade 12, we have to prepare them for the National Exam, so we try to design tests with good questions and benefit from each other experiences especially from those who have long experience in teaching English for the secondary stage. I find it very beneficial because I only have 6 years of experience.

Although having discussions with more experienced teachers might help teachers compensate for their lack of assessment knowledge, experience alone might not be sufficient for teachers to be able to conduct assessment that supports students' learning. Professional training is also needed to equip teachers with the knowledge and training they need to improve their assessment practices.

The questionnaire's results showed that 54% (28) of teachers reported that they had never attended a course or workshop on assessment, compared to 46% (25) who said

they had. 16% (8) of them said that these courses/workshops were about assessment in particular, while only three teachers reported that they had a course on assessment during their study. This result may indicate that there is a lack of assessment training in both teacher programs and professional development programs. This finding may be in line with previous studies in EFL contexts that found that teacher programs may not provide teachers with enough knowledge and training to conduct assessment in their classrooms, as most program courses are related to curriculum and teaching more generally (Coombe, Troudi, & Al-Hamly, 2012; Inbar-Lourie, 2008a).

Thus, teachers who had some training on assessment might have benefited from the knowledge they acquired and used it to improve their assessment practices. In her interview, Seren said that she became more aware of using different types of assessment, such as formative assessment, because of the non-mandatory training course she attended during her teaching career. Studies on the influence of assessment training on teachers' practices (e.g., Koh, Burke, Luke, Gong, & Tan, 2018) found that there were significant changes in the quality of assessment tasks designed by teachers who were involved in professional development programs as the programs can help them recognize the importance of identifying learning goals and aligning assessment tasks with the instructional outcomes.

Teacher Perceptions about Assessment for Learning (Research Question 3)

As discussed in the literature, formative assessment or assessment for learning provides useful information that can be used by both teachers and students in order to better support students' learning. Thus, according to Stiggins (2002), using assessment for learning requires teachers to (1) understand and articulate in advance the achievement

targets of the course, (2) inform students about those learning goals, (3) use assessment to inform and adjust their teaching practices to better support students' learning, (4) involve students in the assessment process through self-assessment, and (5) provide students with feedback to give them insight into how to improve. In addition, peer assessment is identified as an assessment approach which supports students' learning (William & Thompson, 2008).

Teachers' responses to seven questionnaire items based on the principles of assessment for learning (Broadfoot et al., 1991; Stiggins, 2002) are summarized in Table 7 and further illustrated in Figure 7.

Table 7

Questionnaire Results for Teachers' Perceptions About Assessment for Learning

<i>Principle</i>	<i>Mean</i>	<i>Median</i>	<i>Mode</i>	<i>SD*</i>	<i>Strongly agree</i> <i>N(%)</i>	<i>Agree</i> <i>N(%)</i>	<i>Neither agree nor disagree</i> <i>N(%)</i>	<i>Disagree</i> <i>N(%)</i>	<i>Strongly disagree</i> <i>N(%)</i>
1. Teachers should first identify the purpose of the assessment when they design the assessment task/test.	4.31	4	5	.843	26(48.1)	23(42.6)	1(1.9)	4(7.4)	0
2. Students will learn the most and perform their best if they know in advance the criteria that will be applied in judging a performance.	4.22	4	4	.744	20(37)	28(51.9)	4(7.4)	2(3.7)	0
3. Teachers should negotiate the assessment criteria with their students in advance.	4.19	4	4	.803	21(38.9)	24(44.4)	7(13)	2(3.7)	0
4. Assessment results should be used to modify instruction to better support students' learning.	4.28	4	5	.763	24(44.4)	22(40.7)	7(13)	1(1.9)	0

5. Students should be given opportunities to reflect on their learning through peer feedback and self-assessment.	4.24	4	4	.642	18(33.3)	32(59.3)	3(5.6)	1(1.9)	0
6. Students should be given feedback which enables them to know how to improve their work and take their learning forward.	4.39	4	4	.627	25(46.3)	25(46.3)	4(7.4)	0	0
7. Involving students in self- and peer assessment activities can enhance their achievement and motivation.	3.93	4	4	.773	12(22.2)	28(51.9)	11(20.4)	2(3.7)	1(1.9)

*SD = Standard Deviation

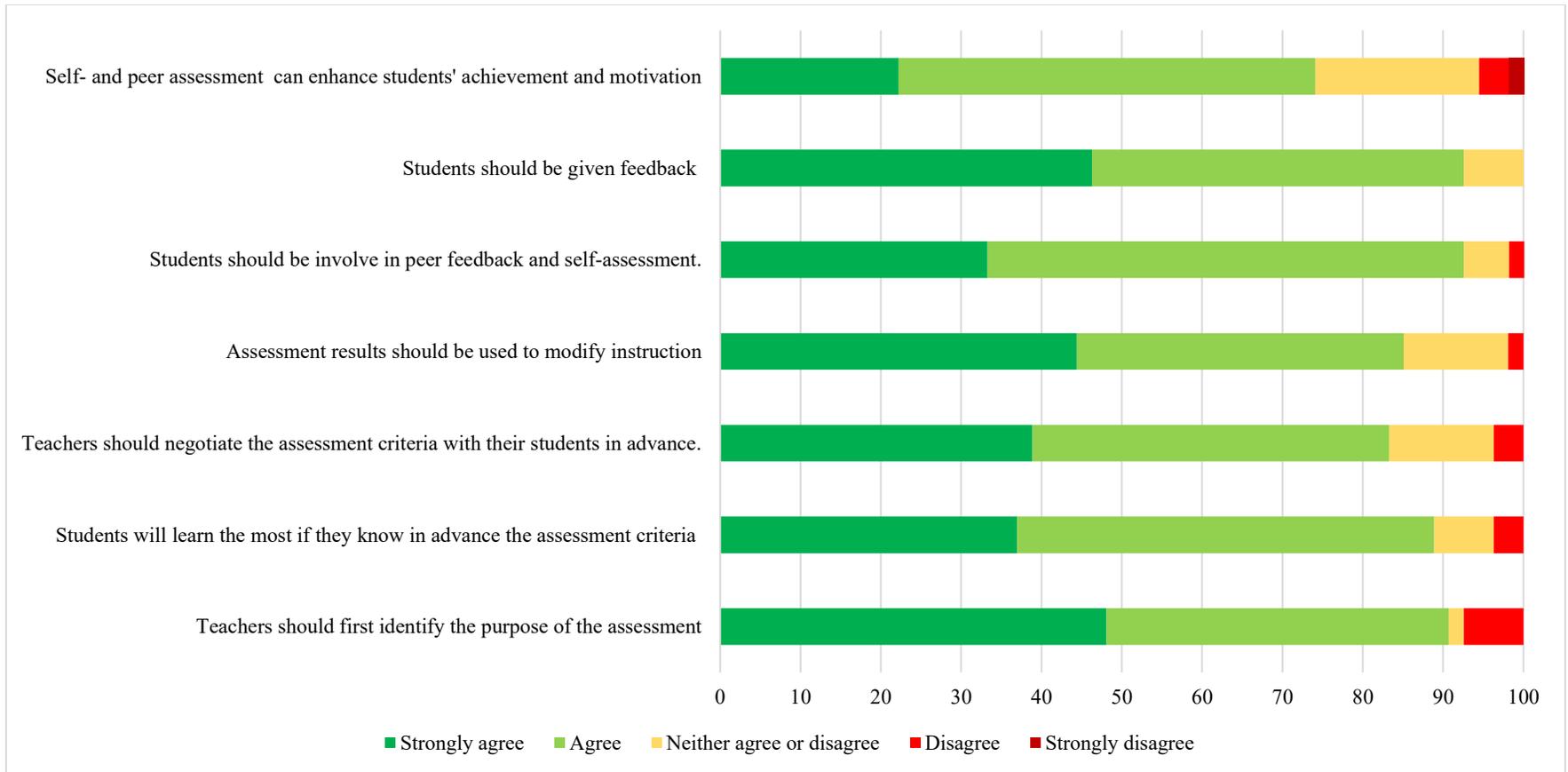


Figure 7. Teachers' perceptions about assessment for learning.

