EFL TEACHERS' BELIEFS AND PRACTICES ABOUT CLASSROOM ASSESSMENT:
A MULTIPLE CASE STUDY IN THE CONTEXT OF KUWAIT

SHAIMA DASHTI

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN LINGUISTICS AND APPLIED LINGUISTICS
YORK UNIVERSITY
TORONTO, ONTARIO

June 2019

Abstract

Classroom language assessment is a recent topic of interest in education research. Yet, few studies have examined teachers' beliefs concerning language classroom assessment or the relationship between teachers' assessment practices and their beliefs. In addition, little research has situated classroom assessment in a specific theoretical approach, especially in the postsecondary English as a foreign language (EFL) context. In this study I investigated the beliefs and practices of EFL teachers regarding classroom assessment, using a social constructivist approach to examine the way contextual factors influence those teachers' assessment beliefs and practices (Shepard, 2000). I also investigated how teachers' assessment practices and beliefs differ between general English (GE) and English-for-specific-purposes (ESP) courses.

This study adopted a multiple-case design using qualitative methods conducted in three data collection stages: I started by exploring teachers' assessment beliefs and practices through initial interviews. I then investigated teachers' assessment practices through classroom observations and document collection. Finally, I conducted post-observation interviews about the teachers' assessment beliefs and practices. Participants included seven EFL teachers teaching GE and ESP courses at a post-secondary institution in Kuwait. I analyzed the data using an inductive approach by analyzing each case individually as well as identifying themes emerging from the analyses.

Results showed that although teachers believed in the effectiveness of classroom assessment and implemented a variety of assessments in the classroom, they only considered summative assessment as a valid means for student evaluation. Most teachers did not identify their practices as formative assessments but considered them part of their teaching practices. The findings also revealed that various contextual factors influence teachers' assessment beliefs and practices. Those factors include the teachers' educational background and teaching experience, their beliefs about students' L2 proficiency level, the local culture, the classroom physical setting, and the assessment policies. Results also showed that teachers' assessment practices did not appear to differ greatly between GE and ESP courses. This study has implications for teachers and policy makers on how to improve assessment practices by encouraging teachers to join, and policy makers to offer, professional development programs that focus on classroom assessment. Recommendations for future research are also discussed.

Acknowledgments

First, I express my deepest gratitude to my supervisor and mentor, Dr. Khaled Barkaoui. I am grateful for your continued guidance, invaluable support, devoted time, and valuable feedback during each stage of this dissertation. Thank you for challenging and encouraging me to develop my critical thinking skills through your extensive knowledge and expertise. I have been extremely privileged to be your student. Without you, this dissertation would not have been possible.

I also extend my sincere gratitude to my supervisory committee members, Dr. Antonella Valeo and Dr. Linda Steinman, for sharing constructive feedback and providing me with warm encouragement. Moreover, I am extremely grateful to the rest of the examining committee members: Dr. Philipp Angermeyer, Dr. John Ippolito, and Dr. Beverly Baker, who made my PhD defense a pleasant learning experience through their knowledge and insightful comments.

I offer special thanks to Rose Frezza-Edgecombe, the graduate program assistant at the DLLL, for her constant guidance, substantial assistance, and valuable advice.

I would like to acknowledge the scholarship and financial support provided by the PAAET. In addition, I express my appreciation to the LC administration for facilitating my data collection procedures.

I am indebted to this study's participants, who have immeasurably contributed to this research and made it such an enjoyable experience.

I must also thank my friends in Kuwait and Canada and my extended family for their emotional encouragement and reassurance.

Many heartfelt thanks to my parents, my brother, and my sisters for their unconditional love and emotional support, which played a major role in my academic achievements.

Finally, yet most importantly, I would like to express my deepest thanks to my soulmate—my husband, Qais. Thank you for your patience, for embracing me with tremendous support, and for everything you did to keep me focused and driven. Of course, I must also thank my daughters, Dalia and Layan; you have always been my inspiration and the driving force behind my academic pursuits. Your unconditional love and understanding have kept me strong even during the hardest moments of this journey. I can't thank you enough; I would not be where I am without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

Administrative Affairs Unit	(AAU)
Beliefs, Assumptions, Knowledge	(BAK)
Certificate in Teaching English to Speakers of Other Languages	(CELTA)
College of Basic Education	(CBE)
College of Business Studies	(CBS)
College of Health Sciences	(CHS)
College of Nursing	(CoN)
College of Technological Studies	(CTS)
English as a Foreign Language	(EFL)
English as a Second Language	(ESL)
English as an Additional Language	(EAL)
English for Academic Purposes	(EAP)
English Language Teaching	(ELT)
English-for-specific-Purposes	(ESP)
General English	(GE)
Initial Interview	(II)
Kuwait University	(KU)
Local Education Authority	(LEA)
National Curriculum	(NC)
Pedagogical content knowledge	(PCK)
Post-observation Interview	(PI)
Teaching English as a Foreign Language	(TEFL)
Teaching English as a Second Language	(TESL)
Technical Affairs Unit	(TAU)
Technical and Vocational Education Department	(TVED)
Test of English as a Foreign Language	(TOEFL)
Testing and Measurement Unit	(TMU)
The Language Center	(LC)
The Public Authority for Applied Education and Training	(PAAET)
United Arab Emirates	(UAE)
United Kingdom	(UK)
United Stated	(US)

CHAPTER 1:

INTRODUCTION

This study aimed to investigate English as a foreign language (EFL) teachers' classroom

assessment beliefs and practices in the context of Kuwait. Classroom assessment plays a

significant role in teaching and learning (Black & Wiliam, 1998; Gipps, 1994). Nevertheless,

classroom assessment is only a recent topic of investigation in the assessment literature.

Furthermore, despite the growing interest in classroom assessment, studies have approached it by

focusing on assessment criteria and standards, validity and reliability, and the influence of high-

stakes mandates on classroom assessment practices (Hill, 2017). Very few studies have

addressed the "actual processes of classroom-based assessment" (Hill & McNamara, 2012, p.

396). In addition, as Black and Wiliam (1998) argued, classroom assessment has not been

grounded in one theoretical framework. Over the last two decades, classroom assessment has

been investigated within various theoretical frameworks, including psychometric, cognitive, and

constructivist perspectives (Yin, 2005).

Classroom assessment has not only attracted the attention of researchers interested in

exploring its nature in general, it has also been viewed as a tool to inform teachers about

students' strengths and weaknesses, and to help students understand their own learning progress

(McMillan, 2013; Smith, 2003). Interest has recently grown in the study of classroom assessment

from the perspective of teachers, particularly teachers' beliefs concerning classroom assessment

(e.g., Davison, 2004; Hill & McNamara, 2012; Moss, 2003; Rea-Dickins, 2001). In the past,

studies about teachers have focused on the ways teachers manage their classrooms and

schedules, plan teaching and learning activities, design assignments, provide feedback, prepare

lesson plans, and assess students' understanding of learning materials (Fang, 1996). Researchers

have placed little emphasis on the beliefs and knowledge that inform such decisions (Fang, 1996; Jia, Eslami, & Burlbaw, 2006). However, this situation has changed, and researchers of teacher development and education have become increasingly interested in what teachers know and believe as well as how their cognition influences their practices and decision-making (Allen, 2002; Borg, 2003, 2006; Clark & Peterson, 1986; Fang, 1996; Fives & Buehl, 2012; Leung, 2005; Nespor, 1987). However, studies on teachers' beliefs have mostly been concerned with teaching or learning practices, or both, rather than assessment practices (e.g., Borg, 2003). Moreover, research on the influence of teachers' assessment beliefs and practices has occurred in other fields such as mathematics rather than in the study of second- and foreign-language teaching and learning. With the growing awareness of teachers as agents in the assessment process (Leung, 2005; Rea-Dickins, 2004), however, researchers have delved into the mental world of teachers to better understand the theories teachers use to define and conduct classroom assessment activities (Leung, 2005). While some studies have found teachers' beliefs to be consistent with their practices, others have identified some lack of cohesion between teachers' assessment beliefs and their practices (Borg, 2006; Brown, 2004). Few studies appear to have established a direct relationship between contextual factors and teachers' assessment beliefs and practices (e.g., Brown, 2004; Davison, 2004). Moreover, to my knowledge, no study has directly addressed how teachers' beliefs influence their assessment practices.

The main goal of this research is thus to explore the assessment beliefs and practices of a particular group of teachers in a specific instructional context, the relationship between these teachers' beliefs and their practices, and the external factors that influence their assessment beliefs and practices. There has been a gap in investigating this topic in EFL contexts and at the postsecondary level in general, and in Kuwait in particular. Accordingly, this study aims to

bridge this gap by investigating seven EFL teachers' beliefs and practices at the postsecondary level within an EFL context: the Language Center (LC) at the Public Authority for Applied Education and Training (PAAET) in Kuwait. The study is situated within a specific theoretical framework: social constructivist theory.

1.1 Literature Overview

This research builds on two fields of language assessment: classroom assessment practices and teachers' beliefs about classroom assessment. Classroom assessment is defined as "any 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 socialization purposes" (Hill & McNamara, 2012, p. 396). Classroom assessment is embedded in teachers' everyday classroom practices, which may sometimes produce difficulty in identifying a certain activity as an assessment activity (Moss, 2003). It is important to note that in classroom assessment, teachers play the most crucial role in all assessment stages, "from planning the assessment programme to identifying and developing appropriate formative and summative assessment activities right through to making the final judgments" (Davison & Leung, 2009, p. 401). In formative assessment, teachers assess students during the instructional phase, when they are engaged with students, and through observations and monitoring during student-to-student interactions and peer activities (Katz & Gottlieb, 2012). On the other hand, summative assessment is "the process by which teachers gather evidence in a planned and systematic way in order to draw inferences about their students' learning, based on their professional judgment, and to report at a particular time on their students' achievements" (Harlen, 2005, p. 247).

Several studies have demonstrated the complex nature of teachers' assessments. Rea-Dickins (2001), for example, based on a qualitative study, provided a model cycle of teacher-based assessment that emphasized four stages: planning, implementing, monitoring, and recording and disseminating. In the first two stages, teachers aim to improve students' learning, and in the last two, they aim to meet the accountability demands dictated by policy (Rea-Dickins, 2001). Rea-Dickins and Gardner (2001) similarly found that teachers' assessments are complex and vary depending on whether they are formative or summative in purpose. Thus, decision-making is an important component of teachers' assessment practices.

Since the 1970s, researchers of instructional development and education have become interested in examining teachers' thought processes and how teachers' cognition influences their practice and decision-making in the classroom (Borg, 2003; Fang, 1996; Leung, 2005). However, such studies have not established a clear relationship between teachers' beliefs and practices. Some studies have found teachers' beliefs and practices to be inconsistent (Borg, 2006; Brown, 2004). Several factors may have contributed to such discrepancies. First, the literature on teacher education, in an attempt to provide a clear description of teachers' beliefs, has defined this concept using a plethora of terms (Borg, 2006; Pajares, 1992), such as *knowledge*, *attitudes*, *beliefs*, *perspectives*, *understandings*, and *conceptions* (Clandinin & Connelly, 1987; Clark & Peterson, 1986). Because these terms have been used interchangeably, Pajares (1992) has criticized this literature for presenting so many definitions of *beliefs*, arguing that "the difficulty in studying teachers' beliefs has been caused by definitional problems, poor conceptualizations, and differing understandings of beliefs and belief structures" (p. 307). Woods and Çakir (2011) urged researchers to conceptualize teachers' beliefs "through a process of interpretation

stemming from a teacher's own experience" (p. 389) rather than conceiving the constructs according to how previous researchers interpreted them.

Second, studies that investigated teachers' beliefs and practices regarding language assessment have examined one or more of the contextual factors that influence teachers' beliefs and practices; one such potential factor is the context (e.g., country, institution) in which the assessment occurs. According to Cheng, Rogers, and Wang (2008), teacher practices come from a combination of knowledge, beliefs, values, and experiences, and these practices are influenced by the interaction of such attributes within the "instructional context in which [teachers] teach" (p. 25). Davison (2004) argued that teachers' practices are also influenced by their beliefs about the cultural, social, and institutional contexts of their assessment practices, such as the purpose of the assessment and its relationship with teaching and learning, as well as the teacher's role in the assessment process. Cheng, Rogers, and Hu (2004) examined teachers' assessment practices in various contexts within English as a second language (ESL) and EFL classrooms. They attributed the complex nature of such practices to several internal and external factors, such as the nature of the course, the instructor's teaching experience and assessment knowledge, and the influence of external testing on instruction and learning. Subject matter knowledge also influences teachers' assessment beliefs and practices. Cumming (2001) found that ESL/EFL instructors' assessment practices varied depending on the specificity of the purposes of the course. He found specific-purpose courses provided instructors with a precise election of assessment tasks, as opposed to the general-purpose courses in which teachers judged broadly and focused on learners' general language skills (Cumming, 2001). Another contextual factor, and a main point of investigation in this study, is the influence of external high-stakes testing on teachers' beliefs and practices (Rea-Dickins, 2008; Tierney, 2006). Rea-Dickins (2008) argued

that assessment policies mandated by external educational authorities have created "an imbalance in the range of assessment opportunities that teachers provide for their learners, leading to an orientation towards 'language display' rather than 'language development' opportunities" (p. 264). Few studies have examined the impact of high-stakes testing on teachers' assessment practices and beliefs. Those studies have demonstrated how teachers' practices and knowledge are inconsistent with standardized policies set by administrators (Arkoudis & O'Loughlin, 2004; Troudi, Coombe, & Al-Hamily, 2009). Davison (2004) has also shown the conflict between how teachers conceive of themselves as assessors and how their assessment beliefs and practices are influenced by institutional standards.

Finally, researchers examining teachers' beliefs are faced with certain methodological challenges (Borg, 2006). Because each methodology has its strengths and limitations, the nature of the cognition that is revealed in any study may be determined by the methodology selected (Borg, 2006). One challenge in observing individual beliefs is the distinction between ideal instructional practices and actual practices. In such situations, self-reported and verbal commentary methods may provide researchers with information about ideal practices, whereas observational methods may provide a better sense of the reality of classroom interactions (Borg, 2006). Speer (2005) demonstrated another significant issue in the relationship between beliefs and practices, arguing that researchers should differentiate between two types of beliefs: professed beliefs (what teachers say) and attributed beliefs (what teachers demonstrate). Speer claimed that the classification of professed and attributed beliefs has affected the design of research and methodology in the field, adding that the inconsistencies found in certain studies may have been the result of "a lack of shared understanding" (p. 370). Speer (2005) added that if the goal of a study is to investigate the role of teachers' beliefs in forming their practices, then

data relating to beliefs must be collected in a manner consistent with the data collected on practices.

1.2 Research Questions

This study investigated EFL teachers' classroom assessment beliefs and practices, guided by the following questions:

1. What are the L2 assessment beliefs and practices of seven EFL teachers teaching GE and ESP courses at a language center in a public post-secondary institution in Kuwait?

2. What factors influence the classroom assessment beliefs and practices of these EFL teachers?

To answer the research questions, this study adopted a multiple case study design using qualitative methods. Qualitative studies are widely considered the best way to investigate teachers' beliefs and the relationships of such beliefs to classroom practices (Borg, 2006; Speer, 2005). I selected the design and methods according to the research questions. I used three stages of data collection: (a) initial interviews, (b) classroom observations, and (c) post-observation interviews. Moreover, during the first two stages, I collected teachers' documents (e.g., course syllabi, teaching plans) for the courses to be observed.

1.3 Significance of the Study

As indicated in the study's theoretical background, few studies have examined teachers' beliefs or the relationship between teachers' practices and their beliefs concerning language classroom assessment, especially in postsecondary education in the EFL context. Those few studies that have looked at this question have found that there are some discrepancies between teachers' assessment beliefs and their practices.

In an attempt to address these challenges, this qualitative multiple case study, situated in a social constructivism theoretical framework, investigated the beliefs of seven EFL teachers regarding classroom assessment and examined the relationships between teachers' assessment beliefs and their practices. It further investigated how EFL teachers' assessment beliefs and practices are influenced by contextual factors inside and outside the classroom (e.g., program and institution policies, number of students, motivation, and culture). The study also compares and contrasts EFL teachers' beliefs and assessment practices in GE and ESP courses. Finally, this research discusses the possible implications of emphasizing the role of teachers as assessment policy makers and will increase teachers' awareness of their own knowledge of assessment and classroom teaching/assessment practices.

It is also important to note that literature on the topic of investigating teachers' beliefs and practices in relation to L2 classroom assessment is significantly limited in the Gulf states countries (i.e., Oman, Qatar, Saudi Arabia, The United Arab Emirates, Bahrain, and Kuwait) in general, and specifically in Kuwait. This could be due to the continued dominance of the old traditional testing and learning culture in this context. The limited literature in the field of assessment in this context would subsequently limit the professional development programs in assessment that are necessary for raising teachers' assessment literacy and result in assessment reform in the educational sectors. Hence, by investigating teachers' assessment beliefs and practices in this specific context, I also aim to increase teachers' awareness and knowledge of assessment and to provide researchers in the field with insights about language assessment at the postsecondary level in an institution specializing in teaching the English language in an Arabic context. I also believe that in Kuwait's educational schools and institutions, especially the postsecondary institution being investigated, policy makers need to pay attention to teachers'

voices because they are an important element for educational and assessment reforms (Troudi et al., 2009).

1.4 Overview of the Dissertation Chapters

This dissertation consists of six chapters, including this chapter. Chapter 2 presents key relevant studies about assessment, classroom assessment, language assessment, teacher beliefs about assessment, and methodological issues in research on teacher assessment beliefs and practices. Chapter 3 presents an overview of the educational context in Kuwait in relation to assessment, and in the PAAET and LC specifically. Chapter 4 describes the research design and methods of the study, the processes of data collection and analysis, ethical consideration, and the participants. Chapter 5 reports the study's findings in relation to the two research questions of the study. Chapter 6, the final chapter, discusses the study's findings in relation to the literature and presents implications for research and practice.

CHAPTER 2:

LITERATURE REVIEW

Classroom assessment plays a significant role in teaching and learning (Black & Wiliam, 1998; Gipps, 1994), and research has grown in this field (Turner, 2010), especially regarding teacher's roles and beliefs in relation to classroom assessment. This chapter reviews the literature related to the focus of the study on two areas of teacher assessment: teacher practices and beliefs about classroom assessment. Although the current study is focused on teachers' beliefs about and practices of second language assessment, the discussion is situated within the general area of classroom assessment. First, this chapter discusses various approaches to classroom assessment and its definition. It then presents studies on teacher various assessment purposes and methods and surveys various approaches to the study of teachers' beliefs. Then, it reviews several studies on teachers' assessment beliefs and practices as well as studies that have investigated various contextual factors that may influence teachers' assessment beliefs and practices and discusses methodological challenges to researching teacher beliefs.

2.1 Second Language Classroom Assessment

To understand classroom assessment, it is important to understand the concept of *assessment*. Assessment has been defined in various ways and in relation to different methods and purposes, all of which refer to the same activity: collecting information that is used for decision-making (Bachman & Palmer, 2010).

Because the role and purposes of classroom assessment have not been well defined and understood by practitioners in the educational field (Campbell, 2013), assessment has often been defined broadly and used in relation to testing and student evaluation in the educational fields (Brown & Abeywickrama, 2010; Campbell, 2013). However, Clapham (2000) stated that

assessment is an umbrella term used to cover all means of assessment and testing. Testing, conversely, is "a measurement instrument designed to elicit a specific sample of an individual's behavior" (Bachman, 1990, p. 20). Bachman and Palmer (2010) defined assessment as the outcome of "the process of collecting information about something that we're interested in, according to procedures that are systematic and substantively grounded" (p. 20). The outcome, they pointed out, could be verbal or written score. Bachman and Palmer's (2010) definition of assessment demonstrates that assessment includes different means of evaluation, which could be tests or other methods of evaluation; thus, decision-making is a salient part in the assessment process. However, Bachman and Palmer's definition does not specify who is responsible for decision-making.

In terms of the general description of language assessment, Bachman and Palmer (2010) defined language assessment as the process of collecting information to make decisions related to aspects of a learner's language ability. Similar to Bachman and Palmer (2010), Chapelle and Plakans (2013) argued that tests and assessment are used to signify the systematic gathering of language behaviors to get information about language ability, and that these terms should be identified based on their contexts. For example, in research on second language acquisition, tests are used to provide evidence for a learner's development, whereas in the classroom, assessment is used to provide information to teachers and to students about the learning process. In the applied linguistics fields, the term *second language assessment* is used among researchers and practitioners (Chapelle & Plakans, 2013). In addition, testing is related to high-stakes testing, while classroom assessment is more frequent than classroom testing (Chapelle & Plakans, 2013).

The previously mentioned definition of assessment by Bachman and Palmer (2010) is inclusive of different means of assessment by any stakeholders and could include tests. Yet the

literature in the field of assessment distinguishes *testing* from *assessment*; in addition, using the

term assessment with other terms such as measurement, evaluation, and testing to refer to the

same activity is controversial. While Bachman and Palmer (2010) believed all these terms refer

to the process of collecting information for the purpose of obtaining information about a

learner's performance and that the distinctions among the terms are unnecessary, other

researchers have made a distinction between assessment and testing (e.g., Black & Wiliam,

1998; Brown & Abeywickrama 2010; Clapham, 2000; Shepard, 2000). Brown and

Abeywickrama (2010) argued that assessment and testing should not be used synonymously.

Assessment is also used in the literature to separate alternative assessment from testing

(Clapham, 2000). Alternative assessment is distinguished from testing by the use of alternative

methods, such as self- and peer assessment, portfolios, observations, and learning logs (Fox,

2013). Alternative assessment complies with Black and Wiliam's (1998) description of

assessment, which links the term *assessment* with formative assessment to include all activities

that teachers and students undertake to obtain information that can help improve teaching or

learning.

Because this study focuses on second language classroom assessment, the term *classroom*

assessment is used in this dissertation and is distinguished from testing. Classroom assessment is

defined as the information collected, noticed, used, and evaluated by teachers to assist them in

improving students' learning and decision-making (Bachman & Palmer, 2010; Leung, 2004;

Lynch, 2001; McMillan, Hellsten, & Klinger, 2011; Yin, 2010). It is embedded in teachers'

everyday classroom practices, and it may be difficult at times to identify a certain activity as an

assessment (Hill & McNamara, 2012; Moss, 2003). It is important to note, however, that no

consensus has been found to describe a clear-cut construct of classroom assessment, which may be due to the diversity of approaches to assessment as discussed in Section 2.2 below.

2.2 Theoretical Approaches to Classroom Assessment

Pellegrino, Chudowsky, and Glaser (2001) argued that every type of assessment is informed by a conceptual or theoretical framework based on the way individuals learn and develop their knowledge and understanding. Moreover, the theoretical framework has a major influence on the teacher's design and use of an assessment. Thus, different approaches to classroom assessment are heuristically related to different theoretical frameworks. One early approach to classroom assessment is the psychometric view, which stems from the behaviorist theory of learning (Brown & Abeywickrama, 2010). In terms of learning, the psychometric approach measures the characteristics of individuals based on their achievements and treats learners as passive rather than active subjects in the assessment process (Gipps, 1994; Swain, Kinnear, & Steinman, 2010). In relation to teachers' assessments, the psychometric approach was prevalent in the 1990s, and the belief was that standardized measurement is the best way to help teachers in the decision-making process (Shepard, 2006). Teachers were trained to construct their own tests based on standardized educational measurement theories, and validity and reliability frameworks were developed to meet the needs of teachers learning about those educational measurements (Shepard, 2006). Little focus, Shepard (2006) argued, was given to teachers' classroom assessment practices. Instead, teachers were taught about assessment only according to standardized testing criteria and measurement test books, which were mainly issued for grading purposes (Shepard, 2006). Teachers' classroom assessments constructed for summative purposes are related to the psychometric theoretical perspective (Harlen & James, 1997; Yin, 2005).

Researchers in the assessment field became more interested in the nature of classroom assessment and earlier works, such as those by Natriello (1987) and Crooks (1988), stressed the importance of investigating the nature of classroom assessment but provided a limited framework to support the argument for the importance of classroom assessment (Gardner, 2006). However, the situation started to change in the 1990s (Turner, 2012). As Black and Wiliam (1998) argued, assessment studies began to pay more attention to the relationship between assessment and classroom learning and teaching and less attention to the psychometric features of tests that were "weakly linked" to students' learning experiences (p. 7). More recent studies have moved toward a broader view of assessment; there has been "a paradigm shift, from psychometrics to a broader model of educational assessment, [and] from a testing and examination culture to an assessment culture" (Gipps, 1994, p. 1). As a result, classroom assessment has evolved and has been redefined in a variety of ways (Gipps, 1994), and researchers have been attempting to find an alternative theory and rationale for an approach to classroom assessment that is distinct from the psychometric approach (Black & Wiliam, 2009; Turner, 2012).

Another approach to conceptualizing assessment is the cognitive approach, which is based on the cognitive theory of learning (Pellegrino et al., 2001; Shepard, 2000, 2006). The cognitive theory of learning argues that "learning is an active process of mental construction and sense making" (Shepard, 2000, p. 6). The cognitive theory was derived from cognitive psychology, which, according to Pellegrino et al. (2001), attracted theorists and researchers from various fields of sciences (e.g., linguistics, anthropology, computer science, psychology, and neuroscience) who were concerned with the study of individuals' minds and their function processes, whether individually or in groups. A major component of the cognitive theory is that

learners "construct their understanding by trying to connect new information with their prior knowledge" (p. 62). Unlike the psychometric approach to assessment, the cognitive approach focuses on how individuals acquire their knowledge. In terms of teachers' assessment, drawing on Popham (2000), Pellegrino et al. (2001) pointed out that in the assessment literature, including educational measurement and testing, assessment is referred to as "a process by which educators use students' responses to specially created or naturally occurring stimuli to draw inferences about the students' knowledge and skills" (p. 20). When relating this theory to teachers' cognition about language assessment, the cognitive theory investigates teachers' mental processes during the learning/assessment activity and accentuates teachers' interpretation and gathering of students' performances to aid in the teachers' decision-making (Borg, 2006; Yin, 2005). According to the cognitive theory, the purpose of assessment is to assess when, how, and whether an individual uses the knowledge he or she learned (Pellegrino et al., 2001).

The social constructivist approach, a third approach to assessment, views assessment as a social and cultural activity in which all learners are engaged. It considers learners as active constituents in their own assessment processes because, according to the model, students learn best by interpreting new knowledge and relating it to their existing knowledge (Gipps, 1994; Lund, 2008). The constructivist approach builds on the cognitive approach to learning (Pellegrino et al., 2001). It suggests that "learning is a process of knowledge construction; that learning is knowledge-dependent; and that learning is tuned to the situation in which it takes place" (Gipps, 1994, pp. 21–22). Thus, classroom assessment includes social, contextual, and collective dimensions as well as cultural and intellectual tools (Lund, 2008).

Shepard (2000) argued that the constructivist model should embrace sociocultural theory, cognitive theory, and constructivism, all of which share core principles. Shepard (2000) thus

elaborated the constructivist model to the theoretical framework of social constructivism to incorporate the cognitive, sociocultural, and constructivist models. In terms of learning, based on the social constructivist model, Shepard (2000) explained that learning is "an active process of mental construction and sense making" (p. 6), and that both the learner's cognitive ability and the context must be emphasized. The role of the teacher in this model is (a) to construct knowledge in which learning development is socially and culturally embedded and (b) to enable learners to self monitor their own learning progress (Shepard, 2005). The social constructivist model requires a complicated assessment process that includes diverse strategies to assess the deep understanding and learning of students as well as the engagement of both students and teachers (Gipps, 1994). Studies that adopt this theoretical framework investigate the nature of teacher–student interaction in the classroom and the complex nature of teachers' assessment, including formative assessment (Leung & Mohan, 2004; Moss, 2003; Rea-Dickins & Gardner, 2000) in all its forms (e.g., performance assessment, alternative assessment, self- and peer assessment, and portfolios). Section 2.3 reviews empirical studies on teachers' various assessment practices.

The different approaches to assessment create a challenge to adopt a specific theoretical background for classroom assessment (Yin, 2005). Empirical studies on teachers' classroom assessment are often situated within a specific theory or rationale in which classroom assessment can be grounded (Black & Wiliam, 2009). Black and Wiliam (2009) urged researchers in the field of assessment to locate classroom assessment within a specific framework to understand its complex nature. Hence, this study looks at teachers' classroom assessment from a social constructivist approach that not only involves the social context of assessment but also sheds light on the cognitive aspect of assessment and/or learning. Another reason for adopting a social

constructivist approach is that this study aims to examine the relationship between teachers' beliefs and their practices. A social constructivist approach may enable me to look at (a) the cognitive side of teachers' thought processes and the social dimension in looking at what shapes beliefs and (b) the nature of teachers' assessment practices in relation to their beliefs and contexts.

Among the many definitions available in the assessment literature and in relation to the social constructivist approach, classroom assessment is defined in this study as "any 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 socialization purposes" (Hill & McNamara, 2012, p. 396). This definition emphasizes the role of the teacher as a decision maker in the assessment process and describes the teachers' assessment practices in terms of the collection of data and evaluation.

The following section discusses teachers' assessment practices and reviews key studies that have investigated the complex nature of teachers' classroom assessment practices. It first introduces the formative and summative purposes of assessment separately. It then discusses the complex nature of teachers' classroom practices fluctuating between the formative and summative purposes of assessment.

2.3 Classroom Assessment Practices: A Continuum of Formative and Summative Assessment

The discussion above has shown a shift from viewing assessment as a testing instrument to assessment conducted to mediate learning and instruction. The assessment literature distinguishes two fundamental purposes of assessment: formative and summative. The definition of summative assessment is straightforward and described generally as teachers or assessors

systematically gathering evidence and summarizing the students' achievement status, based on professional judgment which is usually constructed at the end of a course for the purpose of reporting and certification (Harlen, 2005; Sadler, 1989). The definition of formative assessment is more complex than summative assessment and it has been redefined over the years because of the complex nature of teacher practices (Bennett, 2011; Yorke, 2003).

Formative assessment has existed as long as teaching, but the term was first proposed by Scriven (1967, as cited in Bennett, 2011) in the context of program evaluation as a distinction between formative and summative assessment (Bennett, 2011; Huhta, 2008). Scriven (1967) defined summative assessment as assessment used to provide information to help judge the educational program comprehensively and to facilitate program improvement. Bloom (1969) used the same conceptions of formative assessment proposed by Scriven (1967) but linked it to student evaluation (Bloom, 1969). He stated, "Evaluation which is directly related to the teaching-learning process as it unfolds can have highly beneficial effects on the learning of students, the instructional process of teachers, and the use of instructional materials by teachers and learners" (p. 50). Further, Black and Wiliam (1998) emphasized the importance of formative assessment to teachers and students to improve the learning process. Bloom's (1969) and Black and Wiliam's (1998) meta-analysis served as a platform for the formative assessment literature investigating the integration of assessment in the learning process.

Researchers in the assessment field agree that in formative assessment, teachers assess during the instructional phase, when they are engaged with students, and through observation of and monitoring student-to-student interactions and peer activities (Hill & McNamara, 2012; Katz & Gottlieb, 2012; Rea-Dickins, 2001). Formative assessment is central to providing the teacher with information about what a learner as an individual or learners as groups have learned and

understood (Rea-Dickins, 2001). It further assists teachers in determining suitable classroom activities and in providing feedback about their teaching strategies, thus directing their instructional planning (Rea-Dickins, 2001).

The definition of classroom assessment by Hill and McNamara (2012) used for this study, and presented above, incorporates both formative and summative purposes of assessment. The relationship between the two purposes of assessment is interrelated, which can be attributed to several factors. First, the nature of teacher practices in the classroom is complex and requires amalgamating formative and summative tasks in the assessment process. This is also the case for the studies on formative assessment, which have been conducted alongside summative assessment. Davison and Leung (2009), for example, discussed the nature of teacher assessment practices saying that teachers have the most crucial role in all assessment stages, "from planning the assessment programme to identifying and developing appropriate formative and summative assessment activities right through to making the final judgments" (Davison & Leung, 2009, p. 401).

Second, because the previous literature on educational assessment has primarily focused on traditional summative and standardized testing (McMillan, 2013), the distinction between purposes of assessment (i.e., summative and formative) remains sometimes equivocal (Rea-Dickins, 2007; Teasdale & Leung, 2000), especially in defining each (Bennett, 2011; Davison & Leung, 2009). For example, when describing the purposes of teacher-based assessments, some researchers referred to them as *informal* or *planned* in reference to formative and summative assessment, respectively (Davison & Leung, 2009). Davison and Leung (2009) explained that informal assessment practices, usually conducted while teaching, include observing students' behaviors and language performances and asking students questions to check their

understanding. Planned assessments, on the other hand, include more structured self- and peer assessments, portfolios, and presentations using scales and rubrics to assess language performance. Davison and Leung's (2009) description of informal and planned assessment could be applied to assessment conducted for formative purposes. Teachers may conduct various formative assessment tasks in the classroom (oral or written) that could be planned or spontaneous and formal or informal and conducted with individual students or at a group level (Varier, 2015; Yorke, 2003).

Third, while several studies that described teachers' assessment practices focused on the formative nature of assessment when describing classroom assessment (Brown, 2004; Black & Wiliam, 1998; Leung, 2004), other studies found that formative and summative assessments occur in the same teachers' classroom assessment practices (Hill & McNamara, 2012; Leung & Mohan, 2004; Rea-Dickins, 2001; Rea-Dickins & Gardner, 2000). While summative assessment is often described as tests conducted at the end of a unit, term, or year to judge a learner's achievement for accountability purposes (Bachman & Palmer, 1996; Leung & Mohan, 2004) and formative assessment is designed to enable students to monitor their own learning and to assist teachers in modifying their own teaching (Bachman & Palmer, 1996), Rea-Dickins (2007) argued that the same assessment method can be used for summative or formative purposes:

> Teachers may use the same data obtained from assessments for different purposes at different time intervals, formative in one context (e.g., [learner's] language sample used to inform discussion at a teachers' planning meeting where action is agreed for language support for that individual learner) and summative in another (i.e. where that same language sample is used as part of a [learner's] school Language Achievement Record). (p. 509)

Several studies demonstrated the complex nature of teachers' assessment. For example, Rea-Dickins (2001, 2006) investigated different sorts of teacher-based assessment for both summative and formative purposes, aiming to demonstrate how teachers might be influenced by the assessment purposes (i.e., formative and summative) and other conflicting demands. In her

model of formative assessment with a focus on learners' roles and teachers' roles in assessment orientations, that could occur during the same assessment episode, Rea-Dickins (2001) found that assessment purposes may shift from an emphasis on the teacher-centered assessment role to an emphasis on the basic role of a learner's engagement in an assessment process. All assessment episodes, Rea-Dickins argued, whether formative or summative, may offer language-learning opportunities.

Rea-Dickins (2001) examined the assessment practices of teachers of English as an additional language (EAL). She observed various stages of teachers' assessment practices and decision-making. Using classroom observation, teachers' interviews, video and audio recordings, and lesson transcripts of one mainstream teacher and two support teachers in an English primary school, Rea-Dickins (2001) developed a model cycle of teacher-based assessment in which four stages are emphasized: planning, implementation, monitoring, and recording and dissemination (Figure 2.1).

Stage 1: Planning
- Identifying the purpose for the assessment (why?)
- Choosing the assessment activity (how)
- Preparing the learners for the assessment
- Who chooses/decides for each of the above?

Stage 2: Implementation
- Introducing the assessment (why, what, how)
- Scaffolding during assessment activity
- Learner self- & peer monitoring
- Feedback to learners (immediate)

Stage 3: Monitoring
- Recording evidence of achievement
- Interpreting evidence obtained from an assessment
- Revising teaching and learning plans
- Sharing findings with other teachers
- Feedback to learners (delayed)

Stage 4: Recording & Dissemination
- Recording and reporting progress toward National Curriculum (NC)
- Formal review for Local Education Authority (LEA) or internal school purposes
- Strategies for dissemination of formal review of learners

Figure 2.1. Model of teachers' assessment processes and practices (adapted from Rea-Dickins, 2001, p. 435)

Figure 2.1 describes the processes that teachers implement when conducting classroom assessments. In the planning stage, teachers identify the purpose of assessment, make decisions on how to prepare students for the assessment, and determine what to assess. The second stage is the assessment decisions made in the classroom and the scaffolding that occurs during the assessment process. Feedback is given to learners at this stage. In the third stage, the teacher collects and records evidence of students' performance. The teacher interprets the evidence and revises plans (teaching or assessment plans) and then provides delayed feedback to students. The fourth and final stage includes the process of the teacher's reporting and reviewing students' assessment formally for administrative purposes. The first three stages of the model refer to teachers' cognition processes, while the fourth stage refers to the summative purpose of the assessment. Thus, teacher-based assessment practices, as Rea-Dickins (2001) put it, are "plotted at different points along a more 'formal' to 'informal' continuum" (p.437). The more formal assessments would include stand-alone preplanned assessment activities (e.g., formal language tests) carried out for summative and/or reporting purposes; the teacher would nevertheless scaffold learners during the assessment activities. The informal assessments would occur as part of teaching and learning with "strategic teacher intervention—by questioning, seeking clarification from learners, asking for an explanation . . . [and] pushing some learners' forward in their understanding and language learning" (Rea-Dickins, 2001, pp. 437-438).

Rea-Dickins and Gardner (2000) conducted a study that demonstrates the complex nature of teachers' assessment practices, including assessments conducted for both formative and summative purposes. Rea-Dickins and Gardner's case study investigated the nature of teachers' assessment practices in EAL primary-level classes in nine schools in the United Kingdom (UK). Based on a series of interviews and observations of teachers' classroom practices, the authors

found that the relationship between formative and summative assessment was not as directly explicit as it may appear in the assessment literature. Though teachers were aware of the different purposes of different types of assessments, they sometimes conducted formative assessments for summative purposes and vice versa. For example, students' performance in the classroom assisted teachers in high-stakes decision-making; teachers conducted summative assessments to improve students' performances and language skills in the classrooms and to seek feedback to modify their own teaching. Rea-Dickins and Gardner argued that researchers should further investigate whether teachers are aware of the purposes of the assessments they conduct and the reasons behind their selection of specific assessment criteria.

Hill and McNamara (2012) expanded the model of classroom assessment proposed by Rea-Dickins (2001; See Figure 2.1) to include more intuitive, less visible assessment tasks that could be spontaneous, involving planned or unplanned, implicit or explicit, and embedded assessment. They distinguished between three dimensions of assessment (first delineated by McNamara, 2001): evidence, interpretation, and use, and added four questions to cover the scope of assessment (Hill & McNamara, 2012, p. 398):

1. What do language teachers do when they carry out classroom-based assessment?

2. What do they look for when they are assessing learners?

3. What theory or "standards" do they use?

4. Do learners share the same understandings?

Table 2.1 presents Hill and McNamara's (2012) assessment model, showing the relationship between the dimensions and scope of assessment. The purpose of their study was to build a comprehensive framework for future research on classroom assessment and to benefit educators in the field of assessment.

Table 2.1

Classroom Assessment Scope and Dimensions (Adapted from Hill & McNamara, 2012, p. 398)

Scope	Dimensions	
1. What do teachers do?	Evidence	What activities or behaviors are assessed? Is it planned/incidental, explicit/embedded? Does it target individuals, groups, the whole class?
	Interpretation	Is reflection sustained or fleeting?
	Use	How is assessment used?
2. What do they look for?	Interpretation	What criteria do they apply?
3. What theory or "standards" do they use?	Interpretation	What are the values guiding assessment?
4. Do learners share the same understandings?	Evidence	What are learners' beliefs about
	Interpretation	how assessment is conducted,
	Use	interpreted and used?

Both Rea-Dickins's (2001) and Hill and McNamara's (2012) models of assessment were based on teacher decision-making and a cognitive approach to teacher-based assessment. However, Rea-Dickins's (2001) model could be seen as describing the stages of the assessment process conducted by a teacher in the classroom to identify the cognitive approaches in which such strategies would be situated. Hill and McNamara (2012) argued that previous assessment frameworks, including Rea-Dickins's (2001), emphasize only the first part of their framework, as presented in Table 2.1, asking "What do teachers do?" but do not address the other questions in their framework. Hill and McNamara (2012) further argued that Rea-Dickins's (2001) model is based on her own study in the context of EAL, whereas their framework is based on a study in a school-based foreign language setting.

Hill (2017) has argued that Hill and McNamara's (2012) model of assessment benefits not only researchers but can also increase teacher awareness about the process of assessment. She pointed out that the framework could improve teacher assessment literacy and raise their awareness for further professional development needed in relation to classroom assessment. Hill's (2017) framework is similar to the one in Table 2.1, but she reworded the questions

underlying each scope as more directed to a teacher than researchers to observe his or her own

practices. Table 2.2 presents the frameworks in both Hill and McNamara (2012) and Hill (2017).

Table 2.2

Comparison of Hill & McNamara (2012) and Hill (2017) Assessment Framework

	Adapted from Hill & McNamara (2012, p. 415)	**Adapted from Hill (2017, pp. 5-7)**
1. What do language teachers do?		
1.1 Planning Assessment	Is there planning for assessment? How detailed is planning? What is its intended relationship to instruction? How does it relate to external standards and frameworks?	• How does planned assessment relate to teaching and the intended learnings (including relevant curriculum standards and frameworks)? • How are the learners' existing knowledge, language background, capabilities, and interests taken into account? • How are learners' social, emotional, and psychological attributes taken into account? • What role do learners have in setting learning goals and making decisions about when, how, and why they will be assessed?
1.2 Framing Assessment	Is assessment made explicit to learners? How is this done?	How do learners become aware of when, how, and why they will be assessed?
1.3 Conducting Assessment	What opportunities does the classroom provide for assessment? Does assessment tend to focus on the class, group/pairs of students or individuals?	• Who carries out assessment (teacher, student, peer, others), and whose judgment 'counts' in grading decisions? • What proportion of assessment is planned and formal and what proportion is unplanned and incidental (e.g., observation)? • What evidence of learning is provided by routine classroom activities and interactions (e.g., class discussions)? • Who is the main target of informal (incidental) assessment (the whole class, groups/pairs, individual students)? • Does formal and informal assessment focus on processes and well as products (e.g., are learners encouraged to discuss the basis for their responses)? • Where do formal assessment activities come from (e.g., textbook, self-designed, other teachers), and how well do they fit the intended purpose in terms of nature, scope and level? • Do you use a range of assessment methods and is the method appropriate for the intended purpose?

		• How do you ensure the fairness, quality and reliability (trustworthiness) of assessment activities and processes? • Is assessment conducted in an ethical manner (e.g., preserving student confidentiality)?

1.4 Using Assessment Data	How is assessment-related information used?
	Teaching
	Learning (feedback)
	  Person-referenced
	  Task-referenced
	    • Confirmatory
	    • Explanatory
	    • Corrective
	Reporting
	Management
	Socialization
		• To document growth in learning
		• To judge and grade students
		• To report to stakeholders (students, parents, school, external authorities)
		• To prepare students for exams
		• To inform teaching
		  (a) How is assessment used to diagnose needs and plan teaching?
		  (b) How is assessment used to evaluate teaching?
		• To enhance learning, motivation and self-regulation by providing quality feedback
		  (a) Does feedback focus on features of performance (rather than on innate qualities e.g., intelligence)?
		  (b) Does feedback explain which aspects were done well (e.g., "You used a good variety of vocab and sentence structures.")?
		  (c) Does feedback tell the student how to improve (e.g., "You need to review the work we did last week on the use of the passive form.")?
		  (d) Does the timing (immediate/delayed) and format of feedback (e.g., comments only vs. marks) encourage learner uptake?
		• To manage teaching
		  (a) Is assessment used to discipline learners or to encourage them to work harder?
		  (b) Is assessment used to socialize learners into a new assessment culture (e.g., using assessment rubrics, preparing for high stakes exams)?

2. What do teachers look for? **What information about valued enterprises, qualities, and standards is available?**		
	_____	• What is the balance of skills (listening, speaking, reading, writing), knowledge (vocab, grammar, cultural) and abilities addressed in assessment? Does this reflect the relevant curriculum priorities? • What are the valued qualities (e.g., accuracy, fluency, variety), behaviors (e.g., effort, presentation, attendance), and student-centered factors (e.g.,

		well-being) communicated in written or verbal instruction, written or verbal feedback, and formal reporting? Are these qualities consistent with the intended learnings (including relevant curriculum standards and frameworks)?
2.1. In Advance	In written/verbal instructions and/or assessment rubrics?	
2.2. In Feedback	In written and/or verbal feedback?	
2.3. In Reporting	In reporting deliberations and/or written reports?	
3. What theory or "standards" do they use?		
3.1. Teacher Theories and Beliefs	What does the data reveal about teachers' beliefs about? • the subject or content area, • second language learning and teaching, and • the nature of assessment?	• What are your beliefs and understandings about the nature of the subject (the nature of language; relationship of language and culture), how students learn a second language, how language should be taught, and how language should be assessed (e.g., learner agency, appropriate uses of assessment)? • What is the basis for these beliefs and understandings? • How do these beliefs and understandings influence your assessment practices?
4. Do learners share the same understanding?		
4.1 Learner Theories & Beliefs	What does the data reveal about learners' beliefs about second language learning and the nature of assessment?	• How do you ensure students understand the focus and purpose of assessment? • How do students perceive their role in planning, conducting and judging assessment? • How do you ensure students have understood and engaged with feedback?
5. How does the context for teaching shape your assessment practices?		
	NA	• Who influences decisions about content and methods in your assessment (school, supervisors, students, external authorities)? • What other factors do you need to take into account when planning and conducting assessment (e.g., class size, learner characteristics, external examinations, students, parental expectations)? • What is the impact of testing and assessment practices both locally and in the broader context and what is your capacity to influence change?

Table 2.2 shows that the questions proposed in Hill and McNamara's (2012) model address specific assessment events that may be clear to researchers during the process of observation or through conducting interviews. Furthermore, the questions were based on a study that investigated teacher classroom assessment practices. Hill's (2017) expanded model, on the other hand, addresses the knowledge and beliefs of teachers as well. The final question in Hill's (2017) framework (question 5 in Table 2.2) is an addition to Hill and McNamara's (2012) previous framework and addresses the impact of contextual factors that might influence teacher beliefs and practices—a point addressed in Section 2.7.

2.3.1 Classroom Assessment as Learning Mediation

As can be seen from Rea-Dickins (2001) model (see Figure 2.1), classroom assessment involves different strategies that aim at mediating learning. One of the main characteristics of classroom assessment, represented in stage 3 of the model, is making use of the information obtained from assessment based on classroom interaction (Al-Sawafi, 2014). Teachers use the information obtained from assessment to evaluate their pedagogy, inform teaching strategies, provide feedback to students, and mediate learning. This section discusses one of the main strategies obtained from assessment to promote learning: feedback.

2.3.1.1 Feedback in relation to classroom assessment.

Feedback has been recognized as a basic feature of the teaching and learning processes (Brookhart, 2007; Brown & Hudson, 1998; Davison & Leung, 2009; Gipps, 1994; Tunstall & Gipps, 1996). It is embedded within various kinds of assessment (in tests and classroom assessment) to inform teachers, students, or stakeholders about students' understanding and abilities and to help teachers adjust their instruction based on students' performance (Kunnan & Jang, 2009). In classroom assessment, especially that which is conducted for formative purposes,

a high priority is placed on teacher–student interaction, especially via dialogue in which feedback and questioning are provided (Davison & Leung, 2009). Feedback exists in different forms and for different purposes, and it may be in the form of verbal interaction (e.g., verbal comments after certain activities), or written commentary based on the type of assessment (Davison & Leung, 2009; Yorke, 2003).

While a few studies have focused on types of feedback in classroom assessment, not many have described the nature of teacher feedback discourse and what type of feedback they use (Ruiz-Primo & Min, 2013). Feedback in language assessment has been mostly associated with written corrective feedback, and many studies have demonstrated teachers' and students' roles in written feedback in L1 or L2 (e.g., Barkaoui, 2010; Montgomery & Baker, 2007), or discussed feedback as a general concept related to teacher classroom practices not specifically related to assessment (e.g., Ruiz-Primo & Min, 2013).

Although feedback exists in all kinds of assessment, in this study feedback is concerned with the context of classroom assessment. However, it is important to introduce the purposes of feedback in the testing context. In a large-scale testing context for summative purposes, feedback is used to inform teachers about the gap between the curricular goals and the target students' level being tested; this information is used to monitor student progress, and the feedback is provided to all stakeholders, including students, parents, teachers, and school administrators (Kunnan & Jang, 2009).

Gipps (1994) indicated that formative assessment feedback offered to teachers and students is different from that of assessment for accountability or monitoring purposes: "Performance feedback must . . . emphasize mastery and progress, rather than normative comparison" (p. 41). Gipps stressed two important reasons for feedback within the process of

teaching: "it contributes directly to progress in learning through the process of formative assessment, and indirectly through its effect on pupils' academic self-esteem" (p. 129–130). As such, studies on formative assessment consider feedback as an important element of formative assessment (Gipps, 1994).

Leung and Mohan (2004) compared feedback between summative and formative assessment, implying that feedback conducted in formative assessment provides better opportunities to engage in the learning process than does summative assessment. In summative assessment, feedback is straightforward and is given about whether the answer is right or wrong (Leung & Mohan, 2004). In formative assessment, feedback is not straightforward because it is integrated into teaching and learning activities (Leung & Mohan, 2004). They argued that classroom assessment conducted for formative purposes "carries an enormous potential for providing useful and helpful teacher feedback, so that students can engage with further learning and/or revise what has been learned" (Leung & Mohan, 2004, p. 338). Teachers benefit from feedback obtained from classroom assessment as do students because they have the opportunity to discuss their performance with the teacher or through self-assessment (Kunnan & Jang, 2009; Leung & Mohan, 2004).

