

**Language Teacher Cognition on English Grammar Assessment:
Investigating EFL Teachers' Beliefs and Practices regarding
Classroom-based Assessment in Public Higher-Educational Institutions
in Saudi Arabia**

Masha'el Salem AlSalem

A thesis submitted for the degree of Doctor of Philosophy in Applied
Linguistics

Department of Languages and Linguistics

University of Essex

May 2020

Dedication

To my one and only

My late father: Salem

Abstract

This thesis investigates teachers' beliefs about English grammar assessment (EGA) and how these are linked to their professional practices when writing their grammar exams in their educational contexts, in higher educational facilities in Riyadh, Saudi Arabia. The central focus of this study is threefold: (a) teachers' beliefs and factors which shaped these beliefs, (b) their actual practices and the factors that influenced their EGA, and (c) the relationship between teachers' beliefs and practices regarding EGA.

The study is guided by the theoretical framework of Activity Theory (AT) and more specifically by Engeström's (1999) third generation of AT which was used as the interpretive tool to explore the two systems: how teachers' view EGA and how they actually assess EG in their classes. AT also allows to identify the contradictions that create conflicts between EFL teachers' beliefs and practices with regards to EGA. A mixed-method research design was used and included a questionnaire (N= 94), semi-structured interviews (N= 32), retrospective thinking (N= 20) and document analysis (N= 28) with EFL teachers in four public higher educational facilities.

The study showed both congruence and tensions between teachers' beliefs and practices. Teachers' beliefs were greatly congruent with their practices regarding the purposes of EGA, EFL teachers' role in constructing their grammar exams, preferable items format and the sources from which EFL teachers draw exam question . Conversely, teachers' beliefs were incongruent with their practices concerning how EG should be assessed: integratively vs. explicitly.

In addition, several factors: contextual, personal and conceptual, were identified as to have helped shaped , affected and/ or altered EFL teachers' beliefs and practices with regards to EGA.

This study concludes by providing some implications which could serve more than one purpose by creating knowledge which would be useful for researchers in the field of language teacher cognition and English grammar assessment.

Acknowledgments

All praise, appreciation and gratitude is due to almighty ALLAH for giving me the strength and perseverance to complete this work. I would like to offer my sincere thanks to the many people who have supported me throughout these years and who have provided me with all the help and guidance in completing my research.

First, I am forever grateful to my supervisor Dr. Adela Gánem-Gutiérrez whose support, guidance and encouragement helped me throughout the research process. I am also thankful to Dr. Christina Gkonou, Dr. Sophia Skoufaki and Dr. Tracey Costley from the ELT Research Group for their constructive criticism and prompt feedback during the various stages of my research presentations. Special thanks are also due to King Saud University for sponsoring my research scholarship and to all my fellow EFL teachers who participated in this study.

To my husband: Majed whose love, support and patience helped me endure the hardship of this journey, and to my two daughters: Toleen and Ajwan who encouraged me endlessly to finish my work.

Lastly, to my Mom and dear sisters who were always sympathetic and supportive. Thank you for your prayers.

List of Abbreviations

LTC	Language teacher cognition
EFL	English as a foreign language
EGA	English grammar assessment
AT	Activity Theory
SA	Summative Assessment
FA	Formative Assessment
AL	Assessment Literacy
AfL	Assessment for Learning
NCAAA	National Commission for Academic Accreditation and Assessment

Table of Contents

Dedication	i
Abstract.....	ii
Acknowledgments	iv
List of Abbreviations	v
Table of Contents	vi
List of Figures.....	xiii
List of Tables	xvi
1. Introduction.....	1
1.1 Overview of the Topic	1
1.2 Statement of the Problem.....	1
1.3 Significance of the Study and Research Questions.....	3
1.4 Brief Definitions of some key Terms.....	4
1.5 Thesis Structure	5
1.6 Summary of the Chapter	8
2. Review of Related Literature	9
2.1 Introduction.....	9
2.2 Brief History of Teacher Cognition Research	9
2.3 Language Teacher Cognition.....	10
2.3.1 An overview of LTC and definitions.....	10
2.3.2 Beliefs, practices and theoretical approaches.	17
2.4 Assessment.....	19
2.4.1 Overview and definition.	19

2.4.2 Assessment and theories of learning.....	26
2.4.3 Concepts and forms of classroom assessment	27
2.4.4 Assessment literacy (AL).....	30
2.4.5 Language teacher cognition on assessment.	34
2.5 English Grammar Instruction and Assessment.....	53
2.5.1 Overview and definitions.....	53
2.5.2 Studies on grammar assessment.....	54
2.6 Summary of the Chapter and Conclusion	56
3. Theoretical Framework.....	59
3.1 Introduction.....	59
3.2 Activity Theory (AT).....	59
3.2.1 An overview of AT and definition.....	59
3.2.2 Three generations of AT.....	60
3.3 Activity Theory in Applied Research	67
3.4 Activity Theory in the Present Study.....	70
3.4.1 The activity network.	71
3.4.2 Elements of the AT.....	73
3.4.3 Contradictions.....	75
3.5 Summary of the Chapter and Conclusion	78
4. Research Methodology	79
4.1 Introduction.....	79

4.2 Research Design and Rationale	79
4.2.1 Mixed-method methodology.....	81
4.2.2 Mixed-method design.	82
4.2.3 Mixed-method measures.....	83
4.3 Research Context and Participants.....	86
4.3.1 Overview of the higher-educational context in KSA.....	86
4.3.2 The participants.....	96
4.4 Data Collection Instruments and Procedures for data collection and analysis	98
4.4.1 Questionnaire.....	99
4.4.2 Semi-structured interviews.	110
4.4.3 Retrospective thinking.	120
4.4.4 Document analysis.....	126
4.5 Positionality statement: My role as the researcher	131
4.6 Triangulation.....	133
4.7 Research Quality	134
4.7.1 Validity.	134
4.7.2 Reliability.....	136
4.8 Ethical Considerations	138
4.9 The Pilot Study	139
4.9.1 Context and participants.	140
4.9.2 Instruments.....	143

4.9.3 Main findings.....	154
4.10 Summary of the Chapter and Conclusion	155
5. Results.....	159
5.1 Introduction.....	159
5.2 EFL Teachers' Beliefs about EGA and Factors that have Shaped these Beliefs.....	159
5.2.1 What are EFL teachers' beliefs about how English grammar should be assessed?	159
5.2.2 What are the factors which helped shape such beliefs?	203
5.2.3 Further exploration of the questionnaire data: Beliefs associations/connections with each other.....	212
5.2.4 Summary of Section 5.2.....	220
5.3 EFL Teachers' Practices of EGA and Factors that Influence these Practices	221
5.3.1 How do EFL teachers actually assess grammar in their teaching environments?	221
5.3.2 What are the factors which influence their practices other than their beliefs?	267
5.3.3 Summary of section 5.3.	273
5.4 The Relationship between EFL Teachers' Beliefs and Practices regarding EGA.....	274
5.4.1 Congruity between teachers' beliefs and practices in relation to EGA.	276
5.4.2 Incongruity between teachers' beliefs and practices in relation to EGA.....	278
5.4.3 Factors that influence the relationship between teachers' beliefs and practices in relation to EGA.....	279
5.4.4 Summary of section 5.4.	280
5.5 Summary of the Chapter and Conclusion.	280

6. Discussion	283
6.1 Introduction.....	283
6.2 The Activity Systems and The Activity Network.....	283
6.2.1 Subject.....	286
6.2.2 Object and Outcome	288
6.2.3 Mediating Artefacts (MA)	290
6.2.4 Division of Labour.....	294
6.2.5 Community	297
6.2.6 Rules	303
6.2.7 Summary of section 6.2	305
6.3 Contradictions	305
6.4 Summary of the Chapter	312
7. Conclusion, Implications and Suggestions for Future Research	313
7.1 Introduction.....	313
7.2 Key Findings and Contributions to knowledge	314
7.2.1 EFL teachers' beliefs about EGA and factors shaping these beliefs	314
7.2.2 EFL teachers' practices of EGA and factors influencing these practices.....	315
7.2.3 The relationship between EFL teachers' beliefs and practices with regards to EGA	317
7.2.4 Main contributions of the study	318
7.3 Pedagogical Implications and Recommendations	320
7.4 Limitations of the study	322

7.5 Potential for further research	324
7.6 Summary	326
References.....	327
Appendices.....	340
Appendix A: Project Information Sheet	340
Appendix B: Participant Consent Form.....	342
Appendix C: First Draft of the Questionnaire.....	343
Appendix D: Final Draft of the Questionnaire	348
Appendix E: Interview Questions for EFL teachers who currently teach Grammar (First Draft).....	353
Appendix F: Interview Questions for EFL teachers who currently hold the post as coordinators for the English Grammar Course in their departments (First Draft).....	355
Appendix G: Interview Questions for a) EFL teachers who currently teach English grammar course and b) current course coordinators, if any (Final Draft).....	357
Appendix H: Information about the Digital Voice Recorder.....	359
Appendix I: Dragon- speech to text software @ https://www.nuance.com/dragon/business-solutions/dragon-professional-individual.html	360
Appendix J: Sample of a Complete Interview Transcription.....	361
Appendix K: Document Analysis	367
Appendix L: Sample of MAXQDA Coding Outcome	368
Appendix M: Document Analysis (English Grammar Written Exams)	375
Appendix N: Sample of English Grammar Exam with Marks Allocation	382

Appendix O: Examples of Quantitative Analysis of the Questionnaire Data in the form of SPSS Output.....	384
Appendix P: Summary of Participants' Results from the Qualitative data	390

List of Figures

Figure 1: Structure of the thesis	6
Figure 2. The interrelationship between beliefs and knowledge.	12
Figure 3. Components of the LTC framework (Borg, 2006, p. 283).	16
Figure 4. The relationship among testing, assessment, evaluation and teaching.....	23
Figure 5. Conceptualization of assessment literacy with the cognitive and practical dimensions.	33
Figure 6. Continuum of beliefs and conceptions about the purpose of assessment. (Barnes, Fives, & Dacey, 2015, p. 287)	36
Figure 7. Schematic conceptualization of LTC on assessment.....	51
Figure 8. Variables that determine the importance of grammar (Celce-Murcia, 1991, as cited in Brown, 2000, p. 363)	54
Figure 9. Vygotsky’s model of mediated action. (Engeström, 1999).....	62
Figure 10. Engeström’s model of activity system. (Engeström, 1987, p 78)	64
Figure 11. Engeström activity networks. (Engeström, 2001, p 136).....	66
Figure 12. Structural model of activity network of EFL teachers’ beliefs and practices with regard to English grammar assessment.....	72
Figure 13. The convergent prototype of mixed-method research design. (Terrell, 2011, p. 267)	83
Figure 14. Current research design.	85
<i>Figure 15.</i> Overall representation of the research participants.....	98
Figure 16. Participants’ place of work.....	103
Figure 17. Participants’ experience of teaching and assessing English grammar.	104
Figure 18. oTranscribe template and layout.	117
Figure 19. General overview of the coding system in MAXQDA 2018.	118

Figure 20. Themes and categories derived from the data analysis of the interviews.	119
Figure 21. Coding procedure for the data gathered from the retrospection.....	124
Figure 22. Categories and themes generated from retrospective thinking.....	125
Figure 23. Resultant themes and categories from document analysis.	131
Figure 24. Scholfield’s formula for inter-rater reliability agreement (2005).....	137
Figure 25. Teachers’ general beliefs and attitudes about English grammar assessment.	144
Figure 26. Teachers’ beliefs about the purposes of English grammar assessment.....	145
Figure 27. Teachers’ beliefs about English grammar assessment methods.....	146
Figure 28. Teachers’ beliefs about English grammar assessment formats.	147
Figure 29. Teachers’ beliefs about their roles and the sources used in constructing English grammar assessment tasks.	148
Figure 30. A screen print of how the coding system and the transcript documents are presented in MAXQDA.	150
Figure 31. Themes and sub-themes developed through coding.....	151
Figure 32. Participants’ responses to item no.5 in the questionnaire.....	164
Figure 33. Summary of data results for the purposes of EGA.....	171
Figure 34. Summary of data results for EGA methods.....	185
Figure 35. Summary of data results on teachers’ beliefs about sources that facilitate the construction of EGA items.....	193
Figure 36. Results from the questionnaire and the interviews for who best construct EGA.	201
Figure 37. Teachers’ beliefs about EGA: Themes and categories from the questionnaire and the interviews.	202
Figure 38. Differences in beliefs between Saudi and non-Saudi teachers.	204
Figure 39. Screen plot of the unrotated factors underlying responses to the 46 questionnaire items.....	213

Figure 40. Written exams conducted by the EFL teachers during grammar courses.	223
Figure 41. Section of the exam paper showing the name of the assessment component and assigned a maximum score.	227
Figure 42. The interrelationship between EFL teachers' current practices of EGA and contextual factors.	273
Figure 43. Themes identified to understand the relationship between EFL teachers' beliefs and practices of EGA.	275
Figure 44: Representation of the activity network and themes of the study as AT components	284

List of Tables

Table 1. Realization of formative and summative assessment	29
Table 2. Higher-educational facilities with English language programmes	88
Table 3. Grammar Courses and Assessment Means in Current Research Context	93
Table 4. Data Collection Timeline.....	99
Table 5. Questionnaire Categorical Themes	102
Table 6. Background Information about the Participants in the Questionnaire.....	105
Table 7. Participants' Relevant Classroom Backgrounds.....	106
Table 8. Themes for the Interview Guide and Questions	113
Table 9. Interview Participants' Demographics.....	115
Table 10. Retrospection Participants' Demographics.....	123
Table 11. Participants and Documented Information	129
Table 12. Content analysis of exam papers	130
Table 13. Inter-rater Reliability of the Coding	137
Table 14. English Language Diploma Programme, Female Sector.....	140
Table 15. Grammar Courses Covered in the English Language Intensive Course, Female Sector.	140
Table 16. English Language Diploma Programme, Male Sector.	141
Table 17. EFL Teachers' Demographic Information.....	141
Table 18. Summary of Research Design, Questions, Instruments and Objectives and Means of Data Analysis.....	156
Table 19. General Beliefs of EGA.....	160
Table 20. Results related to the Purposes of EGA.....	166
Table 21. Results for EGA Approaches, Means of Delivery and Assessment Frequency	172
Table 22. Beliefs about Suitable Format for EGA.....	186

Table 23. Beliefs about Sources to Construct Assessment Items/tasks	189
Table 24. Beliefs about Teachers' Role in EGA Construction	194
Table 25. Factor Analysis Loadings of 46 Beliefs about EGA.....	214
Table 26. EGA Item Formats from the Teachers' Exam Samples	229
Table 27. Summary of Data Findings	281
Table 28. Summary of themes as AT components	285

1. Introduction

1.1 Overview of the Topic

Language teacher cognition (LTC) has become an established field of inquiry in educational research since the mid-1990s. LTC is concerned with teachers' mental constructs and how their mental lives inform their teaching practices and decision-making processes (Borg, 2003b, 2003a, 2009). Interest in LTC became evident following the realization that teachers are not just policy implementers, performers or passive instructors who deliver others' ideas, but they are precisely 'active thinking decision-makers' who continuously draw upon their practical and personalized networks of 'thoughts, knowledge and beliefs' (Borg, 2003, p. 81). Accordingly, many studies investigating LTC have been carried out, particularly in the domains of grammar and literacy teaching. However, few (if any) studies, to date, have targeted LTC in relation to teacher-made assessment (Mansory, 2016). Given the considerable importance and impact of this activity in relation to ELT instruction, this study explores English as a Foreign Language (EFL) teachers' beliefs and practices with regard to constructing written English grammar examinations.

1.2 Statement of the Problem

Assessment lies at the heart of many educational practices. In Saudi Arabia, all higher educational facilities – universities, community colleges and institutions – rely heavily on formal assessments, which occur systematically and often take the form of final and midterm exams.

Based on higher-education regulations, undergraduate students must take at least one written midterm exam and a final written exam for each course they take up. Passing these exams not

only serves the important pedagogical goal of ensuring students' content-based knowledge on specific subject domains but is also an essential requirement for graduation. Although this form of classroom-based assessment is not what can be traditionally considered high-stakes, it nonetheless affects the lives of students who take these exams and may suffer serious consequences as a result of failing them.

The setting, administration and scoring methods of these exams are primary responsibilities of a course teacher. This places huge pressure on such teachers, as they have to spend a considerable amount of time (a quarter or a third of their working time) devising their course assessment instruments, scoring the examinees, and reporting the results. Furthermore, this task could become complicated if a teacher is placed in charge of different sections and has to devise different versions of the required exams. To complicate matters, sometimes teachers share the same course but teach different groups; in this case, each teacher might have her/his own views about the exam structures, however, they have to collectively set one final exam. This may, then, result in a lengthy examination writing process, because any teacher may repeatedly seek to modify, delete, or include items in the written exams.

Drawing on my own experience and based on the literature reviewed in chapter 2, I believe that virtually all EFL teachers may have their own ways of constructing their course exams, and that these practices, more often than not, would be influenced by their individual beliefs about the same, their educational and cultural backgrounds, and their experiences and various other factors. In turn, their beliefs may or may not be precisely put into practice, due to a range of other factors, such as the features imposed by the educational sector, what co-constructors of these exams would allow in an examination, what the students expect etc. Thus, in this study, I aimed to explore both the observable and unobservable components

underlying the construction of grammar course exams in four higher-educational facilities situated in Riyadh, Saudi Arabia, at three levels:

1. EFL teachers' beliefs and what affects those beliefs about how English grammar (EG) assessment should be constructed.
2. EFL teachers' practices and factors that influenced those practices of assessing EG.
2. How beliefs and practices are intertwined and what factors in addition to beliefs influence those practices.

1.3 Significance of the Study and Research Questions

As the literature review (chapter 2) shows, there have been some studies on what goes on in the minds of professional testers, especially in areas such as writing assessments.

Furthermore, scholars have studied teachers' beliefs about grammar instruction. However, as stated previously, little research to date has specifically focused on the knowledge and beliefs that underpin ordinary EFL teachers' assessment practices in the domain of grammar.

Furthermore, no related study has been conducted that grapples with both aspects together: teachers' beliefs about, as well as their practice of, grammar assessments through exams. Still less research has been conducted on EFL teachers at higher-educational levels in Saudi Arabia. In addition, although there has been considerable research on teachers' cognition concerning assessment in other respects (section 2.4.4), this study is the first one to investigate EFL teachers who are responsible for constructing their own exams, focusing both on their cognition and how it is transferred into their exam writing practices. Therefore, this research offers an opportunity to explore EFL teachers' beliefs and practices with respect to constructing English grammar examinations. It aims to identify the relationship between their cognition and practices and examine the factors involved in shaping both.

In order to gain a better understanding of language teacher cognition about English grammar exam writing and EFL teachers' actual practices in Saudi higher-educational facilities in Riyadh, this doctoral project addresses the following questions:

RQ1:

- a) What are EFL teachers' beliefs about how English grammar should be assessed?
- b) What are the factors which have helped shape these beliefs?

RQ2:

- a) How do EFL teachers actually assess grammar in their teaching environments?
- b) What are the factors, other than their beliefs, that have influenced their practices?

RQ3:

- a) What is the relation between EFL teachers' beliefs and their current practices?
- b) What are the factors which have led to a convergence or divergence between such beliefs and practices?

1.4 Brief Definitions of some key Terms

Language Teacher Cognition (LTC) is understood as being the thoughts, beliefs and knowledge of language teachers about teaching language (Borg, 2003). In an updated definition of LTC, Borg (2006) adds attitudes, identities and emotions as further unobservable aspects that are also included within LTC.

This study concerns teachers of English as a foreign language (EFL) which is understood to refer to English taught in a country where the language has no regular currency in the day to

day life of inhabitants, so learners get little casual out of class exposure to it unless they seek it on the internet or English medium satellite TV etc.