As shown in Table 7 and Figure 7, most teachers either agreed (69%, n=37) or strongly agreed (31%, n=17) with all of the items. This result may reflect teachers' inclination towards using formative assessment (Black & Williams, 1998), also known as assessment for learning (Gipps, 1994). This inclination is in line with the current trend in CBA that is using assessment to promote students' learning through including both forms of assessment; summative and formative (Stiggins, 2002, 2007).

Looking at each response more closely, firstly, more than 90% (n=49) of teachers agreed that teachers should first identify the purpose of assessment when they design the assessment task/test. Secondly, the majority of teachers agreed that students should be informed in advance about the assessment criteria that reflect the learning goals so they can understand and work toward these goals. This was reflected in the agreement of 83% (n=45) of teachers with the statement "*Teachers should negotiate the assessment criteria with their students in advance*" and 89% (n=48) with "*Students will learn the most and perform their best if they know in advance the criteria that will be applied in judging a performance*".

Thirdly, 85% (n=46) of teachers agreed that assessment results should be used to modify instruction to better support students' learning. With regards to involving students in the assessment process, the majority of teachers agreed that students should be engaged in their learning through peer- and self-assessment as reflected in the following two statements: (1) "*Involving students in self- and peer assessment activities can enhance their achievement and motivation*" with which 74% (n=40) of teachers agreed; and (2) "*Students should be given opportunities to reflect on their learning through peer feedback and self-assessment*" with which more than 92% (n=50) either agreed or

strongly agreed. Finally, regarding teachers' feedback, more than 92% (n=50) of teachers agreed that students should be given feedback that enables them to know how to improve their work and take their learning forward. No one disagreed with this latter statement.

The above results suggest that the majority of teachers believe that assessment is an integral part in teaching and learning since it can be used to inform instruction to maximize students' learning. The interviewed teachers validated this assumption when they indicated that there is a relationship between assessment, instruction and learning. For example, Zeina stated:

Teachers use assessment to figure out the needs of their students and whether they achieved the learning goals, understand the material. Also, to check if there is any need to use other teaching strategies to get the outcomes of the course.

Seren shared this assumption about the role of assessment in improving learning and teaching when she said, "Assessment can help teachers to improve their students' learning as well as their own teaching methods."

It can be said that teachers may hold the belief that assessment can play an important role in improving both teaching and learning. This finding may contradict Bliem and Davinroy's (1997) study on teachers' beliefs about assessment and their relationship to instruction in literacy. Bliem and Davinroy found that teachers' belief systems about assessment were separate from those about instruction. In other words, teachers believed that assessment is used only to document students' performance, and instructional tasks cannot be used for both purposes (assessment and learning).

Since teachers in this study seem to believe that assessment is related to teaching and learning, it might be assumed that their beliefs are in line with their assessment practices as the literature suggests that teachers’ practices are usually driven by their beliefs (Borg, 2006; Woods, 1996). The questionnaire data revealed that among those teachers who strongly agreed or agreed with the statement, “*Assessment results should be used to modify instruction to better support students’ learning*” (85%, n=46), about 74% (n=34) reported that they use assessment more frequently to plan their instruction while 72% (33) said they use it to diagnose strengths and weaknesses in their teaching. These results are shown in Table 8.

Table 8

Frequency of Using Formative Purposes

Purpose	Frequency N (%)					Total
	Always	Usually	Often	Rarely	Never	
To plan my instruction	16 (34.8)	18 (39.1)	10 (21.7)	1 (2.2)	1 (2.2)	46 (100)
To diagnose strengths and weaknesses in my own teaching	12 (26.1)	21 (45.7)	11 (23.9)	2 (4.3)	0(0)	46 (100)

This result may show that is a match between teachers’ belief about using assessment to adjust instruction and the purposes for which they use assessment in their classrooms. However, as discussed in Research Question 1 (purposes of assessment), the questionnaire data revealed that the majority (85%) of teachers use assessment for summative purposes (e.g., to determine the final grade of my students) more frequently than formative purposes (e.g., to plan my instruction and diagnose strengths and

weaknesses in my own teaching). This may show that although assessment was found to be used by teachers for formative purposes, the use of assessment for summative purposes is still dominant.

With regards to teachers' perceptions about using formative assessment methods, Table 7 shows 92% (50) teachers strongly agreed or agreed with the statement "*Students should be given opportunities to reflect on their learning through peer feedback and self-assessment*". This may show that teachers have positive attitudes towards involving students in monitoring their own learning through peer and self-assessment. Table 9 shows the frequency of teachers' use of peer and self-assessment in the four skills, as reported by the 50 teachers.

Table 9

Frequency of Using Peer and Self-Assessment

Method	Frequency N (%)				
	Always	Usually	Often	Rarely	Never
Peer assessment					
Reading	10 (20)	16 (32)	17 (34)	7 (14)	0 (0)
Writing	6 (12)	18 (36)	17 (34)	6 (12)	3 (6)
Listening & speaking	4 (8)	17 (34)	18 (36)	10 (20)	1 (2)
Self-assessment					
Reading	9 (18)	21 (42)	12 (24)	6 (12)	2 (4)
Writing	5 (10)	11 (22)	23 (46)	8 (16)	3 (6)
Listening & speaking	5 (10)	23 (46)	13 (26)	6 (12)	3 (6)

As shown in Table 9, more than 78% (39) of teachers responded positively to using peer and self-assessment to assess their students in the four skills, yet less than 60% (30) reported using these methods more frequently as indicated by “Always” and “Usually” responses. This result may indicate that although the majority of teachers are favorably inclined towards involving students in assessment in order to promote learning, their assessment practices might not be in line with their perceptions. This may be attributed to the lack of assessment knowledge and training since more than half of teachers reported that they never had any course or workshop on assessment. Another reason might be that using peer and self-assessment could be challenging in larger classes where issues of class management may arise. The questionnaire data showed that the average class size as reported by teachers, was 30 students. Moreover, students may be reluctant to engage in such methods of assessment without skilled scaffolding and support. They may not find them as valuable as more familiar assessment approaches which focus on individual performance (e.g., tests and achievement).