Although feedback is important in formative assessment practices to enhance learning, it can be used to serve summative or accountability purposes (Moss, 2003). Moss (2003) indicated that feedback in classroom assessment could be used to serve grading or accountability purposes. In a case study of her own classroom practices, Moss noted that she attempts to engage students with their work to enhance their learning processes. She did this by designing learning opportunities in writing classes involving cooperation among students on different writing assignments. Feedback was not standardized, and although Moss argued against standardized

feedback, she pointed out that setting standardized feedback is sometimes necessary to develop "an analytic rubric for feedback" (p. 16). Thus, commenting on and grading an assignment is a form of feedback that requires a deliberate action from teachers in analyzing student's performance in any type of assessment (Yorke, 2003).

Cheng and Wang (2007) emphasized the role of feedback and grading practices. They interviewed English postsecondary teachers in three countries: Canada, Hong Kong, and China. The majority of teachers designed their own grading criteria, but they used them for different purposes. The majority of teachers in Hong Kong and Canada informed their students about the grading criteria before assessing them to increase learning expectations. Cheng and Wang (2007) found that most teachers in Canada used the marking criteria to inform their teaching pedagogy and student learning, and teachers in Hong Kong were also concerned with the way the marking criteria affect their practices. Chinese teachers did not explain how marking criteria would inform teaching or learning. Their beliefs were influenced by the context—large classes and past experience. In terms of feedback, teachers in all three contexts focused on errors in students' work. Teachers in Hong Kong and Canada, however, provided individual feedback, whereas Chinese teachers provided feedback to the class as a whole as follow-up activities.

2.4 The Role of Teachers' Beliefs in Classroom Assessment

Previous studies that investigated teachers' roles in assessment have focused on their beliefs and practices around summative assessments, or traditional testing (Hill, 2017). However, teachers' roles in classroom assessment vary widely among the various informal verbal and nonverbal tasks that include feedback, questioning, incidental, and embedded forms of assessment (Hill, 2017). As shown in Rea-Dickin's (2001) model (Figure 2.1), classroom assessment can be distinguished from traditional testing in that teachers are the main agents in

the assessment process and in planning, collecting, and analyzing information from various assessment methods and sources (Hill & McNamara, 2012; Rea-Dickins, 2001, 2004). Their role includes determining the strengths and weaknesses of the assessment methods available and selecting from such alternatives based on their understandings of language learning and development as well as their experience (Rea-Dickins, 2004).

Accordingly, recent studies have been conducted regarding teachers' beliefs about their role in classroom assessment. The teacher's role has been investigated not only through one's practices but one's beliefs as well. Fives and Beuhl (2014) argued that teacher beliefs serve as filters, frames, and guides. Beliefs that act as filters affect how teachers understand their knowledge and interpret information based on experiences. Beliefs act as frames during problem-solving tasks and lesson planning, which requires the teacher to consider beliefs about content, students, or making decisions. And the final function is that beliefs guide teacher's actions. As such, teachers' decision-making is a powerful component of their assessment practices. Studies that investigated the teacher's role in assessment examined teachers' perceptions or beliefs about assessment and how they linked their beliefs to their practices. Troudi et al. (2009) claimed that assessment is "an exercise of power that is caught up in an array of issues about testers' and test-takers' voices, roles, and beliefs" (p. 547). The teacher's role can be understood from the discussion on the classroom assessment practices in Figure 2.1 and Table 2.2. However, researchers have also attempted to evaluate what teachers know, think, and believe about assessment in order to understand the complex nature of classroom assessment practices. The following section discusses the place of teachers' assessment beliefs in the assessment literature.

2.5 Teachers' Assessment Beliefs

Since the 1970's, researchers of teacher development and education have been interested in examining teachers' thinking processes and how teachers' cognition influence their practices and decision-making in order to understand teachers' practices (Borg, 2003, 2006; Clark & Peterson, 1986; Fang, 1996; Fives & Buehl, 2012; Leung, 2005; Nespor, 1987). However, studies on teachers' beliefs and practices have not established a clear relation between teachers' beliefs and their practices. Other studies found teachers' assessment beliefs and practices to be inconsistent (Borg, 2006; Brown, 2004). Several factors may have contributed to such discrepancies, including the different understandings and interpretations of the construct of beliefs, the influences of external factors (e.g., high-stakes tests, context) on teachers' beliefs and practices, and the methodological challenges involved in investigating teachers' beliefs in relation to their practices. Before discussing each of these factors, I will discuss theoretical approaches to the study of teacher beliefs.

2.5.1 The Construct of Teachers' Beliefs

In an attempt to provide a clear description of teachers' beliefs, the literature on teacher education has defined this concept with a plethora of terms (Barnard & Burns, 2012; Borg, 2006; Clandinin & Connelly, 1987; Pajares, 1992), including knowledge, attitudes, beliefs, values (Clandinin & Connelly, 1987), perspectives, understandings, principles of practice, conceptions, beliefs and principles, constructs, practical knowledge (Clark & Peterson, 1986), and implicit assumptions (Kagan, 1992). Because these terms have frequently been used interchangeably, Pajares (1992) criticized this literature for presenting so many definitions of beliefs that it results in what he called a "messy construct" (p. 307). Pajares (1992) has argued that "the difficulty in

studying teachers' beliefs has been caused by definitional problems, poor conceptualizations, and differing understandings of beliefs and belief structures" (p. 307).

In a recent literature review, Borg (2006) attempted to provide a systematic description of the concept of teachers' beliefs. Borg reduced the multitude of labels used to refer to teachers' beliefs to a single term: *cognition*. Borg (2003) used cognition to denote "the unobservable cognitive dimension of teaching—what teachers know, believe, and think" (p. 81). Borg (2006) characterized this cognition as the "often tacit, personally-held, practical system of mental constructs held by teachers . . . which are dynamic—i.e. defined and refined on the basis of educational and professional experiences throughout teachers' lives" (p. 35). However, Woods and Çakir (2011) criticized Borg's attempt to synthesize these terms, arguing that there are some caveats in interpreting teacher cognition and that researchers must be careful when selecting a specific term to describe teacher cognition because it might be carried into future studies. According to Woods and Çakir, teachers and researchers should develop their own personal conceptions through experience, which they should subsequently seek to theorize "through verbal articulation, and [to share] them through rhetorical expression" (p. 389). Woods and Çakir urged researchers to conceptualize teachers' beliefs "through a process of interpretation stemming from a teacher's own experience" (p. 389) rather than conceiving the constructs according to how previous researchers interpreted them. Further, Fives and Buehl (2012) argued that the difficulty in understanding teacher beliefs is not the use of terms but in defining and using the term beliefs within the specific fields that describe the construct.

The other challenge in defining beliefs is the fuzzy distinction between beliefs and knowledge, which has caused major confusion in the literature on teachers' cognition (Pajares, 1992). According to Allen (2002), it is difficult to determine where knowledge ends and belief

begins. The distinction is important, however, because these terms are the most commonly used psychological labels in the literature that seeks to define teachers' mental processes (Borg, 2006). The focus of teachers' knowledge emerged in the literature in the 1980s (Borg, 2006). This concept initially appeared with several labels, such as subject-matter content knowledge, pedagogical content knowledge, and general pedagogical content knowledge (Borg, 2006; Fang, 1996), all of which describe what teachers think, know, and believe (Woods & Çakir, 2011). However, the knowledge construct used most frequently in researching teacher cognition, as Borg (2006) noted, is pedagogical content knowledge (PCK). Shulman (1987) developed this term to refer to knowledge that combines content (e.g., math, language) and pedagogy to understand the nature of instruction and students' learning needs. PCK is part of teachers' cognition as reflected in their assessment practices (Scarino, 2013).

In previous studies, beliefs and knowledge were often seen as interchangeable, and it is hard to draw a clear distinction between them (Allen, 2002; Fives & Buehl, 2012). Some researchers have defined beliefs as involving some form of knowledge (Pajares, 1992). Deford (1985; as cited in Richardson, Anders, Tidwel, & Lloyd, 1991) wrote that "knowledge, then, forms a system of beliefs and attitudes which direct perceptions and behaviors" (p. 562). However, Woods (1996; as cited in Woods & Çakir, 2011) contended that both beliefs and knowledge "involve structured and dynamically evolving understandings of phenomena, and that since, in their decision-making practices, teachers' use of knowledge structures is not distinguishable from their use of belief structures, the concept is better seen as a continuum" (p. 384). Woods argued against making a distinction between the terms, coining the acronym BAK (beliefs, assumptions, knowledge) to refer to this continuum. Woods argued that it is "more meaningful to emphasize the spectral nature of the concept in the decision-making and

interpretive processes of the teacher" (as cited in Woods & Çakir, 2011, p. 384). As Borg's term

cognition might include many other terms used to describe the teachers' mental process, I found

that Woods and Çakir's (2011) definition relates more specifically to teachers' beliefs, and thus I

use the term *beliefs* in this study to refer to teachers' *beliefs, assumption,* and *knowledge* (Woods

& Çakir, 2011). I use the term beliefs in accordance with the definitions of beliefs given by

Kagan (1992) and Pajares (1992), who consider beliefs to be the values, attitudes, and personal

knowledge and assumptions about teaching, students, classroom, the taught subject matter, and

the whole educational process.

2.6 Theoretical Approaches to the Study of Teacher Beliefs

Research on teacher cognition began in the 1970s, and, similar to classroom assessment

studies, this research approached teacher beliefs from a number of theoretical perspectives (Borg,

2006; Fives & Buehl, 2012). In the 1970s, such studies tended to view teaching as the reflection

of teachers' behaviors in the classroom, and the process of learning was considered a "product of

teaching" (Borg, 2006, p. 6). However, the field of teacher cognition expanded dramatically in

the 1980s, aligning itself with a constructivist approach that defines teachers as sense-makers

(Borg, 2006). Clark and Peterson's (1986) influential study from this period referred to teachers

as "rational decision-makers" (as cited in Borg, 2006, p. 15). According to Borg (2006), the role

of beliefs in teacher education stemmed from two perspectives: the constructivist theory of

learning and changing belief systems as part of teacher education. According to the constructivist

theory, preservice teachers bring their own beliefs to teacher education, which affects their

learning. This constructivist perspective has three subcategories: personal experience, schooling

and instructional experiences, and formal knowledge experiences. These three subcategories are

discussed in Section 2.7 below. Researchers saw these subcategories as factors that influence the

development of teachers' knowledge and beliefs about teaching.

Negueruela-Azarola (2011) pointed out that, based on earlier research, beliefs can be described in three basic ways: (a) a normative approach that understands beliefs as "opinions or generally inaccurate myths regarding L2 learning and teaching" (p. 360); (b) metacognitive studies that define beliefs as "metacognitive idiosyncratic knowledge or representations characterized by some personal commitment" (p. 360); and (c) contextual research in which "beliefs [are seen] as ideas . . . [that] are interrelated with the contexts and experiences of participants" (p. 360). Studies that adopted the normative approach considered individuals' beliefs about a second language as often erroneous and distinct from those beliefs held by second language scholars (Barcelos, 2003). While studies within the metacognitive approach stated that individuals' metacognitive knowledge assists them to develop potential for learning and reflects what they are doing (Barcelos, 2003), such knowledge about a second language might be incorrect and not always empirically supported, though it does influence the learning/teaching activity outcome (Barcelos, 2003). Negueruela-Azarola (2011), working from the contextual research perspective, pointed out that beliefs are socially rooted conceptual activities. That is, Negueruela-Azarola argued that beliefs can be understood to possess a stable social meaning while being vulnerable to change because of their dynamic nature in specific contexts.

While few studies have approached teachers' beliefs from a cognitive perspective, Yin (2010) adapted a cognitive approach in investigating teacher cognition. He conducted a case study to investigate two English-for-academic-purposes (EAP) instructors' cognitions concerning classroom assessment practices in the UK. Based on classroom observations and stimulated recalls, Yin argued that teachers accessed a number of interrelated cognitions that sometimes overlapped in their practices. He classified the cognitions as (a) strategic cognitions—

assessment approaches and beliefs that are influenced by class size, time, course syllabus, and summative assessments and (b) interactive cognitions—assessment approaches that are influenced by specific principles implemented during interactive assessment (i.e., giving positive feedback, responding fairly to students' responses). With interactive cognitions, teachers are influenced by specific principles, language constructs (e.g., vocabulary, grammar), syllabus aspects, and teacher-related aspects (i.e., what a teacher thinks is important).

2.7 External Factors Influencing the Study of Teachers' Assessment Beliefs and Practices

Fang (1996) argued that teachers may articulate their beliefs outside the classroom, but their practices in the classroom are influenced by several factors. Such factors can be internal, contextual or external (Izci, 2016). According to Izci (2016), internal factors are personal factors related to teachers' assessment beliefs and knowledge (e.g., teachers' experience, training, and knowledge in assessment; Cheng et al., 2004; Hill, 2017) and beliefs about a learner's characteristics (e.g., L1 vs. L2). Other factors could be external factors related to the context and the external policy (Izci, 2016). Contextual factors are those related to the school's assessment culture, students' and parents' beliefs and expectations about learning (Cheng, 2011; Hill 2017), students' motivation levels, students' proficiency levels, classroom management and routines, and context (Cheng et al., 2004; Davison, 2004). Contextual factors could also include teaching or learning environment, such as class size (Cheng et al., 2004). External contextual factors include the high-stakes mandates (Hill, 2017; Izci, 2016; Tierney, 2006). Borg (2003) proposed a framework (Figure 2.2) that identifies four main factors influencing teachers' beliefs: schooling, contextual factors, professional course work, and classroom practices (Borg, 2003).

Studies that revealed the influence of such factors are discussed in the following paragraphs. However, it is important to point out here that often more than one external factor in

Borg's model was examined in any single study (e.g., Cheng et al., 2004).
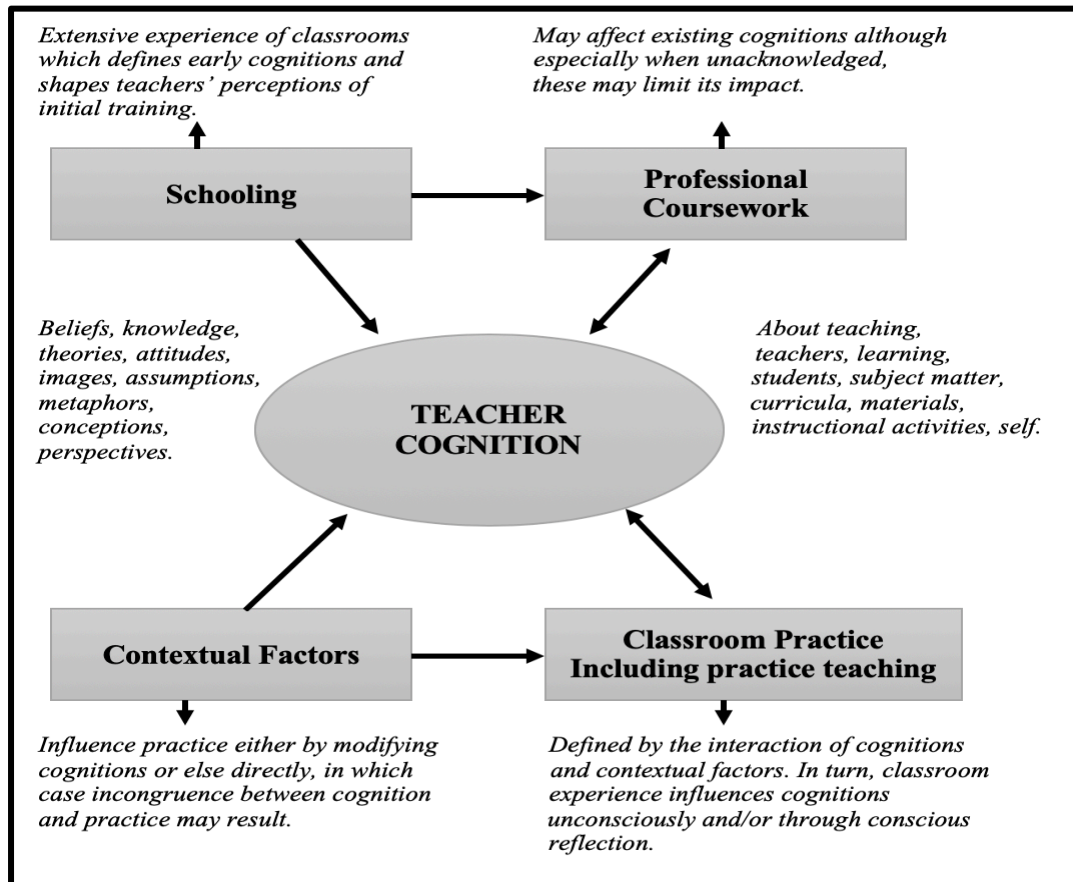
Figure 2.2. Different factors that influence teachers' beliefs (adapted from Borg, 2003, p. 82).

Figure 2.2 shows that teachers' beliefs could involve all aspects related to their work

(Borg, 2003). It also demonstrates the relationship among the main factors related to teachers'

beliefs and their education (i.e., schooling, professional training), teachers' practices and

contextual factors (Borg, 2003). Teachers' experience and education could have a major role in

shaping teacher beliefs and practices. Borg (2006) argued that school experiences have a

significant impact on teachers' beliefs, and past events critically influence teachers'

interpretation of their practices (Pajares, 1992). Nespor (1986) similarly described a belief as an

episodic structure that is stored in the memory and based on prior experience. Fives and Buehl

(2012) emphasized the role of knowledge and experience in teacher beliefs, arguing that a

teacher's understanding of the learning situation is interpreted by his or her own experience. They illustrated that practicing and preservice teachers' beliefs are shaped by the way they learned about teaching. Such beliefs, they stated, act as a filter in recognizing what is important to discuss with students. Fives and Buehl further noted that previous studies have demonstrated that teachers' beliefs about teaching, learning, and subject content (e.g., language, science) affect how they approach the task at hand and how they interpret the learning situation.

Teaching experience is another significant factor that could influence teachers' beliefs and practices, and some studies affirmed classroom experience to have a great influence on teachers' practical knowledge and thus their activities and decisions (Borg, 2006). Kagan (1992) stated that teachers are unlikely to change their beliefs, and they are more likely to gain ideas and construct their beliefs from their practices or their fellow practitioners. Such beliefs could be strongly constructed as early as the preservice stage (Fang, 1996; Kagan, 1992). For example, Kagan (1992) stated that "student teachers are more influenced by their cooperating teachers than by their college supervisors or by university courses" (p. 75).

In relation to the influence of experience on teacher assessment beliefs and practices, Sikes (2013; as cited in Al-Sawafi, 2014) pointed out that experienced teachers held negative attitudes toward changing their assessment practices. Experienced teachers' negative thoughts could result in rejecting assessment reform if they think their current practices are right. Sahinkarakas (2012) also found this high evaluation of self-efficacy in experienced teachers. She conducted a study with 100 participants, comparing experienced teachers to preservice teachers' perceptions of language assessment. She used a single method of quantitative analysis, asking teachers in her study to complete the metaphor, "A language assessment is like . . ." Her analysis revealed that prospective preservice teachers were more motivated and scored higher on self-

efficacy about their assessment practices than experienced teachers. She argued that experienced teachers highly evaluated their practices and did not need to provide evidence of the effectiveness of their teaching. She argued they might also lose their motivation for increasing teaching effectiveness because of their long experience.

Literature on classroom assessment and teacher beliefs also cites assessment literacy and professional development training as one significant factor contributing to classroom assessment implementation (Bennett, 2012; Hill, 2017) and to teacher competence (Xu & Brown, 2016). Bennett (2011) argued such competence may be well sustained by a deep domain understanding and awareness of the fundamentals of measurement. Bennett asserted that teachers need more time to build assessment knowledge with practice, and they should receive training in formative assessment materials, such as getting trained through projects, observational guides, and task sets. Several studies noted assessment literacy and assessment training have a major effect on teacher beliefs and practices (e.g., Al-Sawafi, 2014; Brookhart, Moss, & Long, 2010).

Brookhart et al. (2010) conducted a project engaging six remedial reading teachers in a professional development program on teaching an international language. The study investigated what reading teachers learn about and what methods teachers used in their own formative assessment practices. Through teachers' online responses to questions, face-to-face meetings, notes, observations, and students' work samples, teachers were required to examine their conceptions and identify four items: (a) an area of concern on the use of formative assessment in the classroom, (b) their learning agenda for formative assessment, (c) their assumptions about formative assessment and how such assumptions were changing, and (d) ways in which their focus on formative assessment may connect to their classroom practice. Brookhart et al. found that teacher participants changed their own assessment practices during the project of

professional development. Moreover, they found that teachers changed their own beliefs as they engaged in the professional growth program. As Brookhart et al. (2010) stated, "The fact that [teachers] were able to find ways to modify their instruction, even when faced with a highly scripted context, in order to respond to student learning needs, points to a critical aspect of professional learning—change of belief" (p. 52). Although some authors described beliefs as tenacious (Kagan, 1992), Brookhart et al. (2010) argued that if teachers question their current knowledge and understanding, their beliefs can be modified based on their practices and knowledge. The authors reported that changes in beliefs are positively reflected in teachers' assessment practices and student achievement.

Another area of investigation in this study is the subject matter being assessed, which also influences teachers' practices. Clarke and Gipps (2000) noted that secondary school science and math teachers tend to adopt formal approaches to assessment for end-of-unit or regular classroom tests, whereas English teachers tend to use more formative types of assessment, such as note-taking or self-assessment. Teachers also have different assessment orientations and practices depending on the purpose of the course (e.g., ESP, EAP). In an interview-based study of highly experienced ESL/EFL writing instructors in six different contexts, Cumming (2001) found that ESL/EFL instructors' assessment practices varied, depending on the purposes of the course. The author found that specific-purpose courses provided instructors with a precise curriculum organization rationale and a definite election of assessment tasks. In these courses, instructors were focused in their judgments, as opposed to the general-purpose courses, in which teachers judged broadly and focused on learners' general language skills (Cumming, 2001). However, Cumming (2001) argued that having such a broad focus in the general-purpose approach could be problematic because it requires various methods of assessment and challenges

instructors in the area of teaching writing.

In addressing a number of factors that appeared in Borg's (2003) model, Zhang and Burry-Stock (2003) investigated teachers' assessment practices and self-perceived assessment skills targeting three objectives: (a) finding the relationship between teachers' assessment practices and self-perceived assessment skills, (b) analyzing classroom assessment practices across content areas and teaching levels, (c) and examining the connection between teachers' self-perceived assessment skills and years of training in measurement and teaching. Zhang and Burry-Stock hypothesized that assessment practices are influenced "by content and intensity of instruction whereas self-perceived assessment skills are influenced mainly by teaching experience and professional training" (p. 326). The authors found further that grade and teaching levels affect assessment practices. For example, there was a difference between elementary and secondary-level school teachers in terms of assessment quality and methods; secondary schools relied more on summative assessment, while elementary teachers relied more on formative assessment. Their findings also suggested that teachers' assessment practices differ according to the subject content and that measurement training affects teachers' assessment practices. Knowledge has a crucial role in influencing measurement, teaching, and teachers' assessment skills "regardless of their teaching experience" (Zhang & Burry-Stock, p. 335).

Several studies showed local and external factors relating to context might influence teachers' beliefs and practices (Cheng et al., 2004; Davison, 2004; Fives & Buehl, 2012; Hill, 2017). Context has been a significant issue in determining "the degree to which teachers' beliefs vary or remain consistent" (Fives & Buehl, 2012, p. 475).

Studies that investigated the effects of contextual factors on language assessment have examined one or more of these contextual factors that may have an influence on teachers'

practices. For example, some researchers found that context, experience, and high-stakes testing might have an influence on teachers' assessment beliefs and/or practices (Xu & Liu, 2009; Zhang & Burry-Stock, 2003). One contextual factor, the context in which assessment occurs (e.g., the country, cultural, or institutional context), has been a common topic of investigation for many researchers concerned with teachers' assessment beliefs and practices (Cheng et al., 2008; Davison, 2004).

According to Cheng et al. (2008), teachers' practices come from a combination of knowledge, beliefs, values, and experience, and these practices are influenced by the interaction of such attributes within the "instructional context in which they teach" (p. 25). Davison (2004) also argued that teachers' practices are influenced by their beliefs about the cultural, social, and institutional context of their assessment practices, such as the purpose of assessment and its relationship with teaching and learning, the teacher's role in the assessment process, and the teacher's prior beliefs in regard to the content and the students being assessed. The following paragraphs review studies that have examined how teachers view their role as assessors and the impact of the social, institutional, and cultural contexts and standards on their assessment practices.

Davison (2004) conducted a study among 12 secondary school teachers in two contexts (Australia and Hong Kong) to investigate their various constructions of written assessment and their different interpretations of assessment criteria. Using multiple qualitative methods (i.e., self-reporting, verbal protocols, individual questionnaires, and group interviews), the study revealed that teachers in the different contexts varied in their interpretation of assessment criteria. Australian teachers adhered more closely to standardized criteria in their assessments within the instructional community, trusting their professional assessment judgment. Teachers in

Hong Kong, on the other hand, felt that their judgment would not be respected by the instructional community in particular and the country in general if they did not follow external policy standards. Based on these findings, Davison proposed a framework that classifies teachers "along a cline from assessor as technician, to interpreter of the law, to principled yet pragmatic professional, to arbiter of 'community' values, to assessor as God" (Davison, 2004, p. 324). Davison revealed how teachers viewed their roles in assessment and what factors affected their self-perceptions. At one end of the cline are teachers who are bound by the standard assessment criteria (criterion-referenced assessment); at the other end are teachers whose assessments are more intuitive (community-bound and norm-referenced assessment). Davison urged that future studies should investigate the relationship between teachers' assessment beliefs and assessment practices, how various assessments are influenced by how teachers conceive themselves as assessors, and how both assessment beliefs and practices are influenced by the social and institutional contexts. Such studies will enable teachers and educators to pinpoint problems raised regarding various kinds of assessment.

Cheng et al. (2004) also emphasized the role of context, experience, high-stakes mandates, and assessment knowledge. The authors conducted a comparative 3-year study among 267 teachers to examine their assessment practices in three different contexts. The authors used survey questionnaires through which teachers from Hong Kong, Canada, and China self-reported their decision-making practices on the assessment of students' language skills and abilities. As with Davison (2004), Cheng et al. (2004) demonstrated the complex assessment and evaluation practices present in ESL/EFL classrooms. Cheng et al. attributed the complex nature of such practices to several internal and external factors, such as the nature of the courses, the instructors' teaching experience and assessment knowledge, the level-specific demands of their

students, the teaching and learning environments (e.g., class size), and the influence of external tests on instruction and learning. For example, instructors in Hong Kong used fewer objective assessment scores and fewer assessment purposes than did instructors in Canada and China. Further, instructors in Hong Kong did not experience stress because of the externally mandated tests, which are usually believed to dominate teaching in other contexts. In Canada, ESL/EFL students are required to pass the Test of English as a Foreign Language (TOEFL) to gain admission to a university degree program. In China, every student is required to pass mandated tests to obtain a bachelor's degree. Despite the differences in practices, purposes, and assessment methods, all instructors spent a similar amount of time on assessment, and all had similar course loads. However, teachers from Hong Kong and Canada spent more time judging and scoring than did their counterparts in China. Cheng et al. (2004) attributed this difference to teachers' obedience to the mandated testing policy and class sizes in China.

Cheng et al. (2004) demonstrated the complex nature of assessment practices as a result of the context in which the assessment occurred. However, in a later study, the same authors, Rogers, Cheng, and Hu (2007) obtained similar findings but chose to approach their research from the perspective of teachers' beliefs. The researchers used a questionnaire with 44 instructors in Hong Kong, 95 instructors in Canada, and 124 instructors in China and found these teachers held similar beliefs on the importance of assessment and evaluation in teaching and learning, although they demonstrated different practices from instructors in the previous study. However, the instructors' expressions of their beliefs were uncertain, mixed, and contradictory, particularly concerning the time required for evaluation and assessment, the use of paper-and-pencil performance assessments, and their understanding of and preparation for assessment and evaluation. The participants also demonstrated differences in their confidence levels in terms of

applying what they had learned about evaluation and assessment. However, Rogers et al.'s (2007) findings showed more similarities than differences, and the authors attributed this to several factors: whether the criteria were set by university or departmental policies or by experienced teachers, students' expectations, and the requirements for demonstrating linguistic proficiency.

Xu and Liu (2009) investigated the assessment knowledge and practices of an EFL teacher at a Chinese university. Based on Connelly and Clandinin's (2006) theories of knowledge and through narrative inquiry, Xu and Liu (2009) argued that sociality and location significantly influence teachers' knowledge. With regard to sociality, the teacher was confronted with a dilemma regarding whether to grade students according to institutional standards or to act as a teacher-assessor who could make decisions independently. With regard to location or context, the teacher felt more freedom in making her own assessment decisions and judgments in the classroom than making decisions based on institutional standard criteria. The findings of the study suggest that teachers' assessment practices and decision-making processes are mediated by their society. Teachers' knowledge is also facilitated by the context in which the assessment practice takes place, and context in this study was defined as the teacher's own classroom. Similar to Davison (2004), Xu and Liu's (2009) study demonstrated how one teacher perceived her role as an assessor and what other factors, specifically context, can influence decision-making.

One of the major concerns in the research on the relation between teachers' assessment practices and their beliefs, and a main aspect of the investigation in this study, is the influence of external high-stakes testing on teachers' practices (Cheng et al., 2004; Rea-Dickins, 2008; Tierney, 2006) especially in the higher education context, which relies heavily on summative

assessments (Yorke, 2003). Rea-Dickins (2008) argued that external policy agendas have placed major constraints on teachers' assessment practices. Assessment policies mandated by external educational authorities have created "an imbalance in the range of assessment opportunities that teachers provide for their learners, leading to an orientation towards 'language display' rather than 'language development opportunities'" (Rea-Dickins, 2008, p. 264). In this situation, teachers who perceive themselves as facilitators of language development are viewed as being in conflict with such assessment practices (Rea-Dickins, 2004). These teachers may be led to "mediate the external pressures . . . through the 'filter' of their own professionalism" (Yung, 2002, p. 99) and eschew the standard criteria as the basis for their judgments, instead consulting their knowledge and beliefs (Xu & Liu, 2009). This tends to occur especially when the standardized criteria of assessments directly contradict teachers' beliefs about assessment (Xu & Liu, 2009). Several studies in which external authorities mandated high-stakes tests, such as Australia (e.g., Davison, 2004) and the United States (e.g., Stiggins, 2002), have demonstrated challenges in relating teachers' beliefs to their practices.

Few studies have examined the impact of high-stakes testing on teachers' assessment practices and beliefs. Arkoudis and O'Loughlin (2004) emphasized the influence of high-stakes testing on teachers' practices, whereas Brown (2004) described the impact of high-stakes testing mandates on teachers' beliefs. No studies have directly addressed the influence of external tests on both beliefs and practices. Troudi et al. (2009), for example, aimed to investigate the impact of high-stakes testing mandates on teachers' beliefs and practices, but they failed to use qualitative methods—a significant approach that Borg (2006) believed to be important for investigating teachers' beliefs.

Arkoudis and O'Loughlin (2004) found some inconsistencies in teachers' practices as a

result of high-stakes mandates. They conducted a study of ESL/EFL teachers in Australia, focusing on identifying concerns about the validity and reliability of the teachers' assessment practices and showing how state-mandated assessment policies influenced teachers' internal pedagogical contexts. Through narrative inquiries, Arkoudis and O'Loughlin observed how teachers worked with standard curriculum policies as instruments of assessment, how the teachers evaluated their students' written works based on their implicit knowledge, and how they contrasted such work with external ratings. After finding inconsistencies in teacher practices, Arkoudis and O'Loughlin (2004) attributed these to issues of validity in the ESL standards. The participating teachers expressed frustration when their efforts to rewrite such standards to fit with their view of ESL instruction were not taken into consideration by external education authorities. Arkoudis and O'Loughlin concluded that teachers' practices and knowledge, which had adhered to the standards, appeared to be inconsistent with bureaucratic policy. This bureaucratic policy was shown to have a negative impact on their practices (Arkoudis & O'Loughlin, 2004).

Brown (2004) also addressed the impact of external factors, especially those related to accountability, on teachers' beliefs. Brown first described four major teachers' conceptions about assessment purposes: (a) understanding student assessment as a means to improve teaching and learning, (b) maintaining accountability within the school or other educational authority, (c) maintaining accountability among students, and (d) perceiving assessment as completely irrelevant to teacher practices. Brown (2004) gave questionnaires to 525 primary school teachers and managers in New Zealand to self-report their reactions to the above-mentioned purposes. Brown found that teachers tended to agree that the assessment is a tool for improving teaching and learning and that it is necessary for accountability purposes both for the school and external authorities. Teachers in his study tended to disagree that the assessment is irrelevant to their

practices and that assessment serves only certification or accountability purposes. However, they considered all the above-mentioned purposes as positively related.

To my knowledge, little research has been done in Kuwait about assessment in general or about the external factors influencing teacher's assessment beliefs and practices. My literature search yielded a few studies that investigated teachers' beliefs and practices and the contextual factors influencing them in the Gulf States countries. Troudi et al. (2009), who investigated the assessment philosophies (i.e., beliefs) and practices of EFL teachers in universities in Kuwait and the United Arab Emirates (UAE), used open-ended questionnaires as their sole instrument for logistic reasons. Their findings suggested that the teachers held various assessment beliefs, but there was a gap between their beliefs and practices because they felt a conflict between their roles as language facilitators and assessors. The teachers stated that their decision-making is greatly influenced by institutional policy. Furthermore, teachers expressed frustration and revealed that they have little voice in curriculum and test construction and that their views are marginalized because of assessment policy mandates.

Other studies have been conducted in other Gulf state countries that could be seen as similar to the context of this study. Al-Sawafi (2014) cited a number of studies conducted in Oman that investigated the effects of various contextual factors on teachers' assessment beliefs and practices. Al-Kindy (2009, as cited in Al-Sawafi, 2014) used questionnaires with 52 teachers and a semi-structured observation of two teachers and found a number of external factors that influenced teacher perception, such as lack of teacher assessment training, sudden change in policy to formative assessment, a teacher-centered role in teaching, limited opportunity for peer- and self-assessment, doubts about the efficacy of classroom assessment for learning, and uncertainty of teachers' roles in classroom assessment in a specific context. All these factors

could be linked to teachers' lack of assessment training.

Al-Sawafi (2014), who distributed questionnaires to 237 teachers of English in government schools in Oman, asked teachers to list challenges that influenced their implementation of continuous classroom assessment. The challenges listed were related to curriculum load, large number of students, time-consuming continuous assessment practices, difficulties in dealing with classroom assessment procedures, and the policy maker's interference with the continuous assessment procedures (i.e., school principals). One noteworthy finding in Al-Sawafi's study is that most participants had not received any professional training in continuous assessment. Those who had received training reported that the content of the training course centered on improving the practical aspect of continuous assessment. Al-Sawafi (2014) also used another qualitative method; he interviewed four participants. The participants expressed dissatisfaction with their implementation of continuous assessment because of large class sizes, lack of time for implementing different procedures of continuous assessment, lack of training, learners' lack of commitment to complete the assignments, and policy requirements asking teachers to keep portfolios as the sole assessment procedure.

Alkharusi, Aldhafri, Alnabhani, and Alkalbani (2014), using only questionnaires, conducted a study among 3,557 teacher teaching grades 5–12 from all public schools in Oman aiming at exploring teachers' beliefs, knowledge, and practices. The study revealed that teachers needed further training in educational assessment, especially on how to actively involve students in the assessment process. Alkharusi et al. (2014) added that a lack of training in addition to burdensome teaching loads played critical roles in teachers' attitudes, practices, and knowledge about educational assessment.

Mansory (2016) investigated EFL teachers' roles and beliefs about continuous and

summative assessment practices in a Saudi university. Based on open-ended interviews with 20

participants on their views of their roles, Mansory (2016) found that teachers had no role in

summative assessment practices unless they were part of the assessment committee. Teachers

also expressed negative views regarding their roles in assessment because they had limited roles

in continuous assessment in classroom assessment. Teachers in the study expressed willingness

to become more involved in the assessment processes, but they reported several factors that

influenced their beliefs and practices: lack of student motivation, lack of academic English

teaching materials, unreliable placement tests, very simple tests, and dissatisfaction with the

mark distribution policy (i.e., rewarding students with 25% of the total grade simply for

attendance).

This section presents theoretical approaches to the study of teachers' assessment beliefs

and practices. Then, it presents studies that investigated the complex nature of teachers'

assessment practices, and it highlights various factors that influenced teachers' assessment

beliefs and practices. The following section discusses methodological challenges in researching

teachers' assessment beliefs and practices.

2.8 Methodological Challenges in Researching Teachers' Assessment Beliefs and Practices

Researchers examining teachers' beliefs face certain methodological challenges (Borg,

2006). Several methods are required to attain an understanding of teacher cognition (Borg,

2006), and researchers should carefully select a proper methodological approach to bolster their

findings on the relationships between beliefs and practices. Because each methodology has

advantages and disadvantages, the nature of the cognition to be examined may determine the

methodology to be selected (Borg, 2006). Another challenge in studying beliefs is the distinction

between *ideal* instructional practices and *actual* practices. In such situations, self-reported and

verbal commentary methods may provide researchers with information about ideal practices while observation methods may provide a better perspective on the reality of classroom interaction (Borg, 2006).

Speer (2005) highlighted another significant issue in the relationship between beliefs and practices, arguing that researchers should differentiate between *professed* beliefs (what teachers say) and *attributed* beliefs (what teachers demonstrate). Speer claimed that the classification of professed and attributed beliefs affects the design of research and methodology in the field, adding that the distinction between the two kinds of beliefs in the field is false and that the inconsistencies found in certain studies may have been the result of "a lack of shared understanding" (p. 370). Speer (2005) added that if the goal of a study is to investigate the role of teachers' beliefs in forming their practices, then data on beliefs must be gained in a way that is consistent with the data collected on practices. The author refuted the classification of beliefs as entirely professed or entirely attributed, arguing that such classifications may obscure understanding of this phenomenon.

One of the challenges in understanding how teachers' beliefs influence their practices, is the unclear definition of the term *belief* (Fives & Buehl, 2012; Speer, 2005). Woods and Çakir (2011) similarly argued against the classification of the concept of beliefs, asserting that when researchers use different terms, they are not necessarily referring to different concepts; additionally, when they use the same term, they do not necessarily have the same phenomenon in mind. This may be due to the proliferation of terms describing teachers' mental processes. Another reason could be, as Borg (2006) argued, the notion of *ideal vs. reality* when expressing the understanding of beliefs—how beliefs are expressed in an ideal practice versus how practices are conducted in actual contexts. Another problem occurs when a researcher identifies a

teacher's beliefs based on the observation of that teacher's practices (Speer, 2005). As Speer

(2005) pointed out, if a teacher states a belief about problem solving that is not observed in

practice, researchers tend to conclude that there are inconsistencies between that teacher's beliefs

and practices.

Moreover, Barnard and Burns (2012) argued that teachers can have strong beliefs without

putting them into practice. Thus, a qualitative analysis is seen to be an effective methodology

because researchers can observe interaction in the classroom to understand this complex situation

(Barnard & Burns, 2012). Another concern regarding the inconsistencies between teachers'

beliefs and practices that may create challenges in researching teachers' cognition is that beliefs,

which Kagan (1992) described as "tenacious" (p. 76), are subject to change (Brookhart et al.,

2010; Watkins, Dahlin, & Ekholm, 2005), and are not directly observable (Borg, 2006). Borg

(2006) also argued that theoretically based cognitions differ from those derived from practice.

Within a broader view of institutional and social contexts, instructional practices can also be

understood in terms of the complexity of the interaction between teachers' beliefs and the

contextual factors in the classroom (Borg, 2006). Consequently, as Borg (2006) argued, research

on teachers' cognition should strive for better understanding of the possibility that beliefs can

change over time when practices refute beliefs.

Taking the above challenges into consideration, I used qualitative methods to better

examine the relationship between teachers' assessment beliefs and practices. Furthermore, in

order to understand how teachers' beliefs interact with their practices in reality, I used classroom

observation to achieve this purpose.

2.9 Summary

The literature review above has presented several fundamental implications to guide this study. As observed in the literature, classroom assessment is a complex concept influenced by multiple factors and is influenced by teachers' beliefs about assessment. Thus, it is important to define what constitutes classroom assessment and best practices for situating assessment within a clear theoretical framework. Additionally, a careful examination of classroom assessment requires an understanding of teacher beliefs and the contextual factors that further affect teacher beliefs. As described in Section 2.7 on external factors influencing classroom assessment and Section 2.8 on research methods, most studies were conducted in the government educational institutions. There is a notable lack of studies conducted in the Gulf region, especially at the post-secondary level. Those studies conducted at the post-secondary level used mixed-methods research or depended solely on quantitative questionnaires. This study used qualitative methods, such as interviews and classroom observations, predicted to best examine teacher beliefs and relate them to actual practices.

CHAPTER 3:

CONTEXT OF THE STUDY

This chapter presents an overview of the educational history and context of Kuwait in general, and the history of English language teaching (ELT). Then it presents ELT in postsecondary institutions including the context of the study (PAAET and the LC).

3.1 Education and English Language Teaching in Kuwait

Kuwait is an Arab country located in the northwestern part of the Arabian Gulf[*]. Arabic is the country's first and official language, while English is a foreign language. To understand the importance of English language teaching (ELT) in Kuwait, it is important to first comprehend the trajectory of Kuwait's educational system because education and ELT have been analogous with the country's socioeconomic growth (Al-Rubaie, 2010; Hussein et al., 2002; Mohammad, 2008).

The Kuwaiti Council for Education was founded in 1936 and contributed to the development of the country's education system (Al-Rubaie, 2010). It initiated the Compulsory Act in 1965, making education mandatory for both boys and girls from elementary to high school (Al-Rubaie, 2010). At that time, Kuwaiti education included three 4-year levels following kindergarten: elementary, intermediate, and high school. Postsecondary education was not available; instead, the government offered scholarships to Kuwaiti high school graduates so they could study abroad to obtain bachelor's degrees (Al-Rushaid, 1978). Private schools were established in 1967; some were gender segregated, while others implemented a coeducation system (Al-Rubaie, 2010).

---

[*] The Arabian Gulf is also known as the Persian Gulf.

The English language played a significant role in Kuwait even before the discovery of oil and the globalization era because of Kuwait's geographically unique position that served as an economic center connecting ships from the Middle East to Africa and India, and because it was a British protectorate. Moreover, an American doctor made a modest contribution by teaching private English classes in his house after an American hospital was established in Kuwait in 1911 (Kharma, 1967). English was formally introduced by the government in 1920 but not widely taught until after the discovery of oil. According to Karmani (2005), oil wealth has transformed the Arabian Gulf countries, bringing about increased modernization and industrialization. Discussing the significance of English as a result of oil production and industrialization, Karmani (2005) wrote,

> As Western technologies of every conceivable sort began to pour into the region, there soon emerged a sudden urgency to develop and train an entire national workforce to participate in the region's mass project of industrialisation. It was largely in this context that the potential usefulness of English suddenly seemed to suggest itself (p. 92).

The government began to stress the importance of English in the Kuwaiti education system. In 1956, English began to be introduced at the fourth level of elementary school. The government also created a budget to grant Kuwaiti students scholarships so they could study in the UK and the United States (US) to obtain Teaching English as a Foreign Language (TEFL) diplomas (Kharma, 1967). In addition, in a number of private schools (sponsored by either the government or individuals), English became the primary medium of instruction (Mohammad, 2008). Teachers in private schools were often native English speakers, and students and teachers used English in and outside the classroom.

Even though English is a compulsory subject in public schools, people rarely use it in out-of-school contexts (Mohammad, 2008). Kuwaitis consider English to be a foreign rather than a second language (Osman, 1996), and they need it only to communicate with non-Arab

expatriates or native English speakers. However, English is used in street signs, restaurants and shop signs, street advertisements, and product descriptions (Mohammad, 2008). Furthermore, a high percentage of Kuwaiti citizens travel abroad during the holidays to English-speaking countries, and they are interested in learning the language for communication there. Two official English newspapers are published in Kuwait, and there is an official public TV channel broadcasted in English. Also, the private sector sometimes uses English in official business correspondence, such as in banks (Hajjaj, 1978; Mohammad, 2010).

Since the Iraqi invasion in 1990, English has played a more significant role in Kuwait (Mohammad, 2008). Al-Rubaie noted that after the Kuwait liberation in 1991, "English became perceived as the language of liberation and consensus with the global community" (p. 45). Kuwaitis "become absorbed with foreign culture and parents have encouraged their children to study foreign languages, in particular English" (Mohammad, 2008, p. 3). This dramatic shift toward globalization has altered parents' and the Kuwaiti government's attitudes toward English education (Al-Rubaie, 2010). Both have called for a greater focus than before this era on English in the Kuwaiti school system (Al-Mutawa, 1992; Al-Rubaie, 2010; Osman, 1996). In 1993, after debates in the Kuwaiti parliament and extensive discussions by the Ministry of Education (MoE), the Ministerial Decree 61 was issued in 1993, making English compulsory in public schools from the first year of primary school (Al-Rubaie, 2010). The Public Authority for Applied Education and Training (PAAET), a postsecondary public institution in Kuwait and the context of this study, has emphasized the necessity of promoting the English language in its colleges (Osman, 1996)—a topic explored further in Section 3.2.

3.1.1 English in Postsecondary Education in Kuwait

Kuwait has several recognized postsecondary institutions. Kuwait's higher-education institutions are divided into public and private sectors. Public institutions include Kuwait University (KU) (established in 1966), the PAAET (established in 1982), the Higher Institute for Dramatic Arts (established in 1973), and the Higher Institute for Music Arts (established in 1976). In 2005, private postsecondary educational institutions emerged in Kuwait, some affiliated with international institutions—all of which use English as their medium of instruction (Al-Rubaie, 2010). The private institutions include the American University of Kuwait, the Australian College of Kuwait, the Gulf University for Science and Technology, Box Hill College Kuwait, Kuwait Maastricht Business School, and the American University of the Middle East. The government of Kuwait also provides scholarships to Kuwaiti citizens to study abroad through KU, PAAET, the Ministry of Higher Education, and the Kuwait Institute of Scientific Research (Al-Rubaie, 2010).

Both public and private higher-education institutions in Kuwait offer courses in English as a foreign language. KU is the country's first research university. It has two English departments: the Language Center (LC) and the Department of English Language and Literature. The LC in KU coordinates and supervises English language units (i.e., English divisions) in all KU colleges, offering general English courses as well as French, Farsi, and Hebrew courses. In the Department of English Language and Literature, students can specialize in English literature or linguistics. In addition, English is the medium of instruction in the colleges of Medicine and Dentistry, Engineering, Science, and Administrative Sciences (Al-Rubaie, 2010).

As with KU, English is very important at the PAAET, which has an English department and an LC that functions as a coordinator for general- and specific-purpose English courses. The

colleges of Nursing, Health Sciences, and Technological Studies use English as the medium of instruction. Unlike KU, however, which is generally academic in nature, the PAAET is more concerned with the requirements of the labor force of the country. Thus, it designs all courses in all departments to address the social and economic needs of different sectors. Similarly, in the LC, English for specific purposes (ESP) courses aim to address social and economic English language needs in Kuwait's government and private sectors by offering courses designed for specific purposes in each English language unit (i.e., English for business, English for health sciences, etc.).

3.2 The PAAET

The PAAET is a tertiary-level public institution in Kuwait, established by the Amiri Decree 63/82 in December 1982 to provide the country with a labor force to address the requirements of economic and social growth in the applied technology sectors (e.g., banks, industries, and health and educational institutions). It accepts Kuwaiti youths with high school diplomas, and it has a limited number of seats for expatriates and Gulf Cooperation Council countries (PAAET, n.d.).

The PAAET's roots formed as early as the discovery and production of oil in Kuwait in the 1950s. At that time, Kuwait established training centers and programs to provide a labor force in the oil industry sector (Osman, 1996). Due to the government's increasing demands, the MoE established specialized institutions to train teachers for different subjects (i.e., math and science). Other ministries (e.g., the Ministry of Oil and Ministry of Health) had their own training programs to provide the country with a specialized workforce in the different occupational fields (Hussein et al., 2002). However, the government found it necessary to build a central site to coordinate and sponsor these training programs and institutes. In 1972, it

established the Technical and Vocational Education Department (TVED) to equip students with technical skills and experience and mentor all science majors that the labor market needed, but the PAAET later replaced it. The PAAET includes the applied education and training sectors ("PAAET," n.d.). Figure 3.1 presents the structure of the PAAET sectors. However, I exclude the administrative departments under both sectors because they are not relevant to the focus of this study.

All PAAET colleges and institutes are gender segregated and thus have two campuses: one for males and one for females. However, female and male instructors teach in both campuses. In the education sector of the PAAET, there is a credit-course study system, but the number of course units varies according to the college and program. The period of study per semester for each college is 15 weeks, including a week at the beginning of the semester for orientation and a week off between the last day of study and final exams (Osman, 1996; "PAAET," n.d). Thus, the actual teaching period for each semester is about 14 weeks. The academic year consists of two semesters (i.e., Fall and Spring). The summer semester is optional and lasts for 7 weeks in addition to the periods of registration and examination. Three PAAET colleges offer bachelor's degrees: the College of Basic Education (CBE), College of Health Sciences (CHS), and College of Nursing (CoN). At the CHS and CoN, students can choose to pursue either a diploma or a bachelor's degree. The College of Business Studies (CBS) and the College of Technological Studies (CTS) offer diploma degree. The diploma is similar to a certificate and is offered for students in training colleges.