High-stakes examinations or other assessments are those where significant consequences follow from the performance of students, such as that they do or do not progress to the next educational level, enter university, or get a particular job.

Assessment is a broader term than test or examination: it is seen as '... an ongoing process that encompasses a much wider domain' (Brown, 2003).

Classroom-based Assessment or Formal Assessments '...are exercises or procedures specifically designed to tap into a storehouse of skills and knowledge. They are systematic, planned sampling techniques constructed to give teacher and student an appraisal of student achievement' (Brown, 2003).

1.5 Thesis Structure

Overall, this study is structured into seven chapters as shown in Figure 1. Each chapter systematically discusses essential aspects regarding this research. The following paragraphs provide an overview of each chapter in this thesis.



Figure 1: Structure of the thesis

Chapter 1 presents an introduction, offering a general overview of the study, including the importance of this project, its contribution and the research questions. Chapter 2 focuses on the literature related to the present study. This chapter consists of three main parts: LTC, assessment and English grammar instruction and assessment. The first part outlines the background of LTC in educational research, with subsections discussing definitions, the relationship between beliefs and knowledge and the factors influencing LTC. The second part gives particular attention to assessment, highlighting studies that have focused on both teachers' beliefs and assessment practices. The final part addresses the gap in the existing literature and language education with regard to grammar assessment and teachers' beliefs, especially in Saudi Arabia.

Chapter 3 presents the theoretical framework of Activity Theory (AT), which is used to discuss and interpret the findings in this thesis. It begins with an overview of the development of AT, briefly presenting its three generations and how AT has been implemented in research studies. Six elements – subject, object, tools/artefacts, division of labour, community and rules – are outlined and discussed, allowing the reader to gain a better understanding of how they are used in this study.

Chapter 4 outlines the methodology and the rationale behind the choice of the research design. The chapter then describes the participants, the context, the data collection instruments and the procedures. In addition, the pilot study is presented and explained in detail to provide a better understanding of how this helped shape the main study. Finally, it explains the analytical process through which I interpreted the findings and finishes with issues related to the trustworthiness of research and ethical considerations.

Chapter 5 presents the findings and answers the research questions sequentially. First, findings from the questionnaire and the interviews designed to answer RQ1 are presented. The chapter continues by describing the findings from interviews, retrospective thinking and document analysis to answer RQ2 and RQ3.

Chapter 6 focuses on interpreting the findings using the framework of Activity Theory, in which the practice of writing exams is socially situated within the institutional contexts, explicating the relationship between teachers' beliefs, practices and the various elements of the activity systems: individuals, organizations and artefacts. The chapter then discusses the findings of the research and relates them to the wider literature.

Finally, chapter 7 concludes the study by highlighting the key findings and their contribution to the extant body of knowledge, followed by a discussion of the implications, limitations, and suggestions for future research.

1.6 Summary of the Chapter

This chapter highlighted the need for the present study. It introduced the research aims and the questions to be investigated. It has also outlined the significance of the study along with its potential contribution to its field of study. An overview of the whole thesis was also presented. In order to fully understand teachers' beliefs about and their current practices of English grammar assessment, it is necessary to understand the body of research related to this study and to highlight the gap that has informed this project, which is the subject of the next chapter.

2. Review of Related Literature

2.1 Introduction

In this chapter, I review the relevant literature that has informed and directed my study on EFL teachers' beliefs and practices about grammar assessment procedures, how they develop their beliefs, and how these beliefs and practices interact with each other. I first shed some light on the history of teacher cognition as a research area. Then, I outline my understanding of LTC. Further, I review the research on language assessment/examination, with a particular focus on teachers' beliefs and practices concerning assessment (in general) and English grammar exam assessment (in particular).

2.2 Brief History of Teacher Cognition Research

In earlier years, such as the 1960s, the focus of educational research on teachers was on teacher behaviours and effectiveness (process-product approach). Much research during that time focused mainly on answering one question: 'What do teachers do?' Later, in the 1970s, advances in cognitive psychology highlighted the influence of teacher thinking on their behaviors and led to the birth of research on teacher thinking as we know it today. Gradually, the field of teacher thinking allied more with educational research than with cognitive psychology. The abovementioned question no longer remained the exclusive concern of educational researchers, but other questions came to the fore: 'What do teachers think?', 'What decisions do they make?' Most importantly, 'Why?' As a result, many research studies were conducted throughout the 1980s and 1990s that examined various aspects of the unobservable dimension of teaching. The significant findings of mainstream research generated three main outcomes:

- i. redefining the domain and purposes of teacher thinking research,

- ii. the emergence of the term ‘teacher cognition’, which has become more generally accepted and, in most cases, is used with reference to beliefs, and
- iii. a more sophisticated understanding of the relationships between teachers’ cognition, practices and other aspects of their work (Borg, 2009).

2.3 Language Teacher Cognition

2.3.1 An overview of LTC and definitions.

The field of LTC has become an area of thriving research in secondary and foreign-language education over the last three decades. According to Borg (2003), LTC was established in the mid-1990s and, ever since, has received considerable interest from researchers in different disciplines: applied linguistics, English language teaching and second and foreign language education. The bulk of research carried out in LTC has yielded valuable insights on how teachers’ perceptions (covert lives) affect and are related to their teaching practices (overt lives). The study of LTC has developed a greater understanding of the following:

How language teachers conceive of what they do: what they know about language teaching, how they think about their classroom practice, and how that knowledge and those thinking processes are learned through formal teacher education and informal experience on the job (Freeman and Richards, 1996, p. 1).

2.3.1.1 Beliefs, Thoughts and Knowledge: Conceptual Issues.

All words begin as servants, eager to oblige and assume whatever function may be assigned to them, but, that accomplished, they become masters, imposing the will of their predefined intention and dominating the essence of human discourse.

Pajares (1992, p. 308)

Investigations on teacher cognition in language teaching institutions mainly concern three key concepts: beliefs, thoughts and knowledge (especially practical or pedagogical content knowledge). The terms used in relation to LTC are, however, quite ambiguous, complex and problematic (Kagan, 1990).

One key confusion and complexity arises when drawing a distinction between teacher's knowledge and beliefs (Borg, 2006; Pajares, 1992). Some researches argue in their definitions of teacher's beliefs that these two concepts are different (e.g. Nespor, 1987; Rokeach, 1968). Pajares (1992) stated that the most common distinction used in teacher's beliefs definitions is that 'belief is based on evaluation and judgment; knowledge is based on objective fact' (p. 313). However, a number of researchers consider knowledge as a personal construct, denying the positivist stance that there exists any objective, external truth; for them, teachers' knowledge and beliefs are 'inextricably intertwined' in their minds (Verloop, Van Driel, & Meijer, 2001, p. 446), which makes it impossible for teachers to separate them (Borg, 2003; Woods & Cakır, 2011). Therefore, Woods (1996) and later Woods and Cakır (2011) treat the relation between knowledge and beliefs as a continuum, referring to it as BAK (beliefs, assumption and knowledge). At one end of this continuum, we have beliefs, referring to personal knowledge that can be either explicitly articulated or implicitly embedded in action, while at the other end, we have impersonal and factual knowledge (Woods & Cakır, 2011).

Based on my understanding of the issues, my position is that belief and knowledge are separate but in direct mutual relationship with each other. Sometimes, when we ask someone about what he/she believes about X or Y, s/he may say 'I do not know'. To clarify this relation, I would say that prior knowledge (what we learn from the very early years of our

lives) forms the basis of our beliefs, for we cannot hold a belief about something we absolutely do not know. As we grow up, the knowledge that we receive from various sources (schools, home, society etc.) continues to nourish our beliefs. As we mature, so do our beliefs, and specific belief systems begin to operate. Beliefs, then, filter any current knowledge we come across (outputs) and transfer the same either into input knowledge, which is another piece of information we know about any subject matter, or into intake, which will either add to, refine, or alter our belief systems. Thus, as shown in Figure 2, belief and knowledge operate in a synchronized manner.

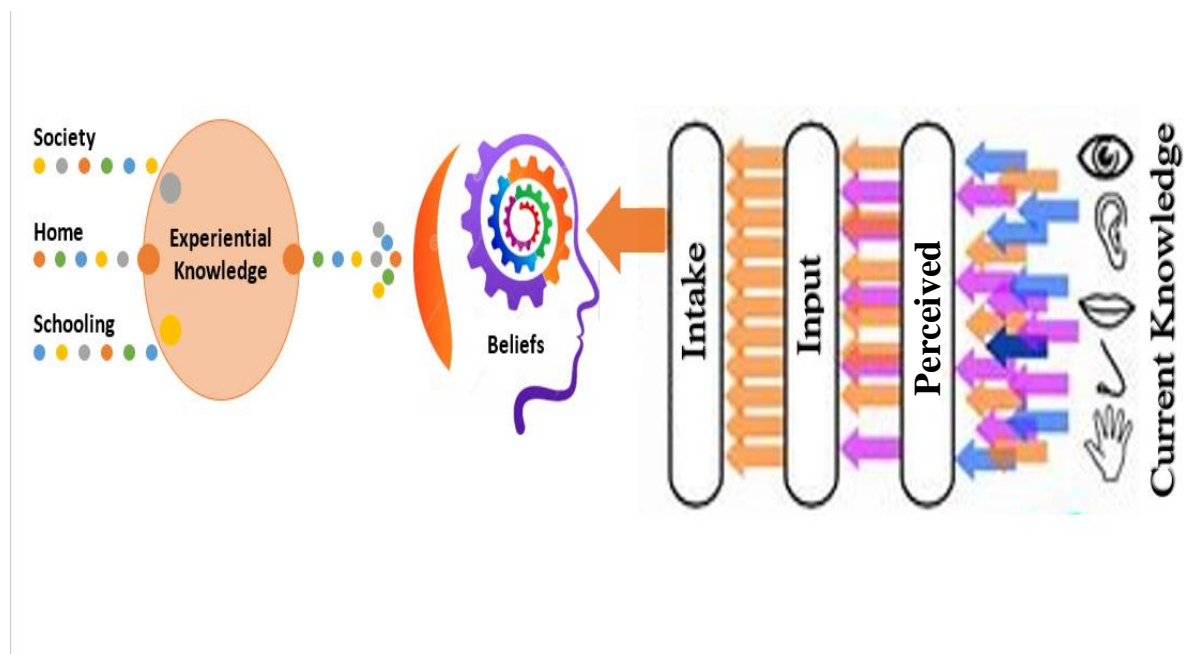


Figure 2. The interrelationship between beliefs and knowledge.

To this end, I define belief as an internal, personal conviction about what is real and true.

Knowledge is factual information about any subject matter. Now, I shall discuss each concept in more detail.

2.3.1.1.1 Beliefs.

The term ‘belief’ in educational research has been used interchangeably with other expressions: attitudes, values, judgments, axioms, opinions, ideology, perceptions, conceptions, conceptual systems, preconceptions, dispositions, implicit theories, explicit theories, personal theories, internal mental processes, action strategies, rules of practice, practical principles, perspectives, repertoires of understanding and social strategy (Pajares, 1992).

Many researchers, including myself, have attempted to make a distinction between beliefs and knowledge (2.3.1.1), others further emphasize the importance of distinguishing between tacit and stated beliefs. Argyris and Schon (1974) state that an individual’s ‘theory of action’ consists of both an ‘espoused theory’ and a ‘theory in use’. Espoused theory comprises a set of stated beliefs, which are defined as ‘statements teachers made about their ideas, thoughts, and knowledge that are expressed as evaluations of “what should be done”, “should be the case”, and “is preferable”’ (Basturkmen et al., 2004, p. 244). Thus, this type of belief articulates teacher’s perceptions of ideal practices that may or may not be reflected in their actual practices (Phipps & Borg, 2007). In addition, this type of belief is usually informed by teachers’ technical knowledge about teaching (i.e., received theory) (Biggs, 1994; Phipps & Borg, 2009) as described in 2.3.1.2. Thus, these stated beliefs can be equated with terms such as perceptions and attitudes (El-Okda, 2005). This is what the term ‘belief’ most often signifies for researchers such as Borg. Hence, in this study, I take the term ‘belief’ to denote teachers’ ideas as to how grammar should be assessed in their educational contexts.

‘Theory in use’, on the other hand, comprises a set of tacit beliefs underlying teacher’s actual classroom practices or, in our case, examining practices. Hence, in many studies, researchers

generally infer these from information about what teachers actually do in terms of practice. They are typically generated from teacher's experiences as learners and teachers as well as from their reflections on these experiences (Borg, 2006). They are primarily implicit and can only become explicit through reflection, a process which may end up changing a teacher's espoused theory (Ellis, 2012; El-Okda, 2005). There are different terms in the literature that have been used to refer to teachers' tacit beliefs: 'personal practical knowledge', defined as 'moral, affective, and aesthetic way of knowing life's educational situation' (Connolly & Clandinin, 1987, p.59); 'practical knowledge' (Elbaz, 1981), defined as knowledge that is 'personal, situated, based on reflection on experience, mainly tacit and content-related' (Meijer et al., 2002, p. 407). 'Ability', defined as a type of knowledge that is 'implicitly embodied, experientially-derived, and unconsciously or "automatically" instantiated' (Woods & Cakir, 2011, p. 348); and 'knowing in-action', defined as actions, recognitions and judgments that professionals carry out both spontaneously and based on their tacit knowledge of situations (Schon, 1983).

2.3.1.1.2 Knowledge.

According to Shulman (1987 as cited in Suwannasom, 2010), knowledge in the teaching context is an amalgamation of two sets of pedagogical knowledge: pedagogical content knowledge and pedagogical craft knowledge. The former, also known as subject content knowledge, refers to the body of information and skills that teachers are expected to deliver to students in a given subject or content area, such as English, mathematics, science etc.

It has been argued that a teacher with a rich understanding of some content is more likely '...to detect student misconceptions...to deal effectively with general class difficulties...' (Shulman & Richert, 1993, p. 109, as cited in Guthrie, 2005, p. 51). In the present study,

relevant knowledge is referred to as assessment literacy (2.3), both in the form of practical performance ability (spontaneous knowledge of how to construct classroom-based assessments) and at the level of metalinguistic knowledge (explicit knowledge or awareness of grammar and the concomitant terminology, rules etc.). Clearly, a teacher needs these kinds of knowledge to set good grammar exams as well as to teach grammar itself.

Pedagogical craft knowledge, on the other hand, refers to ‘the ability of the teacher to present subject information to pupils’ (Gutbrie, 2005, p. 52). This means that teachers must possess or master the skill of communicating knowledge to others. To achieve successful content communication, teachers ‘need a good repertoire of teaching styles, practices, and approaches to enhance content transfer to students coupled with sound management of classroom surroundings’ (Gutbrie, 2005, p. 52–53). In my study, an equivalent of such communication is a teacher’s performance in exam administration and/or her/his repertoire in answering students’ questions, as well as marking and feedback skills.

Together, these two types of knowledge can be used to evaluate teacher competence or professional competence: the level of their command or mastery over content knowledge and their skills, attitudes and experiences that are required to exploit such knowledge pedagogically. In this study, I examine whether teachers’ pedagogical content knowledge – assessment literacy – has any impact on how they perceive English grammar classroom-based assessments, along with how they implement their knowledge as well as beliefs in classroom-based assessment practices (detailed discussion in 2.4.3).

2.3.1.2 Factors influencing LTC.

Within the ambit of studying teachers' beliefs and practices, there has been considerable interest in investigating factors that contribute to shaping teachers' cognition (Borg, 2008). Borg defines contextual factors as 'the social, psychological and environmental realities [institutional, instructional and physical settings in which teachers work]' (Borg, 2003, p. 94). According to Borg (2015), contextual factors can interact with teachers' mental constructs either by changing their beliefs and even knowledge, or by altering their practices without changing the beliefs underlying them (see Figure 3). I regard his view as constituting a model or theory that can clearly be applied to classroom-based assessments even though the captions in his diagram do not explicitly cover assessment; a modified version is provided later, in 2.4.5.4.

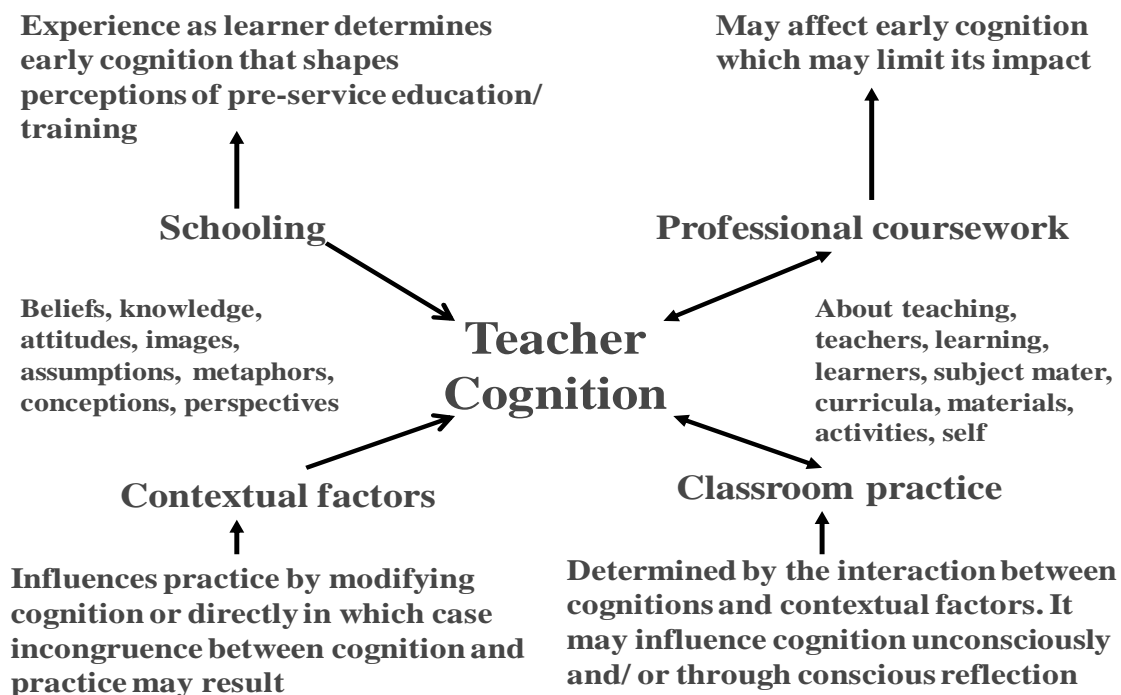


Figure 3. Components of the LTC framework (Borg, 2006, p. 283).

The figure above diagrammatically shows how a teacher's cognition is established early on during her/his schooling experiences, which might continue to be influential throughout

her/his professional life. LTC about teaching, learning and indeed assessment may be altered later as a result of teaching experiences and any teacher training or professional development received. However, when teachers are at work, they may be faced with some contextual factors (for example, curricula, educational system policies etc.) that can impact their practices and dismantle the congruity between those practices and their underlying beliefs. Meanwhile, teachers' ongoing interactions with classroom factors may also begin to influence their tacit beliefs.

Numerous studies in educational research have shown that teachers' knowledge and beliefs provide the basis for their actions and guide their classroom practices (Basturkmen, Loewen, & Ellis, 2004; Borg, 1999, 2003; Richards, Gallo, & Renandya, 2001; Saydee, n.d.). Their findings indicate that this relationship is interdependent and may or may not be congruent (Borg, 2003). This is explored specifically in relation to classroom-based assessments below.