With regards to the effectiveness of peer and self-assessment, teachers may believe that peer and self-assessment activities can be effective in increasing students’ motivation and enhancing their achievement. This may be reflected in the agreement of 74% (40) of teachers with the statement “*Involving students in self- and peer assessment activities can enhance their achievement and motivation*”. Nevertheless, it may be noticed that teachers’ agreement about the potential of peer and self-assessment to increase students’ motivation is relatively low (74%) compared with their agreement about students’ needs to be involved in such type of assessment (78%). This result might be explained by teachers’ lack of assessment knowledge since fewer than half of the

teachers reported having assessment training. It is possible that leads to teachers' inappropriate implementation of this type of assessment, which can influence its effectiveness. Research suggests that peer and self-assessment help develop students' internal responsibility to have control over their learning, which in turn increase motivation and achievement (Black & Wiliam, 1998; Stiggins, 2005).

With regards to teacher feedback, all teachers agreed that students should be provided feedback that helps them know how to improve. Only 7% (4) of teachers were neutral on this topic. This finding is not surprising since the questionnaire data revealed that teachers use a variety of methods to provide their students with feedback during the course. Table 10 shows the frequency of each method as reported by teachers.

Table 10

Form of Feedback Provided to Students During the Course

Method	N (%)
Verbal feedback	14 (26)
Checklist	13 (24)
Written comments	14 (26)
Conference with student	11 (20)
Total test score	12 (22)
A letter grade	2 (8)
Other (e.g., public feedback)	6 (11)

As shown in Table 10, the most frequent methods teachers use to provide feedback are verbal and written comments, checklists, total test scores, and conferences with student. It is thus clear that teachers do not rely only on numeric feedback (e.g., total

test score) that merely summarizes the judgement of students' work. Instead, they primarily use methods that may be considered formative-based (Shute, 2008), as they can provide students with information on how progress can be made and lead to future development.

The match between teachers' conceptions of feedback and their reported assessment practices may indicate that teachers are aware of the importance of feedback in promoting students' learning. In the literature on assessment, researchers have asserted that feedback is an important component of formative assessment and can have positive impact on students' learning, while the lack of feedback can diminish students' motivation and their ability to learn effectively (Sadler, 1998; Black & Wiliam, 1998).

Although Table 10 shows that teachers use various methods to deliver feedback, it is less clear how frequently teachers provide feedback and whether teachers use some methods more frequently than others. The interviewed teachers stated that they usually use written comments. For example, Zeina said, "I usually write some comments on their test's papers". Similarly, Seren stated that she sometimes uses written comments, but it depends on time constraints. She said, "I sometimes write some notes on their test's papers but not all the time because I do not have time to do this, but when I have time I usually do". Rose likewise said that she sometimes writes notes on her students' test papers so they can better understand their mistakes.

Rose also claimed that total test score could be used as a form of feedback since students can know their achievement based on their grades. She stated, "I consider students' grades in the short tests as kind of feedback because their marks tell them how they did in the test and their achievements in the course". Although test scores and grades

show students' performance on the test, numeric feedback not in combination with other methods such as written comments does not help students to know how to improve their performance. According to Black and Wiliam (1998), feedback should identify what students need to revise and provide suggestions that guide them in their own revision.

All three teachers mentioned that they tend to provide feedback for the class as a whole rather than individually, especially when students make common mistakes. For example, Seren said:

I usually discuss the common mistakes/ difficulties they had in the test with the whole class not individually. So, after the test, I discuss the questions and the right answers with the students and explain to them what to do in order to do better in the next tests.

Zeina explained that her favouring this method is related to limited class time. She noted, "I do not actually have enough time to sit with each student and talk about her weaknesses and strengths, but I am always there when they need to ask questions or need explanations".

Teachers' tendency of providing feedback to the whole class may also be attributed to the large classes since the findings from the questionnaire showed that the average class size was 30 students. This tendency was also noted in Cheng and Wang's (2007) study in which they found that EFL teachers in China tended to provide feedback to the whole class instead of to individual students because of the large size of their classes.

The discussion above suggests that teachers may believe that providing students with feedback can help them improve and move towards achieving their learning goals. However, limited class time and the work load (and maybe class size) might interfere with their willingness to use feedback efficiently. Moreover, it seems that teachers' use of feedback is usually related to tests (e.g., written comments or class discussion after tests), which may demonstrate the influence of external testing since teachers tend to focus mainly on preparing students to the Tawjihi using previous tests or similar format so that students can practice for the exam.

With regards to negotiating the assessment criteria with students, although more than 83% (45) of questionnaire respondents agreed that teachers should involve students in discussions about the assessment criteria in advance, the interviews data revealed different views. Seren was the only teacher who stated that she sometimes considers sharing the assessment criteria with her students because she believes that this will help her students do better on the test:

I sometimes share the assessment criteria with my students because I believe that if you tell them how they are going to be assessed, they will try to work hard and prepare themselves better for the test. So, yeah it is good for students to know how they are going to be assessed in advance.

Seren's awareness and perceptions about the effectiveness of such practice may be related to her strong knowledge of assessment, as she had a training course on formative assessment.

On the other hand, Rose and Zeina implied this approach might be absent in some teachers' practices, possibly due to a lack of assessment knowledge, as the interviewed teachers said that they do not use it because they are not aware of such a practice. For example, when Rose was asked if she shares the assessment criteria with their students, she responded, "I think it would be a good way to use with students, but I just do not use it. I do not know, but I might use it in the future!" Zeina also shared this perception. She stated:

I do not know why I do not do that. I just never thought about it, you know what, I should use it! I will try it next term and I will see how it works with my students. I think it may help them to focus on what I want them to learn.

Thus, a lack of assessment knowledge about formative assessment and how to use it in the classroom may contribute to teachers' lack of understanding and implementing of formative assessment.

Overall, the findings related to Research Question 3 revealed that teachers were favorably inclined towards formative assessment or assessment for learning. The findings also showed that teachers' inclination might not be clearly matched by their assessment practices. This result might correspond with previous studies on EFL teachers' beliefs and practices that found that although teachers hold strong beliefs that assessment should be used to improve teaching practices, there was a tendency among teachers to use summative procedures in their classrooms (Muñoz et al., 2012). The discrepancy between beliefs and practices has also been documented in other studies (Cheng et al., 2004; López & Bernal, 2009), where teachers believed they should assess their students based

on their ability to use the language communicatively, and yet their practices showed they mainly focus on vocabulary and grammar tests.

The mismatch between teachers' perceptions about assessment for learning/formative assessment and their reported assessment practices may reveal a lack of knowledge and training on classroom assessment. This finding highlights the need for developing teacher education and teacher professional development programs in Jordan - to better provide teachers with the adequate knowledge and training to better conduct classroom assessment. These programs should focus on using assessment as a useful tool to enhance both; teachers' teaching and students' learning.

Chapter 5 Conclusion

This chapter summarizes the findings of the study, discusses the limitations and implications of the study, and provides some directions for further research.

Summary of Findings

This study explored the practices and perceptions of Jordanian EFL teachers regarding classroom-based assessment (CBA). The findings revealed that the teachers use various assessment methods in their classrooms for varying purposes and with varying frequencies. Teachers conducted assessment in order to achieve different purposes related to (1) students' learning, e.g., achievement ; (2) instruction, e.g., to diagnostic strengths and weaknesses of their instruction and teaching; (3) administration, e.g., to provide information about students' evaluation to school administration; and (4) preparing students to the Tawjihi.