PAAET

The Applied Education Sector

The Training Sector

Colleges

Directorate

Institutes

- Research Directorate
- Directorate of Mission and Cultural Relations

Higher Institute of Telecommunication and Navigation

Higher Institute of Energy
Nursing Institute
Constructional Training Institute
Vocational Training Institute
Secretary and Office Administration Institute
Tourism
The Industrial Training Institute

CBE

CBS

CTS

CHS

CoN

Main Departments

Islamic Studies
Arabic Language and Literature
Science
Math
Social Sciences
Curriculum Development and Design
Principals of Educational
Administration
Psychology
Library and Information Science
Learning Technology
Physical Science
Art
Music
Human Ecology
Interior Design
English
Special Needs Education
Computer Science

Main Departments

Accounting
Business Administration
Banking and Insurance
Computing
Law

Main Departments

Manufacturing Engineering
Mechanical Engineering
Mechanical Power
Electrical Engineering
Electronics Engineering
Civil Engineering
Chemical Engineering
Petroleum Engineering
Applied Science

Main Departments

Pharmaceutical Sciences
Environmental Sciences
Applied Medical Sciences
Medical Records
Food Sciences
Nutrition
Dental and Oral Hygiene

Main Departments

Associate Nursing
(diploma degree)
Nursing Science
(bachelor's degree)
Specialized Sciences

Supporting departments

Statistics
Computer Processing
Office Management
Language Center

English Language Unit

Supporting departments

English Language Unit

Supporting departments

Natural Sciences
English Language Unit

Supporting departments

Biomedical Sciences
General Science Unit
English Language Unit

Supporting departments

Field Training
English Language Unit

Figure 3.1. The organization of the PAAET's sectors

The main objective of the CBE is to prepare preservice teachers to teach in public schools. It has 19 departments designed to meet the needs of public schools in Kuwait, as well as one supporting department—all shown in Figure 3.1. Students must complete 130 credits, including field practice/training, to graduate. The program is a 4-year credit system, at the end of which graduates earn bachelor's degrees.

The College of Business Studies (CBS) aims to fulfill the country's labor market public and private sector needs in the fields of marketing, accounting, business administration, and commerce ("PAAET," n.d.). There are five main departments at the CBS and five supporting departments; all are listed in Figure 3.1. The CBS offers associate degrees and is currently considering whether to offer bachelor's degrees. The supporting departments do not offer diplomas but provide courses related to the main departments. The program is designed to be completed in 2 years, and students must complete 68 credits, including field practice, to graduate.

The College of Technological Studies (CTS) aims to provide the country's technical labor needs in the fields of industry, engineering, and technology. It aspires to develop national vocational capabilities and raise student awareness of technical information. Figure 3.1 presents its nine main departments and two supporting departments. The CTS offers a five-semester credit system program, and students must complete 84 credits, including field training ("PAAET," n.d.). According to Osman (1996), there are two periods within the semester in which students engage in fieldwork to relate theoretical study to industrial application.

Finally, the CHS and CoN share similar objectives: producing professionals for the Kuwait's health sector. The CHS aims to prepare national cadres for hospitals, clinics, pharmacies, and other public environmental/health institutions. As shown in Figure 3.1, the CHS

has seven main departments and two supporting departments. Students must complete 68 credits, including field training. They also have the option of obtaining a bachelor's degree in health sciences, which requires 4 years of study. The CoN aims to prepare students to be technical, professional, and clinical nurse specialists by providing them with medical knowledge through theoretical courses and teaching them to apply this knowledge in clinical training. Students can choose to obtain associate's degrees in nursing (completed in five semesters) or bachelor's of science in nursing degrees (completed in eight semesters). Figure 3.1 displays the CHS and CoN departments.

The focus of this dissertation is the LC at the PAAET, which supervises English language units in each of the five colleges of the applied education sector (CBS, CBE, CTS, CoN, and CHS). English language units in these colleges are considered divisions of the LC—the main department.

As shown in Figure 3.1, the English language units supervised by the LC exist primarily as supporting departments. However, there are two English programs—one offers a bachelor's degree and is a main department in the CBE, while the other functions as a supporting department in PAAET colleges. In the main English department at CBE, students can obtain a bachelor's degree in Teaching English as a Second Language (TESL)/TEFL, linguistics, literature, and vocational courses (e.g., curricula and method) in English. The LC, on the other hand, organizes and supervises general English and ESP courses for all PAAET colleges, and the PAAET considers it a supporting department ("PAAET," n.d.).

As for assessment and evaluation policies, The PAAET's decree 49 states that students must be assessed continuously throughout the semester and must be assigned a final exam

conducted at the end of the semester. The students' overall course grades should be based on the following:

- Students' continuous assessments must constitute 50% of their final grades.

- Students should have final exams worth 50% of their final grades.

The policy does not clearly state the type of continuous assessment that should be conducted throughout the year; therefore, it is unknown whether the assessment rubrics and criteria for student assessment should be based on the teachers' or the LC departments' decisions. According to PAAET assessment policy, the department may amend the above grading distribution (i.e., 50% for continuous assessment, and 50% for the final exam) in a manner consistent with the course type after obtaining the approval of the PAAET's Academic Affairs Committee.

3.2.1 The Language Center

The LC was established in 2007, and it is housed at the CBS. It functions as a coordinator of all English language units in the five PAAET colleges ("PAAET," n.d.). The LC's general role is to design ESP and general English (GE) courses for different majors in each college, determine the specific English courses the labor market requires, and design and administer English placement tests to students before they take any English courses. Figure 3.2 presents the governance structure of the LC. The LC coordinates five English language units, one at each college, and has three administrative units: the Testing and Measurement Unit (TMU), the Administrative Affairs Unit (AAU), and the Technical Affairs Unit (TAU). The LC recruits teachers who hold master's or PhD degrees, and it grants scholarships to teachers with bachelor's degrees to obtain master's or PhD degrees, after which they have to teach at the LC for at least as long as they studied abroad. At the time of the study, there were 73 faculty members and 10

trainers in the LC. Faculty members are those who hold PhD or master's degrees, while trainers

hold bachelor's degrees and are responsible for laboratory training and/or teaching GE courses.
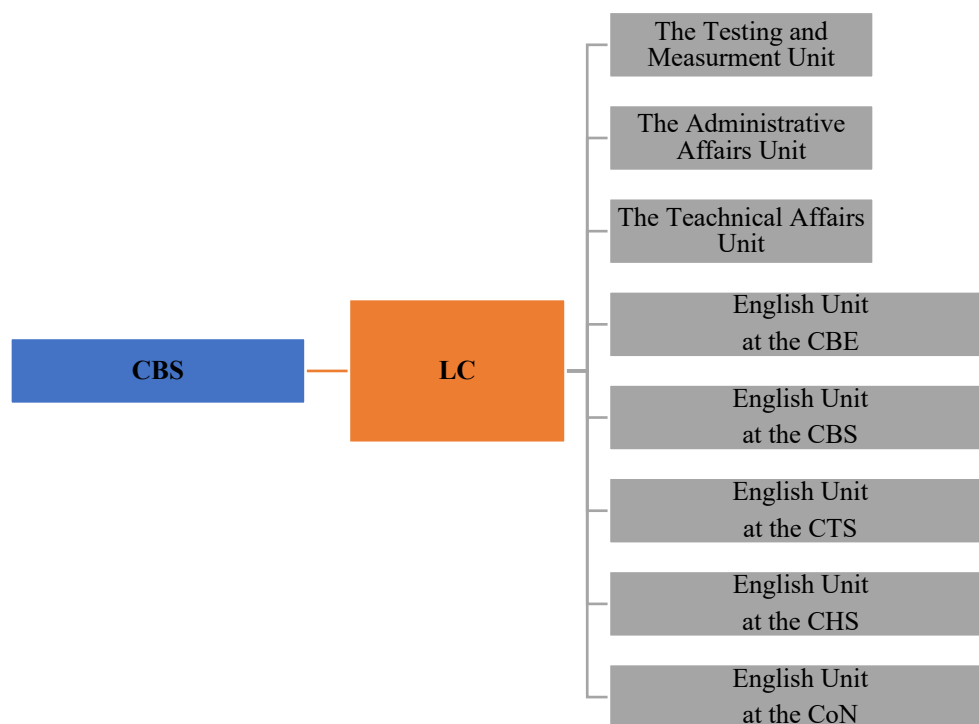
Figure 3.2. The governance structure of the LC

The AAU is responsible for daily procedures including correspondence, conducting

training sessions, managing the LC's budget, and arranging plans and curricula for each English

language unit. The TAU is responsible for adopting training programs to improve instruction,

conducting studies about the LC, contacting reputable educational institutions and universities to

adopt the latest language programs, and proposing techniques to improve administrative and

instructional work, thus creating an ideal atmosphere for a successful language learner. Finally,

the TMU's responsibilities include grading and constructing standardized placement tests

(discussed in the following paragraph) and final tests for remedial and general English courses.

The standardized midterm and final tests that are designed by the TMU are referred to as

"unified tests," which are for GE courses. Standardized GE course tests were mandatory for all English units, but the policy changed in 2012-2013 to allow each unit to decide whether tests must be unified or individual teachers may construct their own exams. Teachers who construct their own exams can request answer sheets from the TMU so their tests can be scored electronically. The TMU provides model answer sheets that are machine scored not only for GE courses but also to ESP teachers who want to construct their tests based on this format.

Each English language unit is expected to offer curricula based on the needs of the main departments and majors of the five PAAET colleges. However, policies are standard in the sense that each college conducts the same English placement test on a day specified in the academic calendar prior to the beginning of the semester. This standardized test is constructed by the TMU, and is printed and securely distributed to each English language unit in the five colleges, on both male and female campuses. Students (in all PAAET colleges) who obtain 60% or more can enroll in an accredited GE course. Students cannot enroll in any ESP courses unless they complete a GE course. Students who score less than 60% on the test take non-credit remedial GE courses. The LC selects the same book for both the Remedial GE and GE courses to be used by all English language units in the five colleges. All GE courses are compulsory. The ESP courses, on the other hand, depend on the student's major.

3.2.1.1 Courses offered by the LC.

The LC offers a variety of courses. Each college must offer two GE courses: Remedial GE and GE level one, and ESP courses. In the CBE, however, there are two GE courses besides Remedial GE. Table 3.1 lists the English courses offered by the LC at the five colleges.

Table 3.1

List of English Courses Taught in the English Units of the LC

College	Remedial GE Course	GE Course	ESP Courses
CBS	5 hours/week (including an hour for laboratory (LAB)) (non-credit)	4 hours/ week (including an hour for LAB)	All ESP courses at CBS are 3 hours/week - 3 ESP courses (business administration, accounting, and banking and insurance) - 2 ESP courses for computer science - English for Law
CBE	5 hours/week (non-credit)	- GE level 1 (3 hours/week) - GE level 2 (3 hours/week)	All ESP courses at CBE are 3 hours/week - English for Computer Science Education - English for Special Needs - English for Library and Information Science - English or Science
CTS	5 hours/week (non-credit)	5 hours/week	- English for Technology (5hours/week)
CHS	- Remedial English I - Remedial English II Both are 5 hours/ week (non-credit)	3 hours/week	The following ESP courses are 3 hours/week - English for Food Sciences - English for Environmental Health - English for Medical Laboratories - English for Medical Records—Level I (10 hours/ week) 5 credit - English for Medical Records- Level II (4 hours/ week) 2 credits - English for Medical Records- Level III (4 hours/ week) 3 credits - English Language Composition (4 hours/ week) - Advanced Reading English (2 hours/ week) 2 credits
CoN	- Remedial English I (10 hours/ week— non-credit) - Remedial English II 5 hours/week (non-credit)	3 hours/week	- English for Nurses I (2 hours/week) 2 credits - English for Nurses II (2 hours/week) 2 credits - Medical English I (10 hours/week) 5 credits - Medical English II (10 hours/week) 5 credits - Medical English III (6 hours/ week) 3 credits

3.2.1.2 Teachers' Role in the LC.

Teachers in the LC are required to teach Remedial GE, GE, and ESP courses. There are no specific criteria regarding who should teach any of the courses. However, more experienced teachers and PhD holders often teach ESP courses, while novice teachers may teach GE and ESP (the beginner level). Teachers select the English courses they want to teach before the semester begins. However, the LC administration distributes courses according to vacancies and teacher preferences. Unlike in public schools, where teachers' voices carry less weight regarding assessment than do those of policy makers, teachers in postsecondary institutions in general and the PAAET in particular have more autonomy to construct their own assessments and/or assessment criteria especially in ESP courses. Although there is a TMU in the LC, its only responsibility is to construct placement tests and standardized midterm and final tests for the Remedial GE and GE courses (upon request from faculty members or heads of English units).

CHAPTER 4:

METHODS

This chapter focuses on the study's methods. It presents the study design and the demographic information of the participant, by first describing the selection criteria for the teacher participants. Second, it describes the data collection procedures, which include initial interviews, classroom observations, and post-observation interviews. Third, it describes the data analysis procedures, and finally, it discusses the validity and ethical issues related to this study.

4.1 Study Design

This study adopts a multiple-case design using qualitative methods guided by the study research questions:

1. What are the L2 assessment beliefs and practices of seven EFL teachers teaching GE and ESP courses at a language center in a public post-secondary institution in Kuwait?

2. What factors influence the classroom assessment beliefs and practices of these EFL teachers?

Researchers widely consider qualitative studies to be appropriate for investigating teachers' beliefs and the relationships of such beliefs to classroom and professional practices (Borg, 2006; Clark & Peterson, 1986; Speer, 2005). A case study is defined as "the study of the particularity and complexity of a single case" (Stake, 1995, p. xi). A case study is helpful to obtain "a thick description of a complex social issue embedded within a cultural context" (Dörnyei, 2007, p. 155), and through qualitative methods employed in case studies, the researcher can undertake a thorough analysis of teachers' beliefs and observe their practices in real-life situations. As discussed in Section 2.8, there are several methodological challenges in investigating teachers' assessment beliefs and practices and their relationships (Borg, 2006).

Speer (2005) pointed out that several studies found major discrepancies between teachers'

professed and attributed beliefs. She urged researcher to select appropriate methods when

examining the relationship between teachers' beliefs and practices. In this regard, a qualitative

research design is particularly appropriate for this study in order to explore teacher assessment

beliefs and practices in-depth through interviews and classroom observation. Qualitative

methods also allow the examination of teachers' practices in natural settings. A justification of

the use of each of the qualitative data collection methods used in this study is given below (see

Sections 4.3.1 and 4.3.2.). According to Yin (2014), the same study can have more than one case,

which means it adopts a multiple-case design. This study examines each teacher as a separate

case, so it adopts a multiple-case design. Yin (2014) pointed out that multiple cases make a study

robust and compelling.

A case study is a multifaceted, bounded system. A case can be a single person or group of

people, an event, an organization, a setting, or a program (Patton, 2002; Stake, 2005). A

qualitative case study aims to describe the unit of analysis, regardless of its type, in detail, depth,

and context (Patton, 2002). There are two types of case studies. An *intrinsic* case seeks to

understand a particular case because of its uniqueness, but that case is not meant to represent

other problems or cases. In contrast, an *instrumental* case study investigates a particular case in

depth to understand a more general external issue. When there is less interest in an individual

case, then a *multiple case study* can investigate a number of cases (Stake, 2005). This study is an

instrumental multiple-case study to understand an issue from an emic perspective.

Although the participants in this study work at the same institution (i.e., the LC in

PAAET), I sought to conduct an in-depth study of each teacher as a separate case rather than

looking at the participants as a group. I treated each participant as an individual case in order to

understand the assessment beliefs and practices of each teacher as well as the various external factors influencing each teacher's beliefs and practices. I also wanted to compere teachers in different colleges. To do this, I first analyzed data from each teacher separately to identify the teacher's assessment beliefs and practices, then I compared and contrasted cases to fully understand the phenomena of classroom assessment in this particular context and the external factors that influence teachers' assessment beliefs and practices, hence achieving the purposes of an instrumental case study.

This study adopts qualitative methods because they are best suited for grasping an in-depth understanding of teachers' cognitions (Borg, 2012; Li & Walsh, 2011). The main methods employed in case studies are observation and interviews. In this study, I selected the design and methods according to the research questions. I also followed three stages for data collection (see Section 4.3): (a) initial interviews, (b) classroom observations, and document collection, and (c) post-observation interviews. Before conducting the study, I applied for ethics approval from the Office of Research Ethics at York University; I received approval to conduct this study on July 09, 2015 (see Appendix A for approval form). After receiving ethical approval, I started the first stage of data collection. During the first stage, I conducted initial interviews with the participants about the themes related to the research questions (assessment beliefs, assessment practices, and the external factors that influence their assessment beliefs and practices). Based on the initial interview of a participant and document analyses, I determined the criteria for classroom observations (the second stage) and areas of investigation for the post-observation interviews (the third stage). The second stage was conducting classroom observation within the same time period as the initial interviews (i.e., in June of 2015), I also collected teachers' documents and materials (e.g., course syllabi and teaching plans) related to the courses that I observed. The post-

observation interviews followed classroom observations to relate classroom practices to teachers' reported beliefs and practices. The collection of documents took place during the three stages of data collection, because not all participants submitted their materials in stage one.

Table 4.1 shows the design and data collection methods related to each of the research questions of the study.

Table 4.1

Data Collection Strategies of the Study

Research Question	Method and Duration
1. What are the L2 assessment beliefs and practices of seven EFL teachers teaching GE and ESP courses at a language center in a public post-secondary institution in Kuwait?	Initial interviews and classroom observations and post-observation interviews **Duration:** June 2015–July 2015 (initial interviews) June 2015–July 2015 (classroom observations) August 2015–November 2015 (post-observation interviews)
2. What factors influence the classroom assessment beliefs and practices of these EFL teachers?	Initial interviews, classroom observations, and post-observation interviews **Duration:** June 2015–July 2015 (initial interviews) June 2015–July 2015 (classroom observations) August 2015–November 2015 (post-observation interviews)

4.2 Selection of Colleges and Participating Teachers

I selected teachers from the five PAEET colleges for this study: the CBS, CBE, CoN, CHS, and CTS. However, only one teacher from the CTS agreed to participate, and she was a part-time teacher in the summer in which I conducted this study. She was originally a teacher at the CHS, but because there was a shortage in teachers willing to teach in the summer, she was recruited to teach an ESP course at CTS part time. I interviewed and observed this teacher but during the interviews I found out that this was the first time she taught at the CTS department. As a result, I decided not to include her in the study. Interestingly, her interview and classroom

observation data yielded similar results to those of the other participants. The final sample, thus, included only four of the five colleges: CBS, CBE, CoN, and CHS. As previously mentioned, education at the PAAET colleges is gender segregated; I included teachers only teaching on female campuses because I thought, as the researcher, I could establish a better rapport with teachers and students on these campuses than I could on male campuses for cultural reasons.

I selected participants for the study using a *purposive sampling approach*. In purposive sampling participants are selected based on specific principles and criteria (Dörnyei, 2007). The sample selection for this study was mainly based on the following three criteria: the course type taught by the participant (i.e., ESP or GE), the number of teachers in each English language unit, and availability. The second criterion means that I aimed to select more than one participant from those colleges which have larger numbers of teachers. For example, because CBS and CBE have the highest number of teachers, I recruited more participants from CBS and CBE. Table 4.2 shows the total number of faculty members at each college for 2014 and their distribution by gender and qualifications.

Table 4.2

Distribution of Faculty Members at the PAAET Colleges

College	Number of Female Teachers	Number of Male Teachers	Number of PhD Holders	Number of Master's Degree Holders	Number of Bachelor's Degree Holders (Trainees)	Total Number of Faculty Members
CBS	24	10	5	29	3	34
CBE	11	3	1	9	4	**14**
CTS	5	7	2	8	2	**12**
CHS	6	1	2	4	1	7
CoN	6	0	1	5	0	**6**
Total Number of all Faculty Members in the LC (April 2014)						**73**

Each year, teachers are transferred across units in the five colleges based on their requests or administrative decisions to fulfill needs. Therefore, the number of teachers in each unit may change every year. However, there are always more teachers in the CBS and CBE than other colleges because they accept a larger number of students each year, so the LC recruits more teachers for their English language units. To reflect the differences in the number of teachers across colleges as shown in Table 4.2, I selected five teachers from CBS, two from CBE, one from CoN, two from CHS, and one from CTS to participate in the study, for a total of 11 teachers. After analyzing the data, I included seven participants in the study because I reached data saturation. Saturation occurs when additional data no longer adds new insights to the research questions but simply repeats what previous participants have conveyed (Creswell, 2014; Glaser and Strauss, 1967). Though this sample is small, as Dörnyei (2007) pointed out, "a well-designed qualitative study usually requires a relatively small number of respondents to yield the saturated and rich data that is needed to understand even subtle meanings in the phenomenon under focus" (p. 127).

Finally, the availability criterion refers to only those participants who were teaching ESP or GE courses in the summer semester and were willing to participate in all three stages of the study. I explained to the participants that the post-observation interviews would follow the classroom observations and that I might ask them to provide further materials even after conducting the post-observation interviews. Although male teachers taught some courses at female campuses, there was only one male teacher teaching an ESP course at CBE at the time of data collection. However, I chose female teachers because, as I mentioned above, for cultural reasons, I could build better rapport with them to ask about their availability than I could with male teachers. I refer to the seven participants using pseudonyms.

Based on the above criteria, the process of selecting the participants went through several phases. The first step was to contact the institution at which I planned to conduct the study for approval. It required the permission of the director of the LC in CBS, which she gave to me to conduct my study in the English units of the five colleges coordinated by the LC. I contacted the director of the LC to ask for the names of all female faculty members teaching on female campuses in the summer semester in which I conducted this study. The administration office of the LC e-mailed me the faculty members' information, including their summer course schedules and their e-mail addresses or cell phone numbers.

Second, I sent an e-mail asking all teachers who were teaching in the summer semester to participate in my study. The e-mail included a description of the topic and objective of the study. I also sent text messages to some teachers who did not have email addresses listed in the faculty members' information sheet. The next step was to obtain consent and signatures from the participants. I sent an e-mail to potential participants to request their participation. All participation was voluntary, and participants had the right to withdraw at any time during the study. I provided them with informed consent letters (see Appendix B), which explained the general objectives of the study, tasks participants were expected to perform, possible consequences or risks of participation, data confidentiality, and their right to withdraw at any time (Dörnyei, 2007).

Eleven teachers volunteered to participate in the study. As noted above, four of these teachers were excluded from the study for the following reasons. One teacher from CHS was excluded because she did not provide usable data about her assessment practices. One teacher from CTS was excluded because she was not familiar with the courses in the department. The last two teachers, both from CBS, were excluded because data saturation was achieved.

Specifically, the themes that emerged from their interview and observation data were similar to those of the other participants. It is important to note here that decisions about excluding some teachers all occurred during the of data analysis stage.

Table 4.3 presents demographic data for the seven participants included in the study. The data are presented according to the course type that the teachers were teaching in summer: pseudonyms, gender, ages, years of teaching experience, qualifications, and colleges. All the participants are Kuwaiti full-time teachers. Based on the first criterion (i.e., the course type), the participants who volunteered to participate in the study were as follows:

- One teacher who taught a GE course at CBS

- Two teachers in CBS and CBE who taught both ESP and GE

- Four teachers who taught only ESP courses at four colleges (CoN, CBS, CHS, CBE)

All the participants were females. Most of them were between 30 and 40 years old. Only two participants (Hajar and Leila) were between 55-65, and had the most years of teaching experience. All participants held MA degrees and attended a teaching conference at least once during their careers, except for Dana. Najla and Mona, however, did not attend any conferences, but held a professional development program certificate (i.e., CELTA).

Table 4.3

Participants' Demographic Data

Courses taught in summer	Pseudonym	Gender	Age	Years of teaching experience in general	Years of teaching experience in the LC	Training Experience	Qualification	LC English Unit
ESP	Najla	Female	35–40	9	7	CELTA certificate	MA	CoN
	Mona	Female	30–35	5	1 ½	CELTA certificate	MA	CHS
	Latifa	Female	30–35	6	6	Attended conference on English language teaching	MA	CBS
	Dana	Female	30–35	4	2	---	MA	CBE
GE	Nadia	Female	30–35	4	4	Attended workshops on teacher education and teaching strategies	MA	CBS
GE & ESP	Leila	Female	55–65	More than 28	27	Attended conference on English language teaching	MA	CBS
	Hajar	Female	55–65	34	22	Attended teacher education conferences	MA	CBE

Prior to conducting the initial interview, I explained to them what the study required in detail and asked them to read and sign a consent form (Appendix B).

A description of the participants' courses is necessary to track their practices in chapter five (presenting the study's findings). Table 4.4 presents the course descriptions, including the course names, types, numbers of students enrolled, credits, and course materials.

Table 4.4

Description of the Participants' Courses

Courses taught in summer	Pseudonym	Name of course	Number of students registered	Credits and hours	Course materials
ESP	Najla	English for the Medical Context Level 3	7	3 credits 6 hours/week	Textbook: *Oxford Nursing*
	Mona	English for Medical Laboratories	14	2 credits 3 hours/week	Reading kit, and external resources (e.g., worksheets)
	Latifa	Business Writing	45	2 credits 3 hours/week	Textbook: *Writing for the Real World* and worksheet taken from external resources (i.e., websites)
	Dana	English for Library Science and Technology	70	2 credits 3 hours/week	Reading kit from various resources
GE	Nadia	General English Pre-intermediate Level	40	Non-credited 5 hours/week	Textbook: Oxford *New Headway Pre-Intermediate*
GE & ESP	Leila	General English Level 1	40	2 credits 3 hours/week	Textbook: Oxford *New Headway Intermediate*
		Business English Level 2	40	2 credits 3 hours/week	Textbook: *Business Basics*
	Hajar	General English Level 1	50	2 credits 3 hours/week	
		English for Computer Studies	50	3 credits 3 hours/week	

Participants were teaching nine courses at the time of the data collection. Two participants were

teaching two courses each, one ESP and one GE. The other five were teaching one course each

(4 ESP courses, and 1 GE). Most GE and ESP courses were 2 to 3 credits, which met for 3 hours

per week. Nadia taught a non-credit remedial GE course, which was 5 hours a week, and Najla

taught ESP for 5 hours a week.

The number of students varied depending on the administrative policy of each college. The CoN has the least number of students, and CBE has the most. The number of students varied also according to the semester. In CoN and CHS, the number of students decreases in summer semesters. On the contrary, the number of students increases in CBS and CBE in the summer semesters.

4.3 Procedures

I used two instruments for data collection in three phases. The instruments were interviews and classroom observations. The first phase consisted of conducting initial interviews with the participants. In the second phase, I collected teachers' teaching and assessment materials, and conducted classroom observations. In the third phase, post-observation interviews were used to gain further insights on beliefs and practices intended to complement the initial interviews and the classroom observations. Table 4.5 shows the timeline for the three stages, which I will describe in the sections below.

Table 4.5

The Data Collection Procedure Timeline

Pseudonyms	College	Course	Number of initial Interviews	Initial Interview (II) Date	Number of Classes Observed	Date of Observation	post-observation Interview (PI) date
Mona	CHS	ESP	1	14/06/2015	2	23/06/ 2015	01/11/2015
						05/07/2015	
Najla	CON	ESP	1	15/06/2015	2	01/07/ 2015	PI 1: 17/08/2015 PI 2: 15/11/2015
						11/07/2015	
Hajar	CBE	GE	2	25/06/2015	1	25/06/2015	01/10/2015
		ESP		01/07/2015	1	01/07/2015	
Dana	CBE	ESP	1	11/06/2015	2	15/06/2015	27/09/2015
						07/07/2015	
Nadia	CBS	GE	1	04/06/2015	2	15/06/2015	16/10/2015
						24/06/2015	
Latifa	CBS	ESP	1	13/06/2015	2	21/06/2015	20/10/2015
						11/07/2015	
Leila	CBS	GE	2	28/06/2015	2	28/06/2015	02/11/2015
				15/06/2015		29/06/2015	
		ESP	1		2	21/06/2015	
						29/06/2015	

4.3.1 Stage One—Initial Interviews

Interviews are the major tool for data gathering in this study. The reason for depending on interviewing as a major method is because the research questions seek to explore teachers' beliefs around assessment in depth. Because no one can observe thoughts, intentions, and opinions, interviews allow the researcher to elicit information about reported behaviors and beliefs (Patton, 2002). Moreover, the interview is a flexible data collection instrument because it enables the researcher to explore thoughts through "multi-sensory channels" including verbal or nonverbal (Cohen, Manion, & Morrison, 2008, p. 349). For example, when the participants defined "assessment," I could draw inferences about their attitudes toward the concept through their expressions (e.g., laughter to show sarcasm or sighs to show disappointment).

There are different approaches to conducting interviews: structured, unstructured (Canh, 2012; Fontana & Frey, 2000)—sometimes referred to as "informal conversational interview" (Patton 2002, p. 342)—and semi-structured. Structured interviews require a researcher to carefully design each question ahead of time. The researcher may write the questions in detail, and they do not change throughout the interviews. One of the main purposes of structured interviews is to ensure the researcher interviews all participants the same way and asks them the same questions (Patton, 2002). Researchers mainly conduct unstructured interviews in fieldwork that requires a natural, spontaneous flow of interaction that occurs when the participant is under observation (Patton, 2002).

I used semi-structured interviews in this study. Although prompts and questions prepare and guide them, "the format is open-ended and the interviewee is encouraged to elaborate on the issues raised in an exploratory manner" (Dörnyei, 2007, p. 136). Dörnyei (2007) added that the semi-structured interview method is useful when a researcher has knowledge and background of the research topic. Such an approach combines structured and informal conversational interviews in that it is more conversational than structured interviews are, but it also allows the researcher to adopt a structured approach by designing the main questions in advance (Minichiello, Aroni, Timewell, & Alexander, 1995). I found the semi-structured interview method to fit this study because as Borg (2006) stated, it is an appropriate method for investigating teachers' cognition and to explore their professional and educational experiences in depth. As I am a teacher and a part of this teaching context, although I designed predetermined, specific questions to guide the interviews, I knew they would be conversational. I consider this to be a benefit because it created a comfortable atmosphere in which the participants expressed their beliefs with ease, knowing I am also a part of their context. According to Canh (2012), this interviewing method in general

has some challenges, such as social constraints, but I did not sense any social constraints,

perhaps because I was able to establish some form of rapport with the participants as their

colleague.

I conducted interviews in two stages. First, I conducted initial interviews prior to

classroom observations. After conducting all classroom observations and collecting materials in

stage two, I conducted post-observation interviews in stage three. I designed the initial interview

questions to address the research questions, so I divided them into four sections (see Appendix

C). The first section concerned participant backgrounds and demographics. The remaining three

sections were all related to my research questions: (a) teachers' assessment beliefs, (b) teachers'

assessment practices, and (c) the impact of educational policies and external factors on teachers'

beliefs and practices. Participants discussed all topics in the order listed in the interview. I

adapted some of the interview questions from the literature (e.g., Rea–Dickins, 2001; Troudi et

al., 2009).

The demographic questions sought the participant's teaching background (e.g., years of

experience, colleges at which they taught), academic background (e.g., qualifications,

professional development programs attended), and the names of the courses they were teaching

in the semester in which I conducted the study.

The second section sought participant's definitions of assessment in general and

classroom assessment in particular. It also investigated their beliefs about what language

assessment should focus on and the uses and purposes of assessment in the classroom. Although

some participants taught only GE or ESP courses in the summer while others taught both

courses, section two sought their beliefs about the differences between the two courses in terms

of assessment. The second section explored teachers' understanding about their roles as well as

student's roles in assessment and the factors that might influence the teachers' assessment beliefs.

The third section of the interview related to teacher assessment practices. Most of the interview questions in this section sought examples of classroom assessment practices. I first asked for instances of assessment practices in general, then required a specific example of assessment practices based on the language focus of the course. Second, I asked participants to provide examples of assessment methods used in the classroom. Third, I asked about feedback and monitoring practices. Next, I asked teachers to compare their assessment practices in the ESP and GE courses. I asked those participants who taught both ESP and GE courses at the time of the study about their practices twice: in the first initial interview, I asked questions about one course, and in the second, I asked questions about the other. However, some questions asked participants to describe their assessment practices in both types of courses in general, regardless of the course they were teaching at the time of the study. I also asked teachers about their assessment rubrics and criteria for both summative and formative assessment. Finally, I asked teachers about their roles in classroom assessment and any challenges that might hinder specific assessment practices.

The last theme of the initial interview concerned the external factors that might influence participants' assessment beliefs and practices. The first set of questions in this section was about the policies of the institution in general (i.e., the PAAET) and the LC specifically, such as grading, tests (for both ESP and GE courses), and reporting assessment results. I also asked participants about their beliefs about the assessment policies and their responses to that policy. I also asked for examples of test construction for the course selected for the study. Some participants described how they constructed assessments for GE and ESP and compared the two.

Finally, I asked teachers about other factors that shape their assessment beliefs and practices. It should be noted here that, while the teachers knew about the main focus of the study when they volunteered to participate in the study, and while I had discussed with them the interview themes before I conducted the initial interviews, they did not know the specific interview questions before the initial interviews.

There were predetermined questions for the post-observation interviews, but it was mainly based on segments from the recordings selected from classroom observations (see example in Appendix D). Prior to conducting the initial interviews, I conducted a single pilot interview with an eligible English teacher at the same institution who was not teaching any courses when I conducted the study to improve the interview.

Based on the pilot, I modified some of the interview questions and eliminated some redundant ones. For example, there were three questions at the beginning of the beliefs theme about the participant's description of "assessment." The first was about the general definition, the second was about classroom assessment, and the third was about language assessment. The pilot participant answered the questions on language assessment and classroom assessment similarly. Consequently, I omitted the question on language assessment in subsequent interviews. Instead, there were other questions in the second section (e.g., teacher practices) that asked about the language focus included in participants' classroom assessments.

The teacher in the pilot study further pinpointed questions with difficult wording. I amended these questions accordingly to avoid misunderstanding. There were also some difficult terms in the third section regarding the external factors that influence teacher beliefs and practices. For example, I explained the policy mandates because the pilot participant did not understand which policy I was referring to. However, the pilot participant was a newly recruited

teacher at the time of the interview. I assumed that participants in the main study would be familiar with the department policies.

In the pilot interview, I asked all the questions I planned to ask in the main study. This took longer than the intended 45–60 minute, so I combined some questions to avoid redundancy. I further eliminated questions if I saw that the participant had already answered them in response to other questions.

4.3.1.1 Initial interview procedures.

Right after the pilot study, I contacted teachers who consented to participate to schedule meetings to further explain the study purpose and methods and get their written consent prior to the interviews. I gave those participants who asked for more details the opportunity to discuss basic interview questions to obtain detailed responses from me. It is also important to note that I conducted interviews during the same periods as classroom observations. For example, on one day, I scheduled an interview with one participant and a classroom observation session with another. The short summer semester required this accelerated timeline. The period of initial interviews lasted for a month.

There were several purposes for the initial interview. First, I wanted to explore teachers' understanding about classroom assessment. I also wanted to identify examples of their assessment practices. I then sought teacher reports about the external factors that might influence their assessment beliefs and practices. Second, because I conducted the initial interviews prior to classroom observations, I planned to use them to decide which classes to observe and to identify themes for the post-observation interviews.

Although all teachers taught English and were fluent in English, I urged them to shift to Arabic, their L1, if they felt they wanted to. There are several reasons for asking them to use

Arabic. First, I wanted them to describe their beliefs in depth without feeling any linguistic restrictions on expressing their views, and I knew that there are some Arabic expressions that could precisely describe their attitudes and thoughts better than English. Second, I wanted them to feel at ease; because Kuwaiti Arabic is the dialect widely spoken in the country, department colleagues use it in daily communication. Third, to ensure clarity in expressing opinions and avoid confusion, I sometimes asked participants to translate what they meant by certain expressions when they described their thoughts. Most of the interviewees used mostly English, but there were many occasions when they resorted to Arabic.

I not only encouraged participants to use Arabic, but also explained some questions in Arabic if I felt the participants did not fully understand the purpose. However, translation created some challenges for addressing questions about assessment. In Arabic, there are many equivalents for the word "assessment," which could be more specific to the notion of testing or program evaluation. I kept using the word "assessment," which could be why some participants thought of the word "assessment" as "testing" in Arabic (see Chapter 5). Code switching also created some challenges for transcribing the interviews and translating expressions because some words have no equivalent in English.

As Table 4.5 shows, I started conducting the initial interviews in May 2015. I audio recorded all interviews for subsequent transcription. The first initial interview was conducted via Skype. I conducted most of the initial interviews face to face in Kuwait. I conducted the second interview at my home to ensure anonymity because the participant shared her office with another colleague, which was not ideal. I conducted the rest of the initial interviews at the college where the participant works in quiet rooms.

I conducted all the initial interviews in one face-to-face session from 40 to 60 minutes, except the first one, which I conducted via Skype and which exceeded one hour. I conducted two separate initial interviews with each participant who was teaching both types of courses. The first focused on their assessment beliefs and practices related to GE courses, while the second focused on their beliefs and practices related to ESP courses. I conducted the two interviews on separate days and asked the same questions the second time, excluding questions where their answers would not differ based on the course.

I audio recorded all interviews using a digital audio recorder. I conducted all interviews prior to classroom observations so they could assist with planning classroom observations. While conducting the interviews, I also collected relevant course documents and materials from the participants. Based on my analyses of the interview recordings and teaching materials, I prepared a set of guidelines for classroom observations. Listening to the interview recordings before the classroom observations helped me determine some areas on which to focus during observations. For example, I determined the language focus of the courses and looked for different teacher assessment practices for this specific language focus (e.g., teacher assessment practices for grammar).

4.3.2 Stage Two

This section presents the second phase of data collection, which involved classroom observation and collecting course materials. This phase took place in June through July of 2015.

4.3.2.1 Classroom observation.

Classroom observation is essential to studying teacher beliefs because it "provides direct evidence of behaviour . . . and allows large amounts of descriptive data to be collected" (Borg, 2006, p. 227). Some researchers have found it necessary to complement the narrative description

of a participant's beliefs with observations from his or her classroom because verbal data is insufficient (Patton, 2002). According to Borg (2006), although observation may provide information about teachers' beliefs, it is "insufficient as a means for exploring [cognitive] processes in more depth and ascertaining the validity of the inferences made" (p. 231). Corbin and Strauss (2008) added that "it is not unusual for persons to say they are doing one thing but in reality they are doing something else" (p. 29). Another reason previously discussed (see Section 2.8) is that teachers may report their practices and beliefs for ideal situations, but the observation enables the researcher to examine them in reality and compare the two (Speer, 2005). The researcher can derive an understanding of a teacher's beliefs from the behavior observed in his or her classroom.

My role in this study was that of a nonparticipant observer. Although nonparticipant observation is preferred for researching teachers' cognition (Borg, 2006), studies on teachers' beliefs have reported that teachers sometimes ask researchers to participate (e.g., give an opinion or answer a student's question) at some point during their observations (Borg, 2006; Yin, 2005). Borg (2006) noted that researchers may find it embarrassing to ignore the request to participate, yet they may not wish to take on the role of teacher. To address this issue, I explained my role to the participants prior to conducting classroom observations, so I did not face any situation in which the participant asked me to get involved. To avoid creating discomfort among students that could alter their practices and reactions, I chose to sit in the back seats.

I chose to collect information during the observation using qualitative unstructured observation (i.e., narrative or descriptive note taking) (Borg, 2006; Stake, 1995; Yin, 2005) for two reasons. First, observation in general is time-consuming (Corbin & Strauss, 2008), and structured observation protocols require significant time to prepare and more time and piloting

than unstructured observations do. Because the summer semester was short, I did not conduct a pilot study for the classroom observation and chose the unstructured approach instead. Because there was only a short time between the initial interviews and classroom observations, I did not have enough time to create an observation scheme based on the participants' reports. I did not plan to have one scheme for all the participants, but rather looked at each participant as an individual case. Second, while structured observation is a unique data collection method associated with classroom studies, I took into consideration a disadvantage that Borg (2006) identified. Based on Evertson and Green (1986), Borg (2006) stated that, in structured observation, the observer may ignore events or behaviors not covered in the structured observational guidelines. Because the participants did not report all their practices in the classrooms, I was willing to explore all aspects of their assessment practices in their classrooms without determining specific categories ahead of time. Some practices may have emerged during observation that did not appear in the observation scheme. I also considered the advantages of unstructured observation cited in Cohen et al. (2008): "unstructured observation provides a rich description of a situation which, in turn, can lead to the subsequent generation of hypotheses" (p. 398).

Although I did not have detailed observation scheme prior to the classroom observations, I determined general criteria for them before the observation took place. The first section included information about the class, such as the time and duration, course name and number, number of students, and arrangement (i.e., class layout). Table 4.5 below provides other general criteria, adopted from Hill and McNamara's (2012) guidelines of classroom-based assessment.

Table 4.6

Guidelines of Classroom-based Assessment (Adapted from Hill & McNamara, 2012, p. 397)

Evidence	Data Approach	What is assessed? How is evidence collected?
	Target	Who is assessed?
	Agent	By whom?
Interpretation	Reflection	Level of attention
	Criteria	Values guiding assessment
Use	Purpose	How is evidence used?
	Agent	By whom?

The guidelines by Hill and McNamara (2012) in Table 4.6, helped me to construct some

observation questions and themes to consider during classroom observation. I also used some

guidelines for directing the observation activities (see Appendix E) from LeCompte and Preissle

(1993) and Spradley (1980) cited in Cohen et al. (2008). The guideline directs the researcher to

certain activities, scenes and events in the observation site through questions such as, how many

people are in the activities? What are their characteristics and identities? How are the activities

being explained? (Cohen et al., 2008). The rest of the sections in the observation schedule

address several basic topics: (a) exploring the teacher's role in classroom assessment, (b)

identifying the assessment activities for different language aspects (e.g., assessing grammar and

vocabulary), (c) looking for teacher feedback and monitoring, (d) identifying sources and tools

of assessment, and (e) investigating some external factors that might be observable other than

those the participants reported (e.g., classroom size and student interaction and motivation).

4.3.2.1.1 Classroom observation procedures.

Classroom observation started in June for a 5-week period. After conducting the initial

interviews with each participant, I scheduled an appointment with her for classroom

observations. The class durations varied according to the college or the type of course. ESP classes ran for 2 hours, 3 days per week, while GE classes were 2 hours, 4 days per week. Teachers do not teach the full hours; they usually end the class about 5 to 10 minutes early to allow students to commute between buildings for their other classes. Also, some classes had delayed starts due to teachers calling names for attendance or setting up equipment. In CoN, the ESP class ran for 3 hours, 4 days per week.

Table 4.5 presents the schedule for the seven participants in this study. To address Research Question One, concerning those participants who taught both ESP and GE courses, I observed four classes for each: two GE and two ESP courses. I selected classes to observe based on discussions with each participant and examination of the assessment materials for each course. As Table 4.5 shows, I conducted two observation sessions with most of the participants. I conducted only one classroom observation for some participants because of their busy schedules, and they agreed to let me observe only one class. However, in summer semesters, classes are longer than during other semesters because summer courses are intensive, and teachers tend to employ more activities to cover the curriculum in a short period of time.

There were several reasons behind selecting two classes per participant. First, as mentioned previously, the summer semester is shorter than regular semesters are, so I was unable to observe more than two classes. Second, one class might not be enough to observe a teacher's assessment practices because teachers could assess language skills differently. I planned to observe more than two classes but considering the time I had for data collection and observation, I confirmed that two classes were adequate—and I gathered a good amount of data for the classroom assessments. Third, most participants allocated each of their classes to no more than two language aspects (e.g., grammar and reading). Therefore, I selected two classes to be able to

observe classroom assessment practices related to different language aspects. Finally, I planned to compare assessment practices between ESP and GE courses, so I observed two classes of each type.

Researchers must record observations in some way—such as through field notes, photographs, or video recording—and with artifacts of teaching materials and students' work (Borg, 2006). Although video recording provides better insight into teachers' assessment practices, most participants refused to allow this except for one participant in CoN (Najla). Because I am familiar with the research context, I expected some teachers would not permit their classes to be video recorded for various reasons. I used audio recording instead. To address the limitations of audio recording, I took detailed observation notes during each session. I audio recorded the classrooms using the same audio recorders I used in the interviews. One recorder was placed on the teacher's table, and the other recorder was placed on the table where I sat. I arrived 10 minutes before the class began to set up the recorders.

To follow the ethical guidelines for the observation, I provided informed consent letters to students (Appendix F) upon their participation and possible voice and video recordings (in the case of CoN). In the case of video recording, I made sure to set the camera in positions that did not show the students' faces. I placed it in the back of the room and recorded their appearances from behind. There was no incident in which any participant withdrew from the study or any student refused to participate.

I used field notes in the form of narrative description. LeCompte and Preissle (1994) stated that field notes contain narrative descriptions of places, people, events, or patterns of interaction. In my study, the notes describe what happened in the classroom. LeCompte and Preissle (1994) added that field notes should answer the questions of "Why?," "What?," "Who?,"

and "How?" The guidelines in Appendix E guided the notes I took. I divided the notebook into two columns (Appendix G). At the top of the page, I wrote the participant's name, course number, and time. In the left column below the title, I wrote descriptive notes of what was happening in the classroom, and in the right column, I allocated a smaller space than the left column for writing any notes or comments that would assist with the case analysis. The observation schedule I used for guidance distracted me during the first classroom observations because I was looking only for answers to the questions in the guideline. In the first observation session, I answered all the questions, but for the rest of the observation sessions, I had some basic questions already in mind while observing, such as types of feedback, assessment methods, and monitoring strategies. The only questions I answered during observation were those related to information about the lesson, such as the class time, number of students, and the classroom physical setting.

4.3.2.2 Collecting documents.

During the first phase, I asked the participants for documents related to the courses selected for this study. Some handed me hard copies of documents such as course syllabi, assignments, policy statements, and reading lists, while others e-mailed them. The process of collecting documents did not end in stage one. The documents included copies of quizzes, midterm and final exams, course descriptions, and worksheets. During stage three, I reminded those teachers who did not share some of their tests to do so. The documents helped me understand the assessment practices that the participants described in the initial interview, including rubrics and paper-based assessments. The course plans helped me understand the objectives and focus of the courses. None of the course plans included details about teacher assessment practices, except for the times of the tests and/ or the course grading scale.

4.3.3 Stage Three—Post-observation Interviews

The third stage of data collection consisted of post-observation interviews. The purpose of these interviews was to ask the participants about specific classroom assessment practices and teaching materials observed in class, as well as elaborate on specific points from the initial interviews. After conducting classroom observations with each teacher, I asked them to schedule sessions for post-observation interviews at their convenience.

The initial plan was to conduct the post-observation interviews immediately after the last classroom observation. However, there were several circumstances that prevented this. First, because the observations were immediately followed by the final exam period and then summer break, most participants preferred to do their post-observation interviews after the summer break. However, only two participants were available and agreed to participate in post-observation interviews in August during the summer break. I conducted face-to-face interviews with both of them off campus. Second, the period between the initial interviews, classroom observations, and post-observation interviews was too short to allow me to listen to the initial interviews and classroom recordings to identify themes related to classroom assessment before conducting the post-observation interviews.

I started the post-observation interviews in August and continued them through November, based on teachers' availability and schedules. There was only one post-observation interview per participant, except for Najla, whom I interviewed twice because her first interview was not clear due to our meeting in a loud café. I was able to transcribe the first post-observation interview with her, but I needed clarification on many points. Table 4.5 presents the timeline for the post-observation interviews. I conducted all the interviews but two via phone calling software (FaceTime and WhatsApp video and voice calls) because I returned to Canada in September. I

only interviewed Najla face to face because her interview, as mentioned previously, took place during summer break. I conducted Najla's second post-observation interview over the phone.

Like the initial interviews and classroom observations, I used two voice recorders. Before conducting the post-observation interviews, I contacted the participants to schedule an appointment for the post-observation interview. Once I scheduled the time with each participant, I went over the initial interviews and classroom observations recordings to select segments related to classroom assessment to play it back to the participants during the post-observation interview. I determined the post-observation interview questions based on findings from the teachers' initial interviews, document analyses, classroom observations, and the literature on classroom assessment. For example, I looked for the general criteria presented by Hill and McNamara (2012) in Table 4.6 while going through the classroom observation audio recordings. If I was not certain whether a specific practice was an assessment or a teaching practice, I created a question asking the teacher to explain what she was thinking when this practice took place. Before playing the audio, I asked the teacher to use the language she preferred to express her thoughts (i.e., Arabic or English). Most of the participants code switched to Arabic most of the time. I also asked them to feel free to stop me at any time if they had a specific question or comment, but none did, except when the audio recording was unclear.

Similar to the initial interviews, the post-observation interviews occurred in a semi-structured manner. However, I created a list of open-ended questions based on the observation field notes and the recording. Appendix D presents some general questions that guided the post-observation interviews. The post-observation interviews were more informal than the initial interviews because we got more acquainted during the data collection period in Kuwait, and some participants were colleagues. Some interviews took longer than others did because some

participants talked about issues not related specifically to the topic of the study. Many times, I stopped recording the interviews because participants talked about personal matters outside the study focus. Additionally, sometimes the connection was poor, which meant I needed to call the participants several times until their voices were clear enough to resume the interview. These challenges occurred during only two interviews, but these two as well as the other interviews were ultimately successful.

4.4 Data Analysis

There is no single agreed-upon method for analyzing qualitative data (Patton, 2002), but the researcher must decide what he or she wants from the data based on the study's purpose (Cohen et al., 2007). This study adopted an inductive approach to data analysis as explained below (Creswell, 2014).

4.4.1 Preliminary Analysis

Preliminary data management and analyses started after I completed stages one (i.e., initial interviews) and two (classroom observations, and document collection). The main method of analysis I used for this step was manual. I wrote down themes and related them to audio segments. Before conducting the post-observation interview with each participant, I listened to the recordings of the initial interviews and classroom observations. The purpose of this step was twofold. First, it established the key points to consider in the post-observation interviews by identifying themes in relation to participants' assessment practices and beliefs. Although I did not intend the initial analysis to be each case's individual analysis, I created a document titled "notes on analysis" with the name of the participant, and I wrote comments about points that would help with the individual analysis. Second, this preliminary analytical step allowed me to get acquainted with the data to determine which participants to include in the final case analysis.