2.3.2 Beliefs, practices and theoretical approaches.

As seen from the account above, there is a widely held view that beliefs play a central role in teachers' decisions, judgments and behaviours. In particular, they heavily influence pedagogical decision-making and thus inform their practices (Borg, 2003, 2006; Farrell & Kun, 2008; Golombek, 1998; Johnson, 1994; Ng & Farrell, 2003; Pajares, 1992; Johnson and Golombek, 2018). Theories and approaches related to teachers' beliefs and practices in educational contexts are, more often than not, affiliated with a cognitive perspective, with beliefs viewed as fixed assumptions that represent teachers' mental lives (e.g. Golombek, 1998; Ng & Farrell, 2003; Nishino, 2008; Pajares, 1992). In this tradition, research tends to focus on the realm of the reality inside teachers' heads. However, it might not be appropriate to view beliefs from a single theoretical stance, because a cognitive-based perspective is too

narrow in its understanding of the contexts or the interactive nature of teachers' daily practices in their teaching environments.

Although research on LTC has, to some degree, acknowledged the influence of context on teachers' beliefs and practices, it tends to operate at a macro-level, associated with issues such as curriculum, students, educational policies and school cultures (e.g. Sato & Kleinsasser, 2004). While such studies are clearly valuable, they may not always provide adequate insights into what happens at the micro-level, when teachers are engaged in specific practices, such as mental activities while constructing classroom-based assessment (pertaining to this study).

As discussed earlier, the predominant cognitive approach to research on teachers' beliefs (e.g. Golombek, 1998; Lee, 2009; Ng & Farrell 2003; Nishino, 2008; Pajares, 1992) does not sufficiently take into account the fact that teachers' beliefs interact with the contexts in which they work (Walsh, 2006). Consequently, this cognitive paradigm has been challenged by other perspectives. For example, instead of viewing beliefs as static traits of a person, which remain constant across situations, from an interactionist perspective beliefs are viewed as entities that may be transformed by or even emerge as a result of teachers' interactions with their respective contexts (Skott, 2001). In this view, beliefs are the products of social interaction rather than the rubric of reality maintained by teachers.

From an interactionist perspective, beliefs constitute complex interactive systems that can be studied through teachers' interactions with their contexts at both macro- and micro-levels. Indeed, teachers can hold beliefs about many aspects, such as learners, curricula, teaching and learning, professional development and the self, etc. – importantly, all of them are

intertwined with each other and are multi-faceted (Breen et al. 2001; Calderhead, 1996; Li, 2008). The interactionist perspective places emphasis on examining beliefs as they relate to the evidence of participating teachers' classroom practices. However, it does not address the methodological issue of achieving a shared understanding of the relationship between beliefs and practice, as it takes little or no account of a teacher participant's interpretation of her/his classroom practice.

In view of this gap, a much more holistic approach is needed to provide the theoretical basis for understanding the relationship between teacher beliefs and practices. A major goal of the present study is to offer a fine-grained interpretation of the relationship between teachers' beliefs and their practices with respect to constructing classroom-based assessments, by drawing on Activity Theory (AT). Within sociocultural perspectives on language teaching and learning, AT is a theoretical framework that can be used to understand teachers' practices as a whole, by investigating them and looking at their development from different angles (e.g. their beliefs, other actors in the system, teachers' prior experiences and so on), all of which can help fully understand their practices (Johnson, 2009). Therefore, and as detailed in Chapter 3, this study uses cultural-historical Activity Theory (CHAT) as an interpretative framework, as it has the capacity to capture significant elements within the broader context of classroom-based assessment, insofar as it allows to 'construct a holistic view of human activities as well as human agency within these activities' (Johnson, 2009, p. 78).

2.4 Assessment

2.4.1 Overview and definition.

Educators and researchers in the field of education consider assessment as a core pillar of the teaching and learning processes (Karim, 2015). According to Cohen (2001), assessment is

one of the least understood areas of teaching and learning. There is a substantial body of research on assessment that examines aspects of assessment design and implementation, practicality, accuracy and availability, rating scale, rating processes and rater training (Brown, 2003; Downing & Haladyna, eds, 2006). However, the majority of this research has focused on large-scale and often high-stakes exams. Little research has been conducted on teachers' cognition with regard to assessment and their role in constructing assessment tasks (Qian & Cumming, 2017). Therefore, it is the aim of this study to investigate teachers' beliefs related to assessment and how these beliefs are connected to their practices, specifically when designing assessment tasks.

As mentioned earlier, assessment is of central importance in education, and yet there is a lack of common agreement in defining this term (Taras, 2005). According to Brown (2003, p. 15), 'assessment is sometimes a misunderstood term in current educational practices'. The terms 'assessment', 'evaluation' and 'testing' seem to be used interchangeably in some educational circles (Brown, 2003; Taras, 2005). However, some researchers in the field of language assessment prefer to make clearer distinctions between these terms (Purpura, 2016; Hughes, 2011; Bachman & Palmer, 2010; Linn & Miller, 2005). Accordingly, there are a number of definitions that differentiate what each term means.

Evaluation:

'(It) involves making value judgments and decisions on the basis of information'
(Bachman & Palmer, 2010, p 21).

This is arguably a very broad category, since such judgments might be based on either objective information or quite subjective and unreliable information, such as personal

recollection of students' class performances. Furthermore, evaluation can be done by a learner by him/herself and does not necessarily have to be done by a teacher/examiner.

Assessment:

‘It is a broad term referring to a systematic procedure for eliciting test and non-test data for the purpose of making inference or claims about....characteristics of an individual’ (Purpura, 2016, p. 191).

‘It refers to a judgement which can be justified according to specific weighted set goals, yielding either comparative or numerical ratings (Scriven, 1967, p. 40 as cited in Taras, 2005, p. 467).

‘...the collection of information, both quantitative and qualitative, obtained through various tests, observations, and many other techniques (e.g., checklists, inventories), that is used to determine individual or group performance’ (Doran, Lawrenz, & Helgeson, 1994 as cited in Wang, 2004).

‘A systematic approach to collecting information and making inferences about the ability of a student or the quality or success of a teaching course on the basis of various sources of evidence’ (Schmidt, 2010, p 35–36, as cited in Mansory, 2016, p. 29).

All these definitions refer to some objective and systematic form of evidence being used as a basis for evaluation, whether it is quantitative or qualitative. Assessment, then, refers not only to formal tests, such as the TOEFL or an end-of-chapter evaluation, but also to other methods of obtaining information about KSAs (knowledge, skills and abilities), such as by observing

L2 performance during pair-work or by asking learners to report their understandings and uncertainties (Purpura, 2016, p. 191). Hence, students can assess each other (peer-assessment) or themselves (self-assessment), instead of being evaluated by teachers or testers.

Testing:

‘An instrument for measuring a sample of behaviour’. (Linn & Miller, 2005, p. 26)

‘Tests are prepared administrative procedures that occur at identifiable times in a curriculum when learners muster all their faculties to offer peak performance, knowing that their responses are being measured and evaluated.’ (Brown, 2003)

‘(It) is a particular type of assessment that typically consists of a set of questions administered during a fixed period of time under reasonably comparable conditions for all students’. (Doran, Laurenz, & Helgeson, 1994, as cited in Wang, 2004. p. 15)

This seems to be the most specific term, as it implies an objective and systematic form of evaluation, normally conducted by a teacher or tester, which is not only quantitative but also involves multiple measurements from each student (a test typically has multiple items to yield a score from each student on a sample of information, and it does not rely on a single response from each).

Based on the definitions listed above, it is possible to say that tests comprise a subset of assessment. The relationship among test, assessment, evaluation and teaching process can be illustrated by the following figure.

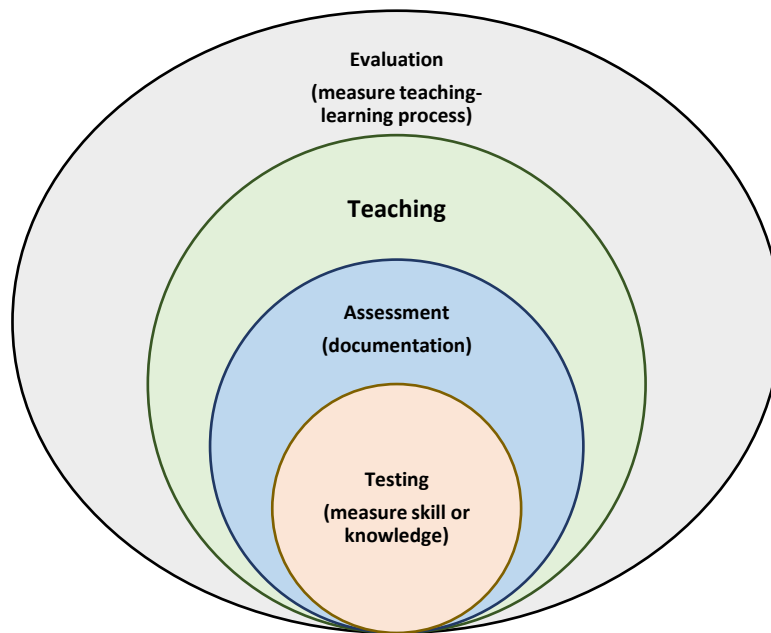


Figure 4. The relationship among testing, assessment, evaluation and teaching.

With the lexical conundrum disentangled by distinguishing among tests, assessments, and evaluations, I now consider some key issues of assessment and their relation to this study.

2.4.1.1 Informal and formal assessment.

According to Brown (2003), assessment falls into two main categories: informal and formal assessment. Informal assessment is usually embedded in teachers' classroom practices, with the purpose of eliciting performance without recording results and making fixed judgments about each student's competence. Forms of informal assessment may include incidental, unplanned comments and responses, along with coaching and other spontaneous feedback to students; e.g. 'Nice job!' Formal assessments, on the other hand, 'are exercises or procedures specifically designed to tap into a storehouse of skills and knowledge' (Brown, 2003). They are often taken periodically in any given educational course; tests are thus formal

assessments. However, not all formal assessments take the form of tests; other forms may include students' journals or portfolios and oral presentations.

2.4.1.2 Norm-referenced and criterion-referenced tests.

Further classifications need to be done to sort out the common terminology related to assessment. It is essential here to make a distinction between norm-referenced and criterion-referenced assessments. The former ranks students based on test scores, usually targets large number of students and takes a long time to be completed (e.g. IELTS or TOEFL). The latter focuses on measuring the skills and knowledge that a group of students has mastered in a certain domain and usually lasts for a class period (e.g. midterms, quizzes and final exams).

2.4.1.3 Formative and summative assessment.

Another useful distinction to bear in mind is the function of an assessment. In the relevant literature, two major functions of assessment are often identified: summative assessment, which is also called the assessment of learning, and formative assessment, also known as the assessment for learning (Büyükkarcı, 2014; Purpura, 2016; Taras, 2005). Both refer to assigning scores or grades to learners after imparting some instruction (i.e., teaching).

However, summative assessment, which is also described as sequential, is where teachers assess how well learners have succeeded in achieving the set educational goals and objectives of a course only after the instruction is completed, hence is assessment of learning. This is typically done to certify to the authorities whether each learner has reached a threshold standard or not, which in turn informs whether a student can progress and ultimately graduate. Formative assessment, on the other hand, refers to the assessment of ongoing learning that occurs at some point during a course (Fulcher, 2010). According to Büyükkarcı (2014), formative assessment provides information and feedback to help inform teacher

instruction and improve student learning over the remainder of a course, hence it is for learning. In other words, the information shared in the course can benefit both teachers and learners, rather than other authorities.

Earl (2003) acknowledged these two approaches and introduced a third approach, 'assessment as learning'.

In this type of assessment, the student is actively engaged in making sense of information and relating it to his or her prior knowledge and in mastering the skills involved. Making sense of the process is called metacognition. It occurs when students personally monitor what they are learning. They use the feedback from this monitoring to make judgments, adaptations and even major changes in what they understand (Earl, 2003, as cited in Thomas, 2012, p, 105).

This type of assessment, however, falls outside the scope of the present study as it does not involve teachers or exams.

The ongoing discussion intends to grasp some of the common assessment aspects and terms. On its part, this study focuses on classroom-based assessment – midterm exams, quizzes and final exams written by English grammar teachers, which take place within the context of a course. Classroom-based assessment has been referred to in the relevant literature by several terms: 'traditional assessment', 'classroom-oriented assessment' (Brown, 2003) and 'in-class assessment' (Neumann, 2010). It is classified as a formal, norm-referenced and summative assessment that is the preferred assessment procedure of the higher-educational institutions mentioned in this study (further details provided in chapter 4).

2.4.2 Assessment and theories of learning.

Historically, educational assessment trends and practices have followed the development of learning theories and teaching methodologies. According to Bloom (1969), when assessment is aligned with the process of teaching and learning, it would yield ‘a positive effect on students’ learning and their motivation’ (p. 18).

In the 1950s, teaching practices were dominated by the behaviourist learning theory, which deemed knowledge as ‘decomposable’, that can be broken down into its component parts without jeopardizing understanding or applicability. Traditional assessment (summative) was associated with this approach to learning and teaching: exams, quizzes and standardized tests with formats made primarily up of multiple-choice, true-false, and short-answer questions. Students focus on identifying the ‘right’ answer, as opposed to developing inquiry skills and deepening conceptual understanding (Doran, Chan, & Tamir, 1998).

The behaviourist approach still plays a dominant role in Saudi higher-educational facilities. Both teaching and learning are heavily reliant on textbooks and the memorisation of factual information. Here, assessment practices are focused on tests which have straightforward right answers, and are formal, summative and norm-referenced.

In contrast to the preceding behaviourism, constructivism or cognitive theory suggests that learning is an active process in which learners construct new ideas or concepts or attempt to understand the world by reflecting on their current/past knowledge and experiences. Formative assessment is therefore particularly associated with this theoretical framework as well as the sort of assessment described by Earl (2003) mentioned previously. Consequently,

learning outcomes can be assessed by students' journals, portfolios and other various forms of keeping records.

Finally, some kind of socio-cultural perspective is often regarded as the most recent development within learning paradigms. As Vygotsky (1978) pointed out, learning needs to be viewed as a social activity that is embedded in authentic, meaningful contexts and is shared by one's peers. From this perspective, assessment approaches include peer- and self-assessment, not only assessment by the teacher and views assessment as tightly connected with the wider context in which it occurs.

The specific sociocultural approach to L2 learning and teaching called Sociocultural Theory has its own particular preferred approach to assessment called Dynamic Assessment which, for example, places emphasis on scaffolding provided along with feedback as part of assessment. In this way, extending the idea of assessment being formative, the assessment is tied very closely to instruction (Herazo et al., 2019). There are even attempts to apply Activity Theory (another sociocultural approach) directly as a method of L2 assessment (Alavi et al., 2019). However, no such approaches are used in this context, and it seems that the participant teachers, in their limited training, would be unlikely to have heard of them. Hence, I do not review these in detail here.

2.4.3 Concepts and forms of classroom assessment

Several scholars have taken up the challenge of conceptualizing teachers' approaches to assessment (e.g., Brown, 2004; Remesal, 2011; Willis et al., 2013; Wolf et al., 1991). The majority of studies (Biggs, 1998; Büyükkarcı, 2014; Davis & Neitzel, 2011; Dayal & Lingam, 2015; Elshawa et al., 2017; Gutbrie, 2005; Karim, 2015; Kim, 2014; Mansory, 2016;

Mehrgan et al., 2017; Mussawy, 2009; Önalán, 2018; Saad et al., 2013; Thomas, 2012; Wang, 2004; Yin, 2017) have identified two distinct approaches to classroom assessment: (a) summative assessment and (b) formative assessment. The former refers to assessment strategies that are considered traditional, formal or teacher-centred and include paper and pencil test, extended responses (essays), exams, quizzes and textbook exercises. The latter is considered an alternative method to traditional assessment. Formative assessment is student-centred and includes concept maps, group work, portfolios, journals and presentations (Rahim, Venville, & Chapman, 2009).

Within summative assessment practices, teachers value measuring and ranking student achievement; teachers not only focus on instrument construction and application but also on the production and use of relative student rankings. In contrast, formative assessment values long-term student development and considers the instructional practices necessary to support students' learning; teachers focus on the relationship between instruction and learning and consider the alignment between assessment, pedagogy, curriculum, and learning.

Teachers' identification with either approach of assessment was argued to have a direct impact upon their classroom assessment practices (Wolf et al., 1991). Similarly, in a landmark article, Shepard (2000) mapped assessment orientations and practices to dominant historical paradigms within educational systems. Specifically, she argued that traditional paradigms of social efficiency curricula, behaviourist learning theory, and scientific measurement favoured a summative approach to classroom assessment, whereas a social constructivist paradigm made provisions for a formative assessment orientation. Table 1 summarizes the various aspects related to formative and summative approaches of assessment as discussed above.

Table 1

Realization of formative and summative assessments (Adapted from Black & Wiliam, 2018 and Wiliam & Thompson, 2008)

	Summative	Formative
Theoretical paradigm	behaviourist	constructivist
Teachers' role	Instructors, leader and sole provider of information	Facilitators who clarify and share learning intentions and criteria for success
Learners	Passive actors, receivers of information and blank slates	Thinkers, active members who acquire the ownership of their learning
Assessment goals	Collection of evidence related to the instructional material	Adapt and improve instructional materials to meet students' needs.
Major feature	Assessment of achievement	Support students' learning
Means of assessment	Standardized tests, teacher-made exams, quizzes and extended essays	Journals, portfolios and scaffolding essays
Scoring system	Numerical values or grade	using rubrics, checklists, and questionnaires

In all, the emergence of formative and summative assessment as two different formats has attracted educators' attention in the current literature. In this section I aimed to provide a brief account of these two different assessment approaches as realized in the wider language assessment fields along with the main aspects that shaped the conceptualization of these assessment formats.

2.4.4 Assessment literacy (AL).

In recent years, educational assessment responsibilities have been largely placed upon teachers (Deluca & Klinger, 2010). It is estimated that teachers spend between 30%–50% of their professional time engaging in assessment activities (Stiggins, 1999). Therefore, the need for professional development, in order to provide teachers with assessment knowledge and competence, has increased phenomenally. In light of this issue, the term ‘assessment literacy’ was coined and has come to refer to the range of skills and knowledge about measurement basics directly related to classroom activities (Popham, 2009).

According to Willis et al. (2013), assessment literacy is ‘a phrase that is often used but rarely defined’ (p. 32). The earliest attempt to define assessment literacy for teachers was made by the American Federation of Teachers (1990); instead of using the term ‘assessment literacy’, they talked about teachers’ assessment competencies. These competencies included selecting assessments, developing assessments for classrooms, administering and scoring tests, using scores to aid instructional decisions, communicating results to stakeholders, and being aware of the inappropriate and unethical uses of tests.

The term ‘assessment literacy’ was first coined by Stiggins (1991) and subsequently taken up by other researchers to elucidate their own definitions of the concept. Interestingly, there are not many definitions of assessment literacy available in the educational literature. I present some of these below:

- AL is ‘... the ability to understand, analyse and apply information on student performance to improve instruction’ (Falsgraf, 2005, p. 6).

- AL is a set of various competencies which enables an individual to ‘understand, evaluate and create language tests and analyse test data’ (Pill & Harding, 2013, p. 382).
- AL is viewed as ‘...a range of skills related to test production, test score interpretation and use, and test evaluation in conjunction with the development of a critical understanding about the roles and functions of assessment within society’ (O’Loughlin, 2013, p. 363).
- ‘The capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the test results’ (Inbar-Lourie, 2008, p. 389).
- The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom-based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals (Fulcher, 2012, p. 125).

Looking at these definitions, it is evident that they mention one or both of two general kinds of ability. By contrast, they take a skills-based perspective on AL; in other words, teachers

have the ability to construct assessments, administer assessment tasks, grade them and give feedback. I argue that this kind of definition represents the practices-focused aspect of assessment, which I also adopt when presenting my own definition of AL below.

Some definitions, however, especially the definition provided by Inbar-Lourie, take a more conceptual view of the abilities of individuals concerned with assessment; that is, teachers are mentally engaged in the assessment procedures and are able to justify why, how and what kind of assessment is done. These are more issues of belief and, in my opinion, Inbar-Lourie's definition touches on topics that are related to my study, such as the purpose for assessment, the efficiency of the tool being used etc. Therefore, I also include this kind of definition as part of my own definition of AL, which is presented later in this section.