The findings also showed that teachers used a variety of methods to assess the four language skills—reading, writing, listening, and speaking. For reading, the most frequently used assessment methods were read aloud and teacher-made tests with primarily short answer and multiple-choice items while the least frequently used method was student portfolios. For writing, teachers frequently used teacher-made tests that involved editing a sentence/paragraph and assessing short essays while methods such as long essays and student journal were used less frequently. Regarding listening and speaking, the results showed that teachers mostly used teacher-made tests that required students to take notes and answer multiple-choice items after listening to a passage. Teachers more frequently used student-centered methods, however, such as providing an

oral description for an event or thing, giving oral directions, and self-assessment. The use of standardized tests was more frequent for assessing speaking than the other three skills.

The findings suggest that although teachers seem to use a variety of assessment methods, their most frequently reported methods were teacher-made tests involving items similar to the Tawjihi exam format, especially in reading and writing. Such tests may not reflect what students know and are able to do with the language, particularly since not all skills are assessed in the Tawjihi (not listening and speaking). Besides the washback of the Tawjihi exam, this tendency may suggest a lack of assessment knowledge that enables teachers to use various types of assessment methods, including summative and formative, since the literature suggests that both types of assessment should be used in a balanced way.

Teachers' decision-making about which type of assessment methods to use seemed to be influenced by several factors. Interviews suggested that the main factor that impacted the type of assessment used by teachers was the National Exam, the Tawjihi, which students take at the end of Grade 12. Teachers reported that they usually try to help their students to practice for the Tawjihi by designing tests that include the types of questions that they can expect on the Tawjihi (e.g., multiple-choice questions and editing a paragraph). Moreover, teachers said that in their instruction and assessment they usually focus on the skills that are assessed in the National Exam, i.e., reading and writing, more so than other skills, i.e., listening and speaking.

A second factor in teacher decision-making about assessment method was students' level of proficiency. Teachers reported that they usually consider their students' proficiency level when choosing and designing assessment tasks. For example, they tend

to design tests that are not challenging for students, especially in classes that have a majority of weak students.

A third factor impacting teacher's choice of assessment method related to the curriculum and the assigned textbook. Teachers reported that the use of a textbook constrained how they planned for their instruction and assessment since they have to follow and cover all of the material in the textbook within the assigned time. This factor also appeared to have a connection to the influence of the National Exam since the assigned textbook is the only required material to study for the exam.

A fourth and final factor influencing how teachers choose their assessment methods was teachers' knowledge about assessment. The majority of teachers reported that they had never attended a course or workshop on assessment and that they had a basic knowledge about assessment from the brief mandatory course which was designed for starting teachers in public schools.

With regards to teachers' perceptions about assessment for learning/formative assessment, the results showed that the majority of teachers either agree or strongly agree with all of the principles entailed assessment for learning including using assessment results to modify instruction, providing students with feedback, sharing the assessment criteria with students, and involving students in self- and peer assessment. The results also showed that teachers' perceptions might not be clearly reflected in their practices in terms of the most frequent purposes and methods that they reported using in their classrooms.

Limitations of the Study

One of the limitations of the current study was related to the difficulty of recruiting teachers. The teachers who participated in the questionnaire were all volunteers who were recruited online using social media networks, mainly Facebook. Although the online questionnaire format helped the researcher to reach the intended group of participants, the number of participants who submitted a complete questionnaire was only 54. Having a larger number of participants would have made it possible to generalize the results to the larger population of EFL teachers in Jordan.

Another limitation was related to the online follow-up interviews. While four teachers volunteered to participate, in the end the researcher was able to conduct interviews with only three of them, as the fourth teacher withdrew before the interview. Although the data from the interviews provided some explanations to the results of the questionnaire, having more participants may have revealed more information and perhaps provided more clarification of the questionnaire results.

Moreover, all three teachers had less than six years of teaching experience. Research shows that teaching experience is one of the factors that may influence teachers' beliefs and practices (Breen, 1991; Borg, 2006). For example, novice and experienced teachers differ in how they plan and conduct their lessons, the decisions they make in the classrooms, and the content they cover (Tsui, 2003). Thus, the results might have differed had interviews been conducted with teachers with more years of teaching experience.

The final limitation of the current study lies in the complexity of examining teachers' practices and beliefs using self-reported methods because what teachers report

about themselves might not correspond with what they actually do in the classroom.

Woods (1996) remarked:

When a belief or assumption is articulated in the abstract, as a response to an abstract question, there is a much greater chance that it will tend more towards what is expected in the interview situation than what is actually held in the teaching situation and actually influences teaching practices. (p. 27)

Therefore, using another data collection tool, such as classroom observation, may have offered a more insights into teachers' practices and beliefs and the relationship between them.

Recommendations for Further Research

This study suggested that Jordanian EFL teachers use a wide range of assessment methods to assess their students' competence in English including both summative and formative; this result was based on what teachers reported about their classroom practices. However, there were some contradictions between what teachers reported in the questionnaire and what was stated in the interviews. For example, some teachers in the questionnaire reported using student portfolio while all teachers in the interviews stated that this assessment method is not used in Jordanian public schools. Thus, responses to the questionnaire should be interpreted with caution and future research could use classroom observation to reveal more information about how teachers use assessment in their classrooms.

Moreover, future research could adopt other approaches, such as case study or ethnography, to better capture the complexities and dynamics of a teacher's practices in

naturalistic classroom settings over time. This may help to obtain a more contextualized, in-depth, and holistic understanding of classroom-based assessment. It could also be beneficial to include other stakeholders. This investigation included only teachers, but it could be valuable to include other stakeholders, such as educational policymakers, to explore their perspectives about educational policies regarding assessment and teacher training in general and assessment in particular.

Future research could also consider including students to investigate their attitudes and perceptions about the assessment methods and procedures used by their teachers in the classroom. This may help to shed light on the role of assessment in promoting students' learning and the impact of assessment on students' motivation.

Finally, the aim of this study was to explore teachers' assessment practices and beliefs regarding classroom-based assessment without focusing on a particular type of assessment, such as formative assessment. Future research may consider investigating teachers' use of formative assessment specifically since this type might be less frequently used by teachers in EFL contexts because of their lack of assessment knowledge, as this study found.

Implications of the Study

In closing, classroom-based assessment plays an important role in teaching and learning a second/foreign language. This study may help to raise awareness among ESL/EFL teachers and other stakeholders, such as educational policy makers, about the importance role of assessment in the language classroom through understanding how teachers conduct classroom-based assessment. Moreover, addressing teachers'

assessment knowledge is essential to improving their practices and considering both summative and formative assessment to better support students' learning.

Finally, the study suggested that there is a need for professional development programs to better equip both pre- and in-service teachers with the knowledge needed to conduct assessments in their classrooms. These programs have the potential to assist teachers to reflect on their practices and beliefs and recognize the impact that beliefs have on their classroom behaviour and decision-making. It may be that these preparation programs should be longer and more in-depth, unlike the current mandatory training course for new teachers in Jordan, which has little focus on assessment. Programs would ideally be long-term, collaborative, and supportive and provide teachers with the knowledge and support to better use assessment tools for different purposes. Integrating assessment with teaching and instruction in teacher professional programs may help to highlight the role of assessment in improving both teaching and learning.