At this stage, for example, it showed that one participant from CHS did not provide enough data regarding assessment, so I excluded her from the study.

4.4.2 Transcription

The first step of qualitative data analysis—especially interviews—is to convert the audio recordings to text because it helps a researcher understand the data thoroughly (Dörnyei, 2007) and prepare for the analysis (Patton, 2002). After completing the initial and the post-observation interviews, the transcription process started. Patton (2002) pointed out that a researcher doing an interview transcription alone, rather than using an external transcriber, "provides an opportunity to get immersed in the data, an experience that usually generates emergent insights" (p. 441). Taking Patton's view into consideration, I did all the transcription myself using an application designed for Macs called *Transcriptions*.

I treated all participants with confidentiality according to the ethical guidelines. I kept all participants' data private, and I discussed participants using pseudonyms. I did not use participants' data for any other purpose than this research. I stored all interview transcripts, field notes, audio or video recordings, and Word documents on my Mac computer, which no one else can access.

The plan I followed was to transcribe all the interviews for one participant at a time. First, I transcribed the initial interview, saving the transcript into a Word document titled with the participant's name. Next, I transcribed the post-observation interview of the same participant and included the transcription in the same document as the first one. The purpose of transcribing the initial and post-observation interviews of one participant at a time was to help recall thoughts and ideas while analyzing the case during and immediately after each transcription. While listening to the participants' interviews, I wrote down comments to guide my case analysis.

After completing each transcription, I listened to the recording again for accuracy. However, I did not follow the same procedures for all participants. Because the transcription of each interview took more than 10 hours, I transcribed a number of them—and when it was the time to analyze the data, I relistened to the recordings to recall my thoughts. I additionally referred to the documents containing the comments I wrote throughout the transcription process. I did not transcribe the classroom observation recordings. Instead, I listened to them carefully, wrote down times at which assessment incidents occurred, and transcribed only those segments that are relevant to the focus of the study.

Because the transcription procedure is time-consuming (Dörnyei, 2007), I did not transcribe every single word, because my main purpose was to obtain the content rather than the linguistic features of the speech. I omitted utterances such as "uh," "um," "you know," and "I mean" because they were meaningless asides by the participants. Yet, some nonverbal reactions informed the analyses, such as pauses, laughter, and sighing. Not all pauses were transcribed, but only those that served the data analysis. For example, when I asked one participant about her definition of assessment, she paused before she answered, which suggested that she was thinking about the concept at the moment of the interview. In this case, I used the code [pause] in the transcription. Table 4.7 shows other transcription conventions that I used throughout the transcription process.

Table 4.7

Transcription Conventions

Transcription Conventions	Description	Example from the study
/	Interruption	*I: Any level/* *P:/ Yeah yeah because when I taught in high school, we did continuous assessment*
[laughter]	Description of verbal reaction	*No, it's more work for instructors [laughter]*
X	Unclear word	*Cause we're doing all X courses*
( )	The word between brackets is uncertain/ not clearly uttered	*Because we don't have enough students so (we're willing to accept) anybody wants to apply.*
—	Incomplete sentences	*The word—when I hear the word assessment, I just—the first thing that comes up to my mind is whether my students are liable to be able to pass or not.*
[translated]	An indication of a researcher's translation of the sentence or word	*I am fortuitous [translated]*
Italicized word	Words emphasized by the participants	Well, I have learned one very important piece of information, which I am applying for a *very* long time
I	Interviewer	*I: And how many years?*
P	Participant	*P: for 11 years.*
[[ ]]	The researcher interpretation of the sentence	*So Maybe because of the policy I would consider this part [[grading]]*
Quotation marks	The participant indicated she says something to others	*I say, "good" for example.*
[sentence/word deleted]	An indication of confidentiality	*I don't know if I should say that, like [content deleted], the people from this specific background*
[pause]	An indication of silence	*Assessment? [pause], for example, quizzes, tests, that's it.*

Mason (2002) noted that because transcription is an inadequate method of recording nonverbal behavior and because some verbal expressions have no written translation in English (such as certain Arabic words), a researcher may use his or her own observation, judgment, interpretation, and experience of the interview when transcribing. I used my own understanding of some Arabic expressions and attempted to find the closest meanings in English. This was

sometimes challenging when an Arabic expression related directly to assessment cannot be translated into English. For example, a traditional expression *Al Shakwa lallah* [the complaint is to Allah (God)] is an idiomatic expression used in the Kuwaiti culture when someone expresses a problem is beyond his or her control. A participant used this expression when reporting her beliefs about the students' low English language proficiency level.

4.4.3 Coding

The period of transcribing facilitated my identification of the initial broad themes related to teachers' assessment beliefs and practices that I used later for coding. In the first case I transcribed, I used manual coding, but for the remaining cases I used NVivo. After completing the initial and the post-observation interviews of each participant, I uploaded the files to NVivo, which is the data analytical software most frequently used in qualitative research (Dörnyei, 2007; Mason 2002).

The first step of coding was to create thematic nodes ("node" is a term that stands for "code" in NVivo) based on my research questions. For example, I created the following nodes for Research Question One: teacher beliefs and teacher practices. Under these nodes, there was a hierarchy of subnodes, referred to in NVivo as "child nodes." The subnodes nodes under teacher beliefs were, for example, "beliefs about learners," "beliefs about teacher role," "purpose of assessment," and "definition of assessment." Some subnodes for teacher practices were, for example, using "worksheets," "grading," "asking short-answer questions," and "feedback" (see Table 4.8 for coding themes in NVivo). In addition to nodes for themes, I created several nodes for demographic information. In the first section of the initial interviews, for instance, there were some demographic questions about educational background. I created nodes for these questions

for reference later on in the case analysis process (e.g., educational background, course description and years of experience).

I created the parent nodes (general thematic codes at the top of the coding hierarchy) and child nodes before analyzing each case separately. I based the parent nodes on the research questions and ideas related to them (see Appendix H for an example of thematic NVivo coding). I developed the subnodes based on: (a) themes that emerged during transcription, (b) questions I developed from the interviews (i.e., interview themes), and (c) Hill and McNamara's (2012) coding scheme for classroom assessment practices. In addition, there were a number of new themes that emerged from carefully reading the participants' interview transcripts. Thus, the analysis adapted an inductive approach, in which themes emerge from the data (Creswell, 2014; Patton, 2002).

The process of coding involved reading through the participant's transcripts line by line and coding segments related to the research questions and themes. There were several segments which belonged to more than one node. For example, when a participant reported her beliefs about student proficiency levels, she linked this point to her assessment practices, such as making tests easy to enable students to pass the course. I coded such segments in terms of more than one code: teacher beliefs about student proficiency level, assessment construction, and external factors related to students. Table 4.8 shows the final coding scheme in the study.

Table 4.8

Coding Scheme

Coding Themes	Description	Examples from the study
1. Assessment beliefs		
a) Teachers' definition of assessment	Teachers' definition of assessment in general, not specifically related to L2 classroom assessment, which is either explicitly reported or induced from their description of assessment throughout the interview	*[Pause] assessment? Exams, and homework, and assignments, all kinds of assignments (Nadia, II).*
b) Teachers' definition of classroom assessment	Teachers' definition of classroom assessment specifically	*Classroom assessment is assessment done inside the classroom to basically informally check students' understanding (Mona, II).*
c) Teachers' beliefs about the purposes of assessment	Participants' reports about their beliefs about the purposes of assessment in the classroom	*I have to assess, because I feel like if I don't assess them on a weekly basis, they will not study, they will not revise. So for me [it is] to ensure that they pass my class and they are studying for the midterm and the final, I have to give them a quiz on a weekly basis. (Dana, II)*
c) Teachers' beliefs about the role of assessment in student learning	Teachers' beliefs about how assessment could improve student learning	*If you're talking about students, you always have different levels, so within one classroom, you can have an A student, B, C, D, or a failure one, so you try to manage it as a teacher, as a successful teacher let's say, and you try to let the whole participate with you. (Hajar, PI)*
d) Teachers beliefs about the importance of the language focus of assessment	Beliefs about the most important language construct to include in classroom assessment	*All of them. The writing is good, but the problem with our students is that our students are not at the level that they are ready to write. You might need plenty of time to teach them writing, which we don't have. This is my belief. (Leila, II)*
2. Assessment practices		
Summative assessment practices	Any reference to "the process by which teachers gather evidence in a planned and systematic way in order to draw inferences about their students' learning, based on their professional judgment, and to report at a particular time on their students' achievement" (Harlen, 2005, p. 247)	

• **Summative assessment methods**		
a) Quizzes	Short, paper-based tests given after a unit or a number of topics	*I give my students 4 quizzes and I take the best 3. So this is 15 (Mona, II).*
b) Midterm exam	Any practice related to exams given in the middle of the semester	*The GE was the midterm, they have a dialog and it was fill in the blanks, and it was misleading for the students. It was even confusing for the teachers (Leila, II).*
c) Final exam	Any practice related to the exam conducted at the end of the semester	*I have the flexibility whether to have it between like 40 or 50 marks, but I can't have 30 or 20 for final (Najla, II).*
d) Assignments	Any type of homework, paper project, or research paper that a teacher includes in summative assessment practices	*Well, I have given them homework, which we collect and we correct collectively or by peer group. Sometimes we do a lot of pop questions in the classroom, some which are credited, and some are not (Hajar, PI).*
• **Processes of Summative Assessment**		
a) Exam construction	The planning and procedures of creating summative assessments (assignments, quizzes, and midterm and final exams)	*There's a standardized format that I have to follow. But the majority of teachers here would do the unified exam. So a lot of teachers never created an exam. I prefer doing my own, but I follow of course the format, I have to follow the format, because someone has to read it, proof read it and approve that I've given them my exam. (Dana, II)*
➢ **Standardized tests**	Reports about unified/standardized tests in the LC; standardized achievement tests intended to measure learners' mastery of the competencies prescribed for specific grade levels (Brown & Abeywickrama, 2010)	*Basically, the exam is like SAT where the students have papers, like where there is a bubble, and they have to fill it in. We take the exam, then we submit it to a committee. The committee sends us an email, telling us what each student got. So I don't correct the exam for the foundation 099. (Dana, II)*
b) Grading	Teachers' reports about the assessment rubrics required by the PAAET or the LC	*For the ESP courses, even their participation you can grade them for participation, cause you do have around 20 marks that you can divide into a participation, quizzes or any other kinds of assessment such as homework, and sometimes I try as much as possible to include their class participation. (Nadia, II)*

Formative Assessment Practices	Any assessment methods or tools used "during instruction in order to give teachers and students a clear idea of how students' performance levels compare with the learning target . . . and how they might close the gap between their current level of understanding and the target" (Brookhart et al., 2010, p. 41)	
• **Methods of Formative Assessment**		
a) Short questions	Any short, verbal questions the teacher asks in the classroom during instruction that could serve monitoring student understanding and assess the teacher's own teaching pedagogy	*One of the examples is concept checking questions. This is when I'm teaching a specific point, and then I ask them again about the point to check whether they've understood it or not (Mona, II).*
b) Group tasks	Assessment tasks conducted by a group of students in the classroom	*I've divided the class into two groups. Group one did the first half of the questions, group two did the second half of the questions and then later on we can check all the answers together (Mona, PI).*
c) Peer assessment	Classroom assessment conducted by peers to assess the level, quality or successfulness of learning outcomes (Topping, Smith, Swanson, & Elliot, 2000)	*Following peer assessment, we check out the answers together, but if you're asking about whether I mark it, like in a quiz and then they mark each others' grades? No honestly (Mona, II).*
d) Self-assessment	The ability to assess a student's own goals without the presence of an assessor (Brown & Abeywickrama, 2010)	*I really think the worksheet allows them to assess themselves to know exactly how much they've acquired (Najla, II).*
e) Worksheets	Any type of assessment conducted in the classroom that assesses students' performance using a paper-based tool, including worksheets or textbooks	*We're doing lot of worksheet in the class. Still, every time I see where the weak points are, the next class I give them more worksheets, we work on them (Latifa, II).*
• **Processes of Formative Assessment**		
a) Monitoring	Any action a teacher takes to check on students' performance in the classroom	*Of course, I can see from her facial expression if she is attentive with me or not. I sometimes call those names of students who are not attentive (Hajar, II).*
b) Planning	Any written or verbal outline for formative assessment in the classroom	*I don't plan my lesson (Hajar, PI).*
c) Scaffolding	"Supports that teachers provide to the learner during problem-solving—in the form of reminders, hints, and encouragement—to	*For those who didn't understand, if a student did a mistake, I would stop her, saying: "Wait, tell me, why did you answer this way?" When she answers, I ask more why*

	ensure successful completion of a task" (Shepard, 2005, p.1)	*questions. And I give her hints, and go with her step by step, till I let her reach the right answer. When I go step by step, she discovers her own mistake. (Leila, PI)*
d) Feedback and follow-up on assessment	Reports to students of written or verbal "information about how successfully something has been or is being done" (Sadler, 1989, p. 120), and reports about the way teachers follow up with students to assess whether they understood lessons	*Usually, I ignore completely the mistakes that the student has on the spot, but … I correct it later on. I try my best not to frustrate them (Hajar, II).*
e) Classroom management	Teachers' reports about overseeing and controlling certain assessment activities in the classroom Any assessment methods or tools used "during instruction in order to give teachers and students a clear idea of how students' performance levels compare with the learning target . . . and how they might close the gap between their current level of understanding and the target" (Brookhart et al., 2010, p. 41)	*Even if it is not an assessment, they need to hear that this assignment is an assessment, because they do not work enough if they knew that it is not graded (Latifa, PI).*
• **Assessment practices between GE and ESP courses**	Teachers' thoughts about how practices would differ between GE and ESP courses	*When I assess GE courses, I don't have to focus on specific language skill. Any language use could be helpful, so yes, I believe students also could do much better in GE because they're not required to use specific language. (Nadia, II)*
• **Teachers' roles in classroom assessment**	Teachers' perceptions about their role in assessment	*[Teachers' role] is 100% central for two reasons: number one, the students—most of them are not educationally mature enough for [classroom assessment], unfortunately. Number two, they're not prepared for this unless we have a more prolonged period to work with them. (Hajar, II)*
Response to the LC policy	Teachers' reports about the way they respond to the assessment policy set by the LC	*Like for example 099 it's unified, 161 is unified. 261 is optional, if you want to make it unified or not. I always refuse to make it unified. What are the consequences? Good question, actually I have no idea (Dana, II).*
3. Focus of assessment	The language aspect that classroom assessment focuses on (e.g., reading, writing, grammar)	
a) **Grammar**		*Another kind of assessment is I always refer to as open-book exam, especially for grammar because I want to know whether the student*

|			
			has comprehended the grammatical rule (Hajar, II).
b) Listening			*When it comes to formal assessment, there are skills that we don't assess students in, but we should. For example, listening and writing, we don't assess them (Mona, II).*
c) Reading			*We all have to read and answer together as a whole, and we fill in the blanks together (Leila, PI).*
d) Vocabulary			*I think vocabulary would be assessed equal to other skills, but if we're talking about an ESP, then I think vocabulary especially technical vocabulary would have higher grades (Mona, II).*
e) Writing			*they have to write an essay, we've done three marked essays that I have to give back to them out of %5 (Najla, II).*
4. External factors affecting teachers' assessment beliefs and practices	Teachers' reports of any factors that influence their assessment beliefs and practices		
a) Context	Any reports related to elements in the context that teachers believe influence their practices, including country, institution, physical class setting, and culture		
• **Cultural context**	The contextual factors related to the culture of students and/ or teachers		*Many say we understand you, but we don't know how to respond. So when I talk in English, they don't respond, what comes to my mind is that they don't understand. So here I switch to Arabic (Latifa, II)* *In our culture, in our learning context, we're very much obsessed with paper-based testing, and I think we should introduce some sort of informal testing. (Mona, II)*
• **Factors related to the students' L2 proficiency level**	The students' English proficiency level		*The challenges unfortunately, most of the students that I taught do not speak English fluently—not fluently I mean they don't practice the language. So when it comes to reading, they have a hard time (Dana, II).*
• **Psychological factors**	Students' psychological factors, which could include, but are not limited to, stress and embarrassment		*All forms of assessment were formal, so they were paper-based. I think they placed a lot of stress on students (Mona, II).*
• **Time constraints**	The time allotted to the class, the course, or the assessment task itself		*So I stopped giving worksheet. And because of the time limits I don't give (Leila, II).*

• **Political context**	Factors related to the educational assessment system in the country	*The last generation was different, even the teaching process at Kuwait educational government schools were different but usually the old type of assessment was also related to evaluating students according to the outcome of the test (Nadia, II).*
b) Assessment policies	The assessment policy set by the institution or the LC	*I love the fact that we have a lot of flexibility compared to other universities. But assessment, I don't like the exams, how they're all multiple choice (Nadia, II).*
5. Internal factors affecting teachers' assessment beliefs and practices	Factors related to the teachers' educational and teaching background	
a) Experience		
• **Teaching experience**	The participant's years of teaching in the same institution or elsewhere	*At the beginning when I started teaching in this college, I felt that I have to follow whatever is given to me, or I'm forced to do certain types of assessments (Nadia, II).*
• **Teachers' schooling**	The participant's educational background and how it influences the participant's assessment beliefs and practices	*Yes, I was very frustrated with teachers who gave us very challenging exams while they did not work hard with us, or a teacher who would give us an exam that needs four hours while the whole exam is two hours. So I still have this in the back of my mind. So I make sure that the students have enough time, this is number one. Number two, that what comes in the exam I make sure that I have really explained it thoroughly. (Hajar, II)*
• **PD programs**	The way assessment professional development programs could influence teachers' assessment beliefs and practices	*Maybe the new teachers would benefit from [PD programs], but for me, I know everything, I won't learn anything new (Leila, II).*

Coding the data using NVivo helped me with case analyses as well. After coding the data for one participant, I conducted a qualitative analysis of the participant's data. During the coding process, I wrote down analysis comments or paragraphs to include later in the report. For example, when I coded the demographic data of a participant, I started writing the first section of the case report (the participant's background) as described below.

4.4.4 Cases Analyses

After I completed all the transcription and coding, I analyzed all sources of data (interviews, document analysis, and observations) from each participant qualitatively. As Patton (2002) pointed out, "case analysis involves organizing the data by specific cases for in-depth study and comparison" (p. 447). Analyzing the data for each case separately was the preliminary stage for comparing cases. In an inductive approach, a researcher can begin by constructing individual cases in order to focus and fully understand each, before such cases are grouped thematically (Patton, 2002). Analyzing each case separately helps search for themes and patterns that interconnect individual experiences (Patton, 2002).

The first step of case analysis was writing a 30–50-page report for each participant that included a rigorous analysis of all data for the participant. All cases reports followed the same structure as shown in Figure 4.1 below.
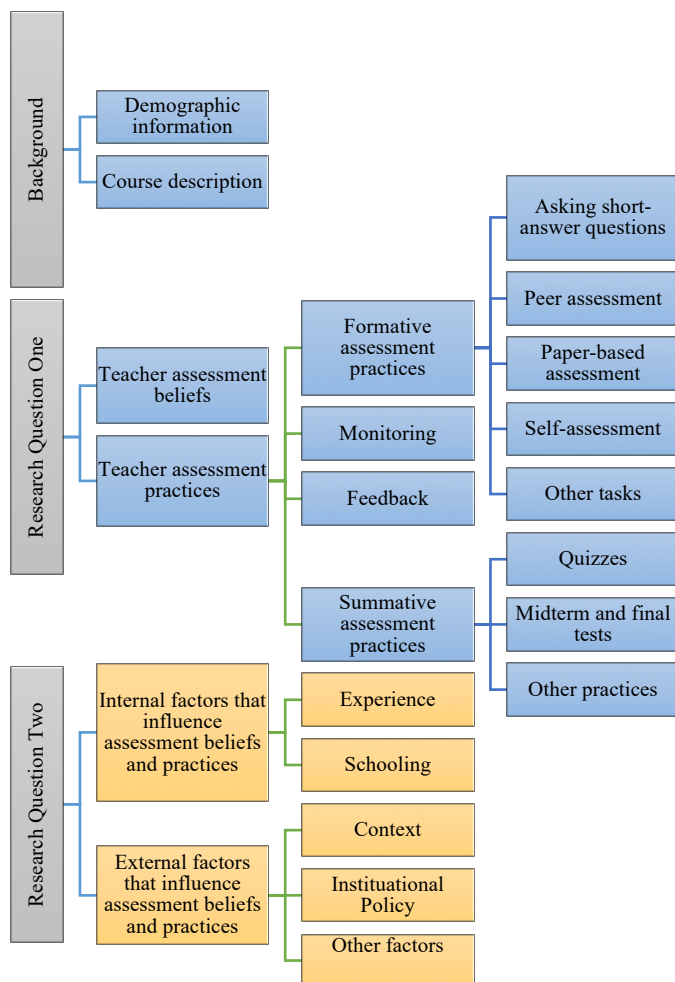
Figure 4.1. Case analysis report organization

      The first section presented the participant's background, which included her pseudonym, age, gender, college, teaching and educational experience, and description of the course selected for this study. The second section discussed Research Question One, which concerns teacher assessment beliefs and practices teaching ESP and GE courses, so it was divided into several themes and subthemes, as shown in Figure 4.1. The informal assessment practices did not apply to all participants, but the practices in Figure 4.1 were examples of the common assessment practices that reoccurred in classroom observations. The examples in Figure 4.1 similarly show the practices that were common across the cases. The third section concerned Research Question

Two. Themes presented in Figure 4.1 were those that were common across cases, but there were other themes than those.

4.5 Validity

The main validity method used in this research is triangulation. Patton (2002) discussed four types of triangulation: methods triangulation, source triangulation, analyst triangulation, and theory or perspective triangulation. Of these types, the triangulation of data sources fits this research best if the researcher uses more than one means of data collection (in this study, observations, initial interviews, and post-observation interviews) (Gliner, 1994). According to Patton (2002), in triangulation of data sources, the researcher must do the following:

- Compare observations with interviews.

- Check for consistency in what people say about the same thing over time.

- Check interviews against other written documents and interview reports.

- Compare perspectives of various individuals who hold different points of view.

Patton (2002) pointed out that finding inconsistencies during the triangulation of data sources does not always mean the data are invalid; it is the duty of an analyst to discover the reasons for these differences. In this study, the initial interviews sought teacher-reported beliefs and practices regarding assessment. Classroom observations focused on teachers' practices in specific contexts, and I followed up with questions about their practices and beliefs in post-observation interviews.

To assess consistencies in the results and ensure a thorough analysis, I used *peer debriefing* (Lincoln & Guba, as cited in Gliner, 1994). Peer debriefing refers to requesting help (e.g., checking analysis and interpretation) from a researcher outside the research project. I conducted this mainly with my supervisor, who commented on several drafts of the data

analytical procedures, and provided thorough feedback on data analysis reports. Beside peer

debriefing, the supervisory committee examined and approved the data collection methods,

which were also based on the literature.

CHAPTER 5:

FINDINGS

This chapter presents findings from the data analysis process. The results were divided into various themes in relation to the research questions. The first research question pursued two main themes: (a) teacher's beliefs and (b) their practices around assessment. For the purpose of data analysis, the two themes of the first research question (i.e., teachers' beliefs and teachers' practices) are divided into multiple subthemes. Through Research Question One, I also planned to compare teachers' assessment beliefs and practices based on course type (i.e., GE vs. ESP courses). Generally, teachers' beliefs were similar across the two types of courses, and the majority of participants did not draw a distinction between their beliefs in ESP and in GE courses. For teachers' practices, I mention at the beginning of the section whether the practices identified in the study were different or similar across course types. I also compared the roles of external factors across course types. Figures 5.1 and 5.2 present the subthemes of each section.

The second research question asks for the external factors that the participants reported had influenced their assessment beliefs and practices. The main external factors were those related to contextual factors, though some educational backgrounds of the participants (e.g., qualification, professional development certification) were a factor as well. Figure 5.3 presents the various themes under Research Question Two.

5.1 Teacher Beliefs about Assessment

The first question is about the various beliefs the participants held around language assessment. It describes the participants' general conceptions about assessment as well as beliefs about assessment practices in relation to the contexts in which they teach. This theme is further

divided into subthemes presented in Figure 5.1: teacher understanding of the concept of assessment, assessment methods and purposes, and beliefs about student learning.
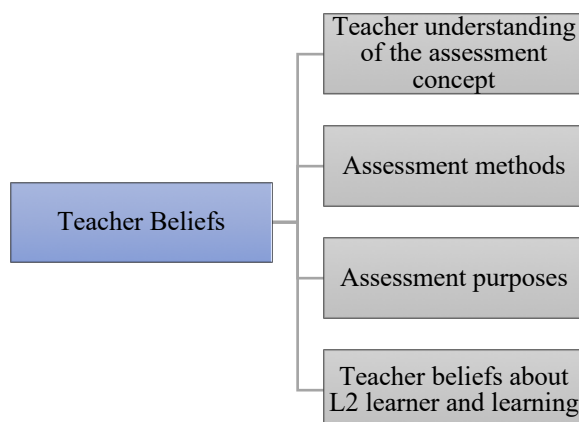
Figure 5.1. Themes of teachers' beliefs

5.1.1 Teachers' Understandings of Assessment

The participants answered two main questions about how they defined assessment. They were asked to define assessment in general and then to define classroom assessment. When the questions were designed, there was a third question about their definition of language assessment, but the participants did not differentiate between language assessment and classroom assessment, and their responses were generally about their courses. Therefore, questions about the description of language assessment were analyzed in terms of another theme, focusing on the language aspects the teachers assessed in their courses.

Although their understandings of assessment varied because of different factors (discussed under the external factor section), the main challenge that surfaced when analyzing the participants' responses was their use of precise terminology to describe different purposes and methods of assessment. Most participants did not use a specific term when describing assessment; they first drew on the first interview question, which directly stated, "What comes to your mind when you hear the word assessment?" This question led to their answering based on

their knowledge of the specific word "assessment." Although their definitions of assessment varied, the prevailing understanding of assessment was associated with paper-based graded tasks. However, they revised their conceptions as the interview continued and gave more detailed answers in relation to their practices. Still, the first time they were asked to define assessment, their answers related directly to graded assignments or tests. Below are examples of the participants' answers about their definition of assessment.

When first asked about assessment, Dana, Nadia and Leila paused for a moment and then expressed a question ("Assessment?"). Nadia's response for example, was: "[Pause] Assessment? Exams, and homework, and assignments, all kinds of assignments" (Nadia, Initial Interview; II). They perceived assessment as a paper-based practice and referred to the assessment methods when first defining assessment regardless of its purposes. Leila, however, equated the term *assessment* with *test* perhaps because in this context, the terminologies *tests* or *exams* were more ubiquitous than *assessment*, when translated into Arabic. But Nadia and Dana included any paper-based tasks as assessment.

Latifa seemed to define assessment according to the prevailing conception in her context. She stated,

> Excerpt 1: What comes to mind? Let me think. This is what we learned: assessment is just like scores; this is what we were raised up and we were taught. Even here, the girls, they think assessment is getting scores. They don't see where the weakness points are, so they can do better; they just want grades to graduate. So it's only numbers. (Latifa, II)

Her statement reveals the influence of her experience on her perceptions; she thought of assessments as graded assignments only. She was hesitant to label segments from the recorded classroom observation as assessments; she considered those to be part of her teaching strategies conducted to check students' understanding. Her beliefs about the purpose of assessment seem to reflect the influence of the context in which she teaches. She asserted her beliefs about the

purpose of assessment after the above statement, saying, "No. I believe assessment would be seeing where your weak points are, so you can do better in that part" (Latifa, II). Again, instead of defining assessment, she explained assessment's main purpose.

Mona expressed a different position when defining assessment as paper-based. In her definition of assessment, she stated, "I think [it's] a paper and a pen [laughs]" (Mona, II). Her expression (i.e., laughter) seems to indicate sarcasm and suggests that she either was not convinced of her definition of assessment or was influenced by the common conception of assessment prevalent among administrators and students in the context in which she teaches (which, as she noted, seems to consider paper-based assessments to be the sole means of evaluating students). It is possible that her statement is a result of the common conception of assessment, because throughout the interviews she expressed a negative attitude toward such assessment beliefs. She stated, for example, "I think that in our environment we are very much obsessed with formal assessment as opposed to informal assessment." Mona also demonstrated her understanding of different methods of assessment even though she did not directly explain the purpose of assessment in the interview; she described assessment as *formal* and *informal assessment*. She defined formal assessment as paper-based tests and informal as any assessment conducted in the classroom.

Other participants (e.g., Dana and Najla) acknowledged that they had not studied assessment and revealed uncertainty when defining assessment. Najla for example, defined assessment as "marks," but she doubted that this was how she *should* define assessment: "I see [assessment] as marks, and unfortunately, I don't know if this is good or bad" (Najla, II). The statement could possibly imply that she was not convinced about the meaning of assessment

because she repeated several times throughout the interview that she had not studied assessment or teacher education.

Hajar defined assessment differently. This was her immediate answer when she defined assessment:

> Excerpt 2: When I hear the word *assessment*, the first thing that comes up to my mind is whether my students are liable [are able] to pass or not. Have they comprehended anything from this course? Have they benefited anything from this course? Assessment is not a paper; assessment is a continuous procedure. It's an every-minute procedure with your students: looking at their [facial expressions], their reflection—reflection is an assessment. It's not only a paper. (Hajar, II)

She perceived assessment as a continuous evaluation process of students' learning and performance. She considered students to be responsible for passing the course, but she saw the teacher as needing to cooperate with students to achieve this goal. When asked about her perceptions of assessment in general and classroom assessment in particular, she said she regarded both as a continuous process that is not specifically related to paper-based tests.

5.1.2 Assessment as Method

The majority of the participants defined assessment as a method consisting of paper-based tests and tend to make a distinction between formal and informal methods. Some participants, such as Latifa and Najla, equated assessment with paper-based assessment methods. Even though they used different types of assessment methods in their classrooms (e.g., asking short-answer questions, providing feedback, and completing tasks along with students), most participants viewed such activities as teaching practices not assessment. Latifa and Najla equated assessment with paper-based methods and reported that the only methods to detect student understanding were paper-based (e.g., tests and assignments). However, they tended to report assessment beliefs that seemed unclear and contradictory when they pointed out that they depended on tests as a sole means of assessment; they were assessing students throughout the class using several methods (e.g., short-answer questions and completing the worksheets). They

seem to think that assessment includes both paper-based and oral methods, but only graded

assignments are truly assessment. Najla said,

> Excerpt 3: [Assessment is] more about what students can do, and how much of it is in their head. [It is] for students to be able to tell their level and what they haven't grasped to work on their weak areas. That is why I give worksheets. I want students to physically see where their weaknesses are. (Najla, II)

Thus, she defined assessment's purpose but indicated her belief that the sole methods to

achieve this purpose were through worksheets. Her summative assessment included not only

paper-based tests but presentations and assignments as well. She did not categorize them as

assessments. She further added:

> Excerpt 4: I see [assessment] as quiz, midterm and final exam. That's what I really assess students for, and also their research or presentation—that's where the marking goes. But my understanding of assessment is that you are able to track yourself on: what you know, how much you know, how much you can work on. That's for me is assessment, and that's why I say conscious learners with these worksheets. (Najla, II)

All the methods of assessment she mentioned here are paper-based. She seems to think that the

teacher role is to mark student performance, while the student's role in assessment is to monitor

their own learning progress (i.e., self-assessment).

Other participants, mainly Leila and Mona, identified the methods of assessment as

formal and *informal* without explicitly using these terms. That is, they did not identify formal

and informal assessment as methods; they simply dichotomized the concept when trying to

clarify their understanding of it. From their description, it seemed that they were referring to the

methods of assessment rather than its purpose. For these participants, all formal assessment are

primarily paper-based tests. All other paper-based assignments that were done in the classroom

were considered informal assessments.

Leila and Mona defined assessment in terms of its method's formality. They

distinguished between informal and formal methods of assessment. They equated formal

assessment with paper-based tests, but described informal assessments as tasks done in the

classroom, including paper-based (e.g., book exercises). For example, Mona's definition of

formal assessment primarily included the method, but she also referred to its purposes. She

reported that only formal methods of assessment (e.g., quizzes, midterm, and final tests) are to be

relied upon for final judgment of students' performance and are the only ones to be reported for

accountability purposes. Any assessment conducted orally, according to her, is informal, as were

the worksheets that she frequently distributed to students during the sessions I observed. When I

asked her about what she meant by formal and informal assessment, she said this about informal

assessment;

> Excerpt 5: Teachers check students' understanding of the lesson without necessarily telling the students that I am assessing you. The reason behind it is just to check their understanding. It's not necessarily graded or marked, and this is one of the assessments that you can use in the lesson so that later on you can check their understanding and then basically build upon it. (Mona, II)

It seems here that Mona is identifying some of the characteristics of informal formative

assessment. For example, for her, informal formative assessment is conducted to track students'

understanding and is not necessarily graded. Formal assessment, according to Mona, is

> Excerpt 6: Any type of formal assessment that you do inside the classroom. For example, it could be a listening test, written, and also could be a formal paper-based test. Or you can also do, like, a formal speaking test, it is also included within the formal area. (Mona, II)

Although Mona thought that paper-based assessment is formal, she stated that informal

assessment could be done in the classroom (e.g., answering worksheets from the reading kit) and

that the purpose of such tasks is promoting students' learning without assessing them formally.

She thought that teachers should integrate both methods of assessment (i.e., oral and written) into

the classroom and that they should not depend on only one method:

> Excerpt 7: I don't believe that students should be assessed 100% formally. I do believe that part of it should be done informally and graded as well, so that you would check their understanding, because this is language use and sometimes there are stress factors. (Mona, II)

It was not always clear whether she considered informal assessment to be a method (e.g., oral

tasks, role play) or a purpose (i.e., formative assessment). But it was clear that she disagreed with

the common conception in her context that informal assessment should not be graded.

5.1.3 Assessment as Purpose

The participants defined assessment in terms of various purposes as well. The majority reported that assessment is necessary for both the teacher and the learner. While it is important for learners to track their level of learning and pinpoint their weaknesses, it is also necessary for the teacher to assess her or his pedagogical methods based on student performance on different assessment tasks. Because none of the participants had taken any courses in assessment, they did not use the terms formative and summative. But their descriptions of the purposes of the assessments they reported using suggest that they designed their assessments to either promote student achievement (i.e., formative) or to evaluate and grade students' performance (i.e., summative).

The majority of the participants considered classroom assessment necessary for tracking student understanding and evaluating their own pedagogy. However, the participants did not report the purpose of formative assessment explicitly; throughout the interview some participants expressed their beliefs using the concept of "teaching" rather than "assessment," possibly because assessment and teaching do overlap in formative assessment.

A number of teachers stated that using assessment is significant for the purpose of evaluating the effectiveness of teaching. Hajar and Dana explained that the importance of classroom assessment is to inform instruction. Hajar for example, identified the purposes of assessment as guiding teacher decision-making based on student understanding:

> Excerpt 8: When you drop a new piece of information and you can see those puzzled faces who do not understand what you're saying, this is my assessment of my work. When I say *assessment*, I put myself before the students because if I give any kind of assessment, whether oral or written, and I find that most of the students haven't done well, that's a very bad score of my [teaching]. (Hajar, II)

Hajar identified the purpose of the assessment as being formative as well without labelling it as such, since her goal seemed to be to assess students' understanding without necessarily assigning

them an assessment task and seeking feedback by looking at students' facial expressions, a very common type of formative assessment during teaching. She further mentioned that she attributed students' not understanding to her teaching strategy. When she explained her practices later in the interviews, however, she outlined other purposes of assessment that included promoting students' learning.

Some other participants, including Mona and Nadia, revealed an overlapping relationship between assessment used to check student understanding and assessment to assess teaching methods. For example, referring to Excerpts 6 and 7 in the previous section, Mona's use of the terms *formal* and *informal* did not indicate whether she was referring to assessment as method or purpose. She implied that informal assessment helped her decide on the level of the students' understanding and her teaching efficiency. Mona believed that assessing students' understanding should inform the teacher's assessment of her own practices: "The most important thing is that you do it to check if the students have understood something, and based on it, I modify my teaching." She additionally described formal assessment as necessary to the learning process, because it enables teachers to judge student performance based on evidence:

> Excerpt 9: In an ESL/EFL context, for classroom assessment, and I mean informal assessment here, you need to check their understanding. You also need to test their knowledge in a non-stressed environment. And then for the formal assessment, you need to do [such types of assessment], so you would actually hold an evidence of their language progress. (Mona, II)

In general, she indicated that the purpose of both formal and informal assessment is to enhance students' learning, and she criticized the practice of relying predominantly on paper-based tests.

Another purpose of assessment was identified by almost all participants, which is that assessment is mainly conducted to assess student understanding. They either presented this purpose within their general definition of assessment or when we discussed some segments related to formative assessment tasks. Najla, for example, reported that the purpose of

assessment is to consolidate students' learning. Her main goal was to enable students to conduct

self-assessment by providing them with worksheets on the subject matter being taught. She

reported that such worksheets should enable students to track their learning progress and identify

areas in which they needed to improve:

> Excerpt 10: [Assessment is] to check how much [students] managed to grasp from the listening or reading for example, or whether they are able to edit their writing from the grammar taught in class. These assessment exercises are to help them see where their weaknesses are, and which areas I need to work on by providing more practice. Also, it will help students to estimate how well they will do in my class. What I'm trying to say is, I encourage them to be conscious learners through these practice sheets that allow me and the students to have a general understanding of the strength and weakness of individuals (and class in total). (Najla, II)

However, by providing students with worksheets in the classroom, she created a situation

that emulates a testing environment. She did not consider any practice other than the worksheet

to be assessment.

Some participants reported another purpose of classroom assessment. They said they

conducted assessment in the classroom merely for the purpose of convincing students to get

engaged in some activities, which suggests the ultimate purpose of some assessments was

classroom management. They complained that their students were passive, and hoped to change

this situation by assigning assessment tasks that are not graded, aiming only to engage students

in the lesson. Dana, for example, said that she sometimes marks participation to make students

"active" in her classes:

> Excerpt 11: If I know that, in some activities, the students are participating, I do take assessments [grade their participation], but my main intention is not for assessments; the main goal is to make them active, as well as do they understand it? And I want them to participate, and, if they don't understand, I try to get them active in the class. So that's the main purpose, and then if they're doing great, that's when I can assess them. (Dana, II)

Dana reported that students make more effort when engaging with graded, paper-based

assessments. Like many teachers at the LC, Dana believed that because English is a mandatory

subject, many students are more concerned about passing the course with a good grade than they

are with acquiring the language. Accordingly, she sometimes informs students that their participation will be marked in order to increase their interaction in the classroom.

Nadia also reported that she uses assessment, specifically short-answer questions, to engage students in the class discussion, but unlike Dana, she said that students' participation is typically ungraded, but it is conducted for class management purposes (e.g., to ensure students' participation in assessment activities). For example, because she had many students enrolled in her class, she asked for participation from everyone to get attention, "Sometimes, I want them to participate just to get their attention, and sometimes I do mention that their participation is [graded]" (Nadia, II). Nadia seems to believe that student participation is more about classroom management than about checking students' understanding. This point can be supported by a practice that I observed: she asked students to participate and placed a mark next to their names in her records when they did. When I asked Nadia about this specific practice, she replied that she sometimes does this to encourage students to participate. Thus, it appears that students' silence reduced their engagement in assessments.

The majority of the participants showed an understanding of assessment as summative perhaps because such purpose carried more weight in their context. The participants indicated that the primary purpose of assessment is to evaluate student performance based on graded tests, such as quizzes, midterms, and final tests.

Dana (in Excerpt 11) and Latifa (in Excerpt 1), for example, believed that assessment is paper-based graded assignments to evaluate the levels of student understanding. Although Dana referred to some classroom informal assessment tasks to check student understanding, she perceived those tasks (e.g., asking short questions) to be part of her teaching practices not assessment, as did other participants. It appears that she did not consider assessment as a

continuous procedure that coexists with teaching, but a procedure performed in the classroom for the purpose of reporting grades, because when asked about how often she would assess students in the classroom, she asserted that she did not assess them on a daily basis: "I do assess them, but not on a daily basis, not on every single activity." She also seemed to view assessment as solely grading because she added, "I [assess] on a weekly basis or overall." Such a frequency of assessment in the classroom contradicts the concept of formative assessment.

5.1.4 Teacher Beliefs about L2 Learners

Data analysis shows that beliefs about classroom assessment reflected the characteristics of the context of the study, and most participants reported beliefs about assessment, relating it to the L2 learners and learning in this institution.

The majority of participants believed that, ideally, students should take part in classroom assessment through self- and peer assessment. Yet they said that students are at very low English language proficiency levels, which hinders teacher–student or student–student interactions and made the teacher role central in assessment. Based on their reports, it was not possible to implement various methods, such as peer and self-assessment, in classrooms.

A number of participants, including Dana and Hajar, indicated that students are passive learners because of their low English language proficiency levels, and that they have to simplify assessments to make it easier for them to understand. Hajar for example criticized the central role of the teachers, saying that it was students who forced this role. She provided several reasons behind learners being passive, including lack of motivation and low proficiency levels. Although she criticized both teachers and students in this situation, she created many classroom assessment tasks that were graded and she called them "bonus assessments"; with these, she believed

students would be motivated to participate, but it appears that she was motivating them to earn grades rather than to create student-to-student interaction:

> Excerpt 12: [Teachers' role] is 100% central for two reasons: number one, the students—most of them are not educationally mature enough for [classroom assessment], unfortunately. Number two, they're not prepared for this unless we have a more prolonged period to work with them. The semester is very short, the students come to you, they don't have anything in their mind except to pass the course, some of them only want to pass, and you know that Shaima. And some of them only want good grades in whatever method, whether attending whether giving you whatever. (Hajar, II)

Dana tried to change the atmosphere of the class, feeling dissatisfied with maintaining the dominant role of the teacher in the classroom. She attributed this situation to the learners, saying that they were "lazy" and "passive." Although she reported in the initial interview her desire to change the situation, she changed her thoughts in the post-observation interview, saying it was not possible to create student-to-student tasks, but she presented other factors that could be a reason for students' being passive: the number of students and the room sizes, which is presented later in this chapter.

Other participants reported that the English proficiency level of the students was related to the type of course and the study program. Najla, for example, reported that students who were enrolled in the diploma programs were at lower levels than those enrolled in bachelor programs. Hajar believed that ESP students have higher English proficiency level than students enrolled in GE courses. Nevertheless, she thought that teachers should not act upon such a belief for every ESP course they teach because, in this summer course, she taught a class that she taught before, and students were at an unexpectedly low level. On this matter, Hajar commented,

> Excerpt 13: I have taught this course before [. . .] It's an English oriented major. It was my mistake because I should not have gone to the class with prior [expectation] that they [the students] are good. So, when I went to the class at the first week, I was frustrated, I thought that they were as good as the one before, which was my mistake again not to compare one class to the other. (Hajar, II)

In the above statement, Hajar appeared to change her beliefs; earlier in the interview, she said that ESP students are at higher levels than GE students. She reported that the ESP course

itself differs. For example, she believed that students taking ESP for technology and math are much higher in level than students in other courses.

5.2 Teacher Assessment Practices

This section explores the second part of the first research question, which investigates teachers' assessment practices. As Rea-Dickins (2001) noted, conducting assessment includes a number of tasks that teachers implement to achieve their assessment goals: identifying the purpose of assessment, choosing assessment methods, and preparing learners for assessment. Thus, this section discusses assessment methods, purposes, uses, and plans of their assessments. The following discussion is divided into two basic sections: summative assessments and formative assessments; each section contains subthemes presented in Figure 5.2. The discussion below presents teachers' practices mainly based on classroom observation, but also document analyses and teacher reports in the interviews. In the initial interview, participants were asked to describe their practices in courses selected for this study. Some participants described assessment they used in general in their classes without distinguishing between ESP and GE courses. Following the initial interview, two classes were selected for each participant to observe their assessment practices. Teachers were then interviewed about these assessment practices.
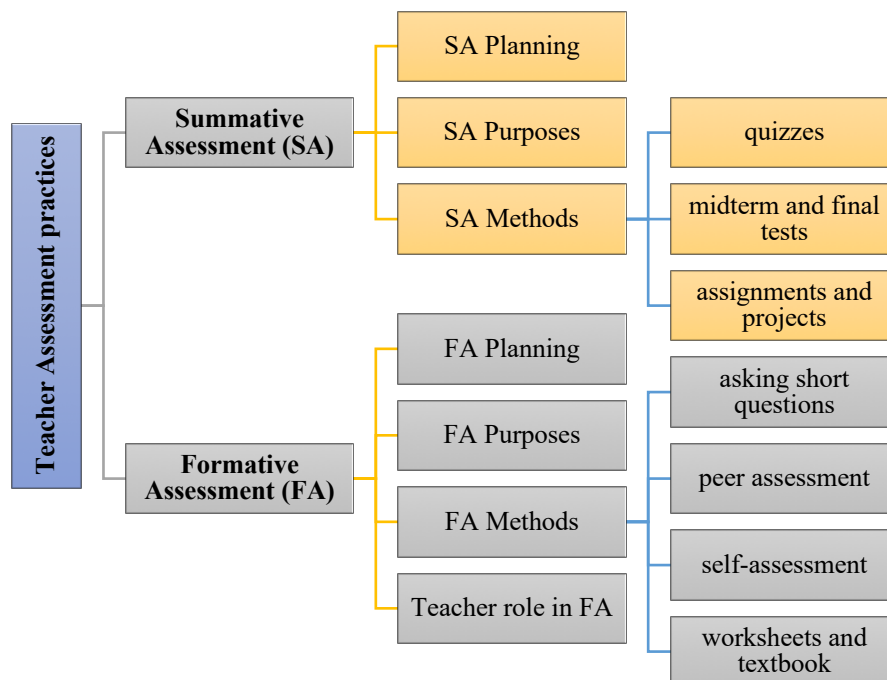
Figure 5.2. Teacher practices main themes

5.2.1 Summative Assessment Practices

As discussed in section 5.1, the majority of the participants implied they saw assessments as paper-based graded assignments. Summative assessment in this study involves tests and paper-based assignments and projects. Unlike formative assessments, which will be discussed below, summative assessment was more structured and constructed based on specific format, rubrics, and criteria, because summative assessment carried higher weight than formative assessment in this context, as the policy mandates. This section presents the participants' summative assessment practices in terms of planning, purposes, and methods.

5.2.1.1 Planning summative assessment.

Unlike formative assessment, summative assessments were structured and planned. Not all participants provided me with course syllabi that show their assessment plans, but for those who did, the planning involved creating specific rubrics and criteria, scheduling the time of the tests, defining the topics to be included, determining grade distribution, and constructing the test

format. The syllabi did not include the dates or the materials that would be included on the tests (see Appendix I). Some participants (e.g., Leila and Nadia) did not have a specific plan but reported the rubrics and their plans of constructing tests in the interviews, such as the mark distribution. I obtained the assessment plan documents either from the participants or the LC website. Teachers did not have a choice to plan the rubrics, but they had freedom to plan the test format. However, a number of participants, especially those in CBS and CBE, explained that they have to abide by the standardized plans for the midterm and final tests for the GE courses.

The plans mainly covered the dates and rubrics of the tests. All participants followed similar plans for the assessment, which were set by either LC or PAAET policy; they included a midterm test, a final test, and quizzes. The number of quizzes and percentages of other tests varied. The tests formed a larger percentage of the assessment rubrics than the percentage of formative assessment. The course syllabus presented the assessment percentage next to each assessment type such that 100% represented the highest achievable grade. Grade weights for various kinds of assignments are presented in Table 5.1, which presents only summative assessments and percentage. (The formative assessment methods which accounted for 10% are presented and discussed in the following section.)

Table 5.1

Participants' Summative Assessment Grading

Participant (Course)	Quizzes	Midterm	Final	Assignments/ Homework	Project	Other
Najla (ESP)	10%	25%	45%	5%	10%	---------
Mona (ESP)	15%	30%	50%	---------	---------	---------
Hajar (GE)	10%	30%	50%	---------	---------	Bonus tasks (did not assign certain percentage)
Hajar (ESP)	10%	30%	50%	---------	---------	Bonus tasks (did not assign certain percentage)
Dana (ESP)	10%	30%	50%	5%	---------	---------
Leila (GE)	10%	30%	50%	---------	---------	---------
Leila (ESP)	10%	30%	50%	---------	---------	---------
Nadia (GE)	10%	30%	50%	---------	---------	---------
Latifa (ESP)	10%	25%	50%	5%	---------	---------

As Table 5.1 shows, the majority of the participants followed the assessment plans set by the LC policy which requires assigning 50% for the final tests. Najla, Dana, and Latifa added written assignment or homework. The policy did not require teachers to include written assignments, but it seems that there is flexibility in the choice of skills involved in the assessment. There is, however, a low percentage left for the other formative assignments. The 10% that the participants said they set for classroom assessment seemed not to be a required percentage. Mona, for example, used the remaining 5% for attendance only, and therefore did not evaluate any formative assessment. Najla seemed to take 5% off the midterm and another 5% from the final test to add it to the final project, which suggests that she preferred to include more summative assessment tasks than formative for the overall student evaluation.

Concerning the column entitled "Other," Hajar was the only one who planned to include written assessments, to which she referred as "bonus questions." This task did not count for the overall final grade of 100% but served as extra credit for those students who wished to add extra marks to their overall grade in case they did not perform well on tests. The type of assessment she used for this category was to ask students to answer questions from the textbook on the whiteboard.