The most comprehensive definition of AL to date has been delineated by Fulcher. Although it combines both cognitive (knowledge) and practical (skills) aspects of AL, it is very broad and involves an ideological system that ambitiously aspires to explain the world and how it is influenced and changed.

Based on the definitions and discussion above, my definition of AL is as follows: AL involves possessing a sufficient theoretical and practical understanding of the building blocks of classroom-based assessment: 1) purpose, 2) objectives, 3) specifications, 4) task selection and item organisation, and 5) scoring, grading¹ and feedback.

Based on the definitions stated above, including mine, it is evident that AL consists of two dimensions: the cognitive or belief domain (an individual's understanding of the fundamental

¹ Scoring has to do with numerical value. Grading is a system of giving letters.

assessment concepts), which is considered rather abstract and can be tapped into by interviewing teachers, for example, and the practical domain of their practices (the procedures employed while constructing educational assessment tasks in specific circumstances), which is more tangible and can be directly observed during the process of production of assessment tasks. Both dimensions are clearly interrelated within educational contexts. Figure 5 represents AL within the cognitive and practical dimensions and the relationship between these dimensions. My study, with its attention to understanding both the beliefs and practices of teachers with respect to assessment, aims to gather data on both those major dimensions.

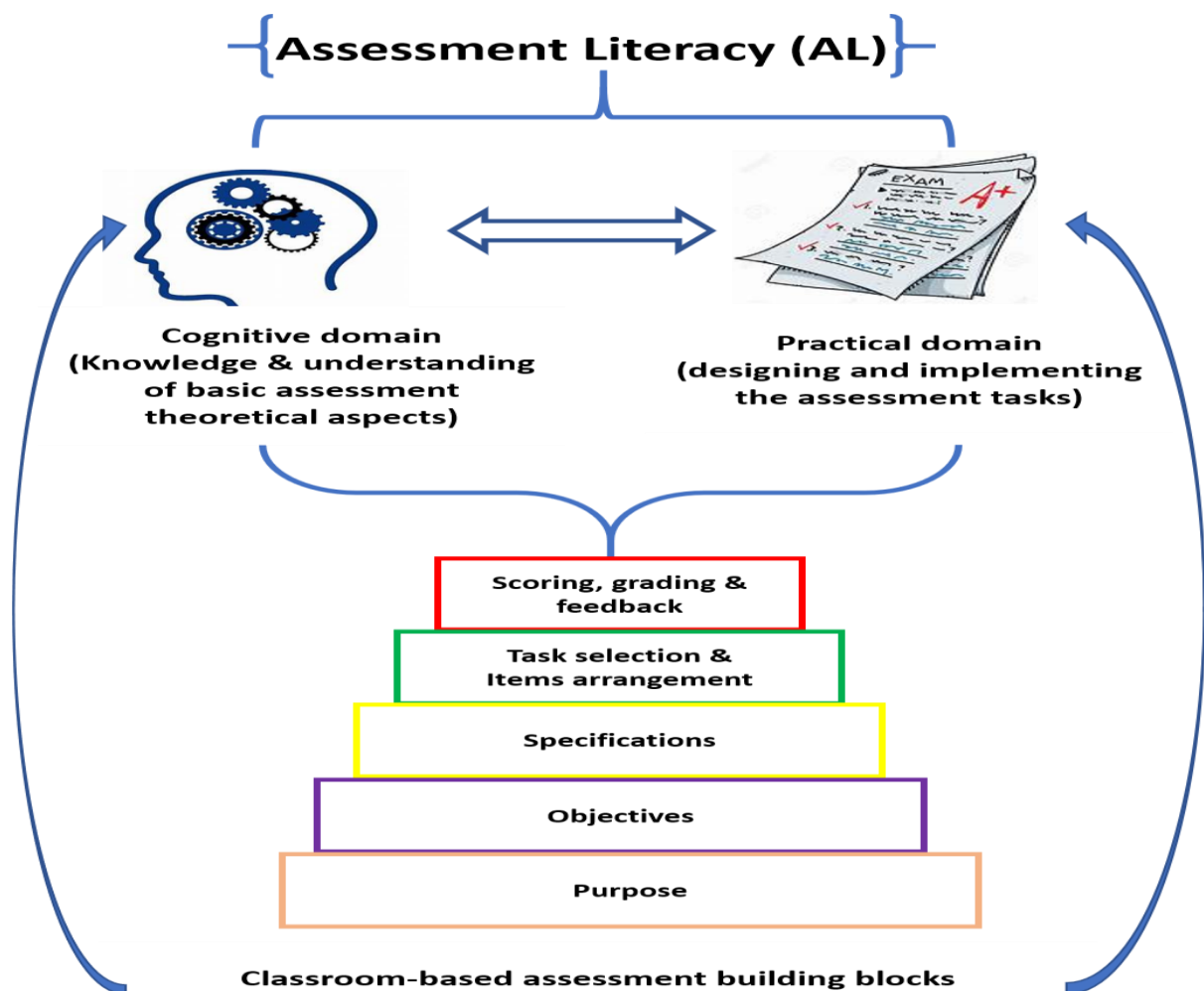


Figure 5. Conceptualization of assessment literacy with the cognitive and practical dimensions.

2.4.5 Language teacher cognition on assessment.

A perusal of the literature on teachers' cognition in relation to assessment reveals a great diversity of research studies carried out from 2003 to 2017. Although none of these studies use the term cognition to refer to their participants' mental representations, most studies resort to other terms such as beliefs (Barnes, Fives, & Dacey, 2015; Büyükkarcı, 2014; Chang, 2006; Elshawa, Abdullah, & Rashid, 2017; Gutbrie, 2005; Karim, 2015; Mansory, 2016; Restrepo & Aristizábal, 2003; Sikka, Nath, & Cohen, 2007; Thomas, 2012), conception (Deneen & Brown, 2016; Opre, 2015) and/or perception (Ağçam & Kaya, 2017; Assunção, 2017; Figueiredo, Alves, & Silva, 2016; Mussawy, 2009; Sahinkarakas, 2012). In addition, studies on teachers' cognition in relation to assessment have been conducted in a number of different contexts, including Turkey (Ağçam & Kaya, 2017; Büyükkarcı, 2014; Sahinkarakas, 2012), Malaysia, Taiwan and Fiji (Chang, 2006; Elshawa et al., 2017; Opre, 2015), America (Restrepo & Aristizábal, 2003; Sikka et al., 2007), Saudi Arabia, Pakistan and Afghanistan (Mansory, 2016; Mussawy, 2009; Thomas, 2012), among others.

It is also apparent that a large body of work in L2 teacher cognition concerning assessment tends to focus on various issues related to classroom assessment in general and are not topic-specific. For example, many studies encourage investigating teachers' beliefs about formative assessment (Büyükkarcı, 2014; Karim, 2015; Mehrgan, Candidate, & Language, 2017; Thomas, 2012), while the majority of work still advocates comparing pre-service teachers' beliefs to in-service ones, considering the role of experience, training and assessment literacy as the main factors that shape their participants' beliefs (Assunção, 2017; Büyükkarcı, 2014; Deneen & Brown, 2016; Mehrgan et al., 2017; Mussawy, 2009; Sahinkarakas, 2012; Sheehan & Munro, 2017).

In this section, I have identified previous studies that are most relevant to teachers' beliefs about assessment. In the following sections, I review the relevant aspects of these studies and discuss them in more detail.

2.4.5.1 Beliefs about the purpose of assessment.

Brown (2002) devised a Teacher's Conceptions of Assessment (TCoA) inventory providing four main purposes upon which the conception of assessment revolves around. These conceptions are as follows:

- It improves teaching and learning.
- It makes students accountable for their learning.
- It holds schools and teachers accountable for their students' learning.
- It is irrelevant, invalid and should be rejected.

Brown's first conception implies that assessment can serve as a tool to diagnose students' learning strengths and weaknesses. Therefore, teachers should employ forms of assessment that allow them to better identify what students learn and what they fail to comprehend. This would enable teachers to target the less-comprehended topics again later in the courses and provide the learners information about what they particularly need to work on regarding learning. The second conception, student accountability, means that the assessment prompts students to be responsible for their own learning and make more efforts by providing them with a measure of what they have achieved so far. The third conception (teacher and school accountability) refers to the use of assessment as evidence to prove that teachers and schools, or in my case the higher-educational sector, complies with the established standards and keeps both sides motivated. Last, 'assessment is irrelevant' means that assessment is seen as purposeless, perhaps because the teachers know their students and the specific curriculum,

and assessment would place pressure, affecting the teachers and the students negatively (Brown, 2002, p. 41).

Using TCoA, Barnes, Fives and Dacey (2015) placed these conceptions on a continuum where, at one end, assessment plays an extreme pedagogical role focusing on students learning and, at ‘the opposite extreme [,] assessment reflects the sole purpose of high-stakes accountability’ (p. 286). Some conceptions as to the purpose of assessment may be located in the centre of the continuum (mixed) because they combine both pedagogical and accounting purposes; how close they are to one end or the other is determined by the way in which an assessment is embedded in a specific context. Outside the continuum, the researchers have taken the irrelevance view, which is associated with assessment being considered as bad and to be ignored. ‘Thus, if teachers believe that assessment is irrelevant then it cannot be used for any of the purposes along the continuum’ (p. 291). To clarify this continuum and the components at each end, I have devised the following figure.

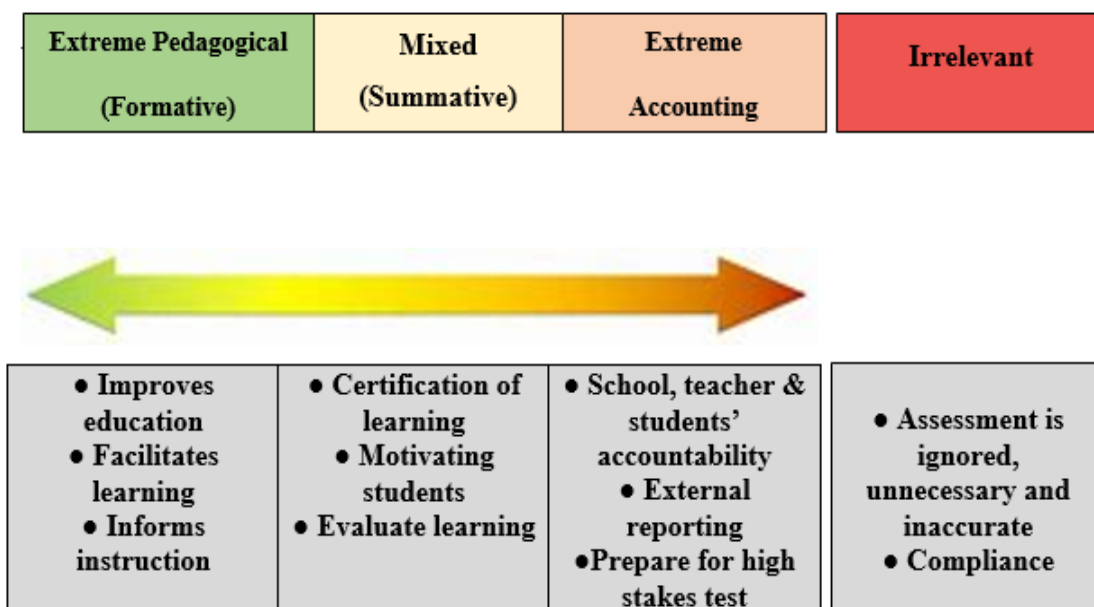


Figure 6. Continuum of beliefs and conceptions about the purpose of assessment. (Barnes, Fives, & Dacey, 2015, p. 287)

While Brown's TCoA inventory provides a useful framework for teachers' conceptions of assessment, it has its limitations. Harris and Brown (2009) argue that the TCoA inventory constrains the purposes of assessment to these four conceptions, while leaving out other significant purposes that may be captured by other studies. In their study, seven major purposes of assessment are identified: (1) compliance, (2) external reporting, (3) reporting to parents, (4) extrinsically motivating students, (5) facilitating group instruction, (6) teacher use for individualising learning and (7) joint teacher and student use for individualising learning (p. 369).

These seven purposes listed above, however, may readily be considered as separate subcategories that fall within the four broad categories of the TCoA (see Figure 6), and they elaborate on the account which TCoA provides. Thus, I believe that Harris and Brown's argument is weakened and Brown's TCoA remains valid at a more general level, as a guide for research studies in the field of teachers' beliefs regarding assessment.

Empirical studies on teachers' beliefs and perceptions about the purposes of assessment point to a degree of agreement among teachers' conceptions across contexts in that they tend to frame assessment as a vital instrument that is employed to achieve various pedagogical and educational goals (Ağçam & Kaya, 2017; Barnes et al., 2015; Chew & Lee, 2008; Elshawa et al., 2017; Mussawy, 2009; Opre, 2015; Sikka et al., 2007).

For example, Elshawa et al. (2017) investigate the assessment beliefs of ESL instructors at a tertiary level in Malaysia, including purposes, methods, feedback and grades reporting. Adopting a cross-sectional research design, the researchers selected their participants purposively. 83 university teachers from six universities answered a four-point Likert-scale

questionnaire composed of 67 items that addressed various aspects of assessment, 10 of which concerned beliefs about assessment purposes. The results revealed that the participant-teachers believed that assessment fulfilled two main pedagogical purposes: informing instruction and improving learning. The focus was, however, subtly different from my own. First, the word ‘assessment’ was not defined and probably was taken by the participants in a broader sense than I intend to use it (just limited to classroom-based assessment). Second, in the text, the researchers refer to measuring teachers’ beliefs and say, for example, that the teachers agreed that ‘assessment should be used for different purposes’ (p. 32), which appears to match my focus on belief in the sense of espoused theory, as described in 2.3.1.1.1. In reality, however, the items offered in the questionnaire for the participant teachers were not worded in that form, e.g. ‘Assessment motivates my students to learn’ (p. 44). Such items try to elicit teachers’ reports about what they think actually happens (i.e., practices), instead of their beliefs about what should be happening. Hence, Elshawa et al.’s (2017) article exhibits some conceptual confusion over which construct it is really studying; this is something that I intend to avoid, as it does not, in the end, precisely fall in my area of interest.

In another study, the analyses of prospective teachers’ perceptions of assessment and evaluation were similar to Brown’s TCoA inventory (Ağçam & Kaya, 2017). Using a questionnaire comprising both open-ended and closed items, the researchers compared two groups of prospective teachers studying classroom teaching at a state university in Turkey. This study included practising teachers who did not actually set assessment tasks themselves but rather were the subjects of assessment. The first group was composed of participants who had not taken the course titled ‘Assessment and Evaluation’, while the second group included those who had taken the same course as part of their second-year studies at university. Preliminary findings showed that the two groups significantly differed in their opinions about

assessment and assessment practices in higher education. Third-year participants acknowledged the necessity for assessment as a tool that provides learners with feedback about their strengths and weaknesses (an extreme pedagogical purpose). The first-year participants, on the other hand, believed that assessment is unnecessary and irrelevant. This attests to the impact of training, through a course which essentially changed teachers' content knowledge of assessment.

Pereira and Flores (2017) report how teachers view assessment in higher-education levels after the implementation of the Bologna Process, an agreement issued in 1999 onwards to ensure the comparability of standards and the quality of higher-education qualifications across European countries. 57 teachers from five Portuguese public universities participated in the study. Data were collected via face-to-face interviews and online open-ended questionnaires. The findings showed that teachers' perceptions of the purpose of assessment remain primarily pedagogical, in that the role of assessment was perceived as to provide feedback that would allow students engage better with their learning processes.

Furthermore, two qualitative studies have identified teachers' conceptions of assessment. Deneen and Brown (2016) interviewed 32 pre-service and practicing teachers about assessment, ranging from understanding and interpreting large-scale test results to the effective use of AfL techniques inside classrooms. Their results revealed polarized views stating that assessment carries both positive and negative consequences. Positive conceptions about assessment indicated that assessment helps students improve their academic capacities and boosts their self-esteem. The negative responses, however, were more related to non-academic aspects. For example, some of the participants mentioned that assessment results may make students feel embarrassed or bad. Moreover, the participants expressed some

concerns about the aspect of assessment fairness. Similar concerns were found in other studies where the participant teachers stressed that assessment should be fair to all students (Saad, Sardareh, & Ambarwati, 2013; Munoz, Palacio, & Escobar, 2012). In addition, Deneen and Brown's study reported that 'none of the participants considered assessment to be irrelevant' (p. 8).

Dayal and Lingam (2015) explored in-service and pre-service teachers' conceptions of assessment in primary and secondary schools in Fiji. Data collected through a reflective exercise with open-ended items were analysed qualitatively. The findings of this study were similar to the results reported in the previous research. Teachers from both groups had distinctive opinions about the primary purpose of assessment. In-service teachers' perceptions about assessment were more inclined towards the extreme pedagogical end of the scale. These teachers believed that assessment is meant to point out the strengths and weaknesses of teaching and learning strategies, leading to improved teaching and learning. Pre-service teachers, on the other hand, focused on the summative function of assessment, such as providing students with scores and measuring how much knowledge students possess subject-wise. In general, all the participants from both the studies were found to hold distinct and contrary beliefs about the purpose of assessment, all of which fall along the continuum of purposes captured in Figure 6.

Other studies have also reported conceptions of assessment at the extreme pedagogical ends of the aforementioned continuum, e.g. the conceptions of 25 lecturers from five public universities in Iraqi Kurdistan, who believed that assessment is used to improve learning and provide feedback on learners' strengths and weaknesses (Karim, 2015). The belief that assessment can play a critical role in adjusting teaching and learning strategies was harboured

by 69 English language teachers in different state primary schools in Adana in Turkey (Büyükkarcı, 2014). This conception was further supported by other participants in relevant studies who agreed that assessment improves teaching and learning by providing constructive feedback (Muñoz, Palacio, & Escobar, 2012; Saad et al., 2013; Sahinkarakas, 2012; Thomas, 2012).

Moving towards the other end of the continuum, cases where teachers' conceptions of assessment reflected accountability purposes were also found in some studies. Davis and Neitzel (2011) interviewed 15 upper-intermediate and middle school teachers of two schools in the south-eastern part of the United States. They found that the majority of teachers' conceptions about assessment was geared towards satisfying external audiences, namely the parents. Indeed, they believed that assessment required them to report to their students' parents and impart information about their children's progress on skills; for them, assessment would primarily serve a purpose of accountability.

Harris and Brown (2009) again reported teachers focusing on a similar function of assessment in relation to reporting to external audiences. 161 five-ten-year-old teachers from 36 schools in the Auckland region participated in their study. 26 teachers were then selected for interviews based on their different perception profiles. After analysing the data, the researchers reported seven categories related to the purposes of assessment, as described earlier in this section: 'Three categories (External reporting, Reporting to parents, Extrinsically motivating students) were all related to accountability' (p. 377). The majority of the teachers expressed agreement that assessment serves as evidence of achievement that can be presented to external organizations (e.g. Ministry of Education, school boards etc.). Reporting to parents was also considered important, although the teachers argued that they

report to parents mainly in order to justify their grading system. The findings of this study then reiterate that assessment is performed especially in order to meet accountability requirements.

Further emphasising the extreme accountability purpose, teachers in Sikka, Nath and Cohen's (2007) research, for example, expressed frustration and concerns with the pressure they would have to endure as assessment was 'mandated by the high stakes assessment requirement of schools' and their school administrators emphasized standardized test practice and a specific format and held the teachers accountable for student performance' (p. 249). This seems to have led to the decrease in teacher morale and the widespread instances of teachers leaving their profession.