References

- Abdo, I. & Breen, G. (2010). Teaching EFL to Jordanian students: New strategies for enhancing English acquisition in a distinct Middle Eastern student population. *Creative Education*, 1, 39-50.
- AlAlili, S. (2014). *Reforming English curriculum in United Arab Emirates: An examination of Emirate teachers' beliefs and practices regarding the adoption of "English continuous assessment rich task" (ECART)* (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing. (3618609)
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297.
- Alderson, J. & Banerjee, J. (2001). Language testing and assessment. *Language Teaching*, 34(4), 213-236.
- Alhabahba, M. M., Pandian, A., & Mahfoodh, O. H. A. (2016). English language education in Jordan: Some recent trends and challenges. *Cogent Education*, 3(1)
- Al-Hazmi, S., (2003). EFL teacher preparation programs in Saudi Arabia: Trends and challenges. *TESOL Quarterly*, 37, 341–344.
- Al-Jamal, D., & Ghadi, N. (2008). English language general secondary certificate examination washback in Jordan. *The Asian EFL Journal*, 10(3), 158-186.

- Al-Khairiy, M. A. (2013). Saudi English-major undergraduates' academic writing problems: Al Taif university perspective. *English Language Teaching*, 6(6), 1–12.
- Al-Khatib, M. A. (2008). Innovative second and foreign language education in the Middle East and North Africa. In N. V. Deussen-Scholl & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education* (2nd ed., Vol. 4) (pp. 227–237). New York, NY: Springer Science.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback—a case-study. *System*, 30(2), 207-223.
- Al Sawalmeh, Y. (2014). Teacher preparation and recruitment policies: The Jordanian case. Tunisia: ALESCO. Retrieved from www.alecso.org/siteold/images/2016files/educationfiles/takwin%20siyeset.pdf
- Assessment Reform Group (2002). *Assessment for learning: Ten principles. Research-based principles to guide classroom practice*. Retrieved from: <https://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf>
- Azadi, G., & Gholami, R. (2013). Feedback on washback of EFL tests on ELT in L2 classroom. *Theory and Practice in Language Studies*, 3(8), 1335-1341.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- 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.

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and Developing Useful Language Tests*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, UK: Oxford University Press.
- Bailey, K. (1985). If I had known then what I know now: Performance testing of foreign teaching assistants. In P. Hauptman, R. LeBlanc, & M. Wesche (Eds.), *Second language performance testing* (pp. 135-180). Ottawa: University of Ottawa Press.
- Baniabdelrahman, A. A. (2010). The effect of the use of self-assessment on EFL students' performance in reading comprehension in English. *TESL-EJ: Teaching English as a Second or Foreign Language*, 14(2). Retrieved from <http://www.tesl-ej.org/wordpress/issues/volume14/ej54/ej54a2/>
- Birenbaum, M., & Dochy, F. (1996). *Alternatives in assessment of achievements, learning processes and prior knowledge: Evaluation in education and human services*. Boston, MA: Kluwer Academic Publishers.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
- Bliem, C. L., & Davinroy, K. (1997). *Teachers' beliefs about assessment and instruction in literacy*. Los Angeles: National Center for Research on Evaluation, Standards,

and Student Testing (CRESST), Graduate School of Education & Information Studies, University of California.

Borg, S. (2006). *Teacher cognition and language education: Research and practice*. New York, NY: Continuum.

Breen, M. P., Barratt-Pugh, C., Derewianka, B., House, H., Hudson, C., Lumley, T., & Rohl, M. (1997). *How teachers interpret and use national and state assessment frameworks*. Canberra, AU: Department of Employment, Education, Training and Youth Affairs

Breen, M. P. (1991). Understanding the language teacher. In R. Phillipson, E. Kellerman, L. Selinker, M. S. Smith, & M. Swain (Eds.), *Foreign/Second Language Pedagogy Research* (pp. 213-233). Bristol, PA: Multilingual Matters Ltd.

Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing*, 18(4), 393-407.

Broadfoot, P. M., Daugherty, R., Gardner, J., Gipps, C. V., Harlen, W., James, M., & Stobart, G. (1999). *Assessment for learning: Beyond the black box*. Cambridge, UK: University of Cambridge School of Education.

Brookhart, M. S. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, 106(3), 429-458.

Brown, G. T. L. (2003). *Teachers' instructional conceptions: Assessment's relationship to learning, teaching, curriculum, and teacher efficacy*. Paper presented at the Joint

Conference of the Australian and New Zealand Associations for Research in Education, Auckland, New Zealand.

Brown, G., T. L. (2002). *Teachers' conceptions of assessment* (Unpublished doctoral dissertation). University of Auckland, New Zealand.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.

Celce-Murcia, Marianne. 2007. Rethinking the role of communicative competence in language teaching. In E. A. Soler & M. P. S. Jordà (Eds.), *Intercultural Language Use and Language Learning*, (pp.41–57). Dordrecht: Springer.

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: a pedagogically motivated model with content specifications. *Issues in Applied Linguistics* 6(2). 5–35.

Cizek, G. (2000). Pockets of resistance in the assessment revolution. *Educational Measurement: Issues and Practice*, 19(3), 19-23.

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3(2), 159-179.

- Chang, C.W. (2005). *Oral language assessment: Teachers' practices and beliefs in Taiwan collegiate EFL classrooms with special reference to Nightingale University* (Unpublished doctoral dissertation). University of Exeter, UK.
- Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In: L. Cheng, Y. Watanabe, & A. Curtis (Eds.). *Washback in language testing: Research contexts and methods* (pp. 147–70). New York: Erlbaum.
- Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Language testing and assessment* (Vol. 7, 2nd Ed.) (pp.1–13). Chester: Springer Science Business Media.
- Cheng, L. (2011). Supporting student learning: Assessment of learning and assessment for learning. In D. Tsagari & I. Csépes (Eds.), *Classroom-based language assessment* (pp. 191-303). Frankfurt: Peter Lang.
- Cheng, L. (2014). Consequences, impact, and washback. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1130–46). Chichester: John Wiley & Sons.
- Cheng, L., & Fox, J. (2013). Review of doctoral research in language assessment in Canada (2006-2011). *Language Teaching*, 46(4), 518-544.
- Cheng, L., & Fox, J. D. (2017). *Assessment in the language classroom: Teachers supporting student learning*. London, UK: Palgrave.

- Cheng, L., Rogers, T. & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods and procedures. *Language Testing*, 21(3), 360–389.
- Cheng, L., Rogers, T. & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, 33(1), 9–32.
- Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, 4(1), 85–107.
- Colby-Kelly, C., & Turner, C. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking format assessment to the next level. *Canadian Modern Language Review*, 64(1), 9-38.
- Coombe, C., Davidson, P., O'Sullivan, B., & Stoyhoff, S. (2012). *The Cambridge Guide to Second Language Assessment*. Cambridge, UK: Cambridge University Press.
- Coombe, C. (2018). Introduction to assessment and evaluation. In J. I. Liantas (Ed.) *The TESOL Encyclopedia of English Language Teaching: Assessment and evaluation* (pp. 1-3). doi: 10.1002/9781118784235.eeltv08
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff (Eds.), *The Cambridge Guide to Second and Foreign Language Assessment* (pp. 20-29). New York: Cambridge University Press.