Planning summative assessment included selecting the language aspects to be included on the tests. All participants indicated that they did not have an active role in selecting the units and the language skills to be covered on the curriculum and the tests. Each course has a coordinator who is responsible for distributing a syllabus at the beginning of the semester that includes the topics to be taught in the course. Latifa noted that she planned to add some grammar topics that were not included because she found that some students did not perform well on it. For example, she added extra materials on the topic of using WH-questions because students did not perform well on the midterm. As for the language focus planned for each course, the participants were required to include all language aspects: grammar, reading, vocabulary, and language function (i.e., communicative daily language). They all agreed that listening and speaking were not required, and they accordingly did not include them.

5.2.1.2 The purposes of summative assessment.

Most participants found it necessary to assess students' levels of performance based on tests. Thus, the purpose of assessment seemed to be to measure learners' understanding of the topics based on what was included on tests. However, the data show that a number of participants seem to have used assessment to help students obtain high grades. It is important to note that the teachers in this context must abide by the assessment plans of the department (i.e., a

midterm and a final tests). However, participants did not indicate that they have to give a specified number of quizzes or assign them certain percentages. Most participants chose to give students more than two quizzes and count quizzes for 10% of the overall grade.

Najla, for example, believed that students should only be evaluated based on paper-based methods, and she applied this belief by distributing the marks over the written assignments but not the formative tasks. Other teachers (e.g., Mona) did not mark classroom tasks at all. The purpose of conducting summative assessment seemed, therefore, a response to the department policy for reporting purposes and to measure students' achievement on specific paper-based tests. The teachers conducted summative assessments because they had to submit the results to the department. They did not appear to attend to monitoring students' learning progress on a continuous basis in the classroom. Although they appeared to rely on summative assessments in response to the department policy, some participants (namely Latifa) criticized relying on summative assessment solely, saying that the mark distribution did not allow her to evaluate student actual performance:

> Excerpt 14: The 50% [for the final] is a lot. I know some girls, they do very good in the midterm, and even very good in the quizzes, but in the final, like I don't know what their circumstances are. For example, you might see an A student, but in the final she does not perform well, so her grade would drop back to C. So I wouldn't prefer to have the final out of 50. It should be the teachers' own rubrics. The final is important, but the marks are too much for it. (Latifa, PI)

Another purpose of summative assessment was assessing teaching effectiveness. At the beginning of one of her classes, Najla commented on the poor performance of students on a quiz. She re-explained the topic on which they did not do well, and announced a make-up quiz. She thought that students did not perform well at the beginning because they needed more explanation, which implied that the test informed her teaching strategies.

5.2.1.3 Summative assessment methods.

This section presents the summative assessment tasks that the participants implemented

in their courses. Unlike formative assessment in which the participants did not describe their

methods clearly, especially regarding whether their methods differed between GE and ESP

courses, summative assessment methods were described in detail in the interviews based on the

course type and the syllabus set by the LC. Teacher role in assessment construction and design

also differed between GE and ESP courses. The summative assessment used by the participants

included: quizzes, a midterm, a final, and written assignments; a final paper was also used in

Najla's class. The following sections discuss the main summative assessment methods used by

the participants.

5.2.1.3.1 Quizzes.

Quizzes are short tests given after each unit or couple of units. Table 5.2 presents a

description of the quizzes used in ESP and GE courses.

Table 5.2

Summary of the Participants' Quizzes Design

participants	Type of Course	Number of Quizzes	Language focus	weight
Najla	ESP	5	Vocabulary and grammar	10%
Mona	ESP	4	Vocabulary	10%
Hajar	GE	2	Grammar and vocabulary	10%
	ESP	3	Vocabulary, grammar and reading comprehension	10%
Dana	ESP	2	Vocabulary and comprehension	10%
Leila	GE	2	Vocabulary and grammar	10%
	ESP	3	Vocabulary and grammar	10%
Latifa	ESP	3	Vocabulary, grammar, and writing	10%
Nadia	GE	2	Vocabulary	10%

Table 5.2 shows that all participants followed the same grading rubrics of assigning 10%

to their quizzes for all their courses. The number of quizzes is not fixed, and quiz design and

number may change each semester. Some participants (e.g., Mona, Nadia and Leila) gave more

than two quizzes but counted only the best two grades for each student. Several participants (e.g.,

Najla and Hajar) also reported that they would amend the content or the number of quizzes based

on the students' levels. Hajar stated that the quizzes were not always on grammar or vocabulary;

if she saw that students needed support in reading, she would assign them a reading

comprehension quiz.

> Excerpt 15: I give them quizzes, for example—especially if I need to train them for reading comprehension, because the students find it very difficult, so what I do is I give them a more challenging piece of passage to take home. They don't know that it's training, but they think it's a quiz. And I say, "Try to do it yourself, ladies, because what you are gonna do, I want to know at what level you are, so when I prepare your exam I will know what level I should give you." So most of them [try their best]. So when I give them the exam, they have been exposed to this way of examination; they are more aware of it. (Hajar, PI)

Most participants did not specify how they calculated the quizzes out of the 10%, but they indicated that each quiz accounts for 10% of the overall grade, and they would then include the students' scores on their best two or three quizzes. Najla's strategy represents other participants' in counting the best quizzes' grades. She said she gives a quiz after each unit:

> Excerpt 16: Every time we finish a unit, I would give them a quiz, and the quiz would be grammar and vocabulary and [sometimes] writing. If we're doing punctuation or if we're doing transitions, things like that, then it would include writing as well. But it is mainly on vocabulary and grammar. (Najla, II)

Najla and Mona stated that the quizzes were planned and designed before the semester started. However, they could amend plans for quizzes depending on the students' English language proficiency levels or the available time. Najla also allowed makeup quizzes if students did not do well on a previous quiz. She believed that quizzes were the main tool for assessing students' understanding on a regular basis, thus, if she felt that students needed further instruction on a topic, she would give a quiz on that topic to gauge their understanding.

Table 5.2 also shows that unlike the midterm and final tests that assessed four language skills, most quizzes focused on specific language skills: vocabulary and grammar. Moreover, the quizzes followed the closed-question format (e.g., multiple choice and matching), which were similar to the questions on the midterm and the final tests. This implies that the participants prepared students through the quizzes for the larger paper-based assessments.

As for the GE courses, at the time of the study, there was a rule set by the English department at CBS that required GE course coordinators to determine the assessment rubrics and tasks. Teachers of GE courses had to assess students by giving them multiple quizzes that made up 10% of their overall grades and was constructed by the coordinator. Yet, the course instructor was responsible for marking the quiz. The quizzes had to be given after each unit, and they had to be comprehensive and include all the language skills in the textbook unit. Leila and Nadia did not agree with this policy because they were forced to teach the lesson quickly to cover what was

included on the standardized quizzes. Before this policy existed, each course instructor was responsible for constructing their own quizzes. Leila added that she first abided by the quiz policy, but then she discovered that the quizzes were too long and time consuming. She stated that when she gave quizzes in her classes, she only assigned 10 to 15 minutes of class time to them and that having long quizzes was time consuming. To address this challenge, she divided the quizzes and had the students take one section of the quiz related to specific language skills one day and another section on different language skills on another day. She did not believe that giving long quizzes was suitable for her students' level.

As for the ESP courses, only one participant reported a different strategy of designing quizzes according to the course type. Hajar used quizzes differently in her GE and ESP courses. Her ESP quizzes were not always divided according to a specific language area or including closed-ended questions as they were for her GE course. For example, one of her quizzes assessed students on both vocabulary and grammar. The vocabulary section included two parts: matching verbs with their collocation and providing definitions. The grammar included one open-ended question that required students to write three sentences using a verb as a gerund in three different ways.

5.2.1.3.2 Midterm and final tests.

Another summative assessment method was midterm and final tests. The midterm tests were longer than the quizzes and involved topics that were covered before the week of the test. The test determined 30% of the students' overall grades. The final test was a final, paper-based assessment conducted at the end of the semester that was worth 50% of the students' overall grades. Although participants agreed that the percentage of the midterm and the final test was

worth 30% and 50% of the overall grades, respectively, some participants (i.e., Najla and Latifa) said that the midterm could be worth 25% and the final worth 45%.

Planning GE tests was not completely done by the teachers. As mentioned in the context chapter, when the LC was established, it was mandatory for all LC English departments to administer standardized midterm and final tests in GE courses. Teachers in all colleges abided by this policy and gave their students tests constructed by the exam committees at each college. Thus, each LC college department had its own test that was different from other departments in other colleges. It was referred to as the "unified test," which was administered at the same time and date, usually on the weekend (i.e., Saturdays), by all GE teachers. When the administration changed in 2013, the director of the LC asked teachers to vote whether to continue to use unified tests or not. Most teachers agreed to keep unified tests as an option within the unit and that teachers should have the option not to abide by it. Teachers at CBE, however, decided to continue with the old policy. Hajar explained that this decision could be attributed to the large number of students in their college; CBE has the largest number of students in the GE courses among all colleges.

At the time of the study, the LC director suggested teachers should abide by the rules of the GE course coordinator. The GE course coordinator sets the assessment rubrics and format of the tests. Nadia and Leila said that they followed these department-set policies in terms of grading. The midterm and the final tests were standardized across all GE courses in the CBS. Nadia and Leila stated that during the summer term, teachers had the choice to decide collectively whether to have unified midterm and final tests or to create their own tests. In the summer semester, only three teachers shared midterm and final tests. Nadia said that her colleagues and she split the task of creating the test, and so did Leila. Nevertheless, the format of

the midterm and final tests followed the answer sheets distributed by the LC for the common placement tests and the midterm and final tests. Those answer sheets were used for machine marking. The testing unit in the LC has a room in which answer sheets are stored and marked by a machine. However, there were also writing sections (e.g., writing a paragraph, answering a short comprehension questions) on the midterm and the final that had to be marked by the course teacher.

The above process is strictly followed in CBE. In CBS, the participants said that they followed a standardized test, but they could give students the test during their regular class times, and it was up to the teachers to follow the coordinator's test or construct a format similar to it. The role of teachers on the test day is to proctor their own test session, distribute answer sheets that are machine scored, provide instruction on how to answer using those answer sheets, collect the tests, and submit the answer sheets to the head of the testing committee. Any member of the testing committee is authorized to mark GE teachers' tests using machine scoring. A member of the testing committee would then email the results of the tests to the teacher.

In CBE, teachers expressed opposing beliefs regarding the final and midterm test practices. On one hand, Hajar followed and supported the unified test policy, she noted that not all teachers in CBE had their GE tests unified. She disagreed with that approach and said she hoped that ESP tests would be unified as well. On the other hand, Dana believed that in the GE courses, she does not have an active role in constructing the test. It is a standardized test for GE and consisted mostly of closed-ended questions established by the coordinators of the course. She believed that a test should include more writing or other types of open-ended questions. Dana said that she has more freedom in ESP to construct her own tests. Similar to Dana, Nadia expressed uncertainty whether the format of the GE test is suitable for students. She reported that

she does not have a voice in the format of GE tests; the coordinator of the course decided on the format and the rubrics.

Leila and Hajar demonstrated opposing thoughts to those of Dana. They reflected a preference of following the rules of standardized tests set by the LC or its units, especially for the GE courses. They said that teachers do have active roles in GE test construction; each teacher in the department submits contribution to some parts of the test (e.g., grammar, vocabulary), but they did not indicate who approves such contributions or constructs those unified tests. Other participants indicated the coordinator sets the format and rubrics of the GE tests. There is a testing committee in each unit, and Hajar and Leila were members of this committee. It appears that they have a voice in the construction of tests.

Participants teaching GE courses reported similar test construction to one another. Tables 5.3 (Leila's GE tests) and 5.4 (Hajar's GE tests) present a sample description of Leila and Hajar's GE tests. They included a main section that included five basic language subsections (reading, vocabulary, grammar, language function, and writing). Although Leila and Hajar indicated that they shared tests with other colleagues, their midterm tests differed in terms of the number of questions and scoring for each section. On the final test, the reading was similar on both tests, but the rest of the sections were different in terms of number of questions and scoring. It is unknown whether the LC had multiple answer sheet formats or if teachers can create their own answer sheets. Neither Hajar nor Leila indicated whether they use the marking machine.

Table 5.3

Leila's GE Midterm and Final Test Items

Type of Assessment	Section	Description	Number of Questions	Total Marks
Midterm	Reading	A full-page reading passage followed by questions	a. Four multiple-choice questions about reading comprehension b. Two short open-ended WH-questions that require students to provide short answers regarding reading comprehension	4
	Vocabulary	Questions related to the vocabulary list taught in the course	a. Completing a paragraph using the correct word from a list b. Four questions on choosing the correct definition c. Completing three sentences with the correct word from the list	6
	Grammar	Questions on grammar	a. Four multiple-choice questions about the grammar of the course b. Three sentences asking the students to write the sentences in the correct grammatical form	5
	Language Function	Questions on daily expressions	a. Two questions asking students to complete the dialogue for a specific situation	5
	Writing	Paragraph writing	One paragraph-writing prompt (about five lines)	5
Final	Reading	One-page reading passage followed by questions	a. Eight multiple-choice questions b. Eight true-or-false questions c. Four multiple-choice questions on finding word references in the corresponding reading passage	14
	Vocabulary	Questions on the vocabulary taught in the course	a. Ten multiple-choice questions on words' definitions b. Twelve multiple-choice questions c. One paragraph requiring students to fill in the blanks with correct words in a sentence	19
	Structure	Questions on grammar	Twenty-four multiple-choice questions	14
	Language Function	Questions on daily expressions	Filling the blanks in one short paragraph	5

Table 5.3 shows that the final test sections followed the format of the answer sheets

provided by the LC. These questions were the closed-ended questions (e.g., multiple choice and

matching words). As mentioned in the previous section, when the LC was established, it created

a standardized answer sheet for the midterm and final tests for the GE courses. Because the

policy of the unified test had changed, it was up to each unit whether to keep following the old policy by constructing their tests in a similar format to that of the answer sheet. This indicates that teachers do have the freedom to construct their own tests. Leila, accordingly, chose not to follow the given test format in its entirety. While she did not follow the format of the answer sheet for the midterm, she indicated that she shared the test with another teacher in the department. They amended a few sections to include open-ended questions that required teacher marking. Furthermore, Table 5.3 shows that writing was only assessed on the midterm. More marks were assigned to the vocabulary section on both the midterm and final, while the other language aspects had a similar mark distribution. The section on the language functions on the midterm was similar to the other language aspects, but on the final, it weighed the least among the other sections.

Table 5.4 presents a sample description of Hajar's midterm and final GE tests. In this case, the format of both the midterm and final followed the answer sheets distributed by the LC for the midterm and final tests. Both the midterm and final tests included closed-ended questions, but there was also a writing section. The course teacher marked the writing section, and it was up to the teacher to give the students the writing section on the same day or on a separate day of the test date. Hajar stated that she preferred to give students the writing section on a separate day because the students are at low English language proficiency level and might need more time on the writing section.

Table 5.4

Hajar's GE Midterm and Final Test Items

Type of Assessment	Section	Description	Number of Questions	Total Marks
Midterm	Reading	About a full-page, 3-paragraph reading passage followed by questions	a. 6 multiple-choice questions b. 6 true-or-false questions c. 2 matching words with their reference in the paragraph	7
	Vocabulary	Questions related to the vocabulary list taught in the course	a. 5 questions on choosing the correct definition b. 5 fill-in-the-blank questions on choosing the correct word c. 6 multiple choice questions on choosing synonyms	8
	Structure	Questions on grammar	a. 2 open-ended questions on changing sentences into negatives) b. 4 multiple-choice questions about? c. 3 open-ended questions on changing sentences into passive voice	8
	Language Function	Questions on daily expressions	a. 4 multiple-choice questions on choosing the correct response for a specific situation b. 4 matching sentences on a dialogue and its response	4
	Writing	Paragraph writing	1 question on writing a paragraph. A short paragraph is required from students (about 5 lines)	3
Final	Reading	One-page reading passage followed by questions	a. 8 multiple-choice questions b. 8 true-or-false questions c. 4 multiple-choice questions on finding word references in the reading passage	14
	Vocabulary	Questions on the vocabulary given in the course	a. 5 multiple-choice questions on choosing the correct word in a sentence b. 5 multiple-choice questions on words' definitions c. 6 questions on filling in the blanks with correct words in a sentence d. 6 matching questions on synonyms e. 6 matching questions on antonyms	14
	Structure	Questions on grammar	24 multiple-choice questions	12
	Language Function	Questions on everyday English expressions	a. 7 multiple-choice questions on choosing the correct response for a specific situation b. 7 matching sentences on a dialogue and its response	7
	Writing	Email writing	1 main question on writing an email	7

As for the ESP courses, the majority of the participants expressed they had the freedom to choose how they planned ESP midterm and final tests, but they said that they had to abide by

the grading policy set by the LC. The final test date, however, is determined by the college administration. There is a specific date for each course, and it is decided based on the class time. For example, all 11:00 a.m. classes are scheduled on a specific date. The scoring process for ESP summative tests differs from the one for the GE. For ESP courses, it is up to the teacher to decide whether he or she would prefer to use the answer sheet provided by the testing unit and have the tests scored electronically or whether he or she wants to score the test manually. If the teacher decides to use machine scoring, then he or she should, similar to what GE teachers do, submit the answer sheets to the testing committee head. Any member of the testing committee is authorized to mark the test and email the results to the teacher. The participants did not clarify their preference or their actual practices in relation to the marking process.

All participants seemed satisfied toward their summative assessment practices in the ESP courses. However, I observed similarities among the participants teaching ESP courses in terms of designing the midterm and final tests. The majority of the participants included four to five sections in the tests: Reading comprehension, vocabulary, grammar, language function, and writing. The format of the tests were mostly closed-ended questions. The following section presents samples of the midterm and final tests of some participants.

Latifa was the only participant teaching a writing course in this study. Her midterm test was worth 25% and the final test was worth 50% of students' overall grades. She said that she followed the standard PAAET policy on grading for the final test. But the midterm test's grade could be worth 20% to 30% and the course coordinator decided the grade percentage and the format of the test. Latifa said there was no clear rule for teachers to follow regarding grade distribution or the test format, which is set by the coordinator, but she chose to adhere to the

policy of the LC. She said that when she plans her tests, she includes the three different language skills: reading, grammar, and vocabulary.

> Excerpt 17: The format is like we have to have reading, grammar, vocabulary. What's different now is that there's a letter writing. They have to write a full letter. This is different from the other courses. The other [courses] all have the same thing; reading, vocabulary, grammar, and language function. But this time it's full writing. (Latifa, II)

Table 5.5 presents a description of Latifa's midterm and final tests. Although this was the only writing course that was included in this study, I observed that both the midterm and final tests followed a similar format as other ESP and GE courses taught in CBS, in that they included four sections: reading comprehension, vocabulary, grammar, and writing. Nonetheless, the writing section is worth the most of the overall test grade on both the midterm and the final. The tests have closed-ended questions, and the only section that required writing was the writing section. Some questions required students to write short answers (e.g., change a sentence from formal to informal). The midterm and the final were not standardized; Latifa wrote both tests. Still, as mentioned previously, she followed the grade distribution for each section in the tests.

Table 5.5

Latifa's ESP Midterm and Final Test Items

Type of Assessment	Section	Description	Number of Questions	Total Marks
Midterm	Reading	About a half a page of letter prompt, followed by questions	a. Three true-or-false questions c. Three multiple-choice questions d. Fill-in-the-blank questions about parts of the letter	4
	Vocabulary	Questions related to the vocabulary list taught in the course	a. Six multiple-choice questions on choosing the correct word to fit in the sentence	3
	Structure	Questions on grammar	a. Six multiple-choice questions on multiple grammar rules	3
	Writing	Different questions on writing	a. Four questions asking students to change a few sentences from informal to formal or vice versa b. A section on writing an email c. A section on filling in missing parts of a letter using the words in the provided list D. A section on filling in an application form	15
Final	Reading	One page of reading, followed by questions	a. Eight multiple-choice questions b. Eight true-or-false questions c. Four matching questions on finding word references in the reading passage	10
	Vocabulary	Questions on the vocabulary given in the course	a. Six sentences matching words with their definitions b. Ten questions on filling in the blanks with the correct words in a sentence	8
	Structure	Questions on grammar	a. Nine multiple-choice questions on different grammar rules b. Three open-ended questions requiring students to change the sentences grammatically	8
	Writing	Writing	a. A section on writing an email b. A section on completing two emails with the suitable provided sentences c. A section on writing a job application letter d. A section on writing a formal thank-you email e. A section on writing an email to a professor/teacher	24

Table 5.5 shows that the writing section was weighted higher than the other sections. For the midterm, the students were required to write an email, and Latifa only provided key themes as a guide. On the final, the writing sections were worth more, and three writing questions required students to write using their own words, but other writing questions detailed directions

on what should be included. As Latifa mentioned, the students are at a low proficiency level, and she was trying to improve their writing. As noted above, there was a shift in the number of writing questions and total marks between the midterm and the final. It is unknown what kinds of assessment Latifa conducted to improve students' writing between the tests because she seemed to be more confident in giving students questions that required more writing on their own. She reported in both interviews that students were at unexpectedly low proficiency levels; perhaps she added more questions on writing because she felt that students were ready to improve their writing as she explained.

> Excerpt 18: Actually I was surprised at this summer course, I expected the girls to be weaker, and it was also Ramadan as you know, but they surprised me, they were perfect. I don't know, like was it the course itself? Because it was writing, so I think it was something different from the other courses I've taught. But they were actually good. (Latifa, PI)

Other language aspects were included on the tests (e.g., reading and grammar), but there were only a few questions worth a low percentage of the grades. Latifa said that the main focus of the course was writing and that she did not focus on other aspects. She seemed to have included the other language portions to follow the format set by the coordinator.

Leila also reported that she had more freedom in designing ESP tests than she had in GE courses. Unlike in GE courses, in which the midterm's and final's responsibilities were shared with her colleagues, Leila created both the midterm and final tests in the ESP course. She also reported that she did not change her tests for the course between semesters; she used the same test. Table 5.6 presents a description of her midterm and final tests. Both the midterm and final tests included only closed-ended questions, and there was no section on writing. It is clear from the description below that Leila planned her ESP test, in a way similar to the GE tests.

Table 5.6

Leila's ESP Midterm and Final Test Items

Type of Assessment	Sections	Description	Number of Questions	Total Marks
Midterm	Reading	About a half-page reading passage followed by questions	a. Six multiple-choice questions b. Six true-or-false questions c. Two questions on identifying words with reference to the paragraph	10
	Vocabulary	Questions related to the vocabulary list taught in the course	a. Ten multiple choice questions on choosing the correct word to fill in the blanks in a sentence b. Six questions on matching words with their synonyms	8
	Grammar	Questions on grammar	a. Sixteen multiple-choice questions on different grammar rules	8
	Business Skills	Questions related to the language of business especially regarding the topics addressed in the course	a. Four questions about matching time b. Four questions on completing the dialogue using the sentences provided	4
Final	Reading	Half-a-page reading passage followed by questions	a. Five multiple-choice questions b. Five true-or-false questions	10
	Vocabulary	Questions on vocabulary	a. Five questions on fill-in-the-blanks using the correct word from the list b. Five questions on matching the words with their synonyms c. Fifteen multiple-choice questions on choosing the correct word to fit the sentences	15
	Grammar	Questions on grammar	a. Twenty-eight multiple-choice questions on different grammar rules	14
	Business skills	Questions on business language	a. Six questions on matching time expressions b. Five questions on matching a dialogue	11

Table 5.6 reveals that there was one section that was different between the GE final and

midterm tests: the business skills section. This section was similar to the language functions

section, but the language in this course was only related to business. The ESP midterm and final used only closed-ended questions and no writing, whereas the GE midterm included a short prompt for writing.

The type of feedback that occurred for the quizzes, midterms, and final was oral. The participants reported they announced the grades, and provided general oral feedback. There are several reasons for the absence of written feedback. It could perhaps be because most participants constructed their tests and practices to include closed-ended or short-answer questions, and many participants did not implement writing assessment. Another reason is the large number of students—a factor that is discussed under the external factor section.

5.2.1.3.3 Assignments and projects.

Another method that some participants used as a summative assessment was graded assignments. However, there is no specific criteria by which teachers should abide in relation to the course type. Only two participants (Najla and Latifa) indicated that they included assignments as a type of assessment. They did not show that this was a response to a specific requirement or policy; it was their choice to include this type of assessment. Najla even indicated that she implements this practice in all her courses regardless of whether it is an ESP or a GE course. Latifa included written assignments in her course because it is a writing course.

For Najla, the course syllabus referred to an assignment as a "final project." Students were required to submit a research paper of 8 to10 pages on a medical topic and present it at the end of the course that counted for 10% of their final grade. Although Najla designed a major plan for this research writing project, she allocated only 10% of the total grade to it. Perhaps she assigned the project such a low percentage because she had divided the rest of marks into different assessment tasks, as the policy mandated assigning 25–30% to midterms and 45–50%

to final tests. Another reason for assigning the project a low percentage was that Najla indicated that this course introduced students to research writing, but other available courses relied more heavily on writing. She explained that the purpose of writing the research paper was to learn how to use APA format and make presentations.

> Excerpt 19: We do a research in our class, but this is really helping them for other classes. So if they can at least know the work citation—If they know the significance of the publisher or the general layout of the work cited, and form in APA . . . I think that's enough for me. (Najla, II)

Here, Najla implied that she evaluated students on the basics of research writing but did not focus on their writing skills. She provided students with a handout that explained research paper requirements, but she did not give them a rubric. She said that she discussed in class how she would mark students' research.

Najla also required students to present on their research topics. She did not elaborate on this assessment, but she provided a document showing presentation assessment criteria. The presentation constituted only 5% of the overall grade. The research paper was also assigned 5% of the final grade. Thus, the grade for the final project was for both the research paper and the presentation. Other than the final project, Najla asked students to write four essays and marked them out of 5% of the overall grade—but she said that the purpose of writing the essays was to prepare students for their final research paper.

Latifa's strategy for these assignments was different. She said that such assignments were worth 5% of the overall grade. Latifa did not provide me any documentation of assignments; she only reported in the interview that she gave students one writing assignment in the summer semester, in which they were to write an email to her about their interest in the course. She stated that many students did not submit the assignment, thinking that it was not graded. When she announced that the assignment was graded, many students submitted it. This demonstrated that students in this specific case cared more about grades than actually learning the material.

According to Latifa, the purpose of such an assignment was to teach students how to write formal emails, but because she reminded those students who did not submit it that the assignment was graded, she seemed to focus on helping students get higher grades. She said,

> Excerpt 20: I mean, even if it was not an assessment, they need to hear that this assignment is an assessment, because they do not work enough if they know that it is not graded. And I was really flexible with submission. Some girls submitted it at the very end of the course. It was okay with me as long as they submitted it. They will only submit if they know it is graded. (Latifa, PI)

Latifa also referred to other purposes of this assessment:

> Excerpt 21: The purpose of such homework is to assess them, but not directly. I mean they have to do it, and then I can see the homework based on my own criteria; I mean, I would provide them with feedback on their weakest points, and on the other hand, such homework assignments are also for me to check on students' understanding, and what I need to focus more on, or they want me to focus on. I mean, I would check their writing letter assignments if one was very weak or a student was struggling with writing. (Latifa, PI)

It seems that she would ask students to do the homework so she could assess her pedagogy based on their understanding. However, because, as she reported, she was flexible with the date of submission, she would not be able to follow up with the students at the same time and assess her pedagogy accordingly. She mentioned that some students submitted the homework at the end of the semester and that practice was acceptable. She mentioned that because of the short length of the summer semester, she did not give more than one writing assignment and that she was flexible with the submission due date, which suggests that she tended to help students get the full 5% grade.

Participants teaching GE courses (i.e., Nadia, Hajar, and Leila) said they never graded homework; it was only given to students to prepare for the next lesson and save time. Leila, for example, stated that she did not grade homework and assigned it to prepare students for the classroom assessments. She stated that in the early years of her career, she used to grade assignments but that she had abandoned that approach: "In the very first year of my teaching, I

used to give [graded] homework, but not anymore. It's been years since I've given homework. I mean I don't give [them] assignments that they submit to me." She added,

> Excerpt 22: I give them homework, but I don't assess them on the homework. I just tell them to do an exercise as homework for the next day, [and] you see that the good students already did it, and those who are not motivated did not. When we come the next day, we answer the exercise along together. But I don't give them homework on extra sheets and I collect it from them. (Leila, PI)

In reference to the GE course specifically, Leila stated that she had given students graded homework during the year prior to this study. This decision was a response to the coordinator's rules:

> Excerpt 23: Last year, I asked students to write homework in the GE course on writing, because the coordinators asked teacher to include writing in the course, and to teach students how to write. So I started with students—I explained everything, and then asked them to write a paragraph on specific topics—not paragraphs, I only asked them to write [simple] sentences . . . I told them to bring the homework. And honestly, when I collected the homework from them, I didn't mark any. It was just for them to do the homework. Then, the next day, I handed them back the [homework], and I told them, "Let's revise it together." I sometimes attended to individual students, asking them for example, "What did you do?" "What did you write?" And I write the correct sentence, or an example sentence on the whiteboard. I mean I select one of the student's answers—if her answer is correct. I write it on the board for those who did not write, they should learn how to write; and those who did a mistake could revise their answers. (Leila, PI)

Although she mentioned that she gave students homework the previous year, I observed her asking students to prepare writing homework for the next class. She taught students the grammar needed to write the paragraphs, and she explained the main question and vocabulary. I did not attend the class after the students did the homework, so I could not observe her strategy of conducting writing assessment tasks and homework in the classroom.

The type of feedback used in this method was mostly oral. Although this method included written assignments, no participant presented any document showing how written corrective feedback was provided. As mentioned previously, Latifa was the only participant who taught a course on writing, yet she did not express her views and practices on providing written feedback. Najla on the other hand, reported using written corrective feedback practices on written assignments, but she did not provide me with a document showing such a practice.

5.2.2 Formative Assessment Practices

All participants implemented different kinds of assessments that could be classified as being formative although none of the participants used the term formative when discussing these assessments, as noted above. This section discusses the participants' formative assessment plans, purposes, and methods. The purposes of formative assessment are discussed with methods because the participants had different purposes for each method.

5.2.2.1 Planning formative assessment.

Most participants did not distinguish between planning assessment for ESP and GE courses. Unless they mentioned that their plans are followed in all courses, for the purpose of comparing course types, I mention in some sections if the participants made a reference to the type of course they were teaching at the time of the study.

None of the participants submitted any plans of their assessments, nor did they report any plans during the interviews. Some of them provided me with their course syllabi, which included a list of the course objectives, the topics to be covered, and an overall grading plan. The syllabi did not contain any plans for assessment. The majority of the participants described in the initial interviews some examples of the formative assessment tasks they typically use in the classroom; I also observed them using other formative assessments during classroom observation.

There are three possible reasons why the participants did not share any deliberate plans. First, in PAAET in general and in the LC specifically, there is no policy that requires teachers to submit any record showing how they plan their lessons or assessments. Second, as discussed above, summative tests carried the highest percentage of the student overall grade. Formative assessment grades were limited to participation marks, and most teachers combined this percentage with the attendance marks, or they just gave full grades for the student if they felt that

the student needed those marks to push their grade higher. Third, the participants seem to have interpreted assessment as consisting only of paper-based graded assignments or tests. Thus, they might have discussed only practices and assessment plans related to summative assessment only. Also, they think of these as pedagogical tasks not assessment.

As for the formative assessment observed in the classrooms, some participants described those as unplanned, while others indicated that they used to plan what tasks to include in the classroom at the beginning of their teaching career, but they do not do so any longer. Both Hajar and Leila, for example, stated that in the ESP and GE courses they taught, assessment tasks were unplanned, but the same assessment tasks had been practiced for years:

> Excerpt 24: [My classroom assessment is] mostly unplanned because sometimes when I teach, something pops up in your mind, and the students sometimes are a bit active, the students sometimes are a bit lazy, the students are a little bit indifferent sometimes, or they're feeling bored, so what I do is—I mean, they control the circumstances that I'm in. (Hajar, II)

Hajar elaborated on the same point in the post-observation interview, noting that her classroom assessment had been "fortuitous"; that is, she decided on the assessment task in the moment. When I played the recorded segments of their assessments, Hajar and Leila noted that they were all unplanned. Leila stated,

> Excerpt 25: I always go with my intuitions—I mean what I think is the right thing [. . .], because of experience probably [. . .] And, when you follow the teachers' book, sometimes it gives you ideas about how to go with your assessment. (Leila, II)

She commented later that her assessments were all unplanned because she believed this was effective for students in her context.

The majority of participants who taught ESP courses reported that they do plan some assessments to be done in the classroom through worksheets. They revealed that their assessments that were based on worksheets were planned, but they did not always plan how those worksheets were used (e.g., peer assessment, group work).

Najla and Mona explained that they only plan their teaching materials but not assessments. Najla, for example, shared a course syllabus with an organized plan for teaching materials. However, the syllabus did not include any assessment plans except for the midterm and final exams. The syllabus did not describe the assessments in detail but listed them under the headings "Course Requirements," while another section, "Instructional Strategies," indicated that students are required to complete assignments and worksheets.

In terms of planning assessments in GE courses, Nadia was the only participant who taught only a GE course in summer. She indicated that she does not typically follow any particular assessment plans. She also does not use worksheets or paper-based assessments. Her practices, as she reported and as I observed, included mainly oral question-and-answer exercises from the textbook. As Nadia indicated, a coordinator provides teachers with the course syllabus at the beginning of the semester that includes course objectives and the grading plan. The syllabus does not include any type of assessment plans. Thus, it is up to the teacher of the course to plan the assessments, but Nadia seems to follow traditional assessment routines (e.g., asking oral short-answer questions). I did not observe any peer or self-assessment in her class. She seemed more concerned with covering the materials in the textbook than planning various assessment tasks to evaluate student performance.

5.2.2.2 Formative assessment methods.

This section presents different classroom assessment tasks that the participants were observed to use and/or reported using. Those assessments seem to be formative in purpose as the teacher would conduct these to improve student learning and/or to monitor their understanding. Most of the participants did not explicitly identify the assessments as being formative. They

include asking short-answer questions orally, self- and peer assessment, completing worksheet and textbook questions, and group work.

5.2.2.2.1 Asking short-answer questions.

One of the most common assessment methods that the participants implemented in their classroom, was asking short-answer questions orally. These questions were almost always created by the teachers and did not come from the textbook or worksheets. These oral questions seem to serve multiple purposes: monitoring student understanding, moving the lesson forward, eliciting information from students, evaluating instruction, and, for some teachers, managing the classroom. The participants used this method in both ESP and GE courses.

As for the participants' understanding of this method, it was challenging to determine the precise functions of short-answer questions based on teacher reports because most participants did not identify this method as part of their assessment practices. In the post-observation interviews, I selected some segments in which the teacher asked short-answer questions. A few participants, such as Mona and Hajar, identified asking such questions as part of their assessment practices. Other participants, such as Latifa and Dana, perceived asking such questions as part of their teaching practices, and they did not label this practice as assessment.

The first and main purpose of asking short-answer questions was to check student understanding and to evaluate teaching. Mona referred to this strategy in the initial interview as "concept-checking questions." She considered this method as the most vital of her practices; she was aware that it was assessment as well as a teaching-embedded practice and used it for twofold purpose: (a) for her to evaluate the effectiveness of teaching and (b) for students to assess their knowledge of a given topic:

> Excerpt 26: [My assessment is] mostly concept-checking questions, that's number one. If I want to make sure that they know a grammar point, I ask questions about it. If [students] know, then I just build upon it, we move on to the other grammar point. If not, then that's a message to me that they need another tutorial

or another extra reinforcement for that specific grammar rule. You can also use it when it comes to instruction giving, because sometimes I give instructions and I'm not sure if they've understood the instructions, especially those related to exams. (Mona, II)

She believed in the effectiveness of this method because this is one of the practices she learned

from CELTA.

> Excerpt 27: Concept-checking questions are one of the things I learned in CELTA. This is a way to check if the students have understood the point that I was trying to illustrate. So instead of repeating again, or moving to the other point, you check whether you need to further elaborate on something, or to move on to another, because you don't want to repeat things over and over again. At the same time you want to make sure that the students understand what you say. So these are very important questions that teachers should ask, to check if students have understood something. You ask during teaching, and you also do it when practicing some of the exercises. It's not just checking what they've written as they [answer], no, you also check during teaching. (Mona, PI)

Unlike Mona, who provided a detailed explanation about the purposes and uses of short-

answer questions, Hajar acknowledged using this method within instruction without identifying

its purposes. She used it to address the whole class rather than to monitor individual student

understanding. Although she found asking questions necessary to track students' understanding,

Hajar did not find using them to be the best method for following up on whether students had

comprehended the lesson. She instead preferred to use paper-based assessments. For example, at

the beginning of one of the lessons I observed, she reviewed the material from the previous

lesson with students, asking the class as a whole about the grammar rule. She justified this

specific practice, saying,

> Excerpt 28: First of all, individual assessment among students is not very effective with us. I mean, if I ask each one in the classroom, well, I would tell some of them how good they are, but knowing how big classes we have, it's very difficult to have oral assessment, very difficult . . . I cannot have 40 questions for 40 students; I cannot have that number of questions unless I divide the assessment for three to four lessons. (Hajar, II)

Having a large number of students thus posed an obstacle to addressing individual students, and

therefore Hajar did not rely on asking short-answer questions to evaluate student understating.

Other participants, such as Dana did not identify asking short-answer questions as an

assessment, but as an instruction embedded assessment. I observed that throughout one of her

classes she used short-answer questions to address the class as a whole rather than asking each individual student. In reference to this specific segment, I asked Dana whether she addressed the whole class using short-answer questions deliberately or not, and she replied,

> Excerpt 29: I guess, because [pause]—honestly, I'm analyzing myself now, I don't know. But I think the only thing is that here, when I was asking and the whole class is actually responding, it means, at that moment, I was hoping that all of them understood what was going on. (Dana, PI)

When Dana listened to a few recorded segments in which she used short-answer questions, she mentioned that the purpose was to ensure that students understood the previous lesson. She added that this was simply part of her teaching strategy.

The second purpose of asking short-answer questions, which I observed all participants using, was to elicit information and move the lesson forward. Hajar, Nadia, and Najla asked short-answer questions when explaining a new lesson. For example, in one session I observed, when Leila came across new vocabulary in the paragraph, she used short-answer questions to check students' knowledge of vocabulary. She did not report on the purpose of asking these questions, but it seems she used them to explain the paragraph and to move the lessons forward. For answers, she depended only on those few students who provided the correct definition and moved on with the lesson. However, if one student provided the wrong answer, she would stop and explain the meaning of the word with examples.

The following example of the use of such questions is taken from Najla's class. She wrote grammar rules on the board and asked students some short questions that seemed intended to check students' understanding of a previous lesson before moving to on a new topic.

> N: We said there are three things that we have to take care of when we are selecting the noun. I said we have to make sure, for example, when we're talking about Sara, we refer to her as . . . [looking at student for the purpose of getting an answer].
> S: Gender?
> N: Gender, yes [writing the answer on the white board] . . . Also?
> S: The number.
> N: The number [writing on the board].
> S: Subject or object.
> N: Yes, and whether it is a subject or object . . . so today we're going to continue on possession.

A number of participants (Nadia and Hajar) seem to have used short-answer questions for class management, particularly in the GE classes I observed. Hajar did not use assessment only for monitoring student understanding or eliciting information during instruction; in both GE and ESP classes, she appeared to use short-answer questions to manage the classrooms. When she was asking students short questions, she appeared to be paying attention to those not participating. For example, when she was asking students some short-answer questions, she called on a particular student to grab her attention asking her, "Fatma, are you following us?" in a friendly tone. In the post-observation interview, she explained that she was monitoring those students who were not participating and called on them to provide answers. When asked about a specific segment in which she directed questions to the whole class, she replied:

> Excerpt 30: Yes, [short-answer questions are] to address the whole class, and I usually ask a person who I think is the least person who was giving me her attention. Because what I was doing was that I was trying to give them a new piece of information, and I have noticed that two or three of them were not giving me their attention. So what I'm doing is that I'm addressing the whole class in order not to put those two or three to embarrass them. I always do that. When, for example, if Mariam is not giving me her attention, I would say, "Yes, class, we have new thing today, all right, Mariam? Could you tell us so and so?" So I'm not addressing Mariam only to avoid embarrassment. (Hajar, PI)

Several participants, such as Leila and Najla, in ESP and GE courses used short-answer questions to scaffold learners during instruction. For example, I observed that Leila scaffolded learners during assessments in her GE and ESP classes using short-answer questions, especially in the grammar and in reading lessons. When she was answering the grammar questions along with students on the compound nouns, if any student gave the wrong answer, she asked several short-answer questions on the same rule to indicate that the student should revise her answers allowing the student to reach the right answer by herself:

> N: *Request* is singular or plural?
> S: [Silence]
> N: [Directing the question to all students] The word *request*. Is it singular or plural?
> S: Singular.
> N: [directing the comment to the student] *Request* is singular, so do you choose *this* or *these*? Let me give you an example. If you say *a chair*, do you choose *this* or *these* before it?
> S: *This chair*.

N: Is it because it is singular or plural?
S: Singular.
N: So, for *request*, will you choose *this* or *these*?
S: This.
N: Yes, *this*, because *request* is singular.

She also asked students who gave the right answer to justify their answers (e.g., "Why did you choose this answer?"). She said that when students provided the right answers, she usually asked them to justify their answers to make sure that the student did not guess the answer, stating, "because some girls bring a used book that has the answers already written, and when they answer, I also want to make sure that it is their own answer" (Leila, PI).

In terms of teacher role in planning and using this method, asking short-answer questions, for example, occurred at the moment of instruction. Those questions often were not planned ahead of time. The teacher held the central role in directing questions to the students for different purposes. Students in all the classes I observed played a passive role and tended to agree with what the teacher asked for or said. None of the classes I observed incorporated discussion; the teachers only asked short-answer clarification questions.

The teacher role in monitoring this type of assessment was the same across all the participants. After asking short-answer questions, they listened to students' answers carefully and when the whole class went silent, this was a sign for them to explain with the same point with more examples and to illustrate her point on the board. Mona reported: "If they're not responding, then I know that they definitely don't know what I'm talking about. So I'd like to modify the teaching and re-teach everything again" (Mona, PI). Hajar tried to monitor individual students during short-answer question assessment and she tried to ensure that the whole class was engaged by noting student facial expressions. She called the names of a few students who were silent, looked confused, and/or were not paying attention. Although she did not direct

questions to those students whose names she called, instead she tried to engage the whole class

by calling some students during the explanation of the lesson. For this segment, Hajar reported,

> Excerpt 31: To address the whole class, this is one way to attract their attention, and I usually ask a person who I think is the least person who was giving me her attention. Because what I was doing, I was trying to give them a new piece of information, and I have noticed that two or three of them were not giving me their attention, so I'm addressing the whole class in order not to put those two or three to embarrass them. I always do that. When, for example, if let's say Sara is not giving me her attention, I would say, "Yes, class, we have new thing today. All right, Sara, could you tell us such and such?" So I'm not telling Sara only to avoid embarrassment. (Hajar, PI)

5.2.2.2.2 Worksheets and textbook tasks.

The other aspect of assessment I observed was the use of tasks taken from the textbook

and worksheets. This kind of assessment occurred several times during observation for all the

participants. The participants selected some activities from the course book and applied different

strategies while handling them. Similarly, they applied different strategies in answering tasks

from worksheets. Worksheets varied; some were constructed by the participants, while others

were handouts of selected and copied materials from different external resources (other

textbooks, materials taken from different websites, or book chapters). For most participants,

assessment activities taken from either the worksheet or textbook were ungraded and conducted

for formative purposes.

Regarding their thoughts of using such type of assessment, Najla was the only participant

who reported that classroom assessment should occur solely through worksheets and tasks from

the textbook. Latifa found it necessary to add supplementary materials beside the textbook only

if the textbook did not cover enough information about a specific topic. The supplementary

materials included worksheets; she said of this, "We're doing lots of worksheets in the class."

Other participants, such as Leila and Nadia, reported that they only depend on the textbook. Both

Leila and Nadia reported that book exercises were sufficient for the GE courses. They said they

rarely provided supplementary materials. They added to the materials in the textbook with their

own exercises, which they illustrated on the whiteboard. Nadia, for example, stated that when students are given handouts, they often stop paying attention to the teacher. They also tend to rely only on the handout, and they think that the handout materials are the only ones that will be included on the tests:

> Excerpt 32: I don't like worksheets. I prefer students to write themselves, so I give them lots of my own sentences on the whiteboard. I have noticed that as soon I start writing one of my own examples on the board, students try to write down the examples. So I believe that giving the students examples on the board is more useful than giving handouts. Students believe that if I give a handout, then this is what I'm going to include in the exam, so they may not attend the lesson. (Nadia, PI)

She could prefer not to give handouts in order to manage student behavior, encourage them to attend class, and divert their focus from the test material so that they could focus on each lesson on its own terms, because she stated that, "students believe that if I give a handout, then this is what I'm going to include in the exam." Nevertheless, during some of the assessment tasks I observed, Nadia reminded students that similar questions would be included on the tests.

Leila pointed out that she used to create extra worksheets, but in the recent GE and ESP courses, which used two books (the main student book and the workbook), the workbook was a supplement to the main book, and it only contained exercises and assessment tasks.

> Excerpt 33: I used to give supplementary materials a long time ago. But now I don't, because I depend on the workbook. The workbook has enough activities that are based on what I explained in the unit. Another thing is that, when I explain something, I write it on the board sometimes, and I would ask [the students] to give me their own examples. So I stopped giving worksheets. And because of the time limits I don't give [worksheets]. (Leila, II)

Leila identified two reasons for depending solely on the textbook. First, she believed that the workbook contained sufficient tasks. Second, she believed that the class's time limit only allowed her to cover the textbook and that spending time on supplementary materials would lead to delays.

In terms of teacher role in this type of assessment, the participants teaching ESP courses reported that they do plan the use of worksheets in the classroom. Najla and Dana said they

planned and constructed worksheets before the beginning of the semester. Their purpose was to cover points that are missing from the course book or to supplement points that are discussed only briefly in the course book. Latifa added that certain topics would occasionally be presented in detail in the course book, but based on student needs in relation to specific topics she would construct supplementary materials to boost student performance. She referred specifically to teaching grammar, reporting that because it was a writing course, grammar topics were briefly discussed in the book. She sometimes knows this ahead of time and constructs worksheets at the beginning of the semester. Otherwise she just uses her judgment during instruction and constructs worksheets based on students' needs.

Those who used worksheets, such as Najla, Latifa and Dana, were mostly on grammar rules, such as preposition, capitalization, and WH-questions. Najla and Latifa, for example, constructed some worksheets, and they took others from sources such as course books or the Internet (see Figure 5.3). The worksheets they constructed focused mostly on grammar, but they also distributed handouts on vocabulary. The grammar worksheets consisted of grammar rules followed by closed-ended questions (i.e., multiple-choice and fill-in-the-blank) and other short open-ended questions addressing topics such as changing the basic structure of sentences. The types of closed-ended questions in the worksheets were similar to those in the tests, but the open-ended questions were different (see Figure 5.4).

Past Simple Practice

A) Complete these sentences in the PAST TENSE, using the correct verb:

* play   * enjoy   * watch   * listen   * talk   * phone

* stop   * walk   * travel   * like   * stay

Example: I watched the late film on TV last night.

1. We really .... the concert last night. It was great!
2. She .... with friends in Brighton last summer.
3. Italy .... very well in the last World Cup.
4. Her parents .... by train from Kiev to Moscow.
5. I .... you four times last night but you were sleeping.
6. We .... along the beach yesterday. It was lovely.
7. She .... the film but she didn't like the main hero.
8. The men .... work at six o'clock.
9. I .... to the new Shakira's album yesterday. It's great.
10. They .... to us about their trip to India. It was very interesting.

B) Complete the story. Use the verbs in the brackets:

Last year I *went* (go) on holiday. I .... (drive) to the sea with my dad. On the first day we .... (look) at the beautiful buildings and .... (eat) in lots of cafes. The next day .... (be) very hot so we .... (drive) to the sea. We .... (leave) our clothes in the car and .... (swim) all day. At six o'clock we .... (walk) to our car, but the car ....(be, not) there. We .... (buy) some clothes and .... (go) to the Police Station. The police .... (be) nice and we .... (sleep) in the police station.

C) Make negative sentences (x) or questions (?) using the PAST TENSE:

Example: he / stay (x) in a hotel
         He didn't stay in a hotel.

         you / learn (?) speak Italian when you were in Italy.
         Did you learn to speak Italian when you were in Italy?

1. she / like (x) the film

2. I / use / (x) the school computer yesterday

3. they / arrive (?) home late yesterday

4. you / ask (?) her for Jane's address

5. he / enjoy (x) the concert in the park

6. we / talk (x) about our winter holidays at all

Figure 5.3. Latifa's worksheet sample on grammar

Exercise Fourteen: Circle the correct answer.

1. Catatonic patients suffer more than those of epilepsy because **(Their, theirs)** episodes last longer than the seizures of epilepsy.

2. The lab. tech. gave **(I, me)** the blood test results.

3. The surgeons are trying to find a donor with a blood group compatible with **(her, hers)**.

4. She gave the insulin injection to **(her, herself)**. She doesn't need anyone's help.

5. The nurse asked **(they, them)** to donate some blood.

6. **(She, Her)** skin suffered the affects of dehydration.

7. The nurse encouraged him to change the dressing **(him, himself)**.

8. **(My, Mine)** has more options than **(him, his)**.

9. This painkiller has no harmful side effects. **(It's, Its)** quite safe.

10. **(They, Them)** donated some blood.