This section provided an overview of research on teachers' beliefs and conceptions with respect to the purposes of assessment. I started by explaining Brown' (2002) TCoA inventory and move on to integrate the same with Barnes, Fives and Dacey's (2015) continuum of assessment purposes. Thereafter, I presented and reviewed the relevant studies by organizing their findings along the suggested continuum. I believe that there seems to be a consensus among most participants in these various studies wherein they perceive assessment as an important tool in education, regardless of the purposes they believe it serves.

Although these studies have been carried out in different educational contexts (tertiary and/or otherwise), there is a lack of pertinent studies which adequately cover Saudi higher-educational contexts. Therefore, this study includes a new context – Saudi Arabia – and examines the beliefs of EFL teachers in four higher-educational institutions in Riyadh.

2.4.5.2 Beliefs about methods of assessment.

Research on teachers' conceptions about the use of various assessment methods suggests a degree of agreement among the participants regarding the most desirable and dependable types of assessment (Büyükkarcı, 2014; Elshawa, Abdullah, & Rashid, 2017; Karim, 2015; Pereira & Flores, 2017; Thomas, 2012).

Elshawa et al. (2017) report the attitudes of 83 English language instructors from six universities in Malaysia on various aspects of assessment methods. In contrast with their treatment of assessment purposes, which I reported in the previous section, in this case they asked questions which did not elicit how much a teacher claimed to use different kinds of assessment, but rather, in general terms, indicated how they evaluated the usefulness of different methods, e.g. 'The language skill can be assessed through true-false items' (p. 46) and 'Self-assessment by the student is a good method of assessment' (p. 44). Here, then, they were accessing teachers' beliefs in light of espoused theory (2.3.1.1.1). The participants reported that many types of assessment tasks were suitable for assessing every language skill they teach, namely reading, writing, listening and speaking. The results showed that the teachers regarded traditional tasks as more capable of serving assessment purposes than alternative ones. For example, the teachers rated multiple-choice items, true-false choices, essay writing and oral presentation as higher than self-assessment, student portfolios, oral interviews and peer assessment. This suggests that teachers' beliefs are more conservative than the purposes for which they claim to use assessments – towards the extreme pedagogical end of the assessment continuum – while the methods they endorse are steered more towards the centre of the continuum.

Karim (2015) used an open-ended questionnaire to determine 25 Kurdish ESL teachers' beliefs about item formats and classroom assessment procedures. The results indicated that all the teachers believed formal assessments (e.g. written exams or oral presentations) provided better evaluation than informal assessments (e.g. incidental, unplanned comments and responses). Also, it was found that 100% of the participant teachers used paper-and-pencil assessments as the primary format to grade students as per the tasks assigned. These findings suggest a preference for summative assessment, but the researcher attributed this to the anxiety and difficulty that teachers encounter while implementing formative assessment practices.

In Pakistan, Thomas (2012) compared the beliefs about selection of assessment strategies by untrained teachers (those who had never received any regular educational training) with those of the trained ones. The research located a pattern of similarities in their responses, indicating the teachers' enthusiasm to use student-centred assessment strategies, although they would refrain from doing so as school policy required the use of summative assessment.

2.4.5.3 Alignment between beliefs and practice.

According to Borg (2003), effective classroom instruction takes place when teachers' beliefs are congruent with their instructional practices. However, a study of the mainstream literature on teachers' beliefs and their instructional practices reveals gaps and contradictions between what teachers say, think or believe and what they do. In this section, I discuss whether or not teachers' beliefs about assessment exhibit any consistencies with their assessment practices. A few studies have addressed this issue, with all of them concluding that there were mismatches between teachers' stated beliefs about assessment and their actual assessment

practices (Büyükkarcı, 2014; Chew & Lee, 2008; Davis & Neitzel, 2011; Karp & Woods, 2008).

Büyükkarcı (2014), for example, investigated the formative assessment perceptions and classroom practices of 69 primary English language teachers. Using both qualitative and quantitative data collection tools, the researcher found differences between the teachers' perceptions of formative classroom assessment and their real assessment practices. The participant teachers were found to garner positive beliefs and attitudes towards formative assessment. However, within classrooms, the teachers reported that they mostly used assessment methods for summative purposes (e.g. exam papers). Büyükkarcı (2014) concluded that such a gap between teachers' beliefs and practices exists because the classes are crowded, the teachers have heavy teaching loads and the amount of time spent in classrooms is limited.

Moreover, a similar mismatch between teachers' beliefs about and practices of assessment was found in Davis and Neitzel's (2011) study. Through classroom observations and semi-structured interviews, the researchers examined 15 teachers in US middle schools to identify their conceptions of assessment and how these conceptions were reflected in their daily assessment practices. In general, the researchers found that the teachers' conceptions of assessment and their assessment practices were incongruent:

These mismatches between the tenets of SRL² and teachers' assessment practices do not reflect deficiencies in teachers' understandings of assessment. Instead, these mismatches occur because teachers are asked to satisfy various assessment audiences, and the interests of these audiences (p. 212).

² Self-regulated learning (p. 202)

In contrast, Chew and Lee (2008) developed an online questionnaire to measure teachers' beliefs and practices concerning traditional classroom assessment. Descriptive statistics were reported together with analyses such as paired t-tests, ANOVA, factor analysis and stepwise regression were carried out on a sample of 148 facilitators³ from a polytechnic in Singapore. The analyses of participants' responses led to the identification of three dimensions of assessment: 1) making learning explicit, 2) promoting learning autonomy and 3) ensuring performance orientation. The results showed no significant differences between the participants' beliefs and practices in classroom assessment for the first and second dimensions. However, for assessment dimension 3, significant differences were found between the participants' beliefs and their practices of classroom assessment. The participants reported that they engaged in assessment practices that would support a performance orientation in their students; for example, using small team discussions and giving students opportunities to negotiate learning using their prior knowledge as well as build upon the knowledge contributed by others in the team, despite their beliefs that these practices might not be important. The possible factors that impacted such a divergence were 'endorsement from leadership, accountability to stakeholders, training and resource support and facilitators' workload' (p. 5).

In sum, describing teachers' assessment practices is a necessary precursor to exploring her/his opinions underlying those practices. The discussion of these aforementioned studies indicates that there are gaps between belief and practice and that various factors contribute to this reported mismatch between teachers' beliefs about assessments and their actual classroom practices.

³ The role of the teacher in a RP (Republic Polytechnic) classroom is, therefore, to serve as a facilitator of student learning (p. 2).

The present study aims to explore EFL teachers' current assessment practices and the relationship between these practices and their beliefs about English grammar assessment (in their teaching environments), and finally establish whether or not their beliefs about assessment resonate with the practices. It is important to point out that much of teachers' classroom practices involve assessment activities, for example, posing questions to learners and judging learners' responses (Jones, 2005); however, this study focused on teachers' assessment practices in the sense of constructing their own assessment tasks (i.e., exams, quizzes and finals).

2.4.5.4 Factors influencing teachers' beliefs about assessment.

Researches such as those discussed in the previous section have provided some evidence on the notion that although teachers may hold entrenched beliefs about assessment, they do not always put these into practice and, moreover, may have to adjust their perspectives and practices according to the inevitable contextual demands (2.3.1.2). This shows that teaching contexts, amongst other things, greatly influence the way in which teachers' conceptions of assessment are transformed into actual assessment practices. In this section, I first present a factor concerning what teachers know about assessment and how knowledge about assessment (or the lack of it) – assessment literacy (2.4.4) – might aid or hinder their assessment practices. Then, I discuss the contextual demands, such as class sizes and time constraints, which might contribute to the divergence or convergence between teachers' stated beliefs and classroom practices.

Sheehan and Munro (2017) investigated whether or not language assessment literacy or knowledge plays a substantial role in shaping teachers' conceptions of assessment, leading to the promotion/prevention of effective practices. 74 teachers participated in this qualitative

study where they were interviewed and observed. The interviews covered the various issues related to and the components of assessment literacy, while the observations focused on the teachers' assessment practices. The findings indicated that assessment literacy is a vital factor that contributes to teachers' conceptions of assessment and, in turn, to assessment practices. Indeed, assessment literacy would allow participant teachers to be critical of the materials they wish to assess and make them aware of the different types of assessment. Similar findings were found among two groups of prospective teachers in a state university in Turkey (Ağçam & Kaya, 2017; see 2.4.5.1). The researchers used an open-ended questionnaire and found that those who had taken a course on 'Assessment and Evaluation' held more positive views about assessment than those who had not taken the course yet and believed that assessment was unnecessary. This suggests that knowledge about assessment can positively influence teachers' conceptions about assessment.

Karim (2015) and Thomas (2012) also reported that due to the lack of assessment knowledge, teachers exhibit misconceptions about and reluctance to use certain assessment strategies (formative assessment methods). This again confirms the impact that assessment literacy might have on teachers' assessment beliefs and practices.

Contrary to the findings reported above, which consider assessment literacy as an influential factor in teachers' conceptions of assessment, Deneen and Brown (2016) find that while teachers' fluency in factual knowledge (i.e., assessment literacy) was enhanced through an educational course given to the participant teachers, conceptions of assessment that might influence assessment practices were not changed via the mediating influence of the assessment course.

In reviewing the studies mentioned in the earlier sections (2.4.5.3 onwards), teachers' assessment practices are not only affected by their beliefs but also by numerous contextual constraints, e.g. class size, time, staff availability, the requirements of audiences or stakeholders to whom the assessment has to provide information (purpose requirements), and the assessment methods imposed by higher authorities (method requirements). Another factor is the ability of the teacher/assessor, who might feel anxious that he/she lacks the necessary knowledge/skills/training to put into practice the different forms of assessment that he/she feels are suitable (Karim, 2015). In Bandura's (1982) terms, this could be seen as a mark of low self-efficacy.

One factor that has been repeatedly presented in various studies is constituted by the practices mandated by high-level stakeholders (e.g. educational policy of a Ministry or a university). For example, after the implementation of the Bologna Process in higher education, some of the participant teachers in the study by Pereira and Flores (2017) reported that they had to change their assessment practices such that they would align with the new vision of their universities: improving the connection between learning and students' future working contexts.

In Pakistan, Thomas (2012) found that even though both the trained and untrained teachers showed enthusiasm towards applying student-centred assessment strategies, they were pressurized by the school or the system's policy or practices, which mandated the use of summative assessment (in the form of formal tests and examinations).

A similar pressure theme emerged in Sikka, Nath and Cohen's (2007) study. All the teachers expressed frustration and concerns regarding the assessment practices that were authorised by

the high-stakes assessment requirements of schools. ‘Their school administrators emphasized standardized test practice and a specific format of tests (mostly multiple choice) and held them accountable for student performance on these types of assessment’ (p. 249). This is somewhat reminiscent of the situation in my context, where the English department imposes some control over the examination methods used by teachers (see an account of this context in 4.3).

Other factors, e.g. teaching load, classroom size, age, gender, experience, educational background and teacher competence, are among the common factors which were identified as shaping teachers’ beliefs about assessment and classroom assessment practices (Figueiredo, Alves, & Silva, 2016; Mehrgan, Candidate, & Language, 2017; Muñoz, Palacio, & Escobar, 2012; Sahinkarakas, 2012).

Based on the studies reviewed above, language teachers’ conceptions about assessment and their classroom assessment practices are influenced by several elements. By identifying these influences and following Borg’s representation of the LTC framework (Figure 6), I propose the following schematic conceptualization of LTC on assessment (Figure 7), which can be shaped, influenced and sometimes altered by a complex nexus of interacting factors.

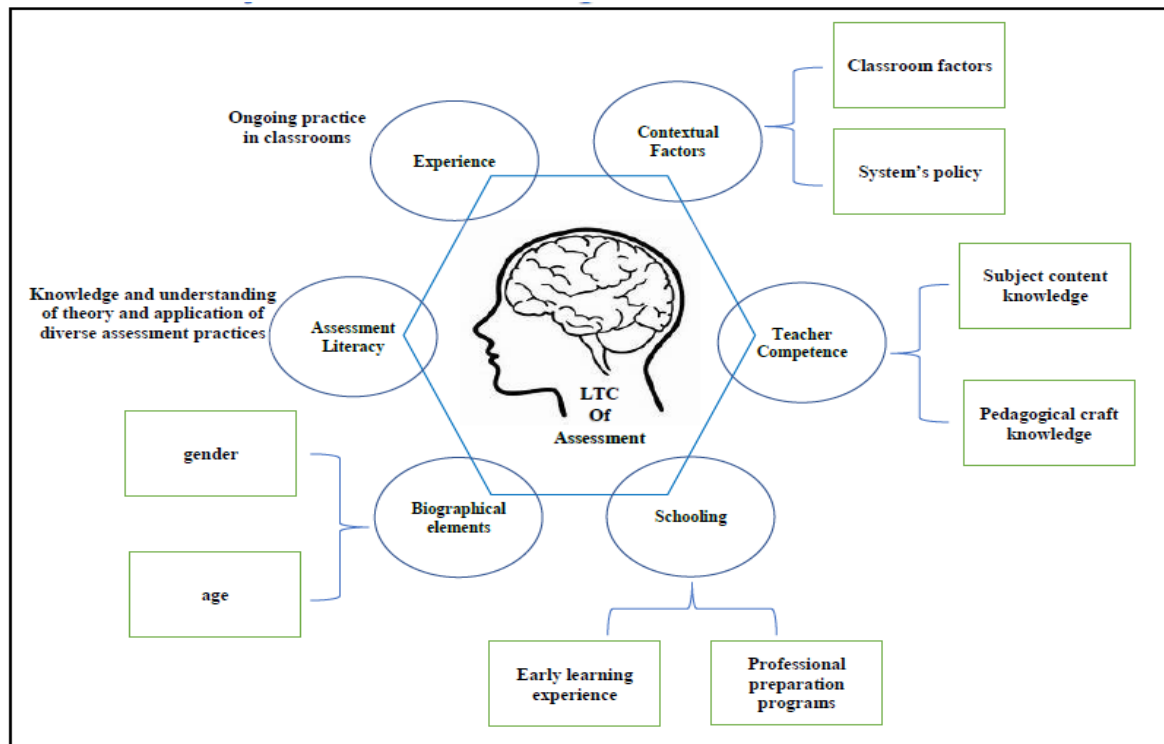


Figure 7. Schematic conceptualization of LTC on assessment.

2.4.5.5 EFL teachers' role in constructing assessment tasks.

When it comes to educational assessment, it is commonly assumed that classroom teachers have the upper hand in designing assessment procedures (Brindley, 1998; Leung & Rea-Dickins, 2007). In some cases, however, they are excluded from the design phase of assignments and are simply assigned the role of administering an assessment task in a classroom (Mansory, 2016; Saad, Sardareh, & Ambarwati, 2013; Sikka et al., 2007). In the context of this study, however, as stated in 4.3, classroom-based assessments are in the hands of the course teachers, albeit being sporadically monitored by department chairpersons and/or course coordinators⁴.

⁴ Course coordinators are teachers who are requested to fill this position either because they currently teach a course and control the majority of the course sections, or because they have extensive experience of teaching a given course. Postgraduate degrees are not mandatory but preferred. A coordinator's job is to validate and proofread the written exams submitted by the course teachers before these exams are signed by the particular departmental chairperson.

In a recent study in Saudi Arabia, Mansory (2016) investigated teachers' roles and beliefs regarding assessment practices in the English Language Institute⁵. The researcher interviewed 20 English language teachers to uncover the role(s) they play in both continuous (formative) and summative assessment practices. The findings revealed that teachers play no role in summative assessment design unless they are members of the Assessment Committee. Also, they play a limited role in continuous classroom assessment.

Similar findings were reported in a study conducted in Iran (Saad et al., 2013). 35 Iranian EFL teachers from different secondary schools all over the country participated in an open-ended questionnaire that was designed to explore teachers' assessment beliefs and practices. The results showed that assessment is centrally controlled and that only a few teachers are involved in the decisions made regarding assessment. Interestingly, both groups of teachers in these studies expressed their desire to become more involved in the assessment design. Unfortunately, they mentioned that their educational institutions were not very receptive of criticisms from teachers, which at times made them reluctant to increase their involvement in assessment or voice their views due to the fear of being labelled in a negative light.

Contrary to what was found in these studies, Elshawa, Abdullah and Rashid's (2017) research in Malaysia reports a more positive and engaging role of participant teachers in designing assessment tasks. In this study, the teachers seemed to experience freedom in preparing their assessment items and also exercised their choice in constructing assessment tasks, either individually or in collaboration. Such an engaged role of teachers in assessment construction is mirrored by the context of my study. However, Elshawa, Abdullah and Rashid's research

⁵ A language unit which provides extensive English courses to undergraduate in their foundational years at King Saud University (in the context of this study).

does not mention whether or not the teachers who construct their assessment tasks are checked by their departments/coordinators.

2.5 English Grammar Instruction and Assessment

2.5.1 Overview and definitions.

Both the teaching of grammar and the assessment of grammatical abilities constitute the most ill-defined domains in ELT (Borg, 1999). The role of instructing and assessing grammar formally and directly remains a perennial area of debate to this day (Borg, 1999). Advocates of explicit grammar instruction and assessment (exalting conscious knowledge of grammatical forms and their meanings) acknowledge the important role that grammar plays ‘as a fundamental linguistic resource for successful communication’ (Ellis, 2006, p. 85). According to Purpura (2004), although the way grammar is currently viewed by education, applied linguistics and language learning and teaching has vastly changed from the traditional perspectives that informed grammar testing in the 1960s, very little development has occurred on the way grammar is assessed in practice. Hence, a lot of the grammar assessment tasks that are currently used, both in standardized tests and in classroom assessment, focus on linguistic structures and discrete-point⁶ measurements.

According to Brown (2000), judicious attention to grammatical forms in adult language classes could prove not only helpful but also might expedite the learning process given that the optimal conditions for overt grammar teaching exist. There are six identifiable variables (age, proficiency level, educational background, language skill, style/register and needs and goals) that determine the role of grammar in language teaching (Celce-Murcia, 1991).

⁶ ‘Discrete-point tests are constructed on the assumption that language can be broken down into its component parts and that those parts can be tested successfully. Such an approach demanded a decontextualization...’ (Brown, 2003).

Figure 8 shows how these variables are placed on a continuum, which runs from less to more important explicit grammar teaching.

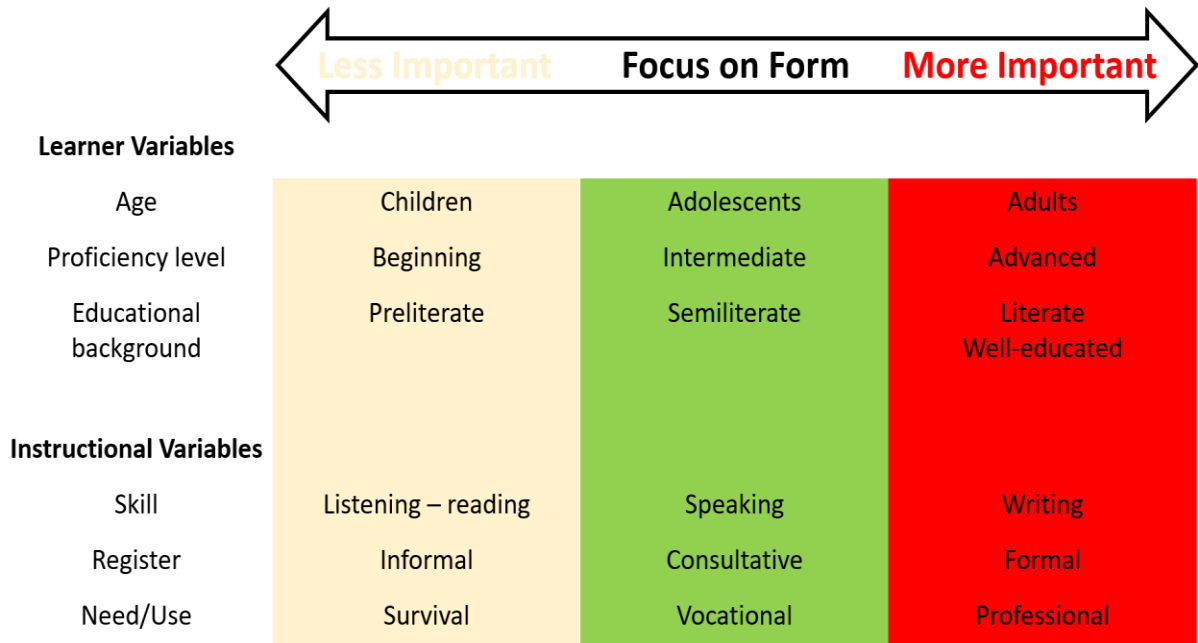


Figure 8. Variables that determine the importance of grammar (Celce-Murcia, 1991, as cited in Brown, 2000, p. 363)

The view of grammar reflected in the discussion above is restricted to the explicit teaching of grammatical forms. As teaching and assessment are so closely interrelated, this type of psychometric-structuralist approach⁷ to teaching gave rise to the assessment of formal patterns of a language structure by means of a discrete-point single sentence format (Purpura, 2004).