- Creswell, J. (2015). *A concise introduction to mixed methods research*. Thousand Oaks, CA: SAGE.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: SAGE.
- Cumming, A. (2001). ESL/EFL instructors' practices for writing assessment: Specific purposes or general purposes? *Language Testing*, 18(2), 207-234.
- Cumming, A. (2004). Broadening, deepening, and consolidating. *Language Assessment Quarterly*, 1(1), 5-18.
- Cziko, G. A. (1981). Psychometric and edumetric approaches to language testing: Implications and applications. *Applied Linguistics*, 2(1), 27-44.
- Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, 21(2), 149-166.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge, UK: Cambridge University Press.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL Teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing* 21(3), 305-334.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly* 43(3), 393-415.

- Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. New York: Oxford University Press.
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). New York: Routledge.
- Falvey, P., & Cheng, L. (2010). Paradigm shift: Teachers' classroom assessment practices in Hong Kong and the PRC. *Melbourne Papers in Language Testing*, 10(3), 29-51.
- Fox, J. (2009). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes*, 8(1), 26–42.
- Fox, J. & Hartwick, P. (2011). Taking a diagnostic turn: Reinventing the portfolio in EAP classrooms. In D. Tsagari and I. Csépes (Eds.), *Classroom-based language assessment* (pp. 47–62). Frankfurt: Peter Lang.
- Fox, J. (2014). Portfolio based language assessment (PBLA) in Canadian immigrant language training: Have we got it wrong? *Contact, Special Research Symposium Issue*, 40(2), 68–83.
- Freeman, D. (1993). Renaming experience/reconstructing practice: Developing new understandings of teaching. *Teaching and Teacher Education*, 9, 485-497.

- Gipps, C. V. (1994). *Beyond Testing: towards a Theory of Educational Assessment*. London: Falmer Press.
- Giraldo, F. (2018). A diagnostic study on teachers' beliefs and practices in foreign language assessment. *Íkala*, 23(1), 25-44.
- Hamilton, J., Lopes, M., McNamara, T., & Sheridan, E. (1993). Rating scales and native speaker performance on a communicatively oriented EAP test. *Melbourne Papers in Applied Linguistics*, 2, 1-24.
- Harmer, J. (2010). *The Practice of English Language Teaching* (4th ed.). London: Pearson Longman.
- Hidri, S. (2018). Discrete point and integrative testing. In J. I. Liantas (Ed.). *The TESOL Encyclopedia of English Language Teaching: Assessment and evaluation*. DOI: 10.1002/9781118784235
- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395-420.
- Hillegas, M. B. (1912). A scale for the measurement of quality in English composition by young people. *Teachers College Record*, 13(4), 1-55.
- Hirai, A., & Koizumi, R. (2009). Development of a practical speaking test with a positive impact on learning using a story retelling technique. *Language Assessment Quarterly*, 6(2), 151-167.

- Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5(1), 8–11.
- Hudson, T., & Lynch, B. (1984). A criterion-referenced measurement approach to ESL achievement testing. *Language Testing*, 1(2), 171-201.
- Inbar-Lourie, O. (2008a). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402.
- Inbar-Lourie, O. (2008b). Language assessment culture. *Encyclopedia of language and education* (2nd ed., pp. 285–300). New York: Springer.
- Jia, Y., Eslami, Z. R., & Burlbaw, L. M. (2006). ESL teachers' perceptions and factors influencing their use of classroom-based reading assessment. *Bilingual Research Journal*, 30(2), 407-430.
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555-584.
- Johnson, K.E (1992). The relationship between teachers' beliefs and practices during literacy instruction for non-native speakers of English. *Journal of Reading Behavior*, 24(1), 83-108.
- Johnson, K. E. (1994). The emerging beliefs and instructional practices of pre-service English as a Second Language teachers. *Teaching & Teacher Education*, 10(4), 439-452.

- Jordan Ministry of Education (MOE). (July 22, 2012). *Beginning teachers' program*. Retrieved from <http://www.moe.gov.jo/ar/node/885>
- Koh, K., Burke, L. E. C., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 22(3), 264-288.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman.
- Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18(4), 351-372.
- Lynch, B., & Davidson, F. (1994). Criterion-referenced language test development: Linking curricula, teachers and tests. *TESOL Quarterly*, 28 (4), 727-743.
- Manjarrés, N. B. (2005). Washback of the foreign language test of the state examinations in Colombia: A case study. *Arizona Working Papers in SLAT*, 12,1-19.
- Madaus, G.F. (1988). The influence of testing on the curriculum. In L. N. Tanner & K. J. Rehaag (Eds.), *Critical issues in curriculum: Eighty-seventh yearbook of the National Society for the Study of Education (Part 1)*. Chicago, IL: University of Chicago Press.
- Malone, M. (2008). *Training in language assessment*. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education (Language testing and assessment (Vol. 7, pp. 273–284)*. New York: Springer.

- McMillan, J. H. (2015). Classroom assessment. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., Vol 3) (pp. 819-824). Oxford: Elsevier. Doi: [10.1016/B978-0-08-097086-8.92074-9](https://doi.org/10.1016/B978-0-08-097086-8.92074-9)
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43.
- Mertler, C. A. (2005). ERRATUM: Secondary teachers' assessment literacy: Does classroom experience make A difference? *American Secondary Education*, 33(2), 76-3.
- Marso, R. N., & Pigge, F. L. (1993). Teacher training in assessment: Overcoming the neglect. In S. L. Wise & J. C. Conoley (Eds.), *Teachers' testing knowledge, skills, and practices* (pp. 129-186). Lincoln, NE: Buros Institute of Mental Measurements.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks: Sage.
- Mitchell, R. & Myles, F. (2004). *Second language learning theories* (2nd ed.). London: Arnold.
- Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile: Issues in Teachers' Professional Development*, 14(1), 143-158.
- Nation, I. S. (2013). *Learning vocabulary in another language* (2nd ed.). New York: Cambridge University Pres.

- Nunan, D. (1999). *Second language teaching and learning*. Boston, MA: Heinle and Heinle Publishers.
- Nurani, S., & Rosyada, A. (2015). Improving English pronunciation of adult ESL learners through reading aloud assessments. *Lingua Cultura*, 9(2), 107-112.
- Obeiah, S. F. & Bataineh, R. F. (2016). The effect of portfolio-based assessment on Jordanian EFL learners' writing performance. *Bellaterra Journal of Teaching & Learning Language & Literature*, 9(1), 32-46. Retrieved from <https://www.raco.cat/index.php/Bellaterra/article/view/306905>
- Oller, J. W., Jr. (1976). Language testing. In R. Wardaugh & H. D. Brown (Eds.), *A survey of applied linguistics* (pp. 275-300). Ann Arbor: University of Michigan Press.
- Pajares, M. (1992). Teacher's beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62, 307-332.
- Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (171-90). New York: Erlbaum
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142-173

- Queen Rania Center for Educational Technology and Information. (2018). *The statistical report of the academic year 2017/2018*. Retrieved from http://www.moe.gov.jo/sites/default/files/ltqryr_lhsyy_llm_ldrsy2017-2018naskh_nhyy.pdf
- Queen Rania Techer Academy (QRTA). (2015). *Teacher education professional diploma*. Retrieved from <http://www.qrta.edu.jo/ited2018.html>
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249-258.
- Rea-Dickins, P. (2007). Learning or measuring? Exploring teacher decision-making in planning for classroom-based assessment. In S. Fotos & N. Nassaji (Eds.), *Form-focused instruction and teacher education* (pp. 195-211). Oxford, UK: Oxford University Press.
- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). New York: Cambridge University Press.
- Richards, J. C., & Schmidt, R. W. (2010). *Longman Dictionary of Language Teaching and Applied Linguistics* (4th ed.). London: Routledge.
- Sadler, R. (1998). Formative assessment: Revising the territory. *Assessment in Education*, 5(1), 77-84.