11. The doctors seem pleased that she has made such good progress since **(her, hers)** operation.

12. Don't touch food with unwashed hands. **(It's, Its)** not hygienic.

Figure 5.4. Najla's worksheet sample on grammar

Latifa and Najla said they planned to provide students with worksheets before the course started, but Latifa explained she would sometimes observe students' performances while various topics were covered. If she assessed that students' understanding of a particular topic was low, she would create some extra worksheets on a specific topic. For example, she found that students did not perform well in grammar exercises, especially WH-question rules, so she made several worksheets on the grammar rules related to WH-questions. She stated that she made this specific worksheet when she found that their performance on this topic on the midterm was poor. This implies that the midterm is a source of assessment that Latifa uses to identify students' weaknesses. Analyzing the final test, only one question required students to rewrite a sentence using a WH-question. It seems that Latifa's purpose of giving students worksheets on WH-questions was more a matter of teaching them the grammar rule for them to understand rather than to cover what was included in the curriculum or what was on the final test.

The majority of participants who used worksheet exercises, especially in the ESP courses, followed similar assessment strategies when conducting such assessments. The following strategy from Latifa's class represents how the participants handled the worksheets with students. Latifa first explained the grammar rule of WH-question and guided students to what was required. She then gave them a few minutes to read and answer the questions. However, students were seated in traditional rows, which discouraged group work, so students worked individually. She went on to answer most of the questions with the students. She did not select individual students to answer, perhaps because of the large number of students. Najla's, on the other hand, was small in student number. So she asked each student to answer questions in order of seating arrangement or, alternatively, whenever a student raised her hand. I did not observe Najla going around the classroom and engaging with students while they completed their worksheets, but she provided feedback for each student. Dana and Mona used group work when conducting this method. For instance, Dana implemented group work to answer one writing task, or she asked students to answer the questions together as a whole class, requested a volunteer to come and write the answer on the whiteboard, or divided the class into groups to share answers.

In terms of the use of this type of assessment, almost all participants mentioned that they did not grade any worksheets or textbook exercises but used this type of assessment for two purposes: for students to track their own learning progress and, based on their responses to the questions, to assess her own teaching plans. Only Dana reported that she sometimes marks the worksheets as a homework. Although she stated that she marked students' homework, she did not collect the homework during the class I observed. Instead, she did the tasks cooperatively with students without taking the names of those who answered. This implied that she may give students homework to increase class participation, and not necessarily for grading purposes,

because she pointed out that students should be informed that their work would be graded so that they do the work. Dana reported that she informed students at the beginning of the course that homework was graded, and I noticed that most students did the homework, so it was easier for Dana to go over the questions quickly.

Several participants used this method to provide feedback to students. However, those participants provided feedback to the whole class during the worksheet activities, except for Najla. She, for instance, provided different forms of feedback proposed by Hill and McNamara (2012). When her students started answering the worksheet questions aloud, the feedback she used for individual students was praise (e.g., "very good" or "excellent"). Najla used other confirmatory feedback too—she repeated and added explanations to the students' answers, and even if a student gave the right answer, she asked more questions on the same point. For example, she would repeat the question while nodding that the answer provided was correct. She also provided corrective feedback when students gave wrong answers, asking the same question again and seeking the correct answer.

The majority of participants scaffolded learners when they used the worksheet method. For example, I observed Mona scaffold learners during assessment activities in grammar. When she was answering the worksheet on present continuous tense along with students, if one or more students gave wrong answers, Mona asked more questions on the same rule in an attempt to help the students understand why the answer was wrong. Even if one student gave the right answer, Mona asked the student more questions on the same point to allow the student to understand how she came up with the answer. Although Mona described scaffolding as an approach to ensure students' understanding, she did not provide scaffolding for individual students. Instead, she

addressed the whole class, and those students who were silent when she asked questions about wrong answers did not demonstrate whether they understood the content.

Some other participants used the worksheets to resemble the actual questions or the question format of the midterm, the quizzes, or the final. This strategy occurred in the ESP classes. Moreover, the strategy of using worksheets in class was used to prepare students for the tests. For example, when Mona asked students to complete a task, she asked them to answer as if it were a test. She reported that she planned the worksheets to be similar in format to the tests, which is perhaps why she asked student to read and answer as if it was a test: "The practices we do are very similar to the way my examinations are designed. I mean it's not fair to give them one type of practice and surprise them in the test with a different type of practice" (Mona, PI). According to Mona, the reason for having students complete a task individually is to make tests easy for them and reduce stress. When she asked students to answer the question as though it were on a test, she reassured them indirectly that the practice exercise would not be graded; therefore, students were prepared psychologically to handle formal tests.

Najla followed similar practices as Mona. She depended on individual assessment; she distributed worksheets first, gave time for students to answer on their own, and then called on students to answer the questions. She did not report whether she intended or planned to construct her worksheet in a format similar to that of the tests. She also did not indicate that she did any particular practices to prepare students for the test. However, through the analysis of the worksheets I noticed the question format in the worksheet is similar to that on the test. For example, she depended mostly on closed-ended questions for grammar.

Monitoring student performance during answering worksheets varied among the participants. Mona and Nadia for example, reported that they usually sit at their desk and observe

whether students are engaged in an activity if the class is small but that she would walk around if the class were bigger. Mona for example, reported,

> Excerpt 34: If it's a large class then I walk around, and this gives extra attention to the students who need me to explain something. Yeah, and during walking, of course it can also allow me to assess their understanding whether they're answering the practice without any difficulty or whether I (need) to provide more practices, or whether the practices suitable for them or not. But if the class is small, I think the class you've attended was only about like 12 to 15 students, like this, and you were sitting very closely. I mean it's OK I can sit next to my desk, or just walk around it, and watch them. (Mona, PI)

However, Mona's class was small and Nadia's was large. Yet, they both did not walk around.

While students were answering the worksheets or textbook questions, for example, Latifa moved around to check their answers, focusing on those who sought help or feedback. The room was small in size, and she set tables in a U-shape format, which made it easy for her to move around to check on each student's work. This physical setting suggested that she intended to create an interactive class with students and teacher as engaged in the various tasks. Even though hers was a writing course, she did not attend to each group when completing different tasks, possibly because she relied on students' facial expressions to detect whether they were engaged or not. She also asked students to raise their hands if they had any questions.

5.2.2.2.3 Peer assessment.

Regarding the participants' understanding of this method, most of the participants defined this method as students reviewing paper-based assessments of their classmates. Nadia, as the other participants, described what she thought were two examples of peer assessment. The first example was "when friends share answers with one another" and she equated this with group work. She described the second example as follows:

> Excerpt 35: Sometimes I inform [students] that there is a quiz, and then at the end of the quiz I ask them to exchange papers, or I assign papers randomly to the students and answer the whole exam with them, and ask them to correct for themselves or for the students. (Nadia, II)

The majority of the participants believed in the effectiveness of peer assessment. Some reported that their practices included peer assessment, and others reported that they could not

implement this method because of various factors that will be discussed later. In terms of the use of peer assessment in GE and ESP courses, it is important to note that the participants often did not indicate that the assessment they use are limited to either GE or ESP courses. It just appeared, perhaps coincidently, that teachers of ESP courses used peer assessment in their classes. However, peer assessment was not planned or conducted regularly perhaps because, as mentioned earlier, the participants did not feel students were ready to engage in this type of assessment. But some participants reported that they do engage their students in a form of pair work that they considered peer assessment.

Several participants such as Latifa, Mona and Leila, believed in the effectiveness of peer assessment, but did not think it could be used in their current context. They explained that students in their context are not trained to respond to such assessment because, as Mona believed, her students could not engage in peer assessment because they are "passive learners," and Latifa reported that most students are graduates from public schools, where English is typically taught through the traditional grammar translation method. Although the curriculum in public schools is moving from a traditional to a communicative approach, many teachers believed in the difficulty of applying communicative approach in their classes. Latifa did not employ peer assessments in her classes, but she said that she would not mind implementing it in the future. Mona highlighted two benefits of peer assessment in the interviews: that peer assessment promotes students' learning as students can learn from each other, and reducing students' stress, as opposed to sharing their answers with the teacher. And, peer assessment is important to switch the roles of the teacher and students. Mona did not grade peer assessment, she repeated that students were not used to it, she could not depend on their marking skills.

During my observation, only Mona and Dana used peer assessment in their classes. However, they both had conflicting views regarding implementing it. In the initial interview, Mona mentioned that students were not used to this type of assessment and that they did not prefer it. But in the post-observation interview, after I played back the recorded peer assessment segment, she stated that students preferred to have their answers checked by their peers because they might "feel threatened" when the teacher monitors them. It is possible that she was referring to students' perceptions about peer assessment in the initial interview, while in the post-observation interview she was speaking about the goal of peer assessment more generally,

> Excerpt 36: This is purely peer assessment. The purpose of that, number one, is to get [students'] self-confidence a little bit up, because once we do answers together, once the teacher checks the answers, some of them feel a bit threatened. And this is an easier way for their answers to get checked. They would rather be corrected by their friend rather than the teacher. As a student and as a learner of a foreign language, sometimes it's a bit threatening to your self-confidence to always speak up the wrong answer. I would rather that one of my friends correct it and then we work up the correct answers together, and then later on when we're both sure of the answers we can check with the teacher. This is also one of the main assessments I've learned in CELTA. (Mona, PI)

Dana reported a different view of implementing peer assessment. In the initial interview, she reported that she planned to implement peer assessment in her course because it would be helpful in large classes in terms of grading and evaluating students. Later, in the post-observation interview, however, she expressed a different view. She stated that it was difficult to manage peer assessment tasks with so many students, and that it could be more manageable in smaller classes. She attributed the failure of implementing such assessments to it being the month of Ramadan when students tend to be less energetic because of fasting and to the large number of students.

Peer assessment did occur during Mona's classes that I observed. Mona asked students to complete a worksheet. She first asked students to answer individually, then she asked them to switch papers and mark each other's papers. Mona reported in the initial interview that this was what she usually does in her classes:

Excerpt 37: When I'm giving them worksheets about a lesson, first of all I ask them to try to test themselves in answering the questions and if they finish they can consult their peers and once they're done, they can either peer check, which is basically, marking their friends' paper. Or I can put them also in a group and they can check in a group. I do believe in that because sometimes they learn better from each other, but you also need to make sure as an instructor that you're there, because they can easily learn something wrong. So you have to supervise, so you have to go around and check. (Mona, II)

Mona's approach to assessment was different from other participants. She preferred peer assessment to be performed under her supervision, because her students were still not ready for peer assessment. For example, in the session I observed while students were completing the worksheet individually, one student asked her friend a question about the worksheet exercise, but Mona asked the student to direct the question to her. Later in the post-observation interview, I asked Mona about this specific moment; she replied that this student usually felt too shy to ask the teacher directly. Mona said that, in this particular case, when students are not instructed to check their answers in pairs yet, she preferred that students ask her so they would get correct guidance rather than get incorrect information from their peers.

Excerpt 38: The teacher is always the correct main source of information. Peer assessment is when they correct each other's exercises or practices that they have done, but the teacher should be the main source of information, because sometimes the students do not know how to explain well. That's why I prefer if she wants to ask a question about something that she didn't understand, she would ask me rather than other students. (Mona, PI)

Mona seems convinced that her students were not ready or able to do peer assessment. As a result, she preferred a structured form of peer assessment where students are not allowed to discuss in pairs until the teacher directs them to do so. This approach seems to contradict the goals of peer assessment as discussed in the literature and could decrease student confidence and autonomy and result in students being passive learners.

5.2.2.2.4 Self-assessment.

Similar to peer assessment, I did not observe self-assessment in the participants' classes, although they expressed their advocacy to apply this type of assessment in their lessons. They believed, however, that students were not ready for self-assessment. The other participants who

reported that they applied self-assessment in their classes seemed to use it for test-preparation than for student learning.

Most of the participants did not provide a clear definition of self-assessment, Hajar, for example, did not provide a clear description of self-assessment, neither did she reflect a clear practice of self-assessment. I asked her in the initial interview whether she conducted self-assessment; she first replied "no," but then revised her answer immediately, stating that she sometimes divided the class into rows and they would exchange papers to mark each other's work:

> Excerpt 39: I have called one of the students to explain something on the board and they assess her—this is one thing. Another thing, I would give them for example a question on the board, and I would tell each one of them to try her own answer, and then when I explain, I say, "Check. Is your answer correct or not?" So she assesses herself; no one knows. Or sometimes they exchange [papers] and [grade each other]. (Hajar, II)

She believed asking students to answer individually and then check their answers with the teacher to be a form of self-assessment, but it was left to the student to perform such an assessment; if a student tried to check her answers and see whether they were correct, she would be engaging in a form of self-assessment.

I only observed Mona and Hajar applied some form of self-assessment. They both applied it in the grammar lessons and seemed to use it for the purpose of test preparation. For example, during one of Mona's grammar lessons I observed, she explained the present continuous tense she distributed a worksheet on the topic, which contained the grammar rule about the present continuous, followed by a set of open-ended questions. Mona appears to complete the exercise using self-assessment. She asked the students to complete the worksheet individually. She explained this strategy as follows:

> Excerpt 40: First of all, I wanted to make sure that they understand everything by themselves, another thing is to promote learner's autonomy and this is when teachers encourage students to depend on themselves in learning and answering questions, and then later on we can check the answers together, so yes I did that deliberately I mean on purpose. (Mona, PI)

Although Mona stated that the purpose of self-assessment was to enhance students' autonomy, when she asked students to complete this task, she created a test-like atmosphere; each student answered on her own, but she did not walk around to follow-up with students. When she asked students to answer the questions as though they were taking a test, she reassured them indirectly that this practice exercise was not going to be graded. She reported that she asks students to answer the questions as if it were a test because she planned classroom assessment tasks to be similar in format to the tests. According to Mona, another purpose of having students complete a task individually like a test is

> Excerpt 41: that at least they know how to test themselves instead of asking their friends for some answers or some help. If they depend on themselves at least it can help in lowering exams' stress. I think, you give them practices now and then and they act as a test, you know at least when the real test comes, it can be a bit easier. (Mona, PI)

For Hajar, she reported that using self-assessment for the purpose of test preparation was not related to the test format but rather the students' awareness of how to review their answers or to identify their mistakes—a strategy that appears to be necessary for all types of assessment. She asked the students to complete an exercise and then answered the questions with them. When she had answered all the questions, she asked whether anyone had gotten all of the answers right. Because she explained that the level of the students was low and that she believed that the teacher role is dominant in her context, Hajar seemed not to apply self-assessment. It may be that she considered this the only appropriate individual assessment that she could apply in her classes.

As indicated above, several participants used short-answer questions for the purpose of scaffolding. Najla used short-answer questions for such a purpose, but she reported that the ultimate purpose of asking short-answer questions was to enhance student self-assessment. For example, she was seen scaffolding learners when she assessed them on grammar tasks. She did

the first question with them, then scaffolding occurred when she instructed them on how to

handle an exercise.

> N: So you have here from 1 to 24. You have nouns and we refer to something in relation to the noun. So *feet* and we want to talk about the swelling of the feet, so we say/
> S: /its
> N: Ok, foot or feet
> S: Feet
> N: One or many?
> S: Many
> N: So, it's a plural, so we say its or their?
> S: Their
> N: Their, OK. So do the exercises 2 to 24 and then we'll do them together.

As she specified in the interviews, she aimed to make students "conscious learners." She wanted

students to assess themselves and understand their mistakes with her help, implying that she was

trying to scaffold students to get them to the point where they could reason by themselves. It is

uncertain whether she considered her strategy to be effective in making students conscious

learners or if she was just following a routine assessment practice that could be aiming at moving

the lesson forward.

5.2.2.3 Teacher role in formative classroom assessment.

Not many participants reported their role, or teachers' role in general, in relation to

classroom assessment. Most of them expressed their beliefs about teacher's roles in relation to

summative assessment relating it to their own contexts. Dana and Hajar acknowledged that the

teacher role in this context is central in classroom assessment. Dana, nevertheless, was against

this situation of holding the central role in assessment. She said that whereas the teacher should

not be dominant in the classroom and that students should be engaged in the learning and

assessment process, she felt it is hard to apply this thought in her context. She stated in the initial

interview,

> Excerpt 42: When I taught last semester, I felt it was teacher-centered most of the time, but I noticed that it wasn't the best thing, especially with 70 students. I noticed girls in back were lazy—I want them to be active, they're passive. So this time, I'm trying as much as possible to make it student-centered. I mean I'm including more activities like group work. Also, for example, if I provide instructions for an activity, I

would let one of the students explain it to the students, as they were teaching themselves, because that's what I've learned from doing my masters; one of the best way to make sure your students understood is to make another students explain it—let them teach. So I'm trying my best, but with 70 students in my class, it is difficult, but I'm trying. (Dana, II)

Dana indicates here that a teacher should not hold the main role in classroom assessment, yet she thought the context was a major factor that would hinder implementing such beliefs in classroom. Although she was seen to apply several group tasks in the classroom, she revised this specific practice and perception in the post-observation interview about the teacher role, saying that it was "impossible" for her to create a student-to-student environment because of the small size of the classroom and the large number of students.

Excerpt 43: Last semester, I was more of the one who's talking and explaining, and I realized it doesn't do anything, like they are considered university-level students, but the way to reach them—because the culture here is different—the way of learning is completely different in Kuwait versus like when I taught in the States. [In the States] when I teach them grammar rules, I'm sure they're gonna study it, and if I give them fun activities, they will participate, and they will study it. Here, it's not the same way. You're not teaching them, and there's 99% of the students are not gonna go home and revise. (Dana, II)

5.3 External Factors

The second research question was: What factors influence the classroom assessment beliefs and practices of these EFL teachers? This section presents what the participants reported as the factors that seemed to influence their beliefs and practices concerning assessment. The main external factors are presented in Figure 5.5. They include factors related to the teacher, and then contextual factors related to students, institution, or the country. The teacher factors include the teacher educational background, and the contextual factors include the teaching context, factors related to the class and students, administrative policies of the LC and PAAET, and the factors related to the country. The following themes are further divided into subthemes to indicate the factor in relation of the type of course.
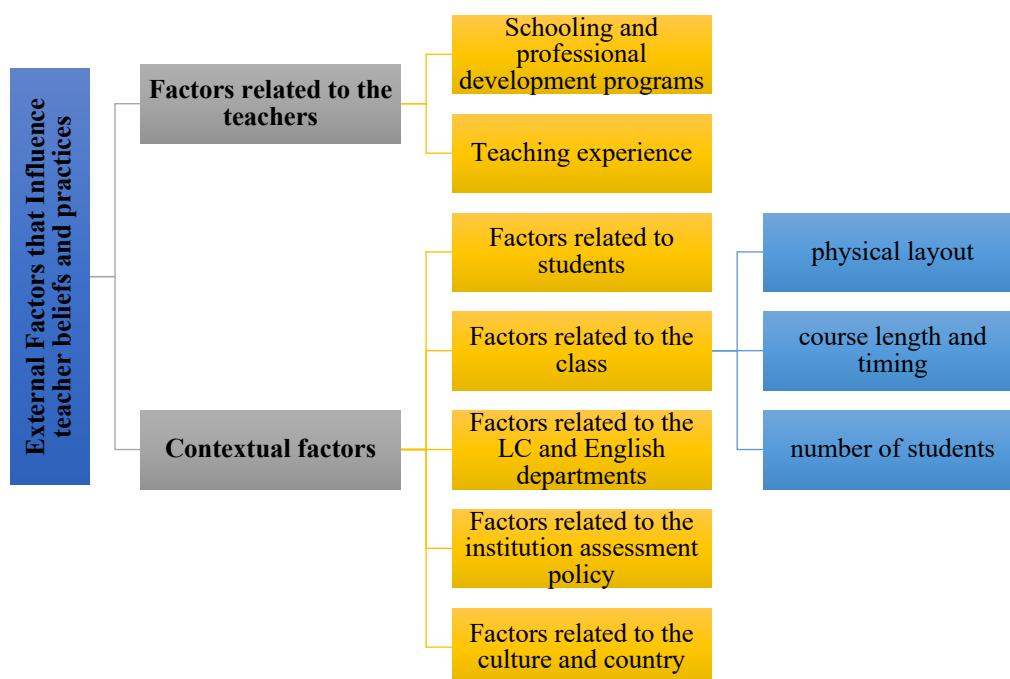
Figure 5.5. Themes of external factors

5.3.1 Factors Influencing Teachers' Assessment Beliefs and Practices

This section presents the factors that the participants expressed as influencing their beliefs and practices around assessment in general. This section is not related specifically to a certain type of course (i.e., GE or ESP). The factors include the educational background and the context, which includes the culture and other factors related to students.

5.3.1.1 Teacher educational background.

Several participants referred to the influence of their previous education on their assessment beliefs and practices. This factor was not related to a specific type of course but was rather about their general assessment beliefs. This section includes the participants' previous education, from undergraduate to postgraduate levels, in addition to the professional development programs, if applicable.

5.3.1.1.1 Schooling and professional development programs.

A number of participants reported the influence of schooling on their beliefs and practices about assessment. Mona, for example, being educated in public schools in which formal assessment constituted a large part of students' grade point averages (GPAs), viewed assessment as tests perhaps because those metrics affected her as a student. Mona, for example, said that such assessments placed major stress on her as a student and that that experience made her consider not depending solely on formal assessment in her practices, but this goal could not be achieved.

> Excerpt 44: Ninety percent of my education—I was taught in public schools and I would say the majority, if not all forms of assessment were formal, so they were paper-based. I think they placed a lot of stress on students, and that definitely made me more aware of this. So I don't believe that students should be assessed a 100% formally. I do believe that part of it should be done informally, and graded as well. So that you would check their understanding because this is language use and sometimes there are stress factors. That's one thing, another thing for my master's, I was actually being tested on formally and informally of my understanding, and I actually scored better, yeah because the stress relief, the stress factor was much lower at that percentage. So yes, it definitely changed my perceptions and beliefs. (Mona, II)

She emphasized stress as a significant factor that influenced her assessment beliefs. Thus, it seems that she had been influenced negatively by formal paper-based assessments when she was a student in public schools; she compared this with her experience in her master's degree phase when she had better familiarity with informal assessments. As such, Mona seems to employ various informal assessment methods in an attempt to introduce students to them.

Latifa also reported the influence of her undergraduate education on her assessment practices. She perceived assessment as grades when she was a student, which is perhaps why she labeled graded tasks as assessments. She referred to her public schooling when comparing the assessment system between her country and the assessment education studies she did for her master's degree in the UK. She stated,

> Excerpt 45: In the UK, it was [a] different [system of assessment] because we didn't have tests. Before you do your thesis, you have courses, and in these courses, you do like ten pages of research, like a report, and that's the only thing that you get graded on. (Latifa, II)

She added that the test construction basics she studied in her undergraduate period had affected her teaching and that she had applied the tips when she wrote the tests. Indeed, her tests contained a variety of questions, yet the format and the questions were similar to those of other participants. For example, her test had four sections, as did those of all the other participants, and the questions were similar (e.g., multiple choice, matching words with references). Thus, the institution's context and policy may have had a stronger impact than her schooling.

Dana also stated that her schooling influenced "what kind of teacher" she currently was in that she used the same teaching strategies as her own teachers had. She pointed out that she studied in private schools in which English was used as the medium of instruction. In private schools, students acquire English through communication and are not taught grammar explicitly. Accordingly, she preferred to teach grammar the same way and she planned to teach students grammar through writing. However, her answers on how her schooling influenced her practice were not related specifically to assessment but teaching. And, as opposed to what she mentioned about the methods of teaching grammar in private schools, the activities I observed did not focus on writing. I did not observe students writing essays or learning grammar through the communicative approach. Her approach to teaching was similar to her colleagues at the CBE. She followed the traditional method in terms of explaining the rules and then answering tasks with students. The students were mostly recipients of information, even though she placed them in groups. And, as mentioned before, she reminded them during most of the assessment activities that these were tasks that would be included on the tests. Her tests also did not include many writing sections, and the sections that required writing were only short-answer ones. It seems that her experience in the LC influenced her practice more than her schooling.

The participants also reported the influence of the professional development programs on their assessment beliefs and practices. Two participants, Najla and Mona, have a certificate from CELTA. Mona described the influence of CELTA in detail, but she related it more to her teaching than to her assessment practices. In the initial interview, she described CELTA as having a huge influence on her teaching practices, although she did not mention assessment specifically:

> Excerpt 46: [CELTA] did highly influence my practice, as well as my beliefs, because for me it felt like it was a very strong basic foundation that I really needed because my BA was purely about linguistics. I haven't had any courses in teaching or in education. So that I believe it was a very good foundation. It talked about so many different aspects of teaching and learning, things like teaching the basic skills, how to avoid the most common mistakes teachers make unintentionally, how to deal with students, how to encourage them, things like very good tips on teaching, how to plan lessons, how to plan and decide on curriculum. Some psychological aspects of encouraging students—when to correct their mistakes and when not. Assessment—we took very basic assessment information, because it was mostly dedicated to teaching and learning rather than assessing. (Mona, PI)

Even though CELTA is a training course in language teaching rather than language assessment, as she pointed out, it appears that most of Mona's assessment practices were gained from CELTA, such as asking short-answer questions and peer assessment. However, in Excerpt 46, when Mona described what CELTA taught her, she included practices that she thought were part of teaching, but in fact could be a part of formative assessment, such as teachers' assessing their own methods (e.g., "how to avoid the most common mistakes teachers make unintentionally") and providing feedback (e.g., "psychological aspects of encouraging students: when to correct their mistakes and when not").

Najla did not describe the influence of the professional development programs in as much detail as Mona, but she indicated the need for more professional programs to be conducted in the LC:

> Excerpt 47: So if [the language center has] a proper way of [instructing us] the steps of standard assessment, by giving maybe a lecture or a class, then I would kind of see what it am I doing in my teaching, and then how is that applicable. (Najla, PI)

When describing the need for more development programs in the LC, she added,

> Excerpt 48: I think that would definitely help us to improve how we're gonna write exams, and also help in trying to focus on the objectives to kind of have better objectives of a course. I can generally speak about the objectives in relation to themes; for example, in reading, and in writing, but again I don't think I'm able [to assess] what kind of skills [students] are leaving my course with. I think to do that, I would need a proper assessment, someone who has really studied assessment. (Najla, PI)

In both instances, Najla stated that assessment training programs could significantly improve teaching. Again, she did not perceive training programs as improving assessment strategies. Nevertheless, she implied that she did not know how to assess language skills and that she needed to learn more about constructing exams. She did not refer to CELTA as having had an impact on her assessment or teaching practices; she stated only that she had learned how to prepare lesson plans.

5.3.1.1.2 Teaching experience.

Another factor that the participants reported was the impact of teaching experience on their beliefs and practices, regardless of the type of course they taught. Hajar and Leila reported that their experience teaching in PAAET for more than 30 years had a major impact on their practices. They stated that most of their assessments were unplanned because they had tried different strategies over the years of teaching. They both agreed that their current practices worked for students in their context. On this point Hajar elaborated,

> Excerpt 49: [My assessment] is unplanned, but I'll tell you something: they're the outcome of experience. They were the outcome of trial and failure, and I tried them many times; I corrected myself many times. It's like experimental periods I have been through, but now I don't plan anything, I just go unplanned to the class, it just pops out, and I know where to go now because I am now more experienced than before. Yes, before I used to plan, but not anymore. (Hajar, PI)

She perhaps had plans in her mind and implemented them routinely in class because of her experience. She was convinced of her assessment methods: "Now I'm a better assessor; I can judge whether this student is good or not as a person. I'm a better evaluator" (Hajar, II). Looking at her classroom practices, bonus questions, for example, appeared to be unplanned; at some

points, when she felt that students were not participating, she gave them a bonus question (an extra credit assignment).

She also stated that her views on assessment were an outcome of students' needs:

Excerpt 50: Dealing with students teaches you a lot. You learn from them. So this is my 34th year of teaching, so I have been exposed to different kinds of levels of students. And I have known if you give them a test that is assessing a higher level than theirs, they're gonna reflect what you have given them. (Hajar, PI)

Some participants' experience had been short (namely Dana and Mona), though they had teaching experience elsewhere. Dana, for example, had taught in public secondary schools prior to teaching in the LC. She pointed out that this experience did not influence her teaching practices, except that she was aware that the educational system in public schools was different from the LC, and she did not want to apply her experience in public schools to the LC because, in the LC, she had more freedom to construct exams:

Excerpt 51: I have zero influence from my experience, and I don't want to take it, but the only thing I do understand that what made my life easier teaching in the LC is that they focus on language function, grammar, set book questions, and reading comprehension, and writing, and sometimes translation. That's the only knowledge that I took, and that I'm using, but nothing else. (Dana, II)

Dana clarified that the way she constructed her tests was similar to the format required by her department. Although she expressed that she had the freedom to construct her own tests in ESP courses, she designed them to be similar to the answer sheets provided by the testing unit in the LC. However, she created the questions. Thus, she followed certain rules, but she did not indicate whether her testing formats were mandated by the department or whether she had the choice to follow them. She only stated that a coordinator for each course set the rules, but she was not certain whether teachers had to follow them. Dana seemed to apply her understanding of the policy regarding test construction. The influence of experience seemed to be more aligned with following what teachers in her department were practicing than implementing her beliefs.

Mona's experience in the LC was short as well. When comparing her experience at LC to the private English institution in which she taught prior to joining the LC, she said that she preferred the policy of the private institution, which allowed using more informal assessments to evaluate students. She said that her experience in the LC affected her beliefs about students in both contexts.

> Excerpt 52: My humble teaching experience made me realize more how obsessed students are with paper-based testing, and I actually do. I am aware that we should definitely try to change that, or maybe alter it just a little bit to cater to students' needs, because for my previous experience when I used to work at the [private college], some of the assessments, we've done them informally in class. Sometimes we would assess projects and presentations. Students had more fun, instructors definitely had more fun, and [students] scored better because […] you can actually cater to the students' different learning styles as well. (Mona, II)

It seems that her experience in the LC negatively affected her assessment practices in that she found that students' evaluations were mainly based on the midterm test, final test, and quizzes. She believed that students would achieve higher marks on informal, rather than formal, assessment.

5.3.1.2 Contextual factors.

One of the major factors that was reflected directly and indirectly in the participants' reports and their observed practices—and that actually cause some discrepancies between their beliefs and practices—was the influence of the context in which they were teaching. The context includes the country, culture, institution, and classrooms, all of which overlapped. That is, a number of participants linked the institutional context to the student culture. The majority of the participants reported the influence of the context without relating it to a specific type of course. The following factors were common among the participants.

5.3.1.2.1 Factors related to students.

This section discusses student educational backgrounds, motivation, English proficiency level and understanding of and role in assessment as factors that influence teacher assessment

beliefs and practices. In order to understand the participants' beliefs around the factors related to students, it is important to note that a large percentage of students in the PAAET colleges consists of graduates from public schools. In public schools, students are taught English as a mandatory subject. However, the participants in this study seemed to believe that their students had graduated from high schools with low levels of English proficiency. They also seemed to believe that it is the MoE who is responsible for students having low English proficiency levels.

The first factor related to students is the low English proficiency level. Najla linked the students' proficiency level to the course type. She was not specific about this course, but she mentioned that in some courses the English proficiency level of the students was low. She thought that some students from certain regions of the country had weak educations, particularly in English:

> Excerpt 53: We have students who are from specific backgrounds; their education is generally low [and] their kind of reception is not so much. So we realize that we have to assess students, for example, on pronunciation, which we did not have. (Najla, II)

The second factor related to students is student reliance on paper-based assessments, which led many participants to familiarize students with the tests. Several participants criticized students' attitudes towards assessment. They believed that students in this context usually wanted to focus on paper-based tests rather than informal classroom assessment. Mona, Leila and Latifa for example, attributed this common belief to the policy of public schools in the country. They said that students in public schools are used to summative paper-based assessments and to situations in which the teacher's role is primarily teaching and assessment. Mona stated that people in her context are "obsessed" with paper-based formal assessment—a situation she looked forward to changing.

> Excerpt 54: I think the fact that most of our students are graduates from the public schools in the country, they are used to very traditional way of seating, assessment, teaching. They're used to the teacher sitting on a chair, with them facing that teacher and basically them being spoon-fed the information. They're not used

to having the teacher as guidance and […] they're not used to other things such as informal assessments. So sometimes even if you want to try to introduce some of those concepts, they won't take it as serious, and they would think that that's not fair, or [they] want to have a proof of it. (Mona, II)

I observed several participants (e.g., Mona, Najla, and Dana) allocated a part of the class time for revision of what would be included in the upcoming test. When I asked Dana about one segment in which she reviewed lessons that would be included on the test and highlighting the main topics for students to focus on, she replied,

Excerpt 55: When I taught another class, not this class. I mean it was the first semester that I start teaching—when I don't tell them what's coming in the exam, it was a disaster, I was shocked with their grades. And if I didn't revise, it was embarrassing. So I was like, "OK, maybe something that I've done wrong, or maybe there is something wrong, I don't know what." So I started revising. (Dana, PI)

Mona also constructed worksheets to be similar to the tests, and she reviewed the content of the upcoming test with the students and provided them with a study guide. She seems to adjust her teaching based on students' needs to achieve higher marks on tests, and she prepared them for the format of the midterm and final tests with worksheets and quizzes. When I asked her about this, she responded,

Excerpt 56: As you know, the nature of our assessments is mostly just paper-based. Students are very obsessive about exams. They always get stressed, and this is a way for me to revise everything with them, check if they want an extra practice; check if they need me to explain something to them, and just tell them what are the things that are included in the test. For example, there are certain terminologies or definitions that if we cover them in the midterm, then they won't come in the final, because they're a lot. (Mona, PI)

Other participants also seemed to address the challenge of low-achieving students negatively by teaching to the test to boost students' grades. For example, Dana reported this factor that affected her assessment and teaching practices; She kept reminding students about the important topics that would appear on the tests. She believed that this practice was necessary for students to do well in the tests and obtain high grades. When I asked her about the policy of testing in the department and the way it affected her practice, she replied,

Excerpt 57: The way the exams, the format, and everything, it does affect my teaching. Like, for example, I know the content of the exam; they get questions on synonyms and antonyms. So, for example, I have to teach it during a class; what is a synonym, what is an antonym. Then, I give them the new vocabulary words; I have to make sure they know this following word has an antonym, and the antonym and so forth,

and I make sure they pay attention to it because it's gonna come in the exam, but my personal beliefs, no, but I have to do this because of what's coming in the exam. (Dana, II)

The majority of the participants reported that they had to lower the level of their assessment to fit student English language proficiency levels. Yet, the analysis shows that teachers faced such a factor by making tests easy so students would pass. Most of the participants expressed their dissatisfaction at having to lower the level of their teaching to meet students' low proficiency levels. However, teachers who were against lowering the level of instruction, examination, and classroom assessment practices would cause many students to fail such courses. It is unknown what consequences a teacher in the LC would face if he or she had students with low grades in his or her classes, but the participants in this study were fearful of their students' getting low grades.

Many participants strongly believed that giving students a difficult high level of assessment would not work not only because students were at a low proficiency level but because their motivation was low as well. Leila stated, "[The students'] motivation is low, and under the factor of motivation there come many issues, such as lots of absences, and they forget to bring their books" (Leila, II). Latifa was disappointed about the situation of teaching and assessment because students are not motivated to learn English for the sake of using it in their daily practice—they take English because they have to: "I came here with high expectations, but, to be honest, every semester I'm more and more disappointed" (Latifa, II). Even in the classroom, she reported, students do not ask for feedback or participate. Latifa said,

> Excerpt 58: Unfortunately, we're talking about here in CBS. The girls are not so cooperative with us, so even if you give them something, like only two or three students are gonna come to you and say, "I have a mistake" or "I didn't know the answer," but the rest, they don't care. Even if you go over it in class, they are not motivated to learn from their mistakes. So we have a huge problem here. (Latifa, II)

Although several participants criticized the low motivation level of the students to learn English, I did not observe a variety of assessments implemented in their classes to address this

factor. For example, Leila, Nadia and Hajar depended solely on the textbook and worksheets for

assessment, and the approach they used in her activities with students was similar to the

traditional grammar translation method. Nadia and Leila only picked students who were willing

to answer, and they did not note the performance of those who were not answering. For other

activities, they simply did the tasks collaboratively with students. I observed students in their

classes were passive learners, and no communicative approach was apparent. Perhaps they taught

many students with low motivation throughout the years and they had adjusted their assessment

routines. On this point, Leila commented:

> Excerpt 59: The students are not motivated. So the student who participates is already motivated and
> willing to learn, and the one who doesn't [participate] doesn't want to learn. So I keep the assessment the
> same, and activities in the book are good; they aren't boring. Also, my way is that when I give them a topic,
> I always ask them for examples about their culture or daily life—something related to them. I ask them to
> create examples of their own, to encourage them to think and participate. Yet most of them do not
> participate. (Leila, II)

Students' low motivation to learn English influenced teachers' practices to provide

feedback. Mona said that her general practice is to give feedback to all students: "So I generally

give them feedback, like, teacher to all students. But if we're doing group work, if the group is

willing to do that, [then] sometimes I give one-to-one as well, if I'm supervising" (Mona, PI).

When asked why she did not provide one-on-one feedback, she replied that she had tried several

times to offer students one-on-one feedback by asking them to see her during office hours if they

had questions or needed detailed feedback. However, not many students took up her offer: "I

noticed that I always offered to students, I noticed that every now and then I would get one or

two students. They're honestly not that motivated, unfortunately" (Mona, PI).

As opposed to Mona, other participants, such as Dana, Nadia and Hajar provided one-on-

one feedback to raise student motivation. When asked about individual feedback, they answered

that they provided individual feedback to increase students' motivation. Hajar added that she

preferred not to show a student directly that she was wrong, ignoring the mistakes so as not to

frustrate students: "Usually I completely ignore the mistakes that the student has on the spot, but

I put it in the back of my mind and I correct it later on. I try my best not to frustrate them"

(Hajar, II). Although her students were adults, she praised them as if they were young learners,

so they would feel happy about it: "When they are doing (group tasks), I move around. I just go

to her paper and [draw] her a smiling face. Although they are adults, they love it. Sometimes I

put stars on their hands, and they love it" (Hajar, PI). I observed that the students in her class

were active, competing to provide correct answers and get her praise feedback.

The low English language proficiency level of the students influenced the way the

participants designed the tests. Hajar commented on the level of the questions she constructed in

her tests as follows:

> Excerpt 60: This is part of the triangle related to students: Their proficiency, if it's high, then the level of
> the exam is high. well, let me tell you something: Sometimes in one class, as in [GE courses], you would
> find—for example, 50 students—you would find maximum 3–4 that are good at English; that is 8%. You
> will find 15–20 who are extremely low. That is 30–40%. The rest are of average level. So what I do is I
> divide the exam; I put 10% or 5% a bit challenging, just to come up with the most distinctive students. I
> cannot give her an A if she's not an A. And I would give very relaxing questions for those who are not
> good at all, so they would feel at ease when they answer the questions. And I usually start with them, so the
> student, when she answers the questions, she knows them—she feels at ease, she doesn't feel frustrated.
> (Hajar, II)

Hajar said that she determines the level of the students by conducting a pilot assessment at the

beginning of the semester. She believed her priority in all types of assessment was evaluating the

students' proficiency level and readiness for assessment, but it seems that she focused on

simplifying assessments to allow students to achieve higher grades. This goal was manifested

through the implementation of various tasks that counted as bonus assessment tasks. Such tasks

were extra credit assignments done in the classroom. She stated that such tasks allowed her to

detect student levels, but she informed students that such tasks were graded and added to their

overall grades, which implies that her focus was raising student grades rather than their level of understanding.

5.3.1.2.2 Factors related to the class.

The other factors that influenced the participants' practices were those related to class. They include, the classroom physical layout, the timing of the class and the semester, and the student number.

The classroom's physical layout.

Some participants pointed to another major factor in reference to context: the classroom layout. The classroom layout affected the way the participants conducted various assessments. As for summative assessment, Nadia and Hajar stated that because the rooms are sometimes small and the number of students large, it was hard to monitor students during tests. Nadia said,

> Excerpt 61: We have a big issue with classes; the tables are very long from the beginning of the class till the end, and they're not separated. So at the time of the midterm, it is really difficult to separate students from each other. So I don't like to give them lots of quizzes, because I know it's difficult for me to change the classroom every time when I have a midterm or a quiz. For example, let's say, for presentations, sometimes students at the back cannot hear the students sitting in front. Because it's a very big classroom, they cannot hear the students who were speaking, or they don't pay attention actually. (Nadia, II)

Nadia did not indicate how this setting created a challenge, but it can be assumed that because of the large numbers of students at the connected tables, the students were at risk of violating the college's ethics through cheating. Although Nadia did not speak on this directly, other teachers in the same context spoke about instances in which students were caught cheating during tests. Hajar for example said that she developed two formats for the test—both included the same questions but in different orders—and this reduced cheating. She used this strategy for the GE courses because it was easier to mark the two forms of the test with the machine rather than manually, but it posed more of a challenge for ESP courses.

As for formative assessment, the participants reported that the classroom physical layout hindered teachers conducting various assessments, such as group assessment, or teachers monitoring the student performance during assessments. I did not observe the majority of the participants moving around the classroom to monitor students' performances during assessments, though it would have been difficult for some of them to do so, especially in CBS and CBE, because there was no gap between the tables and the classroom was small in size. However, Latifa addressed this factor by changing the layout of the class to facilitate the group work.

> Excerpt 62: The room plays a role—like now what I did in my writing class, first of all, there were regular seats, it was just one behind the other, so I put them in groups like five or six. I saw them participating more with each other. So I think they liked working more in groups. So the seating arrangement plays a huge role. (Latifa, II)

The group work did not work for Dana because it was difficult for her to monitor in-class group tasks. She complained about the room sizes because they were not designed to handle such a large number of students: "The room is small—it's very small [in size]. It's impossible for me to get in [among groups,] I have to go around [tables during the process of monitoring.] So things changed a bit. What I do pair work, group work! It's impossible. I can't" (Dana, PI). She mentioned that group work was not effective, and I witnessed her having difficulties moving around the room. She also said that monitoring student interaction was hard in large classes.

Course length and timing.

Another factor related to the classroom was the timing of the class and the length of the semester. In terms of the class length, Hajar mentioned that the classes are 1 hour or 1.5 hours long in the regular fall and spring semesters. She believed that this was not enough for her to conduct assessment "at students' ease" (Hajar, II). This was applicable for her summative assessment: she divided the sections of the tests based on time. For example, she said that,

depending on time, she may give the writing section on a separate day so the students could take their time on it:

> Excerpt 63: If the exam was too long—not too long, but longer than it should be—especially when it's unified, and if the time of the exam is not enough, because sometimes I have classes for 1 hour and a half, and classes for 1 hour. The students of [the 1.5-hour] class, I keep the writing exam separate, while the students of [the 1-hour class], I keep it together within the same week—as I say, to be equal. Number 2, because I find writing very important, and I need the students to write not just to finish . . . because I tell [students] that I am not testing only writing. All the things we learned—for example, the linking words, I want them to use the tenses we have explained correctly, or prepositions correctly. I don't care that much for spelling unless it's too many. I assess paragraphing and punctuation. So the students need to write beautiful writing; they need to have enough time. (Hajar, PI)

Related to the semester timing, several participants noted that the summer semester was short to include the topics in the syllabus. Mona and Najla reported that time was an issue in the summer because the course is taught for 12–14 weeks in regular semesters, while in summer it was only 7–8 weeks. This shift resulted in reducing the curriculum and excluding some content. Najla said, "The summer is a shorter period, so I cannot give as much as the normal course" (Najla, PI). Mona did not relate this factor directly to assessment, and she offered no information about how her assessment in the summer is different from that of ordinary semesters, but the number of summative assessments remained the same as in other semesters (e.g., a midterm and a final test along with three quizzes). Perhaps the content included in the tests was modified, because, as she noted, the curriculum was shorter.

Hajar compared the semester length in PAAET with courses she had taught in public high schools. Hajar said that when she was a high school teacher, she had a longer time for assessment throughout the course, so she had conducted assessment differently:

> Excerpt 64: When I taught in high school, we did continuous assessment differently. We had a long time span. So we had all the time to do all the kinds of assessment we did. For example—and this helped me a lot, this is what I'm using now—I taught advanced writing, [and] the students at the beginning of the semester, they would submit, each one of them, every single day, a piece of writing: a journal. At the end of the year, what they do is that they assess their own work at the beginning of the year. You can't imagine how much they have improved; this has helped me a lot—continuous assessment. (Hajar, II)

Indeed, Hajar specified the ESP course when reporting time constraints. She indicated that students in ESP courses needed more time to grasp new information: "The course needs more time . . . I need each one of the students to finish this course with something new that she has learned; I need more time" (Hajar, II).

Number of students.

Another factor that affects all teachers was the number of students in a course. The participants, especially those teaching in CBS and CBE, mentioned that the number of students enrolled in courses increased in the summer semester. In the CBE, for example, the number of students enrolled in each course exceeds 60 students. Dana and Hajar pointed out that the number of students increased to more than 60 (e.g., 70-80) in the summer and sometimes during specific semesters based on the total number of students admitted to the college. As a result, they agreed that it was difficult to address each individual student. Hajar stated,

> Excerpt 65: Individual assessment among students is not very effective with us. I mean, if I ask each one in the classroom, well, I would tell some of them how good they are, but knowing how big classes we have, it's very difficult to have oral assessment. Something else is that when I ask a question for 40 students in my classroom, I repeat the question. I cannot have 40 questions for 40 students; I cannot have that number of questions unless I divide the assessment for three to four lessons. (Hajar, II)

Both Dana and Hajar addressed this challenge with specific strategies; for example, Hajar asked students to place a sheet with their names on their desks. This technique helped her track and evaluate students' performance. She additionally pointed out that teachers could manage their circumstances. With a large class size, she stated, the teacher can engage the whole class based on the type of assessment she or he chooses. As such, Dana used group tasks to address this challenge but still experienced their ineffectiveness.

> Excerpt 66: As a teacher, you pick up things that sometimes your student doesn't have to say it. You can look at them, they don't understand, you can pick it up, so you can re-explain, but I can't look at ninety students right and left and they're piled up in a class. The first and second lines are normal, but the ones sitting in the back it's not fair for them, unless they speak up. (Dana, PI)

She, nevertheless, asserted that the number of students did not affect the summative assessment, because, at the end, it was one format and it could be machine scored.

The number of students also affected the types of assessments that the participants used in the classroom. Leila and Najla said that students in ESP courses were fewer than in the GE ones, thus they were able to conduct various assessments. Leila mentioned that she taught a class that had fewer than 20 students, therefore, she implemented different tasks than her usual assessment routines:

> Excerpt 67: This semester, I have only 17 students registered in a class. I apply strategies that I have never implemented when teaching big classes. The time is longer; I give them more attention and focus. Although the number is small, many of them are showing me they're willing to learn. I mean, not all the 17 students are at a good level; their grades are really low, but in class, they interact and participate, and they're all following up. So I deal with them differently. I use other methods that I never use with large classes—I deal with them differently than when I have a big number of students. (Leila, II)

Najla added that the small number of students allowed her to mark assignments in a timely manner than she could with a larger number of students.

> Excerpt 68: The number of students—for example—in this class: because it's seven students, it's great. I've been non-stop asking them [to write] "paragraphs… essay… another essay". I have time to [mark], but I am worried, because I am going to be teaching the same course in the fall, and I'm gonna have 20. (Najla, II)

She did not discuss any influence on her formative classroom assessment practices, such as monitoring, short questions, or feedback. She described only the influence in terms of paper-based assignments. She did not mention these because she did not identify formative informal methods as assessment, so she did not reflect on any in her reports.

The number of students in GE courses differed from the ESP ones. For instance, the students eligible to register for GE 099 were those who failed the English placement test set by the LC. For administrative purposes, such courses permit many students to register, so a teacher would have around 50 students. This is especially the case in CBS and CBE, whereas in other colleges the number varies. In CHS and CoN, there tend to be fewer students registered in GE

099 than in the CBS and CBE. Nadia reported that this situation creates unfairness in terms of assessment. She believes that it is unfair to measure students in her context based on standardized tests that have been constructed by the LC because teachers in other colleges tend to have smaller student numbers and often have more time to construct assessment activities. This led her to believe that informal assessments should not be graded; it is arduous, she thought, to keep track of students' individual performances in the context in which she teaches.

Another factor related to the number of the students was that the majority of participants reported or showed through their practices that they do not provide feedback to every individual student but to the class as a whole. They considered it the students' responsibility to ask for feedback if they needed it. Latifa, Nadia, Mona and Dana reflected the same feedback practices; after each task, they reminded students that they should ask for feedback if they do not understand a topic. They depended on students' oral responses to this offer. They did not seem to observe student interactions or facial expressions to detect their levels of understanding; when students gave wrong answers, they simply provided the right answer and moved on to the next point or allocated more time to re-explain the topic. Latifa and Dana elaborated they could not follow up with each individual student, which is one reason they created groups in the class. They further added that it was challenging to recall names in large classes. I did not observe any participant marking students' performance throughout assessments.

Interestingly, although Dana addressed this challenge by creating group work, she acknowledged later on in the post-observation interview that she considered abandoning the group and peer tasks because the number of students was not helpful. She had difficulty monitoring the groups and moving around in the room because of the small size of the classroom and the large number of students.

The number of students influenced teacher assessment practices, as they used several assessment and feedback for managing students' behavior. For example, several participants withheld feedback in certain situations. Leila, for example, told students to answer the homework questions along with her. When she discovered that some students did not answer the homework questions, in a very strict tone, she told students that she would not go over the exercise and provide feedback. Students in this context tend to be passive learners who accept whatever the teacher said, so they may have expected the teacher to go through each book task and answer the homework questions along with them. Therefore, when Leila told them that she would not do the exercise with them, this was a form of punishment. Leila reported that she plans homework to cover the lesson in less time in large classes. When students attended class prepared, it took her less time to cover the lesson and the exercises. She believed that feedback is effective in large classes in that students would know what was expected from them and follow the rules.