2.5.2 Studies on grammar assessment.

There is a substantial body of L2 grammar teaching research that examines teachers’ beliefs, knowledge and practices regarding explicit grammar teaching. What is striking, however, is

⁷ An approach rooted in structuralism and behaviourism that advocates the use of objective items and discrete-point format in assessments.

the relative absence of discussions on grammar assessment consensus on the following: (1) what L2 teachers think about explicit grammar assessments (cognition), (2) what type of assessment tasks are conducted in their teaching environments (practice), (3) what drives L2 teachers to assess grammar or grammatical abilities explicitly (factors) and (4) how the choices made by L2 teachers in grammar assessment influence each other (relationship). In this study, I aim to clarify these issues by exploring EFL teachers' beliefs about and practices related to classroom-based grammar assessment.

There are some studies on L2 writing assessment that include the assessment of grammar or other linguistic aspects (e.g. Barkaoui, 2007; Jarvis, Grant, Bikowski, & Ferris, 2003; Lumley, 2005). The majority of researches, however, have focused on large-scale and often high-stakes writing exams, such as TOEFL and IELTS. Few studies (but see Neumann, 2010; Lee, 2007) on classroom assessment have been conducted with respect to how L2 writing teachers assess grammar in classrooms.

Neumann examined how writing teachers attend to grammatical ability by analysing a case where teachers, assessing their students' academic essays after an ESL programme at a university in Canada, had integrated grammar editing tasks and quizzes into the assessment plans for the L2 writing courses. Using a mixed-methods approach, Neumann employed interviews, a questionnaire and document analysis. The results of the data analyses indicate that while assessing their students' essays, writing teachers mainly focus on grammatical accuracy, the absence of errors and, consequently, the language use that follows grammatical rules.

Relevant to investigating explicit grammar assessment, Hodgson (2017) in his short article reports the effects (on teaching and learning) of imposing separate grammar, punctuation and spelling test (GPS) in primary schools in the UK. The author stated that GPS focuses on ideal and specific grammar structures in a context-free setting. This practice, according to Hodgson, fails to indicate pupils' understanding of language and the grammar they have to grasp, while also contradicting the everyday use of language (communicating messages).

Another study which sheds some light on grammar assessment was conducted by Ahmadi and Shafiee (2015). The researchers aimed to explore both teachers' and learners' beliefs about grammar teaching and learning in language institutes located in Isfahan. The researchers used questionnaires, each of which contained seven sections on various topics regarding grammar teaching. What is relevant here is that the sixth section of the questionnaire (teachers and learners) addressed the issue of how to test grammar skills. The results indicated two contradicting views: the participating teachers strongly emphasised the need for grammar test items at a discourse-level, since it helps in assessing whether the test takers can use grammar correctly in real-life situations. The participating learners, on the other hand, preferred testing grammar at the sentence-level. According to Ahmadi and Shafiee, learners support this stance because the texts given in tests may make comprehending and responding to test items hard, due to the various structures and complicated grammatical features that appear.

2.6 Summary of the Chapter and Conclusion

All in all, the purpose of this chapter was twofold: 1) to outline the main research areas that inform my research thus providing the rationale underpinning it, and 2) to identify the conceptual gaps and methodological gaps that my study aimed to fill. I looked across the

major research works related to the present study and identified the key findings, the themes and the aforesaid aspects of language teachers' beliefs about assessment and their actual assessment practices. An overview of the considerable body of research on L2 teachers' beliefs about and their practices of classroom assessment has provided some invaluable insights:

1. The literature review has focused on language classroom assessment in general, because subject assessment studies focusing on grammar were absent from the research studies available.
2. Almost all teachers in the studies reviewed were found to believe that assessment is crucial in education and serves various purposes.
3. These teachers exhibited various levels of knowledge about assessment types, methods and strategies.
4. Their beliefs about assessment enter into complex, intertwined and mostly incongruent relationships with their classroom assessment practices.
5. Their beliefs about assessment are affected and altered by a number of salient factors.
6. Their role in designing assessment tasks is, in some contexts, marginalised.
7. The majority of research studies employ a mixed-method approach which included questionnaires, interviews and/or observation as data collection tools.

From the first and most prominent observation made above, it is clear that while there have been a number of studies on EFL teacher cognition and practices in relation to grammar teaching and assessment in general, none have been found to be devoted to grammar assessment and only a few are related to language assessment in a more general fashion. Yet, as in my context (4.3), there are situations where EFL teachers have control over grammar

assessments not only in terms of day-to-day classroom evaluation but also in terms of more formal examination methods. Hence, investigations of LTC about EFL grammar assessment (and indeed other areas of language assessment) is required, and this is where my study holds particular relevance. Employing a mixed-method approach to my research design, I explored EFL teachers' 'espoused theory' beliefs about English grammar, examining how they instantiate these beliefs in their practices when they actually set grammar exams.

Furthermore, I examine the factors that shape their assessment beliefs and those that may affect the translation of their beliefs into practice. In addition, I report the teachers' roles in designing English grammar assessment tasks.

To the best of my knowledge, this study is the first to discuss LTC in relation to assessment with specific reference to grammar; specifically, it includes both beliefs and practices, and it is certainly the first of its kind in the context of Saudi Arabia. The findings reported will hopefully help in explaining and understanding the interrelationship between teachers' cognition and their practices, in turn informing educational practitioners, policy makers and teacher practices as well. A further innovative feature of the present study is the involvement of Activity Theory, especially at the interpretation stage. It is to this that the next chapter is devoted.

3. Theoretical Framework

3.1 Introduction

This chapter describes the theoretical framework which guided this research, namely Activity Theory (AT). There are two main sections in this chapter; in the first one, I present a historical overview of the development of AT, including the background, definitions and stages. Following this, I discuss insights into how AT is used in educational research, with a particular focus on how AT is used as an interpretative framework in this study.

3.2 Activity Theory (AT)

3.2.1 An overview of AT and definition.

The history of Activity Theory is complex and interesting. Also known as cultural historical Activity Theory, it originates in the works of three historical perspectives: German philosophy (Hegel), the writings of Marx and Engels, and Vygotsky's cultural-historical psychology (Engeström, 1999). The concept of AT was advocated by Leont'ev (1978) and later expanded by Engeström (1987). Drawing upon Vygotsky's cultural-historical psychology, which emphasizes the role of social interaction in the development of cognition, AT considers individuals as located within an activity system in which individual human minds and behaviours interact with their surrounding structures (social, cultural and historical contexts) and with other humans in any given institutional settings (Nandi & Nandi, 2017). Engeström expanded Vygotsky's model by adding societal and contextual dimensions to Vygotsky's model and linking the activity system to a given context.

To better understand the concept of AT, various definitions are discussed in order to summarise its essence. For example, Kuutti (1996) defines AT as 'a philosophical and cross-

disciplinary framework for studying different forms of human practices as development processes, with individuals and social levels interlinked at the same time' (p. 25). Although Kuutti's definition approaches AT as a theoretical framework, other scholars have considered AT as an analytical framework. Indeed, Johnson and Golombek (2011) state that AT is 'an analytical framework that maps the social influences and relationships involved in networks of human activity' (p. 9). In alignment with Johnson and Golombek's conception of AT, Rantavuori et al. (2016) argue that AT should be considered as a method of systemic empirical analysis. From an analytical perspective, the goal of AT is to provide an effective framework in which one can analyse the complexities of continuous interactions between humans and their surroundings. Overall, AT has become 'international and multidisciplinary' (Engeström, 1999, p. 20). It has been applied in many different research contexts, proving its promise either as an interpretive or analytical base (further details in 3.3).

In this study, therefore, I use AT as an interpretive framework for two reasons: it has the capacity to capture significant elements of the broader context of LTC and EGA and it also fosters an understanding of the human mind (LTC) and practice (EGA) by investigating the interaction among the basic units of an activity system. As I shall argue in this chapter, AT represents a particularly useful theoretical and analytical framework which allowed me to investigate and discuss my subject matter in a holistic, contextualised and theoretically informed manner.

3.2.2 Three generations of AT.

Historically, AT has developed over three stages or generations. The first phase was derived from Vygotsky's perspective of interactions between humans, environments, and goals. The second stage witnessed Engeström's development from Leont'ev of a basic model of AT.

Engeström introduced five central elements to explicate the underlying dimensions and dynamics of AT. The third generation was proposed by Engeström, who argued that the world is too complex to be narrowed down to a single activity. In his re-examination of AT, Engeström highlighted joint activity or practice as the unit of analysis for AT, rather than individual activity (Engeström, 1999). The following discussion elucidates each of these stages.

3.2.2.1 First generation of AT: Mediation.

As mentioned earlier, the first generation of AT is based on Vygotsky's ideas of mediation. According to this perception, humans do not interact with the world by means of stimulus-response reflexes, but they instead make connections between an incoming stimulus and their responses by using various tools available in their environments (Vygotsky, 1978; Engeström, 1999). This connection is referred to as 'mediation'. The idea of mediation through tools advocates a dialectical relationship between humans, society and culture (Lei, 2008).

Vygotsky used the term 'tools' while Cole (1996) employed the term 'artefacts' to refer to 'the fundamental constituents of culture' (p. 144). According to Kozulin (1990), one of the characteristics of Vygotsky's model of mediation is the use of tools which can be resourced via two means: physical objects (e.g. a pen) and/or semiotic systems (e.g. language, sign etc.) (Cole & Engeström, 1993). Moreover, these tools may change or be modified based on the existing contexts and may even play a constraining role, instead of a facilitating one (Cole, 1999).

Another characteristic of Vygotsky's model is its focus on individual activity. Vygotsky saw humans as agents who can regulate their minds. By merging cognitive processes with a whole person's self, Vygotsky surpassed the countervailing relationships between mind and body and presented the human as a whole agent, engaged in activities in specific contexts.

I only want to say...that without man (= operator) as a whole the activity of his apparatus (brain) cannot be explained, that man controls his brain and not the brain the man...that without man his behaviour cannot be explained' (Vygotsky, 1987, as cited in Lei, 2008, p. 219).

Vygotsky's model is usually presented through a triangle, where the subject (agent) is placed at the left end of the triangle base and the object is placed on the other end, while the tools/artefacts are located at the point (Figure 9).

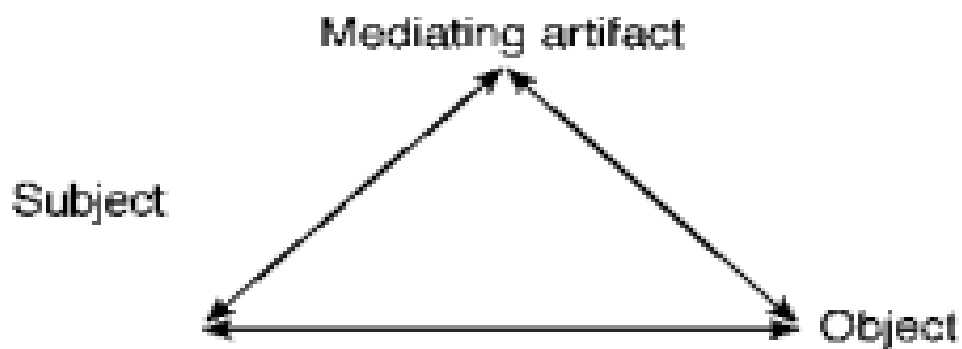


Figure 9. Vygotsky's model of mediated action. (Engeström, 1999)

The problem with this classical representation, however, is that it focused on the individual and failed to consider the societal and collaborative nature of any given action. In other words, the human, the action and the tools do not exist in isolation but are rather intertwined with each other and also with a social world, as events in collective activity system (Cole,

1996). The outcomes, accordingly, would appear limited and situation- bound (Engeström, 1999). To overcome this limitation, the concept of mediation was expanded by Leont'ev, with the notion of collective activity being introduced, which led to the formation of the second generation.

3.2.2.2 Second generation of AT: Collective activity.

As noted earlier, one limitation of the first generation was that the unit of analysis remained focused on an individual. The second generation, related to Leont'ev's work of collective activity, succeeded in overcoming this limitation by shedding light on the important relations between individuals and their environments in an activity system (Engeström, 1987).

The AT developed by Leont'ev (1978) was based on a psychological framework, that emphasised individual motives for specific behaviour (Fujioka, 2014). Leont'ev's model of collective activity consisted of three hierarchical levels: operations – actions – activities. The scholar explained the difference between individual action and collective activity in his famous example of the 'primeval collective hunt':

A beater, for example, taking part in a primeval collective hunt, was stimulated by a need for food or, perhaps, a need for clothing, which the skin of the dead animal would meet for him. At what, however, was his activity directly aimed? It may have been directed, for example, at frightening a herd of animals and sending them toward other hunters, hiding in an ambush. That, properly speaking, is what should be the result of the activity of this man. And the activity of this individual member of the hunt ends with that. The rest is completed by the other members. This result, i.e. the frightening of game, etc., understandably does not in itself, and may not, lead to satisfaction of the beater's need for food, or the skin of the animal. What the processes of his activity were

directed to did not, consequently, coincide with what stimulated them, i.e., did not coincide with a motive of his activity; the two are divided from one another in this instance. Processes, the object and motive of which do not coincide with one another, we shall call "actions". We can say, for example, that the beater's activity is the hunt, and the frightening of the game his action. (Leont'ev, 1981, p. 210)

Although this model contributed to the development of AT, it was criticized for being too simple and failing to show the complex relations between different components in an activity system. Accordingly, Engeström expanded the basic model of AT and systematized it with a graphical representation that would distinguish between the components associated with an activity system. Each component was related to all other components by two-sided arrows to indicate that these components influence each other (Figure 10).

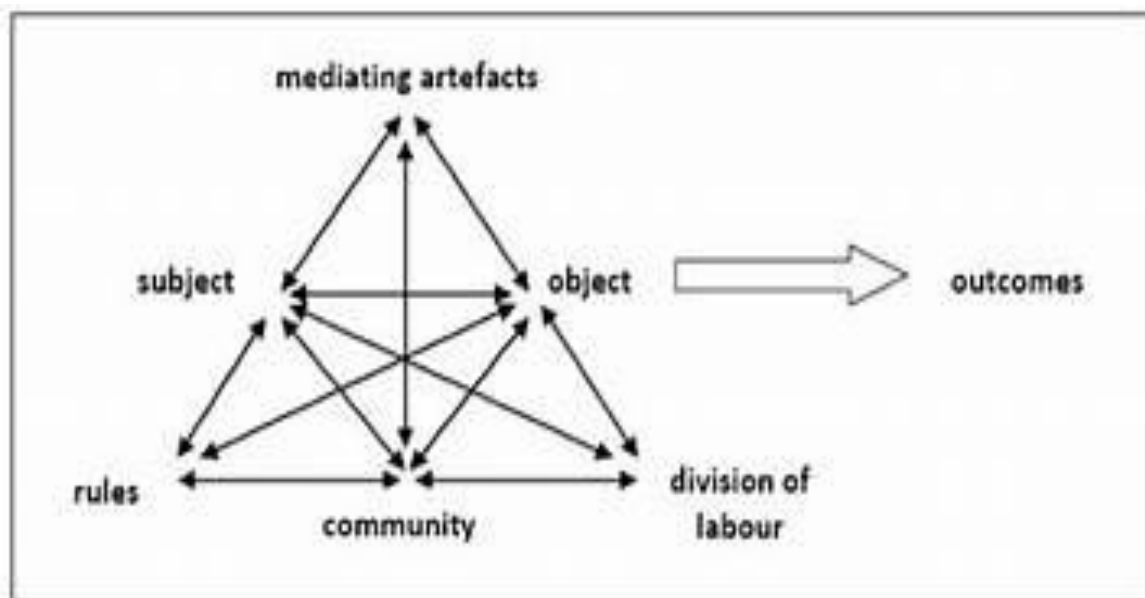


Figure 10. Engeström's model of activity system. (Engeström, 1987, p 78)

Engeström's model, as depicted in figure 10, uses a range of elements, some of which were elicited from Vygotsky's original model, while others were added to provide wider

contextual factors in relation to the actions taken and the methods used to achieve outcomes. The upper section of the model shows the subject (individual, dyad, or group) that is object-oriented (motivated). The object goes through several alterations until it stabilizes as a finished outcome. This process is accomplished by the means of mediating artefacts, which guide and direct the relation among components within an activity system. Many artefacts, either material tools and/or semiotic tools, can be used in different ways depending on the context (Engeström, 2008).

The paragraph above describes the uppermost section of the activity system, represented in Figure 10. The bottom part of the figure focuses on the community in which the individual (subject) is a member. Within the community, the members negotiate or establish their roles, power etc. (division of labour), which are governed by and regulated through explicit and implicit norms and conventions (rules) of the particular activity system.

Although the second generation of AT addressed the overall relationship between the individuals and their environments in any activity and acknowledged the role of the community, Engeström argued that looking at an activity system in isolation would be insufficient to fully explain the complex world we live in today. Accordingly, a new generation of AT was developed to understand the interaction between two or more activity systems.

3.2.2.3 Third generation of AT: Activity networks.

Engeström (2001) noted the following:

When Activity Theory went international, questions of diversity and dialogue between different traditions or perspectives became increasingly serious challenges. It is these

challenges that the third generation of Activity Theory must deal with. The third generation of Activity Theory needs to develop conceptual tools to understand dialogue, multiple perspectives, and networks of interacting activity systems. (p. 135)

Accordingly, the third generation of AT perceives several activity systems; the minimum unit comprises two interacting activity systems that form a network of complex activity systems (Figure 11).

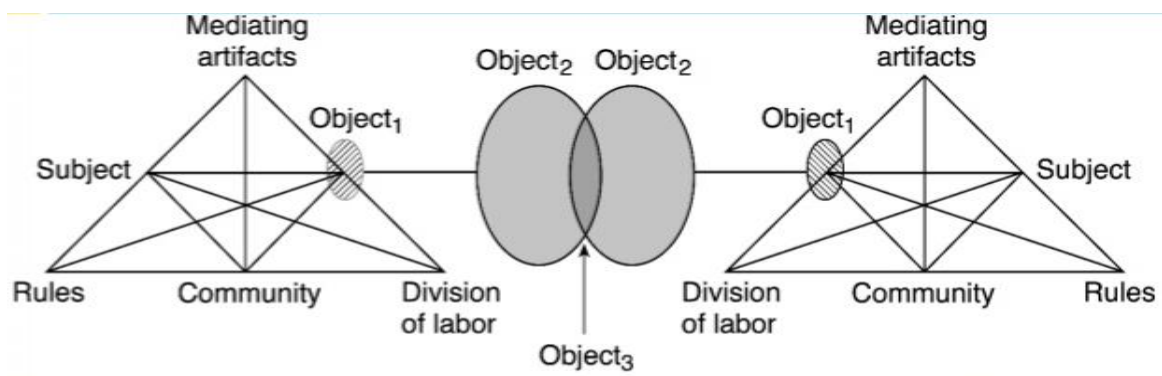


Figure 11. Engeström activity networks. (Engeström, 2001, p 136)

In his latest generation of AT, Engeström introduced conceptual tools to understand the multi-voicedness expressed by the different components within the network as references to their interests and perspectives. Furthermore, the third generation of AT, importantly for us, presented the concept of contradictions: the ‘motive force of change and development’ (Engeström, 1999, p. 9) within and between activity systems. Engeström explained that actions, even well-planned ones, are subject to unexpected innovations, obstacles, or even failures; his model of AT revealed the underlying contradictions that led to this turn of events.