- Savignon, S. J. (1983). *Communicative competence: theory and classroom practice*. Reading, MA: Addison-Wesley.
- Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, 9(3), 15-22.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Shim, K. N. (2008). *An investigation into teachers' perceptions of classroom-based assessment of English as a foreign language in Korean primary education* (Doctoral dissertation). Retrieved from <https://ore.exeter.ac.uk/repository/bitstream/handle/10036/67553/ShimKN.doc..pdf?sequence=2>
- Shohamy, E. (1992). New modes of assessment: The connection between testing and learning. In E. Shohamy, & R. Walton (Eds.), *Language assessment for feedback: Testing and other strategies* (pp. 7-28). Dubuque, IA: Kendall-Hunt.
- Shohamy, E. (1996). Language testing: Matching assessment procedures with language knowledge. In M. Birenbaum, & F. Dochy, (Eds.). *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 142-160). Boston, MA: Kluwer Academic Publishing.
- Shohamy, E. (2006). *Language Policy: Hidden Agendas and New Approaches*. Oxon: Taylor & Francis.

- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing* 13(3), 298–317.
- Song, E., & Koh, K. (2010). *Assessment for learning: Understanding teachers' beliefs and practices*. Paper presented at the 36th Annual Conference of the International Association of Educational Assessment, Bangkok, Thailand.
- Retrieved from https://repository.nie.edu.sg/bitstream/10497/16733/1/IAEA-2010-SongE_a.pdf
- Spolsky, B. (2017). History of language testing. In E. Shohamy, I. Or, & S. May (Eds.), *Encyclopedia of language and education: Language testing and assessment* (3rd ed.) (pp. 375-384). Cham, Switzerland: Springer.
- Spolsky, B. (1985). What does it mean to know how to use a language? An essay on the theoretical basis of language testing. *Language Testing*, 2(2), 180-191.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Stiggins, R. J. (1991). Assessment literacy. *The Phi Delta Kappan*, 72(7), 534-539.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *The Phi Delta Kappan*, 77(3), 238-245.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *The Phi Delta Kappan*, 83(10), 758-765.
- Stiggins, R. J. (2005). *Student-involved assessment for learning* (4th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

- Stoyhoff, S. (2009). Recent developments in language assessment and the case of four large-scale tests of ESOL ability. *Language Teaching*, 42(1), 1-40.
- Tsagari D., Cheng L. (2017) Washback, impact, and consequences revisited. In E. Shohamy, I. Or, & S. May (Eds.), *Encyclopedia of language and education: Language testing and assessment* (Vol. 10, 3rd ed) (pp. 359-372). New York: Springer, Cham. DOI: 10.1007/978-3-319-02261-1_24
- Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing*, 17(2), 163-184.
- Thorndike, E. L. (1911). A scale for merit in English writing by young people. *Journal of Educational Psychology*, 2, 361-368.
- Troudi, S., Coombe, C., & Al-Hamly, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546-555.
- Tsui, A. B. (2003). *Understanding expertise in teaching: Case studies of ESL teachers*. New York: Cambridge University Press.
- Tsushima, R. (2015). Methodological diversity in language assessment research: The role of mixed methods in classroom-based language assessment studies. *International Journal of Qualitative Methods*, 14(2), 104-121.
- Vandergrift, L. (1999). Facilitating second language listening comprehension: acquiring successful strategies. *ELT Journal*, 53(3), 168-176.

- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*.
Cambridge, UK: Cambridge University Press.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study.
Language Testing, 10(1), 41-69.
- Wang, X. (2017). A Chinese EFL teacher's classroom assessment practices. *Language Assessment Quarterly*, 14(4), 312.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination?
Preliminary findings from classroom-based research. *Language Testing*, 13, 318–
333.
- White, C., Sturtevant, E., & Dunlap, K. (2003). Preservice and beginning teachers'
perceptions of the influence of high stakes tests on their literacy-related
instructional beliefs and decisions. *Reading Research and Instruction*, 42(2), 39-
62.
- White, E. (2009). Are you assessment literate? Some fundamental questions regarding
effective classroom-based assessment. *OnCue Journal*, 3(1), 3-25.
- Widdowson, H.G. (1983). *Learning purpose and language use*. Oxford, UK: Oxford
University Press.
- Wigglesworth G., Frost K. (2017) Task and performance-based assessment. In E.
Shohamy, I. Or, & S. May (Eds.), *Encyclopedia of language and education:
Language testing and assessment* (3rd ed.) (pp. 121-133). Cham: Springer.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.

Wiliam, D., & Thompson, M. (2008). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Lawrence Erlbaum Associates.

Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. *Review of Research in Education* 17, 31–74.

Woods, D. (1996). *Teacher cognition in language teaching: Beliefs, decision-making and classroom practice*. Cambridge, UK: Cambridge University Press.

Yin, M. (2010). Understanding classroom language assessment through teacher thinking research. *Language Assessment Quarterly*, 7(2), 175.

Zamel, V., & Spack, R. (2004). *Crossing the curriculum: Multilingual learners in college classrooms*. Mahwah, NJ: Lawrence Erlbaum.

Appendix A

Certificate of Ethics Clearance



Office of Research Ethics
5110 Human Computer Interaction Bldg | 1125 Colonel By Drive
Ottawa, Ontario K1S 5B6
613-520-2600 Ext: 2517
ethics@carleton.ca

CERTIFICATION OF INSTITUTIONAL ETHICS CLEARANCE

The Carleton University Research Ethics Board-A (CUREB-A) has granted ethics clearance for the research project described below and research may now proceed. CUREB-A is constituted and operates in compliance with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS2).

Ethics Protocol Clearance ID: Project # 109815

Project Team Members: Mrs. Malak Swaie (Primary Investigator)
Prof. Janna Fox (Research Supervisor)

Project Title: An Investigation of EFL Teachers' Practices and Beliefs about Classroom Assessment

Funding Source (If applicable):

Effective: November 21, 2018

Expires: November 30, 2019.

Please ensure the study clearance number is prominently placed in all recruitment and consent materials: CUREB-A Clearance # 109815.