5.3.1.2.3 Factors related to the department and the Language Center.

Other factors that had an impact on most participants' assessment beliefs and practices were the assessment policies mandated by the LC, or the department especially and those were mainly the tests format mandated by the GE courses. There was a sense of dissatisfaction among participants about how the tests were constructed, especially those relating to the GE courses. It is important first to discuss the policy of the unified tests to understand the way it influenced the participants' beliefs and practices; because GE 099 is a course taught in all colleges, the LC sets the standardized test for it, although the policy of the standardized tests has been changing since the LC was established. In the first years, the policy dictated that all English departments in all colleges use a standardized test for GE courses. The testing and evaluation unit constructed the test and distributed it on the test day to all of the teachers. The teachers had no previous

knowledge of the content of the test, but they did know the format it would be. However, the

directors of the LC had temporarily discontinued this practice for unknown reasons. In the

meantime, a testing committee has been established in the LC that used a different policy. It no

longer requires all colleges to use the same tests; each department has a coordinator who is in

charge of the GE and the ESP courses and who creates the test formats. Nadia and Dana

indicated that teachers must follow what the coordinator requires for the GE courses, but there is

more freedom in the ESP courses, and a teacher can choose whether or not to follow the

coordinator's requirements.

A number of participants (mainly Nadia, Dana and Mona) said the format of the tests is

fixed in a way that restricts teachers from constructing tests based on their own beliefs. Dana

explained,

> Excerpt 69: I don't like the exams, how they're all multiple choice; how do I really know that the student understands? And I understand why they do it, because we have now like ninety students, so it makes it impossible, and every teacher teaches a certain amount of classes. The problem that we're all facing, is amount of students. (Dana, PI)

Nadia believed that the format was too difficult for students. She stated that in the

semester before this study was conducted, the coordinator asked all teachers to construct tests

using only open-ended questions that required more than one-word answers. She said that the LC

proposed that closed-ended questions are not the best way to measure students' language

competence:

> Excerpt 70: A year ago or more, for example, let's talk about the midterm and final of the GE courses; it used to be a mixture of writing, multiple-choice, and true-false questions. But last year, the coordinators of the course were encouraging us not to use multiple-choice questions. They believed that they're not a suitable way to assess the students, and even in the course description, it's not allowed for teachers to use multiple choice, and true/ false questions in their exams (midterms and finals), for both ESP and GE courses. They're trying as much as possible to limit the number of these types of question. (Nadia, II)

Nadia disagreed with this format. She believed that a test should include both closed- and

open-ended questions because both require students to think critically. She also stated that

having a test that contained only closed-ended questions is not always easy for teachers and students:

> Excerpt 71: Sometimes, multiple-choice questions are more difficult than writing questions, because in the end, the teacher is marking the writing questions. There are certain rules to correct writing, which are general rules, but in the end, each teacher could correct the writing question according to his or her teaching in class, or to his beliefs. But the multiple-choice questions are graded the same way, like sometimes it's more difficult for students to answer multiple choice, it's not always the case that multiple choice are easier for students to answer, or at least in my classes, I've noticed that they're doing much better in writing parts, which means multiple-choice is not always easy for students [to] answer. (Nadia, II)

It seems that she believed that the LC administrators wanted teachers to use more open-ended questions because closed-questions are too easy for the students and that they hope to raise students' performance by adding more open-ended questions. Nadia stated this is not necessarily the case. Both kinds of question can be challenging. She perhaps also preferred closed-ended questions because they are more convenient for teachers to mark when they have a large number of students. However, her statements described a policy that was no longer mandated. She supported the idea that the coordinator gives teachers freedom to act on their ideas and experiences around the policy of assessment and its effectiveness in the classroom.

Some participants (e.g., Leila and Nadia) criticized the policy about the number of GE quizzes required by the course coordinator. Nadia reported that the assessment policy in the department required teachers to give many quizzes in any course—a point she disagrees with:

> Excerpt 72: [The administrators] force us to give a really big number of quizzes for a GE course, which doesn't leave us much space to move on [with the curriculum], or to do our own assessments. Students need different kinds of assessment, so it's really difficult to force all classes and all teachers to use a certain type of assessment. We try to share thoughts and ideas with the LC head. Sometimes they need to try different kinds of assessment, and different ways to see the more suitable ways. But some other times, we're forced to do what they require. (Nadia, PI)

Nadia's reference here is to the policy of quizzes mandated in regular courses, not in the summer during which this study was conducted. Also, at the time of the post-observation interview with her, this policy no longer existed.

The participants acted differently upon the policy. Some were not certain whether this is a policy or if it is a norm followed by teachers in the department. This uncertainty implies that the LC's policies about testing criteria and regulations are unclear to teachers. Dana and Mona also expressed their uncertainty about the summative assessment requirements, and they reported that new teachers usually follow what previous or more experienced teachers have been doing. New teachers are willing to change, but they feel they are not encouraged by the administrators.

> Excerpt 73: We [teachers] are not really encouraged to follow our own beliefs when it comes to language testing-assessment. There is a format that is commonly shared among instructors in my college, and we do have a bit of a flexibility, but generally speaking, you have fixed sections, and even the type of questions like I wanted to change few but then I wasn't encouraged when it comes to that. (Mona, II)

Other participants, such as Najla did not criticize the assessment policy perhaps because in the college in which she taught, teachers met to construct test formats, and although they agreed on having unified tests, no teacher had to abide by the policy of unified tests. She stated that teachers had autonomy in her department and could construct their own tests. Finally, at the time of the interview, Najla held a temporary position as head of the English unit. It is possible that because of her position, she was uncomfortable criticizing something related to the LC's policy mandates.

Leila also did not view test construction, especially for GE courses, as an issue that influenced her practices. She thought that, because she had been teaching in the college for many years, she felt she did not have to abide by the standardized assessment policy. She has chaired several testing committees and was experienced with how tests are constructed and administered in the department. However, she could not change the percentage assigned for the midterm and final tests because these were set by the college and not the department. She expressed satisfaction with how tests were constructed and administered in the LC.

One of the factors that had an impact on teacher beliefs and practices and that is related to the department policy is the curriculum content. Several participants stated that they have to follow the content of the course presented in the syllabus, which mainly includes the language skills. They said that their assessment lacks some language skills, such as listening and speaking assessment. Mona, for example, said she did not include listening and speaking skills in her assessment and teaching practices. She thought that teaching these two skills was important, but students had a very low level of English—a point that could be a factor related to context as well.

> Excerpt 74: There are things that I want to do but they're not commonly known to be used. So sometimes you would feel discouraged […] For example, the listening, I would love to teach more listening; it's very important, but the general policy does not encourage that, because we don't have labs. Even the teaching books, we don't have those specifically designed for listening (Mona, II).

"The general policy" can be interpreted in two ways. On one hand, the PAAET did not provide laboratories for the English unit at CHS. On the other hand, the English unit did not update the course materials. However, Mona said that CHS was planned to move to a new campus for the next semester. It is unclear whether the new campus includes labs for the English unit. Moreover, because the language focus was mainly on vocabulary, Mona constructed her quizzes to only cover vocabulary. She also concentrated on vocabulary on the tests; she gave the same weight to both reading and vocabulary (15%).

Hajar and Leila also commented on the curriculum saying that the books were at higher level that the actual student proficiency level. Hajar, for instance, said the curriculum is important to assessment because when she assessed students, she took three factors into consideration in designing her assessment: the level of the students, the level of the book, and the level of the course. She stated, for example, that the level of the book in the GE course does not reflect the actual level of students. She considered the book higher than their level and said that

books should be assigned after careful examination of students' English language proficiency

levels:

> Excerpt 75: Dealing with students teaches you a lot; you learn from them. This is my 34th year of teaching, so I have been exposed to different kinds of levels of students. And I have known if you give them a test that is assessing a higher level than theirs, they're gonna reflect what you have given them. So what I do is the following: When I put the exam, I bear in mind three things: the level of the course, if it's remedial course 1 or 2 or ESP; the level of the students, which I would have covered through the assessments I have given them before the midterm; and the level of the book they are learning, because I cannot give them a very easy exam while they are an ESP student with a very challenging book [. . .] It has to be divided among these three factors: the level of the students, the level of the course they're learning, and the level of the book itself. Sometimes we change the book if we found it too easy or too difficult. (Hajar, PI)

If the results of the tests showed that most students did not achieve well, she believed that the

assessment itself may be invalid. Hajar was not clear about how she would use pilot assessment

to measure student level. The course objective was to promote student skills in various language

aspects, but Hajar seems to anticipate student level to be low and acted accordingly. She

mentioned that if students were not assessed according to their level, they would not be able to

comprehend the course or and therefore achieve the curriculum objectives. At the end of the

quote above, she used the pronoun "we"; it was not clear whether she was referring to teachers in

the department, members of the curriculum committee, or teachers in the LC in general. In other

parts of the interview, she criticized some LC teachers who think that, to improve students'

levels, assessments must be given at a higher level than the actual student English proficiency

level.

In another instance, Leila reported that when learning materials are simple, students can

comprehend topics better and learn more than when they are stressed because of the curriculum

load and high difficulty level of tests:

> Excerpt 76: The student level is very low. I can't raise the level. They would not feel comfortable with me if I do so. I mean according to me—I don't lower my level of assessment to their actual level. I mean, I just simplify the subject so that when they pass the course, they have acquired some information and understood it. But if I had students with high proficiency level, I would have to increase the level to fit their level and culture. But for our students, I give them according to their level. (Leila, II)

Dana also assessed the curriculum as "outdated." The course she taught was library science, and part of it involved topics concerning technology, but the topics in the reading kit included some tools that were not used anymore (e.g., floppy disks). Thus, she made her own reading kit and used supplementary materials, choosing not to depend on the book. Dana stated that she had expressed her opinions in meetings regarding curriculum development, but no changes had yet been made.

5.3.1.2.4 Factors related to the institution (PAAET) assessment policy.

A number of participants believed that the assessment policy of the institution is the main factor that shaped their teaching practices in general and their assessment practices in particular. The main factor related to the PAAET policy was the summative assessment grading system. The majority of the participants criticized the grading system required by the PAAET and the department saying that summative assessment is assigned the highest grades. As such, the participants relied heavily on tests and quizzes when evaluating students' performance.

Many participants believed that relying on tests was not ideal, and it influenced their assessments. They preferred to have a higher percentage for formative assessment rather than tests and quizzes, but the grading policy did not provide much freedom for teachers. Hajar explained:

> Excerpt 77: It is improper to have the final out of 50 [when a student] has worked all the semester for half of the grade and only one exam would affect her grades. She could have been sick, she could have been, for example, absentminded with a problem, and she has lost 50%, which I don't like. I would love if I would give the final only 20. And I would give the continuous assessment 50: 30 for [the] midterm, 20 for the final, 50 for the continuous assessment. (Hajar, PI)

Latifa agreed adding that:

> Excerpt 78: 50% is a lot for the final. I know some girls, they do very good in the midterm and the quizzes, but in the final—like I don't know what are their circumstances, but you might see that an A student does not perform well in the final, so her grade would drop back to C. So I wouldn't prefer to have the final out of 50. It should be the teachers' own rubrics. The final is important, but the marks are too much for it. (Latifa, PI)

The LC mandated assigning 25–30% to midterms and the PAAET policy mandated assigning 40–50% to the final test. Teachers were required by the LC to give students quizzes, though PAAET did not mandate a specific percentage for that component. This grading policy by both the PAAET and the LC left limited room for teachers to design their own rubrics or grading criteria. The participants were not convinced about the percentage that formative assessment contributes to the final grade (20%). Based on interviews and classroom observations, it appeared that the participants did not use any specific rubrics and depended on their own holistic evaluations of students. I did not observe any teacher actually mark any tasks done in the classroom. This grading factor led many participants to teach to the test, or conduct extra assessment to boost student grades if students needed help, but this intention was not stated explicitly.

Most participants showed their desire for students to get high grades as they shared a common sense of guilt about the high test grading criteria. Many teachers, such as Hajar and Latifa, decided to add bonus marks to address this challenge. The bonus tasks were too easy, so the purpose of such assessments was clearly to help students pass the courses with satisfactory grades. Hajar gave several bonus assessments for students to raise their marks if any of the factors she mentioned occurred (e.g., stress, psychological issues). She further added that she considered students' level of engagement in assessment tasks when determining students' overall grades. Other Participants, such as Leila, Mona, and Dana gave the full 10% and Najla gave 5% for attendance and participation in both GE and ESP courses. They seem to give all students full marks to help them achieve higher overall grades to pass the course, because they had large classes with more than 50 students, and it was obvious that it was challenging to mark each

student's performance or participation during class. I also did not observe any participants marking student performance during the classroom assessments.

The grading policy influenced other participants' practices, such as Najla and Dana, by having them familiarizing students with the tests. Such a practice was evident through the feedback practices. The majority of the participants drew student's attention to aspects from the worksheet or textbook that would be on the test. They reminded students that the point they were explaining would appear on the test. Dana, for instance, repeated many times that "this would be on the test," or "this is how you would answer on the test." Hajar also spent more time on scaffolding and feedback practices in her ESP class when topics would be included on the upcoming tests.

During Najla's classroom observation, when students answered the worksheet, she reminded them that the test would contain similar questions. She reported that her intention of such a practice was to enhance student learning, but admitted that students in her context cared more about their grades than about learning. It seemed that she had adjusted to this reality and had tailored her practices to this demand. I also observed that when she handed back the midterm test and offered some feedback, she expressed dissatisfaction with students' low grades on the test and therefore, she decided to give a makeup quiz so that they could raise their marks. Najla said, "If they do well with the first quiz, then I don't give a second quiz. But if they don't do well, I could do up to three quizzes" (Najla, PI). Najla also revisited the aspects that they had not yet mastered, and she instructed learners on what would be expected of them for the upcoming test. She added: "I do know what's coming in the exam; I will at least [want that to] stick in their head" (Najla, PI). Najla's response indicated that she wanted students to achieve high grades on their tests.

Some participants said that the administrative admission policy contributed to having many students with low academic level. Leila, for example, said that students admitted to CBS are those who have low GPA in high school and whose English proficiency level is low. Moreover, Leila and Hajar criticized the college administration practices, saying that the administrators were lenient with students. They believed that administrators did not support strict teachers who failed students. They, similar to other participants, stated that they should pass many students, because the administrators plan every year to admit a large number of freshmen. If students fail the course, their graduation would be delayed. Moreover, because English does not fall under one of the major courses in the college and the LC is a supplementary department, they believed that administrators treat the LC as peripheral. The administrators, they thought, do not want to have strict teachers in supplementary departments because they think students should focus on their main majors.

Another factor related to the administration factor was reported by Latifa. She said teachers may feel pressured to assist students in obtaining good grades because there is a form provided to teachers when they finish each semester on which they report their grades. This form is completed for statistical purposes, but it is submitted under the name of the instructor. Also, during each semester, students are required to evaluate the instructor, and instructors with good reviews are acknowledged. (However, no instructor is penalized for low evaluations.) These factors could comprise an additional explanation for why teachers attempted to increase students' grades.

5.3.1.2.5 Factors related to the culture and country.

As can be seen above about the factors related to students, almost all participants believed that students in this context usually had to focus on paper-based tests rather than

informal classroom assessment. The participants attributed this factor to the English assessment policy in public schools criticizing also the curriculum content in schools and the English teaching methods. Mona, for example, said that students in public schools are used to not only summative paper-based assessment but also to the teacher's role being the main source of teaching and assessment. She repeated several times that people in her context are "obsessed" with paper-based assessments.

> Excerpt 79: I think the fact that most of our students are graduates from the public schools in the country, they are used to very traditional way of seating, assessment, teaching. They're used to the teacher sitting on a chair, with them facing that teacher and basically them being spoon-fed the information. They're not used to having the teacher as guidance and […] they're not used to other things such as informal assessments. So sometimes even if you want to try to introduce some of those concepts, they won't take it as serious, and they would think that that's not fair, or [they] want to have a proof of it. (Mona, II)

Based on the classroom observations I conducted, although some participants, including Mona, did employ various methods of informal assessment, they did not grade any of them. This strategy could be related to the way she looked at the context.

Other participants believed that their assessment practices would vary depending on the institution in the country. Several participants such as Leila and Dana reported that if the context were a university, they would have changed their assessment practices. Mona reported that she did practice assessment differently when she taught in a private university. She considered her experience at the English private college to have been better than her experience at LC with reference to classroom informal assessment. At the private college, teachers planned various types of formative and summative assessments that made up the students' overall grades. Mona reported that this factor influenced her beliefs. That is, her practices in the LC contradicted her beliefs. She did not mention assessment specifically on this point:

> Excerpt 80: My beliefs are, I would say, more diverse than my practice, but I think because I've taught at the LC, I had to adopt some of my own beliefs. Some of my practices do not reflect my beliefs because they simply work with students' [demands]. (Mona, II)

The other factors that attributed to the low English language proficiency level is, according to several participants, the status of English in some regions of the country. Latifa and Dana believed that students who come from certain regions in the country, have a poor English because it is not spoken or used in their communities. These factors influenced their practices because it necessitated their use of L1 in assessments. Neither believed this was effective, but they said it was necessary for them to understand what was required in assessment. Dana said that she tried to use English in her first year of teaching, but she thought that many students did not understand. Dana stated,

> Excerpt 81: The challenges unfortunately, most of the students that I taught do not speak English fluently—not fluently I mean they don't practice the language. So when it comes to reading, they have a hard time, even though they've been learning English from the first grade. So lot of them don't know how to form sentences. My biggest challenge is, when I ask them something they basically read sentences from the book, they don't know how to form a sentence. That's one of the challenges I'm facing. So I have to translate sometimes, actually the majority of times, I have to translate. (Dana, II)

Leila and Najla added that most of the students in this institution came from specific regions, which she believed have low education levels, especially in English. Leila reported that, because the college accepts students with low high school GPAs, most students from such regions apply to CBS. In this context, she believed that she lowered the level of her assessment to fit students' actual proficiency levels. She thought that students in other universities in the country were better academically and that if she taught in a local university, she would increase the strictness of her assessment approach.

Another factor related to the culture, according to some participants, was that students felt shy or embarrassed to take a role in the classroom assessment tasks. Latifa and Nadia, for example, reported that students in their classes were shy and did not participate because they were not confident that they would provide the correct answers. Latifa stated,

Excerpt 82: I tell them, "It's OK if you don't know. You might learn from your own mistakes, and that we're here to learn, so don't be afraid to ask." They're very hesitant to ask, or they're embarrassed maybe, they're shy. So I'm trying to make it a friendly atmosphere, but they're still hesitant. (Latifa, II)

A factor related to the culture, and that was reported by few participants, was that the semester in which this study was conducted was during the month of Ramadan. In this context, teachers usually try to be flexible in planning curriculum and assessments during this month, given that most students fast and are consequently believed to be less energetic than in other months. Mona's classes, for example, fell within the fasting period, and some classes fell near break-fast (i.e., *Iftar*) hours. She said that she took this situation into consideration and therefore she taught the most important course content in the 2 weeks of summer preceding Ramadan.

Excerpt 83: I think I just had to focus on the main points, and I think in planning. I just planned it better as I focused on the two weeks before Ramadan, so I gave like a really good proportion of the curriculum, and I even made them attend few extra hours. I've already informed them. I said "look you know I think it's better if we cover like a large proportion of the curriculum," and that because once Ramadan starts—I mean it's not gonna be easy for me to teach like two full hours every single day with students' complete attention. They won't be able to do that with Ramadan. I mean without Ramadan it would be OK, but with Ramadan, we all know it, it's not possible. (Mona, II)

Because I was not able to attend her classes before Ramadan because of my own schedule, I could not draw a comparison between her assessment plans before and during Ramadan.

Several participants spoke about the context in general without referring to whether they meant the student, institution or the country. The context had a major impact on their beliefs. For instance, in Excerpt 43, Dana's perception about her context was negative and different from what she experienced before as a student and as a teacher. Based on her few years of experience and the small number of colleges at which she had taught, she may have constructed a stereotype concerning students' proficiency levels and that perhaps her assessment practices that focused upon summative assessment and the importance of grades were also affected by her context. She would likely have different beliefs in a different context, yet she did not report any shift in vision. She stated that, "I was more of the one who's talking and explaining, and I realized it

doesn't do anything," but, in her practice, Dana attempted to address this challenge by forming

groups in most of her sessions. Still, I noticed that it was challenging for Dana to monitor all the

groups and provide feedback. Interestingly, when I played the segments from her classes in the

post-observation interview, she commented,

> Excerpt 84: Listening to [the segments] again, I'm not actually happy [laughs]. I am glad that I do the
> handouts, because it's different. Like at least I get to do what I believe in and what I studied, but no, I hate
> the context. I don't think this is gonna be beneficial for them when they graduate. (Dana, PI)

Her statement and expression (i.e., laughter) reflected a sense of sarcasm that confirms her

pessimistic view of this learning context. It can also be inferred that the above-stated beliefs and

practices were mostly limited to this context. It is clear that if Dana were interviewed and

observed in another situation, she would give different replies. She nevertheless expressed her

preference for assessments that are paper-based, when she referred to "handouts," as a practice

that she learned, and she believes is effective.

CHAPTER 6:

DISCUSSION

This chapter relates the study findings to previous studies on language classroom

assessment and teacher assessment beliefs and practices to gain better insights and draw

conclusions about the field. The discussion is presented based on the research questions, which

investigate the classroom assessment practices and beliefs of seven EFL teachers teaching GE

and ESP courses, as well as the influence of various contextual factors on teachers' assessment

beliefs and practices. Next, following a discussion of the study limitations, this chapter presents

implications for further studies in the area of L2 classroom assessment, and teacher assessment

beliefs and practices, as well as implications for teachers and policymakers.

6.1 Summary and Discussion of Findings

This section relates the main findings to the literature. It first discusses the findings

concerning teachers' assessment practices, followed by findings on teachers' assessment beliefs

in relation to previous studies. Then, I relate findings concerning the influence of the various

internal and external factors on teachers' assessment beliefs and practice to previous research.

6.1.1 Teacher Classroom Assessment Practices

One of the main findings in this study is that, based on classroom observation, almost all

the participants implemented a wide range of classroom assessments for both formative and

summative purposes unlike the limited range of assessment practices they reported in the

interviews. The participants did not identify many of the formative assessments they conducted

in the classroom as assessments. In the post-observation interviews, after listening to the

recorded segments, the participants reported that the purposes of such tasks were to track student

understanding, elicit information to move learning forward, and/or assess their own teaching

methods. As a researcher, I often was able to identify specific classroom assessment practices based on my own understanding of assessment practices as described and discussed in the classroom assessment literature. The formative assessment methods the participants employed included asking short-answer questions, using worksheets for group work, and self- and peer assessment. The participants often reported that they did not plan the assessments that occurred during instruction.

The participants' formative assessments can be viewed from a social constructivist perspective, in which classroom assessment involves contextual, cultural, and social dimensions involving both teachers and students (Lund, 2008), as well as the cognitive perspective of both the learners and the teachers (Shepard, 2000). From a cognitive perspective, the participants in this study seemed to shape their assessments based on their beliefs, but student needs and the cultural and social factors surrounding them affected their practices. Thus, the social constructivist perspective aided the understanding of the participants' assessment practices with the context being the main factor (Shepard, 2000).

Findings regarding the participants' classroom assessment methods in this study are consistent with the studies discussed in the literature review chapter, especially with what Rea-Dickins (2001), Hill and McNamara (2012), and Hill (2017) represented in their models about the complexity of teacher assessment practices, which involve a range of assessments conducted for formative and summative purposes. Such assessments, as the literature reveals, involve the teacher as the main agent and learners as passive participants in the assessment process (Rea-Dickins, 2004).

In Rea-Dickins's (2001) model (see Figure 2.1 in Chapter 2), teachers' assessment practices involve four main stages. The first stage is planning. In this study, because the majority

of the participants could not identify their classroom practices as formative assessments, they did not show any deliberate plans for their classroom assessments. Rea-Dickins (2001) stated that the planning stage involves identifying the purpose for assessment, and then choosing and preparing the assessment tasks. The majority of the participants said they do not plan any assessment conducted in the classroom. Only a few participants (e.g., Najla and Dana) said, for example, they constructed worksheets ahead of the start of the semester and used them in the classroom without specific planning in mind. Such assessments, as the Findings Chapter shows, were conducted using group work and asking short-answer questions. One class used peer assessment.

The reasons behind teachers not planning their classroom assessment could be attributed to various factors. Data show that the participants did not receive any professional development programs in the field of assessment. They did not identify the characteristics of formative assessment and, as such, may not plan formative assessment to achieve its goals. As Bennett (2011) stated, formative assessment is complex and its definition has been redefined in the literature for many years. The majority of participants often reported the purpose of the formative assessment they used, but they did not recognize some of the formative evaluations they used during classroom observations as assessments. Several studies (e.g., Bennett, 2011; Brookhart et al., 2010; Hill, 2017) emphasized the importance of teacher professional development programs in assessment to sustain understanding of the domain and showed that assessment literacy has a major influence on teacher beliefs and practices. Because the participants in this study did not receive such training, they would not be able to plan assessments to achieve formative purposes. I observed that all the participants implemented formative assessments, but most of their assessments involved asking short-answer questions, providing feedback, and monitoring student performance. Such methods do not require

deliberate planning but can be embedded within the teaching (Hill & McNamara, 2012), which also fulfills one of the purposes of formative assessment in the literature.

The participants may have unintentionally planned classroom assessment to serve summative purposes. For example, some participants (e.g., Najla, Hajar and Mona), said they plan worksheets to resemble the format of the final tests, and that they aim to prepare students for the tests in the classroom through those worksheets. Rea-Dickins and Gardner's (2000) study investigated teacher practices in EAL school context. They found teachers were aware that formative assessment could be used for summative purposes or vice versa. Formative assessments could, for example, inform decision making for high-stakes testing. However, in this study, the participants did not indicate that they were aware that they could use worksheets for summative purposes, and they did not use the worksheets for decision making. This finding also aligns with Hill and McNamara's (2012) assessment model that stated that teachers could collect evidence from learning activities and urged researchers to observe whether such activities were planned or incidental. The incidental classroom assessments in this study were using short-answer questions, providing feedback, and reminding students that certain exercises would appear on the test. Conversely, planned classroom assessment consisted of worksheets and textbook exercises.

The second stage of assessment in Rea-Dickins's (2001) model states how teachers implement and monitor assessment, and scaffold learning. This stage appeared in Hill and McNamara's (2012) model as well, which they referred to as framing assessment. The participants in this study implemented various assessment methods in the classroom. The most common methods were asking short-answer questions. The participants were still not aware that they were assessing students; they said they did it to track student understanding. The other

common methods were textbook and worksheet exercises, in which I observed the participants going around the room, monitoring student work and providing feedback to learners. The majority of the participants, however, did not introduce the methods clearly to their students or elaborate on their classroom practices in the post-observation interviews. For example, I observed peer assessment and group work, but when I asked some participants whether they implemented this strategy in a specific way, they replied that they did this at the moment of teaching, and that they did not plan or think about implementing certain assessments in a specific way.

Most of the classroom assessments aligned with what Hill and McNamara (2012) called instruction-embedded assessment. Conducting assessment, based on Hill and McNamara's (2012) model, ranges from explicit, planned assessment to less planned, instruction-embedded assessment. They described the assessment conducted in the classroom as planned and evident when used in teaching activities, and unplanned (instruction-embedded) when used in unstructured observation. During my observation, the participants who used worksheets as an assessment method said they planned this activity, but they did not prepare other methods, such as asking oral questions.

The findings show that the majority of the participants demonstrated the second and third stages of the models of Rea-Dickins (2001) and Hill and McNamara (2012), which is implementation and includes monitoring and scaffolding during assessment to aid decision making. The findings show that when students were engaged in certain tasks, some participants monitored student progress. Yet, some participants reported that they monitored student performance for the sake of managing class time or to provide feedback.

Moreover, almost all participants demonstrated feedback practices posited by previous studies (e.g., Brookhart, 2007; Davison & Leung, 2009; Hill & McNamara, 2012). Hill and McNamara (2012) stated that feedback can be divided into two categories: person-referenced (e.g., approval, disapproval, and punishment) and task-referenced. Task-referenced feedback relates to the quality of a student's performance on a specific task. Task-referenced feedback may be explanatory, corrective, or confirmatory. Explanatory feedback emphasizes positive performance (e.g., saying, "very good" or praising students' interactions). Corrective feedback involves further illustration of a specific point. Teachers give confirmatory feedback, the third type of task-referenced feedback, following a correct response. It can take the form of repeating correct answers, nodding, or making checks on a student's paper. I observed that the majority of the participants used all three types of task-referenced feedback, which is a basic element of assessment (Hill, 2017; Rea-Dickins, 2001; Hill & McNamara, 2012). The explanatory feedback during observation was typically praise when students gave correct answers. Participants also provided confirmatory feedback in similar ways, such as repeating the correct answer when students provided it. They also used physical gestures (e.g., nodding) to reinforce correct answers. Teachers also provided corrective feedback when students gave wrong answers, asking the same question again and seeking the correct answer. Sometimes, when a task required an individual answer from a student, they corrected the individual directly by indicating that the student's answer was incorrect and providing the correct answer (i.e., person-referenced feedback).

Although the participants used feedback, the majority did not identify it in the interviews. They integrated feedback within their teaching practices, which aligns with Kunnan and Jang (2009), who stated that feedback is embedded within assessment to inform teachers on

instruction and student performance. However, several participants in this study reported that they use summative assessment to inform instruction and to monitor student progress. Kunnan and Jang (2009) also referred to this point, indicating that in summative tests, teachers use feedback to address the gap between course objectives and student level. The feedback in this case is for students, parents, and administrators (Kunnan & Jang, 2009). It is important to mention here that in this study, the participants relied heavily on summative assessments, and this is perhaps why they used feedback in summative assessment to inform their instruction and student progress. According to Gipps (1994), feedback in summative assessments is different from formative assessment, and feedback within instruction contributes to learning and teaching progress and therefore to formative assessment. And because the assessment policy affected the majority of the participants' practices, they relied on formal feedback and did not perceive their classroom feedback practices as an element of their formative assessment, as the literature suggests (Gipps, 1994; Leung & Mohan, 2004). However, Moss (2003) noted that standardized feedback, such as in the form of grading and commenting, is necessary in analyzing student performance.

Also, data show that some teachers scaffolded learners especially through the use of short-answer questions. The purpose of scaffolding aligns with the literature because participants who used scaffolding said they did so to help students complete certain tasks. Scaffolding is when a teacher provides support to learners to complete a given task that learners would otherwise not be able to achieve (Bataineh & Obeiah, 2016; Van de Pol, Volman, & Beishuizen, 2010), and such support could be in the form of questions, hints, reminders, and/or encouragement (Shepard, 2006).

The last stage of Rea-Dickins's (2001) model (grading) seems to shape the participants' main assessment practices. According to most of the participants, the main purpose of their classroom assessments was to assign grades and for reporting and dissemination purposes, and this was done through conducting summative assessments. These summative assessments carried the highest weight in students' final grades. As a result, these summative assessments were more structured and planned than formative assessments, and used paper-based tests and assignments almost exclusively. The participants also indicated that grading students based on summative tests aligns with the institution's policies, which mandate that the final tests account for 50% of the overall grade. As for grading practices, almost all the participants followed the same grading policy and included the same evaluation methods (e.g., 25-30% for the midterm test, 50% for the final test, and 10% for quizzes). The last stage also seems to have an impact on their beliefs and practices, as discussed in the following sections.

One of the findings related to teachers' assessment practices was using classroom assessment for the purpose of classroom management. Brookhart (2004) noted that classroom assessment exists at the intersection of three teaching functions: instruction, assessment, and classroom management. Assessments employed for classroom management occurred in various forms in the current study. For example, several participants indicated that they use assessment and grades to manage student behavior such as getting students to participate in the classroom and/or complete assigned work to gain additional marks. That is, some participants (e.g., Nadia) used short-answer questions when they noticed students were not active in the lesson, and this assessment was used to manage students' behavior in the classroom. Other participants (e.g., Hajar) used extra-credit worksheets in her large class to motivate students to participate in the assessment tasks because, as she reported, students in this context rely on graded assessments.

Hill and McNamara (2012) also pointed out that assessment could be used for the purpose of classroom management. In this study, the participants tried to regulate student behavior and motivated students take part in the assessment process by announcing that certain assessments were graded. On the other hand, some participants used assessment for lesson management by using oral questions to move the lesson forward and save time in the classroom.

6.1.2 Teacher Assessment Beliefs

Similar to teacher practices, the investigation of teachers' assessment beliefs in this study utilized a constructivist approach. Recent studies postulated that teacher beliefs are best understood within a constructivist approach, which involves several external factors such as schooling and instructional experiences, knowledge, and personal experience (Borg, 2006). Many studies have shown that such factors affect teacher assessment beliefs and their implementation of assessment (Al-Sawafi, 2014; Cheng et al., 2004; Mansory, 2016; Yin, 2005).

One key finding regarding teachers' assessment beliefs was that the participants tended to provide partial definitions of "assessment." I divided the participants into two groups: those who defined assessment by listing its purposes (i.e., promoting student understanding and assessing teaching efficacy) and those who defined assessment as a method (e.g., paper-based). The participants' partial definitions of the concept of *assessment* in terms of purpose or method could have several causes. First is the variability in the definitions of assessment presented in the literature. As discussed in Chapter 2, scholars have defined assessment in many different ways, though all refer to the process of collecting information for decision making (Bachman & Palmer, 2010). However, although assessment conveys two purposes—formative and summative—the participants seemed to identify summative assessment more clearly than formative assessment. The participants in this study identified the purposes of formative

assessment, but the majority believed collecting information for decision making is done mainly

through tests. In the assessment literature, summative assessment has been clearly linked to

behaviorism, whereas formative assessment has been redefined over the years (Bennett, 2011;

Yorke, 2003). Hence it is important for teachers to receive assessment training to better

understand this concept (Bennett, 2011). This finding aligns with previous studies on educational

assessment that focus primarily on summative assessment, and, as a result, the distinction

between summative and formative assessment is sometimes equivocal (McMillan, 2013; Rea-

Dickins, 2007; Teasdale & Leung, 2000). Because the participants in this study did not engage in

any training courses or education in assessment, the distinction between formative and

summative remains unclear for many of them.

Almost all participants were able to report their understanding of the purposes of

assessment. They believed the purpose of assessment is to check students' understanding and, for

some participants, to assess their own teaching methods. Their beliefs were consistent with the

purpose of formative assessment in the literature, which identifies formative assessment as

central to providing teachers with information about learners' progress and to assess their own

strategies (Hill & McNamara, 2012; Katz & Gottlieb, 2012). I observed that the participants in

this study may have a belief system that Phipps and Borg (2009) referred to as core and

peripheral beliefs. Phipps and Borg (2009) stated that beliefs can be divided into core and

peripheral. Core beliefs are "stable and exert a more powerful influence on behaviour than

peripheral beliefs" (p. 381). Fives and Buehl (2014) added that core and peripheral beliefs "are

organized in systems based on the centrality of the belief to the person's sense of self or self-

concept" (p. 435). The core beliefs of the participants in this study seem to relate to beliefs about

valid student-evaluation methods through summative assessment, while peripheral beliefs seem

to relate to their perceptions about the necessity to implement formative assessment. As the findings show, although all participants implemented various formative assessment methods (peripheral beliefs), several participants said they were not graded, and that the only way to measure student performance was through paper-based assignments and/or tests (core beliefs).

Moreover, the participants' lack of assessment literacy was clear in the first three stages of Rea-Dickins's (2001) model. The first three stages (i.e., planning, implementation and monitoring) involve teacher conceptualization of assessment alongside their practices as discussed above. These stages include the categories of beliefs that Yin (2010) referred to as strategic cognition (e.g., teacher beliefs about the test policy requirements, class parameters, syllabus and course time), and interactive cognition (i.e., teacher beliefs about assessment, the level of students, and stereotype). In this study, teachers reflected on both categories of beliefs within the three stages of assessment practices in Rea-Dickins's (2001) model. Strategic cognition seems to dominate in the implementation phase, while interactive cognition seems to prevail in the monitoring and planning processes. The three stages overlapped in the classroom, but it seems that during the implementation process, the participants' conceptualization of students' proficiency level shaped their assessment practices—a point discussed in the following section.

Teachers' conceptualization about students' L2 proficiency levels demonstrated teacher interactive cognitions in this study. Most of the participants believed that the students' English language proficiency levels and motivation to learn English were low. As such, they believed students were not at a level to complete various formative assessments. For example, several participants stated they do not believe students could conduct peer assessment by themselves. The findings indicate that participants who employed peer assessment supervised and directed

the process. Teachers played a central role as assessors. This observation contradicts how the literature defines peer assessment. In the literature, peer assessment aim to develop students' cognitive skills through reasoning and to help them improve relevant skills enhanced by social interaction (Cheng & Warren, 2005). It should only assist teachers in assessing students' efforts within group work (Matsuno, 2009), but this was not clear in the classrooms in this study. This suggests that the participants' conceptualization and approaches to assessment can affect student engagement and confirms teachers' perceptions that students depend on summative assessment.

Moreover, interactive cognition seemed to lead participants to focus on certain language aspects they found necessary for students to learn in this context. According to Yin (2010), interactive cognition could result in teachers being influenced by specific language constructs. Most of the participants believed in grammar as the main focus of instruction. Based on the analysis of their tests, most participants constructed quizzes only on grammar. They also perceived vocabulary as necessary. The participants' beliefs about these specific language aspects made them stick to the old, traditional summative assessment design. For example, they divided tests into separate sections, each on a different language aspect (e.g., a section on reading, a section on grammar). Studies on assessment have stated that, in the past, tests included separate items and measured student understanding by the number of correct answers (Fox, 2013).

Yorke (2003) pointed out that the focus on summative tests creates pressure in higher education, which could result in assessments shaped to fit a behaviorist approach rather than a social constructivist approach. All the participants' beliefs were consistent with Brookhart's (2007) description of formative assessment, in that its main purpose is to provide teachers with information for pedagogical decisions and give students information to enhance their learning.

However, when it comes to their practices, most of the participants implemented various assessment methods but only recognized and counted those assessments that were conducted for summative purposes (e.g., paper-based assignments and tests). These findings seem to be linked to the psychometric theory prevalent in the 1990s. A few participants, such as Latifa and Najla, reported that summative assessments provided valid means to measure student performance, based on test scores. In this study, teachers seemed to follow this approach only because the policy required it. It also seemed to be the approach they applied due to their lack of assessment training. Previous studies stated that this approach was predominant in the 1990s because there was little focus on teacher assessment practices. The situation, however, in the study seemed to be the contrary. The teacher's role was central, and almost all participants reported they do have active roles in constructing classroom assessments and no standard policy exists, especially in ESP courses.

The participants in this study believed the focus should be more on formative assessment than on summative assessment, but the pressures of the summative assessment demands seem to lead to the dominance of a behaviorist paradigm in this context. The following section discusses factors that shaped the participants' assessment beliefs and practices in relation to the literature.

6.1.3 Factors that Influenced Teacher Assessment Beliefs and Practices

Several internal and external factors influenced the participants' assessment beliefs and practices. Internal factors included their education (i.e., schooling and professional development programs) and teaching experiences, while the external factors were context-related and include those concerning students, the class setting, and the administrative policies of the LC and the institution, in addition to broader cultural factors. This is common in studies of teacher beliefs and practices. Borg (2006) stated that various factors could affect teachers' beliefs (see Figure

2.2. Chapter 2), which could include experience, knowledge, the nature of courses, the institutional policies, and the teaching and learning environment, including the local culture (Cheng et al., 2004; Davison, 2004; Xu & Liu, 2009; Zhang & Burry-Stock, 2003).

In this study, the participants reported that several factors affected their assessment beliefs and practices. Internal factors included prior teaching experience and schooling. Previous studies have emphasized the role of teacher experience in their practices (Borg, 2006; Fives & Buehl, 2012; Nespor, 1986), including prior education. Several participants in this study (e.g., Mona, Latifa and Najla) recalled the way their teachers assessed them when they were students and said that the psychometric methods affected their understanding of assessment and the way they assessed their own students. As for their understanding of assessment, almost all the participants equated assessment with tests based on how their own teachers assessed them, or perhaps because this is the prevailing conceptualization of assessment among teachers, administrators, and students in this specific context. The main assessment methods in public schools, many participants recalled, were summative tests. Although they believed that such methods may be outdated and should not be implemented in the classroom, many participants used these methods perhaps because they felt that such assessment practices are still prevalent in the schools in which they (and their students) received their education. These participants also tended to believe that their students are not used to other forms of assessment, such as peer assessment, as indicated above. This is consistent with Borg's (2006) view that past events and school experience could significantly influence teacher practices.

In terms of experience, the study found that, unlike newer teachers, participants with more teaching experience tended to believe that their assessment practices are effective. As a result, the experienced teachers seemed less willing to change their assessment methods. I

grouped the participants based on their views of the influence of teaching experience on their assessment beliefs and practices. The first group, the experienced participants (Hajar and Leila), believed that their teaching experience had taught them a lot about assessment, so their assessment practices already fit their context. Leila, for example, considered her practices valid, and thought she did not need to attend any professional development programs. Her attitude toward assessment confirms Sikes's (2013; as cited in Al-Sawafi, 2014) statement that more experienced teachers will reject assessment reforms if they think their practices are correct based on their own experience. Those experienced participants did not show willingness to change their assessment practices. This could be due to several external factors, such as context and the absence of professional development programs—a point explained in the following sections.

In contrast, the participants with less teaching experience believed that they need to gain more experience with assessment methods and training in the field of assessment in general in order to improve their assessment practices, perhaps through participating in professional development programs. This point is consistent with the findings of Sahinkarakas (2012), who conducted a study among 100 participants comparing preservice and experienced teachers. She found that preservice teachers were more motivated and scored higher in self-efficacy. She argued that experienced teachers might lose their motivation for increasing teaching effectiveness because of their long experience.

One main finding related to the internal factors in this study was that the teachers did not receive training in assessment. All the participants did have degrees in teaching English but none of them participated in any professional development program or studied assessment prior to teaching. This lack of training influenced the participants' knowledge and beliefs, as can be seen in the way they defined and viewed assessment. In terms of the effects of the lack of assessment

training on their assessment practices, some participants stated that they did not follow any standard framework or receive any training in constructing assessments; many often explicitly acknowledged their need for training in assessment, especially in formative assessment. Yorke (2003) supported this point, arguing that in a higher-education context, it is important for teachers to develop their disciplinary and assessment knowledge repertoire by pursuing professional development programs to learn about student progress and adjust teaching strategies accordingly.

It is important to note here that none of the participants explained why they did not receive any training or professional development in assessment. This could be due to financial reasons related to the budget of the institution, and/or their choice not to attend workshops on this topic. There is no evidence that budget constraints limit the ability of the LC to offer professional development. However, the PAAET in general, and the LC in particular, do have a budget for teachers to attend conferences. They also host some professional development programs and announce off-campus conferences or programs to teacher regularly. Although they do not provide funds for hosting professional development programs, they do provide funds for teachers to attend and participate in conferences abroad. Teachers, thus, can choose to attend conferences on any topic related to teaching, linguistics or TESOL, including assessment. However, it seems that the teachers in this study chose not to attend any conferences or professional development events specifically on assessment, perhaps choosing to attend only events on English language teaching in general. Furthermore, those participants who reported attending professional development events (Mona and Najla) explained that they joined these programs at their own expense, and that these events focused on teaching strategies, rather than assessment. One of these participants (Mona) showed a more detailed understanding of the

purposes of assessment, and seemed to believe that the institution's summative assessment policy should be changed and replaced by a more formative assessment approach.

The lack of professional development programs and its consequent lack of teacher assessment literacy has been discussed in several studies conducted in the Gulf area. Mansory (2016), for example, stated that in his study (conducted at a university in Saudi Arabia), professional development programs are expensive and considered a privilege only offered to certain teachers who are on the assessment committee. As a result, teachers in his study lacked the assessment literacy needed to design quality tests or have an active role in the assessment process. Coombe et al. (2012, as cited in Mansory, 2016) also stated that assessment resources in the Gulf countries are scarce and that creating assessment materials can be complicated for teachers who have a heavy workload or who are not interested in assessment. In another study, Al-Sawafi (2014) indicated that teachers in his study (conducted at a university in Oman) did not receive training in assessment, and argued that teachers need such programs to be able to interpret their practices, especially in continuous assessment. Bennett (2011) suggested that teachers need professional development programs, especially in formative assessment, to build assessment knowledge. Finally, a study by Brookhart et al. (2010) suggested that teachers who attended professional development programs were able to modify their beliefs and change their assessment practices accordingly.

Other external factors concerned contextual factors, mainly factors related to students, courses (e.g., subject matter), policies, the classroom setting, the local culture, and the context (e.g., the institutional context or the country). One of the main contextual factors was related to students, specifically teachers' perceptions of the student L2 learning. Almost all participants believed that students lacked motivation to learn English and had low English language

proficiency. This belief was clearly reflected in their assessment practices for both summative and formative purposes. For example, the majority of the participants reported that they do not provide feedback or implement various formative assessments because of their perceptions of the students' low motivation.

The majority of the participants also believed that students focus on summative paper-based assessments to earn high grades and seem to act on such beliefs as facts. Furthermore, the participants appeared to depend on paper-based assessment not only to evaluate students but also to familiarize students with the final tests in order to boost students' grades. This factor seemed to shape the participants' practice of using paper-based assessments as a sole means of grading and not using formative classroom assessment for student evaluation. The participants not only relied mainly on summative assessment for student evaluation, but they also used paper-based assignments to boost students' grades. This process is similar to what the literature calls grade inflation. Grade inflation occurs when student grades are boosted irrespective of their academic achievement (Kassahun, 2008). Although some participants indicated that extra-credit assignments increased student grades, they admitted that students needed higher grades to pass the course or gain high GPAs. The participants who used assignments as extra credit to boost students' grades used very simple questions. Other participants said they would give all students full marks for attendance.

Relying on summative assessment as the main method to grade students also relates to the context of higher education. Yorke (2003) stated that in higher-education, teachers tend to rely heavily on summative assessment. A study conducted by Zhang and Burry-Stock (2003) also shows that study level affects assessment practices. For example, elementary school teachers focus more on formative assessment than on summative assessment, unlike teachers in secondary

schools. Broadbent, Panadero, and Boud (2018) stated that large classes in higher education cause teachers to use summative assessment because they think it is fairer to evaluate a large number of students based on rubrics and "clear and shared standards and scoring systems" (p. 319). Several participants also believed that the large numbers of students make it challenging to employ various formative assessment methods.

Broadbent et al. (2018) added that the reliance on summative assessment requires "shared assessment practices and high inter-rater reliability in marking to ensure fairness, to avoid tensions with students, such as wanting to change to the 'easiest' teacher's classes, or receive complaints about unbalanced workloads derived from different assessment methods" (p. 319). Not all participants reported those factors listed by Broadbent et al. (2018), but in terms of using summative assessment to ensure validity, only those participants who taught GE courses mentioned the advantages of machine marking for GE course exams. A few participants (e.g., Nadia) did not agree with this process, saying that exams with open-ended questions would show more critical thinking skills than tests with closed-ended questions. However, almost all participants constructed their tests in both GE and ESP courses to include closed-ended questions, perhaps because, as Broadbent et al. (2018) mentioned, they are believed to ensure fairness and because this saves time given the large number of students.

In relation to course type (i.e., ESP vs. GE courses), previous studies have shown that subject matter knowledge impacts teachers' assessment beliefs and practices (Borg, 2006; Cumming, 2001; Yin, 2010). In this study, the participants teaching ESP and GE courses did not report major differences in their assessment practices, especially in classroom formative assessment, but with regard to summative assessment, some participants said they had more freedom to design their own exams in ESP courses, whereas in the GE courses, some participants

reported that they had a minor role in constructing tests. In GE courses, the course coordinator

constructs the tests. Cumming (2001) found that teachers varied in their assessment practices

depending on the purposes of the course. He found that in specific courses (e.g., writing),

teachers focused on their judgment, unlike in GE courses where teachers focused on students'

general language skills. This finding was not replicated in this study. The participants teaching

ESP and GE courses followed similar routines for assessment for formative and summative

purposes. For example, although Latifa taught a writing course, she did not employ peer

assessment or other assessment tasks differently from her colleagues who taught GE courses. Her

summative assessment of writing, however, differed from other courses as she required students

to write a letter, and included different questions for writing. This finding could again relate to a

lack of assessment literacy among the participants.

Findings related to the external factors include the influence of the assessment policies of

the institution and the language center. As for the LC, the participants presented two views. A

number of participants opposed the tests, especially those of the GE courses, saying that the tests

were constructed above the students' actual English proficiency level. The other participants,

mainly senior teachers, accepted the tests, and advocated for the policy of standardized tests,

saying it is fair to assess students based on standardized tests supervised by the course

coordinator or by the LC administrators. Nevertheless, almost all the participants were against

the institutional grading policy of having 50% of the overall grade based on a single, final, paper-

based test. Almost all the participants wished to distribute this percentage among multiple tests

and/or formative assessments. Yet, examining the institution grading policy reveals that the

policy is not clear as to whether the 50% must be a paper-based test or could be any other

assignment.

Studies have shown that assessment policies restrict teacher assessment practices (Arkoudis & O'Loughlin, 2004; Rea-Dickins, 2008; Tierney 2006; Xu & Liu, 2009). A few studies conducted in higher education in the Gulf countries (similar to the context of this study) also revealed the influence of institutional policies on teacher assessment practices (e.g., Mansory, 2016; Troudi et al., 2009). However, Yin (2005) mentioned that previous studies have focused on government-mandated assessment policies rather than the post-secondary or private institutional policies on teacher assessment cognition. He found that teachers in higher education are influenced by institutional policy but not governmental policy. Although the policy of the institution shaped teachers' assessment beliefs and practices in a weak way in this study, it did affect their beliefs around summative assessment in terms of both test design and rubrics and teacher group decisions on assessment design.