Contradictions are defined as ‘historically accumulating structural tensions within and between the activity systems’ (Engeström, 2009, p. 57). Engeström (1987) describes four

kinds of contradictions. Primary contradictions occur within an activity element (e.g. tools). Secondary contradictions exist between elements (e.g. tools and rules). Tertiary contradictions occur between older and newer versions of activities. Quaternary contradictions generate between adjacent activities in networks. The role played by contradictions in this study is detailed in 3.4.3.

In sum, how AT is framed, whether via mediated action, activity system, or activity network, depends on the focus of research. In AT, human minds and practices are perceived as activities produced when individuals or groups (subjects) accomplish their goals (objects) using mediational means (physical, conceptual and/ or semiotic tools), by following norms and regulations (rules) and adopting specific roles (division of labour) within their communities. A further elaboration of how these elements are presented in this study is given in 3.4.2.

3.3 Activity Theory in Applied Research

The interest in the use of Activity Theory in research has grown rapidly over the last three decades. Ample work has been conducted on the applicability of AT in diverse disciplines and research areas across professional and academic fields; for example, health and patient care, information systems, psychology, management, information systems and education (Hashim & Jones, 2007). For the purposes of this study, I refer to studies in the field of education which employ AT either as an analytical tool to disintegrate themes emerging from data, or as an interpretive framework which aids us in understanding how the results obtained from data interact and develop in a given context.

To begin with, Grigoryan (2018) uses AT as an analytical tool to investigate teachers' perspectives towards implementing iPads versus textbooks in the language classrooms of a

tertiary-level institution in the UAE⁸. The qualitative study identifies the pros and cons of using iPads in language teaching and learning by analysing 24 reflective journals submitted by four teachers. Using AT, two conceptual models are formed: iPad-based and textbook-based models (third generation of AT). The results show that ‘the language achievement is easier, faster and more enjoyable when iPads are used as a means of learning’ (p. 39).

Grigoryan’s research indicates that AT was ‘well suited for the analysis process and activities involving significant components in higher education’ (p. 33).

Binjimah (2017) conducted a mixed-method study to explore the perspectives of pre-service science teachers involved in a preparation programme at a university in Saudi Arabia. In this study, AT (third generation) is used to analyse the relationship between two academic activity systems, universities and schools, and to identify the contradictions between them. Through the analysis, some contradictions emerge, namely the gap between the actual outcomes and the intended outcomes of the preparation programmes. These contradictions are perceived as opportunity to change and improve the activity systems. This study further confirms the usefulness of AT as analytical tool.

Anastasiou (2017) explores teachers’ beliefs regarding inclusion and dyslexia and how these are linked to their current practices when working with dyslexic learners in classrooms within two cultural contexts: those of Cyprus and North West England. The research utilizes an AT analytical framework (second generation) within a qualitative research methodology to present teachers’ personal interpretations in the context of their schools and how teachers view their roles and responsibilities in supporting dyslexic learners in classrooms. Guided by the lens of AT, the unit of analysis becomes the teachers’ professional practices as a

⁸ United Arab Emirates

collective activity. The findings suggest that even though the teachers come from different cultural contexts, they seem to share similar concerns and present congruity between their understandings and professional practices. This study; therefore, just like the previous ones, provides an additional example of the successful implementation of AT in the interpretation of qualitative findings.

From an interpretive perspective, Hirsh and Segolsson (2019) apply AT in qualitative research to understand the leadership practices in teacher-driven school development projects in Sweden. AT (second generation) was used as an interpretive framework to provide a way in which to conceptualize the actions of individuals and their contexts, which not only shape but are also shaped by these actions. The study finds that there is a need to promote a systematic collaborative work among teachers in an activity system. This study is one of the few researches which commends the use of AT as a framework to understand transformations in collective practices.

Another recent study which is relevant to my research context is Alkader (2018), which explores teachers' perceptions and practices regarding social media use as a tool in English language education in the context of Saudi women's higher education. Adopting a case study design, the research uses AT (second generation) as an interpretive framework to explain the activity system 'as a dialectical process in which the teachers interact with the environment' (p. 188). According to Alkader, AT assists in clarifying the relationship between teachers' beliefs and their sociocultural contexts as well as that between teachers' beliefs and their practices. Therefore, this study serves as another contribution toward the benefits of employing AT to conceptualise teachers' beliefs within an activity system, which is something that I endeavour in this study.

To sum up, this section has shown how AT has been applied in previous studies. It has proved to be a significant theoretical framework for the analysis and understanding of human mind and actions. As an analytical tool, AT provides an effective approach that dismantles the complexities of continuous interaction between humans and their surroundings in an activity system (Nandi & Nandi, 2017). As an interpretive foundation, AT successfully supports a holistic understanding of human cognition and actions and explains how they interact and influence each other within an activity system. Overall, it is a researcher's discretion to employ AT as it fits the theoretical and practical purposes of his/her study. In the next section, I discuss how AT has informed and grounded my research study as an interpretive framework.

3.4 Activity Theory in the Present Study

In this research, we draw upon Engeström's AT of the third generation in interpreting different activity systems, involving EFL teachers' beliefs and practices (with a common object) and classroom-based grammar assessments. A key advantage of AT, as an interpretive framework, is its capacity to capture significant elements of the broader research context of LTC, along with assessment practices and other salient features that influence this context. Understanding the human mind (cognition) and activity (practice) is achieved by investigating the interactions within and between the elements present in each activity system in an activity network. Another important characteristic of AT is that contradictions within and between the activity systems potentially open the space for an individual's mental state and actional change/development. Overall, the use of the theoretical framework of AT to interpret and understand the current research topic can be summarized in the five basic principles of AT formulated by Engeström, 2001:

1. Teachers' beliefs and practices are activity systems within an activity network that is contextually situated;
2. Each system hosts multiple perspectives, interests and traditions;
3. These elements interact within and between the activity systems;
4. Contradictions are a driving force of change and development;
5. Transformation in the system is possible when contradictions are experienced and individuals begin to challenge the established norms.

3.4.1 The activity network.

Two activity systems are investigated in this study, namely the systems of cognition and action. The first system refers to EFL teachers' beliefs about English grammar assessment. The second system details EFL teachers' practices of assessing English grammar. Each system is represented graphically as a triangle, with a set of arranged elements linked via bidirectional arrows to indicate the interrelation among these elements. The two systems are then combined together in accordance with the third generation of AT to allow the visualization of the context under investigation and also the conceptual lens through which data are interpreted (Figure 12).

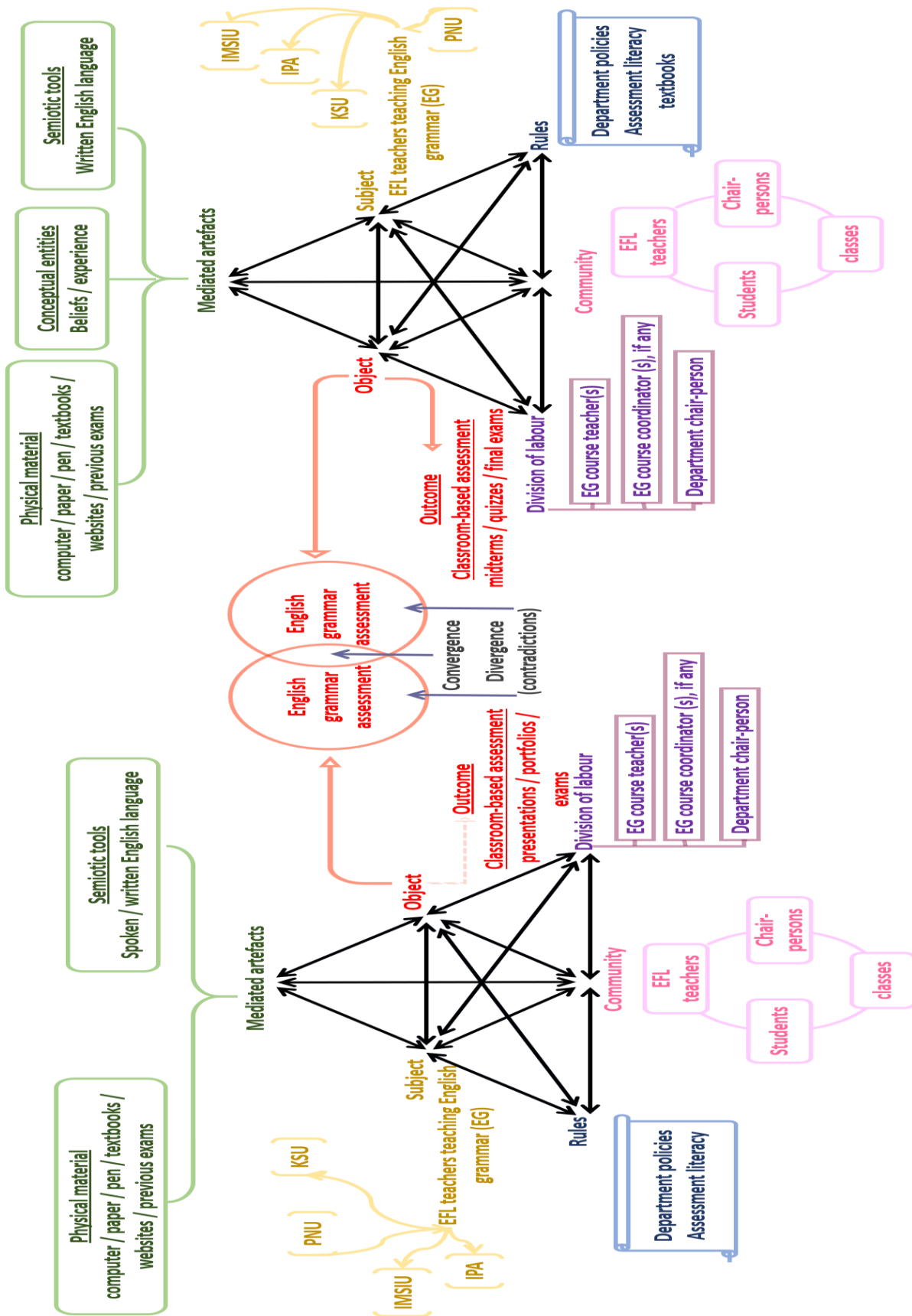


Figure 12. Structural model of activity network of EFL teachers' beliefs and practices with regard to English grammar assessment.

3.4.2 Elements of the AT.

In figure 12, each activity system is represented as a triangle with a set of elements arranged independently and linked by bidirectional arrows. The upper section of the triangle shows three elements: subject, object and mediated artefacts. In the lower section, division of labour, community and rules are displayed.

3.4.2.1 Subject, object and mediated artefacts.

The subject refers to the participants (an individual or a group of people) engaged in an activity to achieve an object, which leads to an outcome. In this study, EFL teachers are the subjects who have been placed in charge of teaching English grammar courses in four public higher-educational facilities.

Activity theorists⁹ agree that an activity is defined by its object. The object indicates the goal(s) of the activity, which motivates the subject to participate in that activity and seek to accomplish a desired outcome. Waite (2003) states that ‘The perceived difference between the current state of the object and the desired outcome provides the motivation for the subject to develop goals and actions to transform the object into the desired outcome’ (p. 3). Within an activity network, this object is potentially shared between the two activity systems, which may subsequently result in tensions and contradictions within and between the systems (detailed discussion in 3.4.3). In this study, the EFL teachers engaged in the activity of classroom-based grammar assessment (object), motivated by the need to assess students’ grammatical abilities. Obtaining the outcome, then, involves assessing students through written midterms, quizzes and final exams. Working towards this object, EFL teachers’ beliefs and practices may prove to be either congruent or conflicted. In a nutshell, classroom-

⁹ Centre for Cultural Historical Activity Theory and Developmental Work Research, 2004

based assessment is related to the need of testing students' grammatical abilities, which EFL teachers aim to achieve through the use of mediated artefacts that exist within an activity system.

Mediating artefacts, in other terms tools or instruments, can facilitate a subject's task towards the object and convert this object into an outcome. The tools used in my activity network are of three types: physical, conceptual and semiotic. Physical tools are concrete and visible materials that include, for example, computers or laptops, which participants use either to type their written exams or to surf the internet and seek the question items suitable for their exams. Conceptual tools refer to abstract entities that could be literally translated through action (e.g. beliefs and knowledge). Here, EFL teachers' beliefs about classroom-based grammar assessment may be seen as a process of mediating their practices through their experiences, as Cole (1996) suggested: '...mind [is] a process of mediating behaviour through artefacts' (p. 143). Semiotic tools can be languages, signs, symbols, images etc. In this study, the semiotic tool involves the English language, in either spoken or written formats.

3.4.2.2 Division of labour, community and rules.

Division of labour specifies the role and responsibility of each participant in an activity system and within an activity network. The reason why I perceive division of labour as separate for each system is because the teacher participants, with respect to their belief systems, may talk about their roles in terms of expectations, what they should do and how freely they can do it, while in a practice system their roles might be minimized and responsibilities downsized/expanded according to their working contexts. This situation is referred to as role expectation and role demarcation (Leadbetter, 2008). Division of labour

may also include how work is shared and why, which leads to the discussion of ‘division of authority and status’ (Roth & Tobin, 2002, p. 114). A good example in this regard could be the power relationship between a department chairperson and her/his departmental teachers or between two teachers, who share the same course, but one of whom teaches the majority of the course group. This may result in tensions and contradictions with regard to the division of labour in classroom-based grammar assessments.

Community is defined as ‘multiple individuals or subgroups who share the same general object’ (Roth & Tobin, 2002, p. 114). In the context of this study, the relevant community includes human and non-human agents, for example, teachers, students, classes and culture. The community governs the division of labour and determines the rules to be obeyed/disregarded by the participants in an activity system.

Rules are the explicit or implicit regulations which govern actions and interactions within an activity system (Roth & Tobin, 2002). In the case of my study, rules might be related to general departmental policies or even mandate specific guidelines with regard to writing exams; they could either be implemented or ignored. The latter may result in clashes within the community and become a source of tensions and contradictions within and between activity systems.

3.4.3 Contradictions.

The exploration of teachers’ beliefs and practices in educational settings often reports gaps between teachers’ practices and teachers’ intellectual stances underlying those practices. In this study, the incongruity between beliefs and practices is approached from an AT theoretical perspective as an exemplar of contradictions. The concept of contradictions

constitutes one of the five basic principles of the third generation of AT; contradictions have been described as ‘misfit within elements, between them, between different activities, or between different developmental phases of a single activity’ (Kuutti, 1996, p. 34). They have also been characterised as conflicts (Dippe, 2006), as tensions (Basharina, 2007; Berge & Fjuk, 2006) and, more precisely, as historically accumulating tensions (Engeström, 2001).

Although ‘contradictions’ have negative connotations, in AT, they are not considered problematic but rather as a source of change and development (Fleer, 2016). Engeström and Miettinen (1999) emphasise contradictions as ‘the motive force of change and development’ (p. 9). Engeström (2001) explains how contradictions can lead to innovation and transformation in an activity system:

As the contradictions of an activity system are aggravated, some individual participants begin to question and deviate from its established norms. In some cases, this escalates into collaborative envisioning and a deliberate collective change effort. An expansive transformation is accomplished when the object and motive of the activity are reconceptualized to embrace a radically wider horizon of possibilities than in the previous mode of the activity. (p. 137)

Additionally, Engeström (2001) talks about internal contradictions that could occur within an activity system and the external contradictions which exist between activity systems in an activity network. Internal contradictions include tensions or breakouts in an element in an activity system (primary contradictions). For example, within the space of community, clashes may take place between or among community members. A related example, from my research context, would be an English grammar teacher disagreeing with a course coordinator over exam formats, scoring systems, or question items. Secondary contradictions may also

occur within an activity system but involve misfits between two elements in an activity system. For example, the rules imposed by an institution to assess grammar through written exams may generate tensions between teachers who might wish to incorporate oral presentations in the assessment procedure and the university rules. The third or tertiary level of contradiction exists 'between a newly established mode of activity and remnants of the previous mode of activity'. For example, while teachers can be requested to use portfolios in assessing students, some teachers may lack knowledge and/or experiences of using this method. This may cause tensions and even lead to teachers' resentment toward the new ways of assessment. The fourth level escalates the conflict such that it occurs between an activity system and its neighbouring activities (external scope). For instance, teachers' beliefs can divert from their actual practices. However, contradictions within the study are not represented via the classification discussed above, as contradictions are not empirically evident but interpreted as they manifest in the study data, precisely between teachers' beliefs and current practices (incongruity).

To sum up this concept and link it to my study, first I argue that, based on what is already known from teachers cognition literature, it is to be expected that this study will show that contradictions result from the misalignment of EFL teachers' beliefs about English grammar assessment with their actual practices of assessing English grammar. Second, the investigation of contradictions and tensions provides a lens through which to understand how and why deviance between EFL teachers' beliefs and practices towards English grammar assessment occurs. Third, the recognition of contradictions delivers insights into the potential that exists for change and development of such activities. In all these areas the AT perspective should prove valuable when it comes to examining the research findings.

3.5 Summary of the Chapter and Conclusion

This chapter discussed the theoretical framework of AT. First, I presented a detailed account of the origins and definitions of AT. Then, the development of this theory was discussed, starting with the first generation (Vygotsky's concept of mediation), followed by the second generation (based on Leont'ev's work of collective activity) and, finally, Engeström's third generation (activity network).

The chapter then moved on to justify the use of AT as a theoretical framework in the current study by discussing some studies which have employed AT either as an analytical or an interpretive framework. Finally, I discussed how the third generation of AT is employed in this study to help provide an interpretive framework, with particular attention to the elements of AT: subject, object, mediated artefacts, division of labour, community and rules and the role of contradictions within and between the activity systems as a possible source for transformation and development.

4. Research Methodology

4.1 Introduction

This chapter provides details of the research methodology adopted in this study. First, the research design and rationale are explained. Next, an overview of the research context and participants is discussed. Then, the data collection and analysis procedures are presented. The chapter concludes by addressing the issues of data triangulation to establish the research's validity, the ethical dimensions and the pilot study process.

This is an exploratory research designed to answer the following research questions:

1. a) What are EFL teachers' beliefs about how English grammar should be assessed?
b) What are the factors which have helped shape those beliefs?
2. a) How do EFL teachers actually assess grammar in their teaching environments?
b) What are the factors which influence their practices, other than their beliefs?
3. a) What is the relationship between EFL teachers' beliefs and their current practices?
b) What are the factors which lead to the convergence/divergence between their beliefs and practices?

4.2 Research Design and Rationale

The present study investigates EFL teachers' beliefs about English grammar assessment and their current practices of assessing English grammar, along with factors that influence each and the relationship between EFL teachers' beliefs and practices. The study employs a mixed-method research (MMR) design, which Creswell (2006) defines as follows:

Mixed methods research is a research design with philosophical assumptions as well as methods of inquiry. As a methodology, it involves philosophical assumptions that guide the direction of the collection and analysis of data and the mixture of qualitative and quantitative approaches in many phases in the research process. As a method, it focuses on collecting, analysing, and mixing both quantitative and qualitative data in a single study or series of studies. Its central premise is that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone. (p. 5)

In MMR, integration is considered a key concept (Moseholm & Fetters, 2017), defined as a purposeful process by which a researcher combines quantitative and qualitative data together in one study (Creswell, 2015a). Integration could be achieved within MMR's three major principles, methodology, design and methods, which I inspect closely and relate to the present study in the following sections.