Restrictions:

This certification is subject to the following conditions:

1. Clearance is granted only for the research and purposes described in the application.
2. Any modification to the approved research must be submitted to CUREB-A via a Change to Protocol Form. All changes must be cleared prior to the continuance of the research.
3. An Annual Status Report for the renewal of ethics clearance must be submitted and cleared by the renewal date listed above. Failure to submit the Annual Status Report will result in the closure of the file. If funding is associated, funds will be frozen.
4. A closure request must be sent to CUREB-A when the research is complete or terminated.
5. During the course of the study, if you encounter an adverse event, material incidental finding, protocol deviation or other unanticipated problem, you must complete and submit a Report of Adverse Events and Unanticipated Problems Form, found here:
<https://carleton.ca/researchethics/forms-and-templates/>

Failure to conduct the research in accordance with the principles of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans 2nd edition* and the *Carleton University Policies and*

Appendix B

The Jordanian Ministry of Education Approval (Arabic)



وزارة التربية والتعليم



الرقم ٥٤٥٥٨١١٠/٣
التاريخ ١٥ صفر ١٤٤٠
الموافق ٢٠١٨/١٠/٢٥

السيد مدير التربية والتعليم

الموضوع: البحث التربوي

السلام عليكم ورحمة الله وبركاته، وبعد؛

فأرجو العلم بأن الطالبة ملاك تركي سواعي تقوم بإجراء دراسة عنوانها "دراسة استقصائية حول أساليب الاختبار والتقييم لدى معلمي اللغة الإنجليزية والاعتقادات المتعلقة بها في المرحلة الثانوية"، استكمالاً لمتطلبات الحصول على درجة الماجستير في تخصص علم اللغويات التطبيقية من جامعة كارلتون/ كندا، ويحتاج ذلك إلى تطبيق أدوات الدراسة على عينة من معلمي اللغة الإنجليزية في المدارس الثانوية التابعة لمديرتكم. راجياً تسهيل مهمة الطالبة المذكورة، وتقديم المساعدة الممكنة لها، على أن تتم مطابقة الأدوات المطبقة مع الأدوات المرفقة، وألاً تُستخدم البيانات والمعلومات المتحصلة إلا لأغراض البحث العلمي.

واقبلوا الاحترام

وزير التربية والتعليم

الدكتور ياسر العمري

مدير البحث والتطوير بالوكالة



نسخة/ لمدير إدارة التخطيط والبحث التربوي
نسخة/ لمدير البحث والتطوير التربوي بالوكالة
نسخة/ لرئيس قسم البحث التربوي
نسخة/ الملف ١٠/٣
المرفقات: (٦) صفحات

Appendix C

Questionnaire

PART 1: Background Information

- **Gender:** Male Female Other
- **Age:** 26-30 31-35 36-40 41-45 46-50 over 50
- **Years of teaching experience:**
 - less than 5 years
 - 6-10 years
 - 11-15 years
 - 16-20 years
 - more than 20 years
- **Educational qualifications:**
 - BA in English language and literature
 - BA in English Language/Field Teacher
 - Higher Diploma
 - MA/ MEd
 - PhD/ EdD
 - Other _____
- **How many classes are you teaching per week at this time?** _____
- **What is the average size of your classes?** _____
- **Have you attended a course or a workshop on assessment and evaluation? *If yes, please specify the type of course (undergraduate/graduate/special) or the workshop (who provided it) and how many hours.***

- **How do you perceive your knowledge about language assessment and evaluation?**
 - Excellent
 - Very good
 - Good
 - Poor knowledge
 - No knowledge

PART 2: Purposes of Assessment

I use assessment in my classroom for the following purposes/reasons:					
	Always	Usually	Often	Rarely	Never
	100%	75%	50%	25%	0%

1. To group my students for instruction purposes in my class.					
2. To obtain information on my students' progress.					
3. To plan my instruction.					
4. To diagnose strengths and weaknesses in my own teaching and instruction.					
5. To provide feedback to my students as they progress through the course.					
6. To motivate my students to learn.					
7. To make my students work harder.					
8. To diagnose strengths and weaknesses in my students.					
9. To formally document growth in learning.					
10. To determine the final grades for my students.					
11. To provide information to the school administration.					
12. To provide information to an outside funding agency.					

PART 3: Methods of Assessment

A-Reading Assessment

I use the following methods to assess reading:					
	Always 100%	Usually 75%	Often 50%	Rarely 25%	Never 0%
1. Read aloud/dictation					
2. Teacher-made tests containing					
a. cloze items					
b. sentence completion items					
c. true-false items					
d. matching items					
e. multiple-choice items					
f. short answer items					
3. Student summaries of what is read					
4. Student portfolio					
5. Peer assessment					
6. Self-assessment					
7. Standardized reading tests					

B. Writing Assessment

I use the following methods to assess writing:					
	Always 100%	Usually 75%	Often 50%	Rarely 25%	Never 0%
1. Teacher-made tests containing					
a. true-false items					
b. matching items					
c. multiple-choice items to identify grammatical error(s) in a sentence					
d. editing a piece of writing such as a sentence or a paragraph					
e. short essay					
f. long essay					
2. Student journal					
3. Peer assessment					
4. Self-assessment					
5. Student portfolio					
6. Standardized writing tests					

C. Speaking and Listening Assessment

I use the following methods to assess speaking and listening:					
	Always 100%	Usually 75%	Often 50%	Rarely 25%	Never 0%
1. Oral reading/dictation					
2. Oral interviews=dialogues					
3. Oral discussion with each student					
4. Oral presentations*					
5. Teacher-made tests asking students to					
a. Give oral directions					
b. Follow directions given orally					
c. Provide an oral description of an event or object					
d. Prepare summaries of what is heard					
e. Answer multiple-choice test items following a listening passage					
f. Take notes					

g. Retell a story after listening to a passage					
6. Peer assessment					
7. Self assessment					
8. Standardized speaking tests					
9. Standardized listening tests					

PART 4: Beliefs about Language Assessment

	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
1- - Assessment should be always associated with marks.					
2- Assessment results should be used to modify instruction to better support students' learning					
3-Students should be given opportunities to reflect on their learning through peer feedback and self-assessment.					
4- Students should be given feedback which enables them to know how to improve their work and take their learning forward.					
5- Teachers should negotiate the assessment criteria with their students in advance.					
6- Teachers should first identify the purpose of the assessment when they design the assessment task/test.					
7- Students will learn the most and perform their best if they know in <i>advance</i> the criteria that will be applied in judging a performance.					
8- Involving students in self- and peer assessment activities can enhance their achievement and motivation.					

-When you give feedback to your students during the course, how do you provide that feedback? Please check all that apply.

a. Verbal feedback

b. Checklist

c. Written comments

d. Conference with student

e. Total test score

f. A letter grade

g. Other _____

Appendix D

Interview Questions

- 1- In your opinion, what is the main goal of classroom assessment?
- 2- What types of assessment methods do you use to assess students' reading competence in English language in your classroom? What is the most frequent one you use and why?
- 3- What types of assessment methods do you use to assess students' writing competence in English language in your classroom? What is the most frequent one you use and why?
- 4- What types of assessment methods do you use to assess students' speaking competence in English language in your classroom? What is the most frequent one you use and why?
- 5- What types of assessment methods do you use to assess students' listening competence in English language in your classroom? What is the most frequent one you use and why?
- 6- When designing an English language test, is there any skill you focus on more than others? If yes, why?
- 7- Do you think that students should be assessed based on what they know about English language (i.e. grammar) or what they can do with language (i.e. speaking)?
- 8- Do you think that you have sufficient knowledge about how to design and use evaluation tools? Why?
- 9- Do you think that providing teachers with course and workshops on assessment can help them better conduct classroom assessment or do you think they learn that through their own experience in teaching? Why?
- 10- When designing an English language test, what are the sources you use to choose test items? Which one do you use the most?
- 11- What percentage of your assessments and evaluations are based on coursework and on tests and examinations?
- 12- In your opinion, how important is it to provide students with feedback during the course? What methods do you use for that?