In this study, the majority of the participants stated that the department requires a certain format for the GE tests, and, in CBE specifically, tests have to be marked using scoring machines and have to be unified. The participants in CBS said that the course coordinator oversees the testing policy that GE tests do not have to be administered on the same date and that GE tests can be shared among certain groups of teachers. Several participants also criticized the test format for the unified tests, especially in GE courses. They said that they play a minor role in the construction of the GE tests, unlike in ESP courses.

As for the summative assessment policy, in this study, no clear guidelines or policy seems to exist concerning test construction in terms of design or rubrics. PAAET has a grading policy for the final test to make up 50% of the grade. The teachers therefore could construct other assessments using their own rubrics. Nevertheless, the majority of the participants agreed that the grade requirement set for the final exam (i.e., 50%) is high and may not be fair. As for

test design, participants teaching GE and ESP courses were found to follow similar test designs and rubrics. The midterm and the final tests consisted of four sections and mostly closed-ended questions. The final tests' percentages follow PAAET's assessment policy, but the midterm grade percentage and design appear to be similar across the LC teachers. In other colleges, the participants mentioned they do not have to follow a specific format, and teachers can decide if they want to share their tests with other teachers or not. Even though all participants reported that they are free to design their own tests in ESP courses, they followed similar test formats in all their courses. This suggests various factors influence the participants in addition to the institutional testing policy.

Although the participants did not express this point explicitly, it appears that their belief that the institution puts more emphasis on summative assessment influenced their practices. There is no indication that tests are strictly regulated by a standard assessment policy. In addition, because participants appear to have not received any training in assessment, they designed assessments based on the traditional testing formats that include discrete multiple-choice questions on different language aspects (Fox, 2013). Furthermore, these language aspects were tested separately. Although the PAAET handbook asserts that 50% must be reserved for the final test, it does not specify how the rest of the 50% should be distributed. The participants believed it is the department policy that set the 30% for the midterm and 10% for the quizzes, but when asked about the source of the distributions, participants did not agree whether the LC or the course coordinator set this grade. This conflicting response suggests that teachers in this context would have a voice if they had general assessment training and were confident about assessment. This aligns with several studies on teacher beliefs and practices in the Gulf region that found that

there is a need for more professional development offerings (Alkharusi et al., 2014; Al-Sawafi, 2014; Mansory, 2016).

Finally, other external factors influencing teachers' assessment beliefs and practices include the classroom parameters both in terms of student number and classroom size. These factors were mentioned in studies that investigated teachers' assessment beliefs and practices (e.g., Cheng et al., 2004). Cheng et al. found that teacher assessment practices in China were influenced by the mandated policy and classroom size. In this study, aside from the policy discussed above, classroom size was a factor that many participants believed hindered their implementation of formative assessments. For example, participants in CBE and CBS mentioned that the size of classrooms is small compared to the large number of students enrolled in the course. This lack of space affected their feedback and monitoring practices. As Yorke (2003) mentioned, in higher education, the number of students may be high, which would cause the teacher to rely on summative assessment rather than formative assessment. Yorke (2003) added that in a higher-education setting, the increased number of students in a class would lead to less assessment of individual students. Large class sizes place additional demands on teachers as they could be already engaged in research activities, intra-instructional administration, and the generation of funding (Yorke, 2003).

6.2 Limitations of the Study

According to Borg (2006), understanding teacher cognition and bolstering the relationship between teacher beliefs and their practices requires several research methods. Speer (2005) also urged researchers to differentiate between professed beliefs and attributed beliefs, that is, distinguish between what teachers say and what they demonstrate in the classroom to avoid inconsistency in studying the beliefs and practices of teachers. Accordingly, in this study I

selected two qualitative methods, because as Barnard and Burns (2012) argued, qualitative methodology is necessary to understand the complex relation between beliefs and practices. To understand teachers' beliefs, I used interviews, and for teacher practices, I used classroom observation and document analyses.

However, the main limitation I encountered in this study was that I was not able to conduct stimulated recalls, which many studies have employed to better understand teachers' assessment beliefs in relation to their assessment practices (e.g., Yin, 2005). This was due to the study time limits. This study was conducted during the summer term, which is shorter than regular semesters. I conducted post-observation interviews instead, but several participants could not immediately recall their practices. The use of the audio recording was successful in this situation, helping them to recall some of their practices. Nevertheless, using the recording was actually a limitation. The majority of the participants did not feel comfortable, and/or they reported that their students would not feel comfortable if the class was video recorded. The video recording may have helped in the process of stimulated recalls and in the follow up interviews. Only one participant agreed to have her class video recorded, but her voice in the recording was not clear. I did another post-observation interview in which I used the audio recording, and it was paired with the video recording. This participant was able to recall her practices assisted by document analysis at the time of the post-observation interview.

In addition, some participants noted that their students were shy and hesitant to participate perhaps due to my presence and the recording, which further limits the validity of the recordings. I did not notice this factor being at play in all my observations, yet I deliberately sat at the back of the class, taking the role of a non-participant observer. It may take several sessions in order for students to become used to my presence and more actively participate in class.

Additionally, I made every effort to reassure students that the audio recording was only for the purpose of research and would be accessed by no other person than me. Yet, it seemed the recording may have placed some stress on students. Still, the participants reported that students were passive learners and felt embarrassed to participate in front of their peers, even without my presence.

Furthermore, because the study took place in the summer term, once it was over, the teachers went on vacation and could not provide me with additional documentation or conduct an immediate post-observation interview. For example, for some participants I needed copies of more quizzes or samples of their grading practices, but they could not provide them as they were away. For the same reasons, I had to conduct the post-observation interviews via phone due to the distance; though in this case the participants were very supportive and helped the process to go smoothly.

Another limitation was the homogeneity of the student population; all students were female students and a large number of them in this institution came from certain regions that are known to have weak education in English, and as discussed in the previous section, student characteristics can strongly influence teacher conceptualizations and practices. The findings cannot then be generalized to institutions out of this context, as other higher education institutions in Kuwait may have students from different genders, backgrounds, regions, and English language proficiency levels. The same factors apply to the participants who were all female Kuwaiti teachers. As mentioned in the methods chapter, I was unable to include teachers on male campuses because of cultural restrictions. Teachers on male campuses may have to deal with factors that may be similar to or different from those on female campuses. These are important factors to address in future research. Participation by teachers of different genders or

backgrounds could have broadened the findings. Because I selected a qualitative methodology, I could not employ questionnaires to include a larger sample of teachers from the LC. Yet, my primary goal was to acquire rich insights and a deeper understanding of teachers' assessment beliefs and practices, which was well-suited to individual case studies.

This study was able to highlight many external factors and investigate differences in assessment beliefs and practices in relation to ESP and GE courses. Yet, additional factors could exist that teachers did not report. For example, the participants did not mention the issue of job security. Teachers in this institution do not feel threatened that their jobs will be terminated at any time. In the public sector in Kuwait, teachers are generally well paid, and student evaluation reports do not affect job security but they can help them get promoted to higher positions. The sense of job security perhaps led some participants to accept to deal with the low proficiency levels of the students. In other words, the teachers may want the situation to change, but it does not affect their position if it remains the same.

6.3 Implications

With these limitations aside, this study highlights several implications for teaching, policy, and further research.

6.3.1 Implications for Teaching

One of the major findings in this study is that teachers need to increase their assessment literacy so that they could increase their knowledge of classroom assessment methods, purposes and uses, and enhance student learning. It can also help them understand various methods of assessment, design effective assessment and scoring schemes, use feedback to improve learning (Brookhart, 2011), and understand assessment in relation to subject matter (e.g., GE vs. ESP courses; Yin, 2010). Thus, teachers need to attend professional development programs, which

may lead to innovation in assessment policy and practices. It seems teachers in this context have a significant role in guiding their own practices. Although some participants criticized the assessment policies of the department, especially experienced teachers (e.g., Hajar, Leila), it appears that such policies are malleable based on teacher discussion and the recommendation of the coordinators. This study could be used as an educational resource on assessment, from which teachers in this context could benefit by getting a deeper understanding of the institution policy requirements, and by reflecting on their own assessment practices.

6.3.2 Implications for Policy

Because teachers are the primary agents and decision makers in the assessment process (Giraldo, 2018; Rea-Dickins, 2001, 2008), policy makers need to work with teachers to improve teacher knowledge of assessment and improve language assessment practices by offering professional development programs. Popham (2009, pp.9-10) suggested 13 possible topics to include in any assessment literacy professional development program. Five of those topics seem particularly relevant to this study:

- The basic functions of educational assessment, especially the evidence from which inferences can be made about the students' knowledge and skills.

- The improvement and construction of constructed-response and selected-response test items.

- Scoring and development of portfolio assessments, performance assessments, peer assessments, and self-assessments.

- Designing and implementing formative assessment procedures along with both experience-based insights and research evidence.

- Ways to collect and interpret evidence of students' interests, attitudes, and values.

This study could also help administrators, stakeholders in the LC and the PAAET better understand the nature of teachers' beliefs about assessment. The findings could provide insights into which areas policy makers and teachers need to address to inform assessment reforms. Based on the findings of this study, implications for policy reforms can be made in relation to the different factors reported by the participants. These factors included student number and curriculum design. As several participants pointed out, room size is small compared to the number of students registered which hindered the teachers' use of formative assessments. One recommendation is to reduce the number of students per course. In terms of curriculum design, a number of participants believed that the English language courses offered by the department do not cover all English language skills (e.g., listening, speaking). Thus, in addition to offering training courses and professional development programs, one implication is to broaden the focus of courses to include all language aspects and engage teachers in the selection of course assessment materials. The variety of language skills addressed in the curriculum may enable teachers to implement different assessment methods depending on the language aspects being assessed.

6.3.3 Implications for Further Research

This study contributes to the assessment field in general and the study of teacher assessment beliefs and practices in particular. As discussed in the literature review chapter, a gap exists in the study of this field in the Gulf region and in Kuwait. To the best of my knowledge, the only study that investigated teacher assessment beliefs in Kuwait utilized questionnaires as a single method in one university in Kuwait (i.e., Troudi et al., 2009). Conducting research in a context that has not been investigated can provide important insights of teachers' assessment beliefs and practices in different contexts especially at the post-secondary level.

One of the key findings of this study was that teachers needed professional development in order to increase their assessment literacy. Further studies could examine teachers' assessment beliefs and practices after teachers join professional development programs. Studies such as that of Brookhart et al. (2010) found that teacher assessment beliefs may change as a result of professional development programs. Borg (2006) also urged researchers to examine teacher beliefs longitudinally as such beliefs can change over time due to experience and practice.

As can be seen from the findings, summative assessment is very dominant in this context, and the influence of assessment policy appears to shape teachers' practices towards this purpose of assessment. A number of studies about the impact of summative assessment, accountability, and high-stake testing mandates on teachers' classroom assessment have been reported in literature (e.g., Arkoudis & O'Loughlin, 2004; Brown, 2004; Troudi et al., 2009). Yet, there has been a dearth of research on the influence of such assessments at the university level, particularly in terms of the washback effects of summative assessments on both students and teachers. Further research is needed to examine the influence of summative assessments at the post-secondary level including teachers, students' and stakeholders' beliefs and practices concerning summative assessment, and the external factors influencing their summative assessment beliefs and practices.

References

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2014). Educational assessment profile of teachers in the Sultanate of Oman. *International Education Studies*, *7*(5), 116-137.

Allen, L. Q. (2002). Teachers' pedagogical beliefs and the standards for foreign language learning. *Foreign Language Annals*, *35*, 518-529.

Al-Mutawa, N. (1992). *The education pathway after the Iraqi invasion: Where is it leading to?* Proceeding of the Conference of the English Language Curriculum in Public Schools. Plaza Hotel Kuwait.

Al-Rubaie, R. (2010). *Future teachers, future perspectives: The story of English in Kuwait* (Doctoral dissertation, University of Exeter). Retrieved from Open Researcher Exeter (ORE).

Al-Rushaid, Y. (1978). *Tareekh Al-Kuwait* [Kuwait history]. Beirut: Dar Maktabat Al-Hayat.

Al-Sawafi, O. S. (2014). *Investigating English teachers' beliefs and practices in relation to the continuous assessment reform in the Sultanate of Oman* (Doctoral dissertation, University of Leeds). Retrieved from White Rose eThesis Online.

Arkoudis, S., & O'Loughlin, K. (2004). Tensions between validity and outcomes: Teacher assessment of written work of recently arrived immigrant ESL students. *Language Testing*, *21*, 284-304.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford university press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford, UK: Oxford University Press.

Barcelos, A.M.F. (2003). Researching beliefs about SLA: A critical review. In: P. Kalaja, & A.M.F Barcelos (Eds.), *Beliefs about SLA: New research approaches* (pp. 7-33). Springer, Dordrecht.

Barkaoui, K. (2010). Do ESL essay raters' evaluation criteria change with experience? A mixed-methods cross-sectional study. *TESOL Quarterly, 44*(1), 31-57.

Barnard, R., & Burns. A. (Eds.). (2012). *Researching language teacher cognition and practice: International case studies*. Bristol, UK: Multilingual Matters.

Bataineh, R. F., & Obeiah, S. F. (2016). The effect of scaffolding and portfolio assessment on Jordanian EFL learners' writing. *Indonesian Journal of Applied Linguistics*, *6*(1), 12-19.

Belcher, D. D. (2006). English for specific purposes: Teaching to perceived needs and imagined futures in worlds of work, study, and everyday life. *TESOL Quarterly*, *40*(1), 133-156.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5-25.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*, 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*, 5-31.

Bloom, B.S. (1969). Some theoretical issues relating to educational evaluation. In R.W. Tyler (Ed.). *Educational evaluation: New roles, new means. The 63rd yearbook of the National Society for the Study of Education, part 2 (Vol. 69*; pp. 26–50). Chicago, IL: University of Chicago Press.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, *36*, 81-109.

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

Borg, S. (2012). Current approaches to language teacher cognition research: A methodological analysis. In R. Barnard & A. Burns (Eds.), *Researching language teacher cognition and practice: International case studies* (pp. 11-29). Bristol, England: Multilingual Matters.

Broadbent, J., Panadero, E., & Boud, D. (2018). Implementing summative assessment with a formative flavour: A case study in a large class. *Assessment & Evaluation in Higher Education*, *43*(2), 307-322.

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

Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 43–62). New York: Teachers College Press.

Brookhart, S. M., Moss, C. M., & Long, B. A. (2010). Teacher inquiry into formative assessment practices in remedial reading classrooms. *Assessment in Education: Principles, Policy & Practice*, *17*, 41-58.

Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*, 301-318.

Brown, H. D., Abeywickrama, P. (2010) *Language assessment: Principles and classroom practices.* White Plains, NY: Pearson Education.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, *32*(4), 653-675.

Campbell, C. (2013). Research on teacher competency in classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 71-84). Thousand Oaks, CA: Sage.

Canh, L. V. (2012). Interviews. In R. Barnard & A. Burns (Eds.), *Researching language teacher cognition and practice: International case studies* (pp. 11-29). Bristol, England: Multilingual Matters.

Chapelle, C.A., & Plakans, L. (2013). Assessment and testing: Overview. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics*, (pp. 240-244). Oxford: Blackwell/Wiley.

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing, 21*, 360-389.

Cheng, L., Rogers, W. T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education, 33*, 9-32.

Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, *4*(1), 85-107.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, *22*(1), 93-121.

Clandinin, D. J., & Connelly, F. M. (1987). Teachers' personal knowledge: What counts as 'personal' in studies of the personal. *Journal of Curriculum Studies*, *19*, 487-500.

Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics*, *20*, 147-161.

Clark, C. P., & Peterson, B. P. (1986). Teacher's thought processes. *Handbook of research on teaching.* New York: Macmillan.

Clarke, S., & Gipps, C. (2000). The role of teachers in teacher assessment in England 1996-1998. *Evaluation & Research in Education*, *14*, 38-52.

Cohen, L., Manion, L., & Morrison, K. (2008). *Research methods in education*. London: Routledge.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438-481.

Cumming, A. (2001). ESL/EFL instructors' practices for writing assessment: Specific or general purposes? *Language Testing, 18*, 207-224.

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing, 21*, 305-334.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, *43*, 393-415.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, *38*, 47-65.

Fives, H., & Buehl, M. M. (2012). Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? *APA Educational Psychology Handbook*, *2*, 471-499.

Fives, H., & Buehl, M. M. (2014). Exploring differences in practicing teachers' valuing of pedagogical knowledge based on teaching ability beliefs. *Journal of Teacher Education*, *65*(5), 435-448.

Fontana, A., & Frey, J. H. (2000). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 645-672). Thousand Oaks, CA: Sage.

Fox, J. (2013). Language assessment. In C.A. Chapelle (Ed.). *The encyclopedia of applied linguistics (pp. 1-6)*. Oxford: Blackwell/Wiley.

Gardner, J. (2006). *Assessment and learning*. UK: Sage Publication Limited.

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. UK: The Falmer Press.

Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile Issues in Teachers Professional Development*, *20*(1), 179-195.

Glaser, B. and Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Gliner, J. A. (1994). Reviewing qualitative research: Proposed criteria for fairness and rigor. *Occupational Therapy Journal of Research*, *14*(2), 78-90.

Hajjaj, A. (1978). Recent trends in educational media for ELT in Kuwait. *Programmed Learning & Educational Technology*, *15*(1), 57-63.

Harlen, W. (2005). Trusting teachers' judgment: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, *20*, 245-270.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education*, *4*, 365-379.

Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment*, *6*(1), 1-17.

Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, *29*(3), 395-420.

Huhta, A. (2008). Diagnostic and formative assessment. In B. Spolsky & R. M. Hult (Eds.), *The handbook of educational linguistics* (pp. 469-494). Malden, MA: Blackwell.

Hussein, A., Al-Saleh, F., Al-Khaled, F., Al-Ghunaim, A., Al-Zebin, D., Al-Hajji, Y., Qasem, Q. (2002). *Tareekh al-taleem fi dawlat al-Kuwait: Derasa tawtheeqiya* [The history of education in Kuwait: A documentary study]. Kuwait: Center for Research and Studies on Kuwait.

Izci, K. (2016). Internal and external factors affecting teachers' adoption of formative assessment to support learning. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 10*(8), p. 2541-2548.

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*, 407-430.

Kagan, D. M. (1992). Implication of research on teacher belief. *Educational Psychologist*, *27*, 65-90.

Karmani, S. (2005). Petro-linguistics: The emerging nexus between oil, English, and Islam. *Journal of language, Identity, and Education*, *4*(2), 87-102.

Kassahun, D. (2008). Standardization techniques for grade-inflation problems at higher educational institutions of Ethiopia: The case of Addis Ababa. *Assessment & Evaluation in Higher Education*, *33*(1), 33-44.

Katz, A., & Gottlieb, M. (2012). Assessment in the classroom. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-8). Oxford: Blackwell/Wiley.

Kharma, N. (1967). The teaching of English in Kuwait. *ELT Journal*, *21*(3), 253-260.

Kunnan, A., & Jang, E. E. (2009). Diagnostics feedback in language testing. In M. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 610-25). Oxford, UK: Blackwell Publishing.

LeCompte, M. D., & Preissle, J. (1994). Qualitative research: What it is, what it isn't, and how it's done. *Advances in Social Science Methodology*, *3*, 141-163.

Lee, S. H. (2014). *Korean middle- and high-school math teachers' understanding of formative assessment: An interview study* (Doctoral dissertation, University of Illinois at Urbana-Champaign). Retrieved from Illinois Digital Environment for Access to Learning and Scholarship (IDEALS).

Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly*, *1*, 19-41.

Leung, C. (2005). Classroom teacher assessment of second language development: Construct as practice. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 869-893). Mahwah, NJ: Lawrence Erlbaum.

Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing*, *21*, 335-359.

Li, L., & Walsh, S. (2011). 'Seeing is believing': looking at EFL teachers' beliefs through classroom interaction. *Classroom Discourse, 2 (1)*, 39-57.

Lund, A. (2008). Assessment made visible: Individual and collective practices. *Mind, Culture and Activity,15*, 32-51.

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, *18*(4), 351-372.

Mansory, M. (2016). *EFL teachers' beliefs and attitudes towards English language assessment in a Saudi university's English language institute* (Doctoral dissertation, University of Exeter). Retrieved from Open Researcher Exeter (ORE).

Mason, J. (2002). *Qualitative researching*. Thousand Oaks, CA: Sage.

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, *26*(1), 075-100.

McMillan, J. H. (2013). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 3-16). Thousand Oaks, CA: Sage.

McMillan, J. H., Hellsten, L., & Klinger, D. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction*. Ontario: Pearson Canada.

McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, *18*(4), 333-349.

Minichiello, V., Aroni, R., Timewell, E., & Alexander, L., (1995). *In-depth interviewing: Principles, techniques and analysis (Second ed.)*. Melbourne, Australia: Longman.

Mohammad, E. A. (2008). *The place of writing in first grade Kuwaiti English education: A sociological case study* (Doctoral dissertation, Queensland University of Technology). Retrieved from QUT Digital Repository (26369).

Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing*, *16*(2), 82-99.

Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, *22*(4), 13-25.

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, *22*(2), 155-175.

Negueruela-Azarola, E. (2011). Beliefs as conceptualizing activity: A dialectical approach for the second language classroom. *System*, *39*, 359-369.

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, *19*, 317-328.

Osman, F. H. (1996). *ESP students' underachievement: Possible reasons and solutions with special reference to the Public Authority for Applied Education and Training (PAAET) in Kuwait* (Doctoral dissertation, University of Bath). Retrieved from Electronic Thesis Online Service (ETHOS) (300490).

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*, 307-332.

Patton, M. Q. (2002). *Qualitative research and evaluation methods.* Thousand Oaks, CA: Sage.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, D.C.: National Academies Press.

Phipps, S. & Borg, S. (2009). Exploring tensions between teachers' grammar teaching beliefs and practices. *System, 37*(3), pp.380-390.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental?. *Theory into Practice*, *48*(1), 4-11.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, *18*, 429-462.

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, *21*, 249–258.

Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning-focused interpretation. *International Journal of Applied Linguistics*, *16*, 163-188.

Rea-Dickins, P. (2007). Classroom-based assessment: Possibilities and pitfalls. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 505-520). Springer, Boston, MA.

Rea-Dickins, P. (2008). Classroom-Based language assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education*, Vol. 7: *Language testing and assessment* (2nd ed., pp. 255-270). New York: Springer and Business Media LLC.

Rea-Dickins, P. & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing, 17*, 215-243.

Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The relationship between teachers' beliefs and practices in reading comprehension instruction. *American Educational Research Journal*, *28*, 559-586.

Rogers, W. T., Cheng, L., & Hu, H. (2007). ESL/EFL instructors' beliefs about assessment and evaluation. *Canadian and International Education/Education Canadienne et Internationale*, *36*, 39-61.

Ruiz-Primo, M. A., & Li, M. (2013). Examining formative feedback in the classroom context: New research perspectives. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 215–232). Thousand Oaks, CA: Sage.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119-144.

Sahinkarakas, S. (2012). The role of teaching experience on teachers' perceptions of language assessment. *Procedia-Social and Behavioral Sciences*, *47*, 1787-1792.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, *30*, 309-327.

Scriven, M. (1967). The methodology of evaluation. *Curriculum Evaluation, 1,* 39-83.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*, 4-14.

Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, *63*(3), 66-70.

Shepard, L. A. (2006). Classroom assessment. *Educational Measurement*, *4*, 624-646.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*, 1-23.

Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, *22*(4), 26-33.

Speer, N. M. (2005). Issues of methods and theory in the study of mathematics teachers' professed and attributed beliefs. *Educational Studies in Mathematics, 58*, 361-391.

Stake, R. E. (1995). *The art of case study research*. London: Sage.

Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 443-466). Thousand Oaks, CA: Sage.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*, 758-765.

Swain, M., Kinnear, P. & Steinman, L. (2011). *Sociocultural theory in second language education. An introduction through narratives*. Bristol: Multilingual Matters.

Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing?. *Language Testing*, *17*(2), 163-184.

The Public Authority for Applied Education and Training (PAAET) (n.d.). Retrieved from: www.paaet.edu.kw

Tierney, R. D. (2006). Changing practices: Influences on classroom assessment. *Assessment in Education*, *13*, 239-264.

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education*, *25*(2), 149-169.

Troudi, S., Coombe, C., & Al-Hamily, 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.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, *22*(4), 389-404.

Turner, C. E. (2012). Classroom assessment. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 65-78). New York: Routledge.

Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271-296.

Varier, D. (2015). *A case study examining formative assessment in a postsecondary English language program* (Doctoral dissertation, Virginia Commonwealth University). Retrieved from VCU Scholars Compass.

Watkins, D., Dahlin, B., & Ekholm, M. (2005). Awareness of the backwash effect of assessment: A phenomenographic study of the views of Hong Kong and Swedish lecturers. *Instructional Science, 33*, 283-309.

Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F.K. Lester Jr. (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1053-1098). Greenwich, CT: Information Age Publishing.

Woods, D., & Çakır, H. (2011). Two dimensions of teacher knowledge: The case of communicative language teaching. *System*, *39*, 381-390.

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149-162.

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, *43*(3), 492-513.

Yin, M. (2005). *A progressively focused qualitative study of teacher thinking in English for Academic Purposes (EAP) classroom language assessment* (Doctoral dissertation, University of Bristol). Retrieved from Explore Bristol Research.

Yin, M. (2010). Understanding classroom language assessment through teacher thinking research. *Language Assessment Quarterly*, *7*(2), 175-194.

Yin, R. (2014). *Case study research: Design and methods.* Thousand Oaks, CA: Sage Publications.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, *45*(4), 477-501.

Yung, B. H. W. (2002). Same assessment, different practice: Professional consciousness as a determinant of teachers' practice in a school-based assessment scheme. *Assessment in Education: Principles, Policy & Practice*, *9*(1), 97-117.

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*(4), 323-342.

Appendix A: York University Ethics Approval

YORK U
UNIVERSITÉ UNIVERSITY
OFFICE OF RESEARCH ETHICS (ORE)

Certificate #:	STU 2015 - 092
Approval Period:	07/09/15-07/09/16

Memo

To: Shaima Dashti, Applied Linguistics - Graduate Program,

From: Alison M. Collins-Mrakas, Sr. Manager and Policy Advisor, Research Ethics
(on behalf of Veronica Jamnik, Acting-Chair, Human Participants Review Committee)

Date: **Thursday, July 09, 2015**

Re: Ethics Approval

EFL Teachers' Beliefs and Practices about Classroom Assessment: A Mulitiple Case Study in the Context of Kuwait

_____

I am writing to inform you that the Human Participants Review Sub-Committee has reviewed and approved the above project.

Should you have any questions, please feel free to contact me at: or via email at:

Yours sincerely,

Alison M. Collins-Mrakas M.Sc., LLM
Sr. Manager and Policy Advisor,
Office of Research Ethics

RESEARCH ETHICS: PROCEDURES to ENSURE ONGOING COMPLIANCE

Upon receipt of an ethics approval certificate, researchers are reminded that they are required to ensure that the following measures are undertaken so as to ensure on-going compliance with Senate and TCPS ethics guidelines:

1. **RENEWALS:** Research Ethics Approval certificates are subject to annual renewal. Certificates **must be current** in order for research activities to continue.
   a. Researchers are required to submit a request for renewal to the Office of Research Ethics (ORE) for review and approval prior to the expiry of the certificate.
   b. **Failure to renew an ethics approval certificate or** (to notify ORE that no further research involving human participants will be undertaken) **may constitute a breach of Senate Policy on research involving human participants.**

2. **AMENDMENTS:** Amendments must be reviewed and approved **PRIOR** to undertaking/making the proposed amendments to an approved ethics protocol;

3. **END OF PROJECT**: ORE must be notified when a project is complete;

4. **ADVERSE EVENTS**: Adverse events must be reported to ORE as soon as possible;

5. **AUDIT:**
   a. More than minimal risk research may be subject to an audit as per TCPS guidelines;
   b. A spot sample of minimal risk research may be subject to an audit as per TCPS guidelines.

FORMS: As per the above, the following forms relating to on-going research ethics compliance are available on the Research website:
   a. Renewal
   b. Amendment
   c. End of Project
   d. Adverse Event

Appendix B: Informed Consent Letter for Teachers

Study name: EFL Teachers' Beliefs and Practices about Classroom Assessment: A Multiple Case Study in the Context of Kuwait

Researcher: Shaima Dashti
                  Doctoral Candidate
                  Linguistics and Applied Linguistics, York University, Toronto, Canada
                  Email:

Purpose of the research: The study aims at investigating English-as-a-Foreign-Language (EFL) teachers' beliefs and practices about classroom assessment in Kuwait. It further investigates how EFL teachers' assessment beliefs and practices are influenced by contextual factors inside and outside the classroom. This will be done through initial interviews, classroom observation, and follow-up interviews.

This research will be presented to York University as fulfillment for the degree of doctor of philosophy. Findings from the study will be submitted for publications to a number of peer-reviewed journals after the dissertation defense (summer 2016), such as: *Language Assessment Quarterly and Language Testing*. The findings will be presented at relevant conferences such as Language Testing Research Colloquium (LTRC), American Association for Applied Linguistics (AAAL), Teaching English as a second Language (TESOL) international convention, and TESOL Arabia.

What you will be asked to do in the research: You are kindly asked to answer interview questions about your beliefs and practices in regard to second language classroom assessment. The interview will take about 45-60 minutes. You are also kindly asked to allow the researcher to observe 2-3 of your classrooms. After classroom observations are complete, the researcher will ask you for follow-up interview(s). The estimated time for the follow-up interview is 45-60 minutes. If more information is needed, the researcher will ask you to conduct another follow-up interview.

Risks and discomforts: The researcher does not foresee any risks or discomforts from your participation in the research.

Benefits of the research and benefits to you: The researcher hopes that this research can enable you to become more aware of your various beliefs and practices with regard to language classroom assessments. Because your beliefs and practices are powerful components of assessment policy improvement, the findings will enable you to work together with policy makers at the Language Center to improve assessment policies and set professional assessment programs, thereby improving assessment, teaching and learning in the EFL classroom. You will also be given an appreciation gift upon your participation in this study.

Voluntary participation: Your participation in the research is completely voluntary and you may choose to stop participating at any time. Your decision not to continue participating will not influence the nature of the ongoing relationship with the researcher

or with staff of York University either now or in the future.

Withdrawal from the study: You may stop participating in the study at any time, for any reason, if you so decide. Your decision to stop participating, or to refuse to answer particular questions, will not affect your relationship with the researchers, York University, or any other group associated with this project. In the event that you withdraw from the study, all associated data collected will be immediately destroyed wherever possible.
You will still receive the appreciation gift for your decision to participate in this study even if you decide to stop participating at any time for any reason.

Confidentiality: All information you supply in the interview and in the recording will be confidential and your name will not appear in any publication or report of the research. Data will be collected using audio/video recorders, and handwritten notes. All your data will be safely stored in the researcher's personal computer (PC) and on a flash memory device saved in a locked cabinet, and the access to these files will be limited to the researcher. The records will be kept on the researcher's PC and the flash memory device will be kept in the locked cabinet until the successful defense of the research, at which time files on the PC will be deleted and the flash memory device will be destroyed using the hammer method.
Confidentiality will be provided to the fullest extent possible by law.

Questions about the research? If you have questions about the research in general or your role in the study, please feel free to contact me or my Graduate Supervisor- Dr. Khaled Barkaoui either by telephone at _____ or by email _____. You may also contact my Graduate Program- <Linguistics and Applied Linguistics, _____. This research has been reviewed and approved by the Human Participants Review Sub-Committee, York University's Ethics Review Board and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines. If you have any questions about this process, or about your rights as a participant in the study, your may contact the Senior Manager and Policy Advisor for the Office of Research Ethics.

Legal Rights and Signatures: I _____ , consent to participate in *EFL Teachers' Beliefs and Practices about classroom assessment: A Multiple Case Study in the Context of Kuwait*, conducted by *Shaima Dashti*. I have understood the nature of this project and wish to participate. I am not waiving any of my legal rights by signing this form. My signature below indicates my consent.

<u>**Signature**</u>_____   <u>**Date**</u>_____
Participant
<u>**Signature**</u>_____   <u>**Date**</u>_____
Principal Investigator

Additional consent

☐ I _____, consent to have my class video/ audio taped.
☐ I _____, consent that all information reported by me in the initial and follow-up interviews will be audio recorded.

Appendix C: Initial Interview Questions

Date of the interview: _____

Place of the Interview: _____

Length of the Interview: _____

The Interview will start with demographic questions about the participant. Each interview

will be 45-60 minutes.

A. Background and Demographic Questions

Name:

Qualification and major of undergraduate and graduate study:

Experience and Teaching of English:

1. How many years have you been teaching English?

   a. Did you start your career as a teacher of English in LC/this College? Have you

   taught somewhere else? If yes, where and what grade levels?

   b. Have you taught in any other LC colleges than xx? If yes, where?

   c. Have you taught in male campuses? If yes, where? What courses?

2. Do you teach general courses?

   a. What are the GE courses you are teaching/ taught?

3. Do you teach/ have you taught ESP courses?

   a. How many ESP courses do you teach/ have you taught?

   b. What are the ESP courses you are teaching/ taught?

   c. What language skill(s) does each (ESP) course focus on?

4. How many classes are you teaching now?

   a. How many students are enrolled in this class?

b. What is the level of this course?

c. What is the level of students?

5. Have you attended any professional training program about language assessment? If yes:

a. What was the program specifically about? (Lee, 2014)

b. Was it related to informal assessment or test, and measurement?

c. What did you learn from this professional program? (Lee, 2014)

d. Did this program contribute to your beliefs and knowledge about language assessment? How?

B. Interview Questions

The following interview questions will be asked for all participants. However, for those participants for whom I will observe both their GE and ESP courses, I will conduct two interviews asking the same questions but on one time in regard to their assessment practices in the GE course, and on the other in regard to their ESP course.

Teachers' assessment beliefs

1. What comes to your mind when you hear the word assessment? (Lee, 2014)

2. How do you define classroom assessment? Provide examples.

3. In your opinion, why should you assess students?/ In your opinion, what are the purposes of assessment?

4. How do you identify the purposes of assessment? (Rea-Dickins, 2001)

5. In your view, in what ways does language assessment influence students' learning? Please explain.

6. What kind of assessments do you believe are most effective for measuring students' performance in xxx (in reference to a specific language construct)?

7. What do you think is the most important language construct to assess in GE courses?

8. What do you think is the most important language construct to assess in ESP courses?

9. What do you use assessment for? What are the kinds of informal assessment you use? What are the formal assessments?

10. Do you believe that your assessment practices are similar in ESP and GE courses? (This question is for those teachers who will be observed for both GE and ESP courses).

11. How do you follow up with whether students learned what you taught? (Wiliam, 2007)

12. Do you believe that students should take part in assessment activities? How? Why (not)?

13. Please discuss how the following factors contributed to your beliefs and knowledge about assessment.

    a. Your Schooling

    b. Your teaching experiences

    c. Your students' proficiency level

    d. Your culture and context

14. What do you think is the teacher's role in relation to assessment in LC? How do you see your role in assessing students' language? For example, central or not? (Troudi et al., 2009)

15. What do you think of the current teachers' assessment practices in LC in general? (Troudi et al., 2009)

Teachers' practices

1. How often do you assess your students in one class? Why?

2. How do you assess your students' performance? (Wiliam, 2007) Why do you follow such types/ strategies of assessment?

3. What are some examples of the assessment strategies that you use regularly in your daily practice? Please describe your process of applying these strategies. (Thompson & Wiliams as cited in Lee, 2014)

4. What type of assessment activities do you prefer to use in the classroom? Please elaborate.

5. Do you allow students to assess themselves and do you use peer assessment? How? Why (not)?

6. How do you respond to students during an assessment activity? What kind of feedback do you use?

7. What are some challenges to administering assessment strategies/ practices in your classroom(s)? (Lee, 2014)

8. Are your assessment practices similar across language constructs/aspects (e.g., listening, reading, grammar)? How? Can you explain how you assess each language construct/aspect in xx course?

9. On what language construct do you focus the most when you assess in GE/ESP courses? Why?

10. Are your assessment practices similar in male and female campuses? How do you assess in both campuses? (This question is to be asked for those teachers who teach at male campuses).

11. Do you rely more on informal/unplanned assessment or on formal/ planned assessments? Why?

12. How do you record students' performance during assessment?

13. Do you have specific rubrics or criteria for assessment? What are they?

14. Do you use informal assessment for the end of term evaluation of a student (accountability purposes)? Why?

15. What criteria do you use to assess students in general? Who sets the assessment criteria in your college? Do you follow standard measurement criteria set forth by the LC (if any)?

16. How do you interpret evidence obtained from assessment? (Rea-Dickins, 2001)

17. Do you share assessment results with your colleagues or other teachers? How? Why (not)? (Rea-Dickins, 2001)

Impact of the educational standard policy and external factors on teachers' beliefs and practices

1. Please describe the assessment and grading policies in your institution. (Lee, 2014; Troudi et al., 2009)

2. Please describe how students are assessed in your institution. Please provide specific examples (Troudi et al., 2009)

3. How are you responding to the assessment policies of your institution?

4. Please describe the grading standards set forth by the LC. Do you follow the LC grading criteria? Do you have your own grading criteria?

5. Are there any standardized tests used for each course? Please describe them.

6. Are exams unified in GE/ ESP courses? Please explain.

7. Please give some examples of the exams you construct for your courses. Do you follow specific grading criteria for those exams? Explain.

8. What factors are influencing your assessment beliefs and practices? Explain how each is influencing your assessment practices.

In case the teacher did not include one of the following factors, I will ask the following questions:

9. How do the following factors influence your assessment beliefs and practices?

    a. Context (the classroom context, the cultural context)

    b. Subject matter (ESP vs. GE courses)

    c. Accountability/ policy standards

    d. Students' proficiency level

10. What other challenges do you encounter when assessing your students? How do you react and cope with the challenges you mentioned? (Troudi et al., 2009)

11. Do you think that your assessment practices have changed/ are changing due to teaching experience or other factors? How? Why (not)?

Appendix D: Post-observation Interview Questions

Sample of the Post-observation Interview Questions (Taken from Dana's post-observation interview)

General Interview Questions

1. Did you notice that your assessment practices were different during the summer? Why or why not?

2. What challenges do you usually encounter when you assess your students? How do you cope with these challenges?

3. You mentioned in the first interview that you are considering doing peer assessment with seventy students. Did you do it in this course? Why do you think it would help with seventy students?

4. Do you rely on informal or formal assessment? Why?

5. You mentioned that exams are to be unified for the ESP as well. Do you agree with having unified exams for your courses? Why or why not?

6. You mentioned that you have to abide by the assessment format, such as synonyms and antonyms, multiple choice, and so on. Why? Who sets these rubrics? Are you satisfied with the rubrics? Why or why not?

7. What suggestions do you have to improve assessment in the LC?

Classroom Observation Part

1. At the beginning of the class, you reviewed some skills with students and mentioned that these skills would be tested on the quiz. Why do you follow this strategy? Do you believe that it is important to revise the test materials? Why or why not?

2. (Min11:14): You went over the homework that you assigned to them in the previous class. Why do you answer the homework with them? Do you still collect and grade their homework even after revising it with them?

3. Most of the activities I observed were done in groups. Did you plan to do this? Did it work the way you wanted?

<u>CO1</u>

Segment 1 (punctuation from the handout): min.16.25-18.00, skip to min. 28.52-30.00

1. What language skills are you focusing on here?

2. What assessment(s), if any, did you do during this session?

    a. What type of assessment?

    b. How and why was it done?

3. Was the assessment planned? How? Why or why not?

4. Why did you select this assessment activity?

5. Did you provide feedback in this lesson? Do you consider this feedback as part of the assessment? How? Why or why not?

6. What is your strategy of monitoring the groups in this exercise?

7. How was students' interaction?

8. Did you record students' performance in such exercises?

CO2

Segment 4 (Reading comprehension) 5.11 then skip to minute 18.07-20.00

1. Do you do groups in all your classes? In ESP and GE?

2. What language skills are you focusing on here?

3. What assessment(s), if any, did you do during this session?

    a. What type of assessment?

    b. How and why was it done?

4. Was the assessment planned? How? Why or why not?

5. Why did you select this assessment activity?

6. Did you provide feedback in this lesson? Do you consider this feedback as assessment? How? Why or why not?

7. What is your strategy of monitoring the groups in this exercise?

8. Did you mark students' performance in such exercises?

Appendix E: Classroom Observation Guidelines

Classroom Observation Guidelines

Date and Time: _____

Course: _____

1. **Space: The physical setting** (Spradley, 1980; as cited in Cohen et al., 2008)
   - Table and seating arrangement
   - Class time

   The people situation (Spradley, 1980; as cited in Cohen et al., 2008)
   - Student number
   - Student identities

2. **What is taking place in the lesson?**
   - Test
   - Review
   - Explaining a new lesson
   - Type of activities
   - Types of assessment

3. **What is assessed?** (Hill & McNamara, 2012)
   - **Language construct focus**

     i. Grammar
     ii. Vocabulary
     iii. Listening
     iv. Reading
     v. Writing
     vi. Speaking

4. **Who is assessed? Who is assessing? By whom?** (Hill & McNamara, 2012)

5. **Who is making decisions, and for whom?** (LeCompte & Preissle, 1993, as cited in Cohen et al., 2008)

6. **How is assessment conducted? What resources/ methods are being used in assessment?**
   - Formative assessment
   - Summative assessment

7. **Description of the teacher and students' interaction during assessment** (LeCompte & Preissle, 1993, as cited in Cohen et al., 2008)
   - What appears to be the important issue that is being discussed
   - Who is talking and who is listening?
   - Who is making decisions, and for whom?
   - What is being discussed frequently/infrequently?
   - What non-verbal communication is taking place?
   - What are the roles of the participants (the teacher and students)?

8. **When is the assessment taking place?**

9. **What type of feedback is being used?**

10. **How does the teacher monitor assessment tasks?**

Additional notes:

Appendix F: Informed Consent Form for Students

Student Consent Form

I am a doctoral student at York University, and I am doing a study about teachers' language classroom assessment. Your teacher has agreed to participate in this study and will be the focus of my research. As part of this research, I will observe and video-/ audio-tape the classroom, and discuss the results and segments of the observation with the teacher. I would like to ask your consent to being videotaped and perhaps discussed by the teacher and myself.

The data will be used for research purposes only. Your name will not appear in any part of the research.

As part of the rules of the university, I have to provide a paper record of your consent. Please check the appropriate box below, complete the rest of the form and return it to me. Also, please feel free to discuss with me any questions or concerns before you give your consent.

Your cooperation is truly appreciated.
If you have any questions or concerns, please contact me at:

Please tick the appropriate box below and sign and date below.

I consent to being recorded by Shaima Dashti in relation to this course, and I consent to such data being analyzed and used for research purposes. I understand that as far as possible, anonymity will be preserved if extracts are included in research publications or reports.

I do not consent to being recorded and discussed by Shaima Dashti and the teacher in relation to this course.

NAME:

SIGNATURE:

DATE:

Source: Yin, M. (2005). *A progressively focused qualitative study of teacher thinking in English for Academic Purposes (EAP) classroom language assessment* (Doctoral dissertation, University of Bristol). Retrieved from Explore Bristol Research.

Appendix G: Sample of Classroom Observation Notes

At the beginning of the class.
- M distributed the assignments and MT
- Students looked at their assignments and grades.
  M waited for questions.
- M revised worksheet 1+2 and noted that the materials
  in the worksheet will be included
  " Your MTs was a mess" MQ ————→ did she revise
  for students to get
  better grades? What's
- Worksheet ③ 8.39 (video recording) Minute   the purpose of reminding
  • Open questions were given to   Students that the material
  students and M gave them 10 mins   will come in the test?
  to answer the questions.
  • Then M collected the paper ————→ Is it graded?

min
9.44
- Students Looked

12.30 (what does ___ mean?) ————————→ voc assessm.

- M was writing a number of vocabulary } the voc. (words are
  included in todays lesson.   not included in the
      students were writing down the voc. book

✱ Kind of voc. assessm:
    - What does ___ mean?

  21.35 M testing students prior ————→ Only one student
        knowledge.   was interacting.
                                        Is it only this time
                                        because of the
recording or everytime

Appendix H: An Example from NVivo Coding
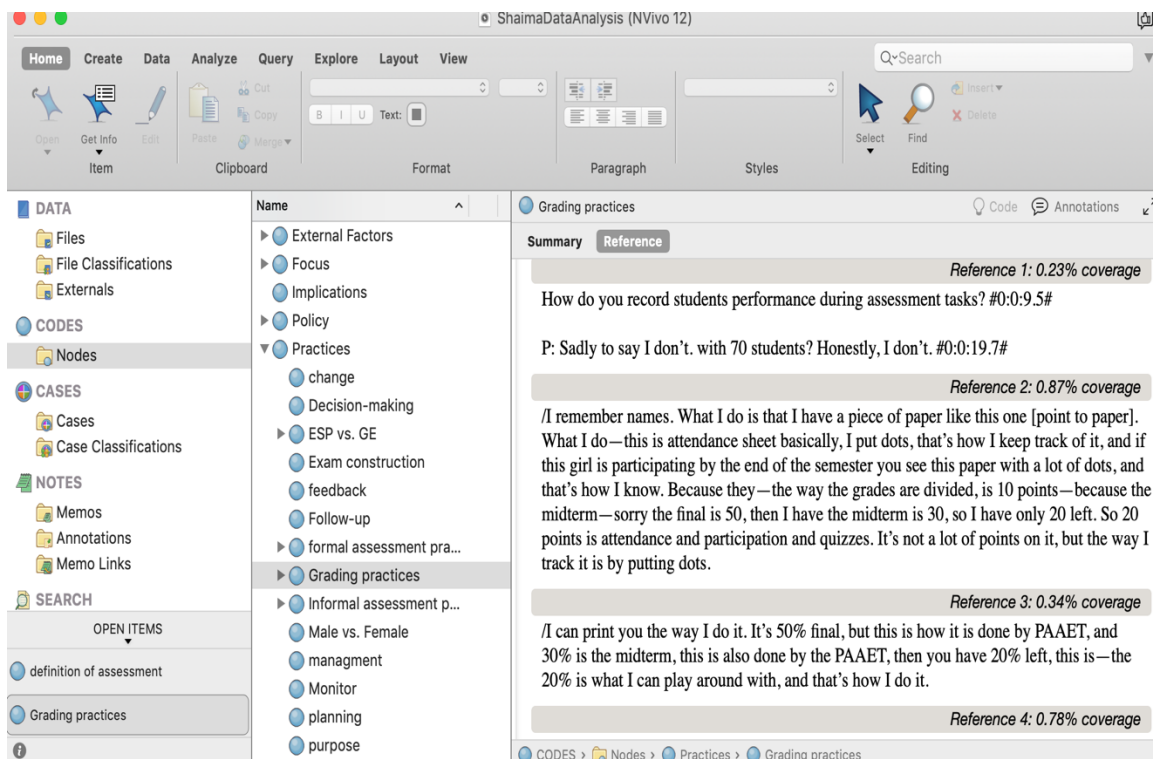
Appendix I: Sample of Najla's ESP Course Plan

PAAET
College of Business Studies
Language Center

The Public Authority for Applied Education and Training
College of Nursing

(B.S.N. Program)

Course Title	:	ESP-1
Course Type	:	Compulsory
Course No.	:	180
Credit hours	:	5
Contact hours / weekly	:	10
Prerequisite	:	EPT pass score or course # 98

Course Description:

This is the first of three ESP courses meant to equip students with the necessary skills required for medical English. It is an integrated course which focuses on the use of English in scientific and medical fields. It is designed to help students communicate effectively in job related activities in the target situation.

Course Objectives:

1. Listen and comprehend medical topics on audio-cassettes and take notes.

2. Practice reading authentic scientific texts, comprehend their gist, and extract details (covering the strategies for efficient reading, skimming, scanning, identifying main supporting ideas, and developing vocabulary).

3. Promote class discussion related to medical topics.

4. Identify and use medical terminology and word affixes.

5. Use Taber's cyclopedic medical Dictionary.

6. Reinforce the grammatical structures and idioms e.g. parts of speech, advanced sentence structures and functions.

7. Understand and fill in medical charts, tables etc.

8. Make summary of reading passages.

9. Write a guided coherent and cohesive paragraph of different types e.g. giving instructions, cause and effect, compare and contrast etc.

Course requirements:
1. Class attendance.
2. Class participation.
3. Examinations and quizzes.
4. Assignments and work sheets.

Instructional strategies:

1. Lectures
2. Assignments
3. Exercises, individual, pair and group work activities

Assessment:

1. Assignments and participation   10%
2. Quizzes   15%
3. Midterm exam   25%
4. Final exam   50%

Required textbooks:

1. Mazyad, S.S. English for Health Sciences: Lower Intermediate Level. 2nd Edition. Riyadh: QEH Publishing, 2009.

2. Mazyad, S.S. Academic Writing for Health Professions: Elementary level. 3rd Edition. Riyadh: QEH Publishing, 2009.

3. Mazyad, S.S. Understanding and Using Medical Terms: Elementary. 2nd Edition. Riyadh: QEH Publishing, 2009.

4. Suleiman S. Mazyad; Medical Consultant 2009. Pharmacy Terms in Common Use. 2nd Edition. QHE Publishing. Riyadh.

Supplementary Materials:

Related supplementary teaching material.

<u>Suggested Books:</u>

1. Prentice – Hall, 1992 <u>Grammar and Composition, Level 3.</u> Prentice-Hall, Inc, Englewood Cliffs, New Jersey.

2. Davis, F.A. 2005. <u>Taber's Cyclopedic Medical Dictionary</u>, F.A. Davis & Co.