It should be noted from the outset that Activity Theory does not prescribe any single method of study (Kaptelinin, 2013). However, its emphasis on a varied range of entities and levels that apply to any actor or activity clearly is consistent with the use of multiple methods. Work done within this theory often uses multiple sources of data, especially qualitative, as explained in the previous chapter. In the present study, as mentioned in chapter 3, AT is brought in to assist the interpretation of the findings rather than to guide the initial design of the data gathering, or indeed the way in which the data was initially analysed (whether statistically for the quantitative data or in terms of coding for the qualitative data).

4.2.1 Mixed-method methodology.

A methodology refers to the philosophical assumptions underlying a research, for example, positivist paradigm, constructivist or interpretivist paradigm and pragmatism (Creswell, 2006). The positivist paradigm is mostly related to quantitative research and proposes that knowledge is gained through tangible and measurable facts. This philosophical stance demands that researchers be emotionally detached and uninvolved with regard to the object of study (Johnson & Onwuegbuzie, 2004). On the other hand, constructivism argues that reality is socially constructed. This paradigm is more popular in qualitative research. Then comes pragmatism, which guides most MMR studies. Pragmatism does not advocate for one research's approach over another or compare between them, but it rather brings together the two previously discussed approaches to form a compatible paradigm. Pragmatism emphasizes on creating a shared meaning and joint action through complementarity; that is, quantitative and qualitative approaches are combined in order to complement the strengths and compensate for the weaknesses within each approach (Shannon-Baker, 2015).

The methodological stance adopted in this research project is pragmatism. The rationale behind a pragmatic approach is that this study contains certain elements of both positivist and constructivist paradigms. Following the former, it gathered quantitative data from a closed-response questionnaire, where it is determined in advance (top-down) what questions the participants would answer and what range of answers they could give. Following the latter, the study also collected qualitative data via interviews, retrospection and analysis of written exam papers composed by the EFL teacher participants, as I aim to investigate what teachers say or do without much prompting, bottom-up. The integration of qualitative and quantitative research approaches allowed me to draw broad generalizations with a detailed understanding of EFL teachers' beliefs and practices related to English grammar assessment.

4.2.2 Mixed-method design.

‘Research design refers to the plan of action that links the philosophical assumptions [see previous section] to specific methods’ (Creswell, 2006, p. 4). Mixed-method designs are broken into two categories: Basic and Advanced. According to Creswell, 2015, basic designs are considered to be at the heart of mixed-method design. Once a basic mixed-method design is encased into a larger theory, for example, experimental or social justice orientation, it becomes advanced. In this section, I discuss the basic design level within MMR.

MMR may follow one of three core designs: explanatory sequential, exploratory sequential and convergent designs (Creswell, 2015a; Fetters et al., 2013). An explanatory sequential design begins with the collection and analysis of quantitative data, followed by the collection and analysis of qualitative data. The primary focus of this design is to explain quantitative results using more detailed qualitative results, i.e., qualitative results help in understanding the quantitative results.

The exploratory sequential design, on the other hand, begins with the collection and analysis of qualitative data, followed by those of quantitative data. This design is used primarily by a researcher who wishes to explore a phenomenon first, followed by the development of instrumentations. Convergent design, also referred to as convergent parallel design, relies on two concurrent data collection phases. Here, the quantitative and qualitative data are collected and analysed during a similar timeframe. The purpose of this design is to compare, contrast, or cross-validate the data findings within a single study (Fetters et al., 2013).

This study employs a convergent design for two reasons: first, data collection is conducted within a restricted timeframe, and hence collecting qualitative and quantitative data at the

same time helps to obtain the required data within a short time. Second, integration within convergent design occurs during the data analysis phase, which proves to be challenging due to the difficulty of comparing results from different forms of data, to provide more accurate and reliable findings and, hence, of understanding the research phenomenon in a more holistic manner. Figure 4.2.2 illustrates how qualitative and quantitative data are integrated in a convergent mixed-method design.

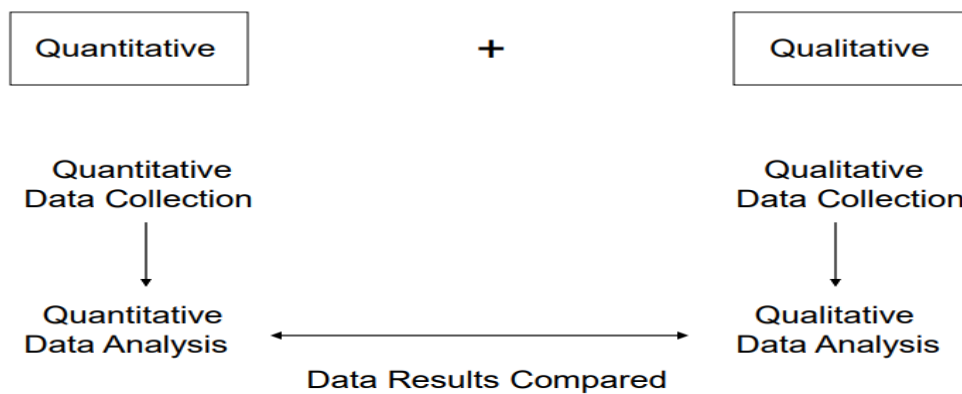


Figure 13. The convergent prototype of mixed-method research design. (Terrell, 2011, p. 267)

4.2.3 Mixed-method measures.

Methods refer to the techniques or instruments of data collection and the procedures of analysis (Creswell, 2003). The concept of integration at the level of methods is classified as follows:

...connecting (linking through sampling), building (findings from one strand inform development of data collection tools or procedures for the other strand), hypothesis generating and testing (using one type of data to generate hypothesis and another type of data to test that hypothesis), matching (reflecting the intent to have themes/constructs match on a domain by domain basis), diffracting (using cuts of data to understand a

phenomenon), embedding (the addition of qualitative data into a multistage study at multiple points). (Moseholm & Fetters, 2017, p. 2–3).

In this research, I use a questionnaire and semi-structured interviews to collect data about EFL teachers' beliefs related to English grammar assessment and to identify the factors which influence these beliefs (RQ 1). I also used semi-structure interviews, retrospection and document analysis to investigate EFL teachers' practices of assessing English grammar in their current teaching contexts, along with the factors which affect these practices (RQ 2). Data are analysed statistically and thematically, and the results generated by all the research tools are combined and compared to establish the relationship between EFL teachers' beliefs and actual practices and identify the factors which shape this relationship (RQ 3). The complete account of the research instrumentations and data analysis procedures is provided in section 4.4. All in all, MMR is suitable for my research aims and questions because it allows me to gain a holistic and more detailed understanding of EFL teachers' beliefs and practices with regard to English grammar assessment.

Through a convergent research design, quantitative and qualitative data were collected independently of each other but were merged at the analysis stage. The quantitative research method included a questionnaire to elicit EFL teachers' beliefs about English grammar assessment. The quantitative data provided broad, generalizable findings on the research topic. Qualitative research methods included semi-structured interviews, retrospections and document analysis that would serve to report on EFL teachers' beliefs about and practices of assessing English grammar and identify factors which influence both. The qualitative data generated in-depth, comprehensive information about EFL teachers' beliefs and practices. In addition, the combination of qualitative and quantitative methods allowed the triangulation of

the data that led to the achievement of research validity. Figure 14 shows the overall research design of this study.

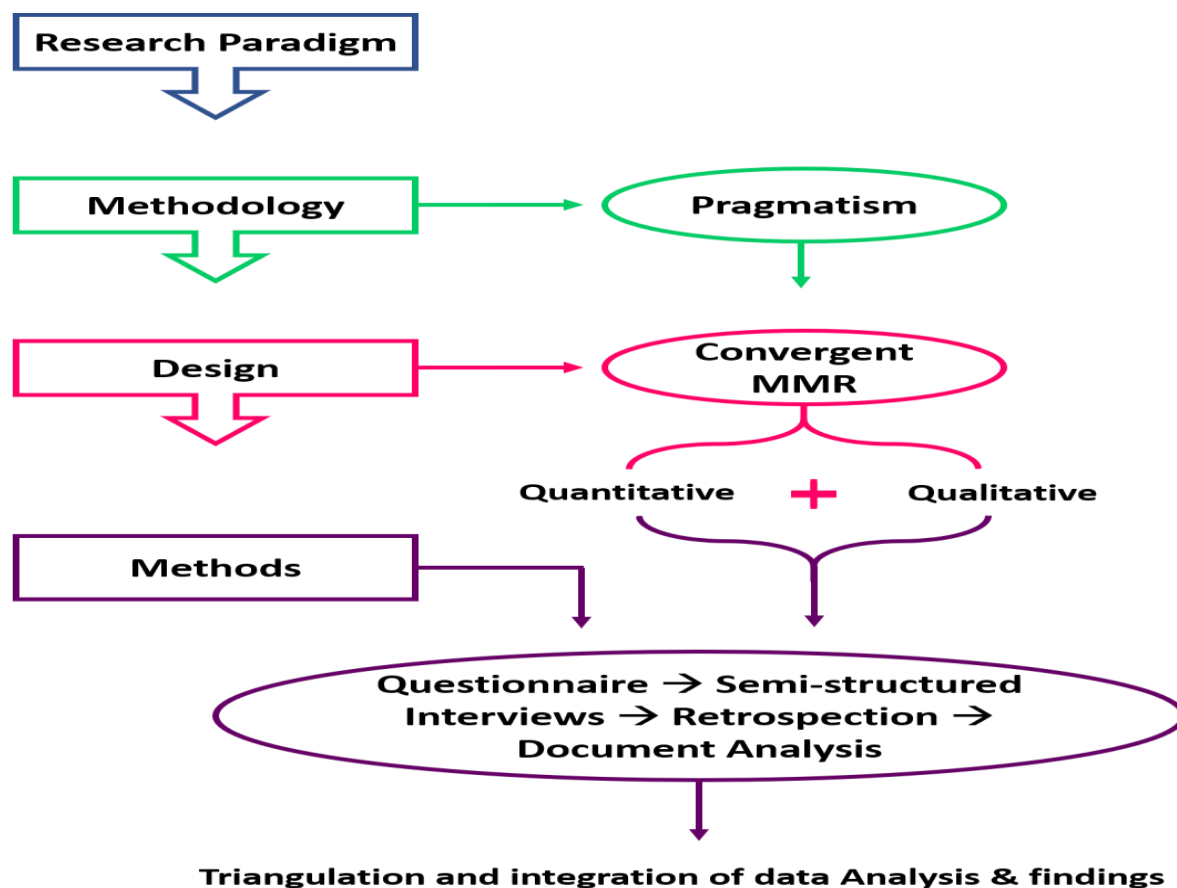


Figure 14. Current research design.

The figure above illustrates the research paradigm adopted in this study. Pragmatism was preferred due to its philosophical stance, which allowed me to adopt what would work in my specific research context. A convergent MMR design was chosen because the field of inquiry involved exploring respondents' beliefs and investigating their practices of English grammar assessment. Four data collection instruments were employed: questionnaire, semi-structured interviews, retrospection and document analysis. Finally, integration was conducted during the analysis and while reporting the findings.

4.3 Research Context and Participants

4.3.1 Overview of the higher-educational context in KSA.

In Saudi Arabia, the higher education sector in the Ministry of Education was established in 1975. It is responsible for developing, implementing and supervising various undergraduate and postgraduate programmes that fulfil the KSA's need for national cadres who are scientifically and professionally qualified to serve the national development objectives.

Over the past 40 years, higher education in the KSA has fostered over 60 educational establishments, including government and private universities and educational institutions. High-educational facilities in KSA are gender-segregated, which means that men and women study at separate campuses and do not meet, except those targeting medical specialities. In addition, all these educational facilities are linked to the Ministry of Education (higher education) but exercise a great deal of administrative and academic autonomy.

That being said, these higher-educational contexts devise their own teaching and assessment methods, which are imposed by the respective academic faculties (professors, associate professors, assistant professors, lecturers and teaching assistants), including recently quality assurance units, as they see fit. For 30 years, almost all higher-educational facilities have favoured a teacher-centred approach¹⁰ in their teaching practices and employed formal traditional assessments (written exams) even in low-stakes mini-assessments termed 'quizzes'. These were commonly carried out by the teaching members, as the expertise of the experienced members was passed on to novices through classroom observation and imitation. Also, these practices were not dictated specifically by the administration facilities but were rather mentally internalised by the members of these academic facilities.

¹⁰ A teacher is actively involved in teaching, while the learners remain passive in receiving information.

However, in the past ten years, higher-educational facilities have become familiar with and aware of the importance of academic accreditation¹¹. Many universities have strived to achieve institutional and programme accreditations both nationally and internationally, through the enhancement of quality and excellence in academic evaluation (NCAAA; <https://www.ncaaa.org.sa/Pages/default.aspx>).

Accordingly, there has been a great increase in the bundles of documents assembled, and forms filled in, to be submitted regarding the programmes offered by universities or institutions and the courses taught within each programme in the KSA. Two kinds of documents have acquired great value in relation to teaching and assessment practices: course descriptions and course specifications. The former intends to inform a particular audience – teachers, students external and internal examiners of the accreditation sector – about the subject matter, approach, breadth, and applicability of a course. The latter states the expected outcomes (objectives and topics week by week) of enrolling in a higher-education course and the means by which these outcomes are achieved and demonstrated through assessment or coursework of all students successfully completing the given course.

Everything therefore has become documented, with the aim of being transparent about what goes on, and faculty members have become required to comply with what entities such as the Quality Assurance Unit think should be written in these documents with regard to the methods of teaching and means of assessment. All these documents, namely the course specifications and course descriptions, are now managed by the Quality Assurance Unit of each faculty in the educational facilities.

¹¹ Academic or educational accreditation is a type of quality assurance process under which the services and operations of educational institutions or programmes are evaluated and verified by an external body to determine if the applicable and recognized standards are met. (Lenn, 1992)

Overall, then, there has been increased action in recent years towards educational accreditation, which resulted in, most notably, the following: 1) supervision and documentation of the educational process, and 2) the syllabus and assessment activities becoming more specified and uniform.

4.3.1.1 Departments/programmes of English language: The status of grammar teaching and assessment.

In KSA, higher-educational facilities provide a wide range of English language programmes. These programmes offer a study plan for students who wish to obtain bachelor's or post-secondary diploma degrees in the English language. There are seven universities in Riyadh that offer pathways to English language bachelor's degrees (BA) in linguistics, literature, or translation. Table 1 outlines these institutions and the English language programmes they offer. For ethical purposes the institutions are anonymized, and letters (e.g., A, B... etc.) were used to refer to each different institution.

Table 2.
Higher-educational facilities with English language programmes

Institutions	Type	College/ Faculty	Programme(s)
Institution A	Public	College of Arts	English Language and Literature
		College of Languages and Translation	English Language and Translation
Institution B	Public	College of Languages	Applied Linguistics, English Translation, English Literature

Institutions	Type	College/ Faculty	Programme(s)
Institution C	Public	College of Languages and Translation	English Language and Translation
Institution D	Public	Foundation course	English language
Institution E	Private	College of Human Sciences	English Translation, Linguistics
Institution F	Private	Faculty of Language Studies	English Language and Literature
Institution G	Private	College of Science and Theoretical Studies	English Language and Translation

The syllabus of most of these programmes includes courses in the core areas of English.

Lower-level courses in most plans are designed to enhance students' own listening, reading, speaking and writing skills, in addition to grammar and vocabulary courses. Upper-level specialised courses provide students with a practical understanding of theoretical and applied linguistics and polish their skills of interpreting, translating, analysing, and critically engaging with spoken and/or written English in various genres.

Grammar courses are essential to English language programmes. Being assigned as an independent course, grammar is taught and assessed explicitly. Grammar courses in English language programmes focus mainly on basic grammatical rules and linguistic terminologies which are devised from the assigned textbooks and detailed in the course specifications and

descriptions. Grammar courses are assessed in a summative framework through classroom-based assessments such as midterms, quizzes and final exams. The following section discusses the grammar courses within the English departments chosen in this study.

4.3.1.2 The current context.

The present study was conducted in fall 2018 in four high educational contexts in Riyadh: Institutions A, B, C and D. These facilities were chosen for two reasons: first, it was convenient for me to conduct the study in Riyadh, Saudi Arabia, as it is my hometown and because I work in one of these facilities. Second, all these institutions allowed me to visit their campuses and contact the teaching faculty of the English language programmes.

Institution A:

As shown in Table 1, English as a major subject is provided by two faculties, the College of Languages and Translation (COLT) and the College of Arts. Grammar courses within these programmes are almost similar. In the Department of English Language & Translation (DELT), there are two grammar courses: Grammar 1 (ENGT 113) and Grammar 2 (ENGT 213). The former course aims to increase students' knowledge of the target language and its structures. Students are trained to make use of grammar and produce simple sentences and also use tenses, nouns, modals and verbs. The latter focuses more on the parts of speech, the clauses and the sentence structures: simple, compound, complex and compound complex, in active and passive voices. In the Department of English Language and Literature (DELL), there are two grammar courses: Grammar 1 (ENG 106) and Grammar 2 (ENG 211). The aims of each grammar course in DELL corresponds with the courses in DELT. All the courses use the same textbook¹², and involve summative classroom-based assessment, namely in-terms,

¹² Understanding and Using English Grammar (4th Edition), Betty Azar (2009), Pearson ESL

quizzes and final exams. The in-term exams and the quizzes or any other classwork (e.g. participation or assignments) constitute 60% of the overall grade of students (100). The final exam makes up the remaining 40%, such that the total of marks adds up to 100. Moreover, a course teacher(s) is responsible for constructing, administrating and correcting these assessment tasks. The results of the in-terms should be announced and discussed with the students within two weeks, while the final results are to be reported within three days (Regulations of Study and Examination for undergraduate, the approved executive rules; https://council.ksu.edu.sa/sites/council.ksu.edu.sa/files/attach/_ldrs_wlkhtbrt_tdy1_shhr_jmd_lakhr_1433h.pdf).

Institution B:

There are three departments of English: English Translation, English Literature and Applied Linguistics. All these pathways share one grammar course (Grammar LIGT 131) that is offered by the Department of Applied Linguistics, and the faculty members from this department are the only ones who teach this course in all the departments in the College of Languages at institution B. This grammar course is usually taught by one teacher, which was the case in my study, to students in level 2 (year 1). It aims to build students' knowledge of various grammatical aspects, namely tenses, clauses, passive and active voices etc. Grammar is assessed in a summative framework, through classroom-based assessments based on the given course textbook¹³. According to the regulations of study and examination for undergraduates at B, assessment procedures should include at least two exams during a semester and one final examination. Just like in A, the course teacher(s) is responsible for constructing, executing and marking these exams

(http://www.pnu.edu.sa/ar/Faculties/Science/Documents/لائحة_الدراسة_والاِختبارات.pdf).

¹³ Understanding and Using English Grammar (4th Edition), Azar & Tracy, 2009.

Institution C:

The College of Languages and Translation (COLT) offers one pathway in English major via the Department of English Language and Literature (DELL). There are three different grammar courses: ENG 118 (year 1, level 1), ENG 119 (year 1, level 2) and ENG 220 (year 2, level 3). The names of the textbooks used are ‘Grammar Sense 1’, ‘Grammar Sense 2’ and ‘Grammar Sense 3’, consecutively. The aim of these courses is to gradually build students’ knowledge of English grammatical rules and help them progress from a basic level (first grammar course, ENG 118) to intermediate (ENG 119) and then advanced (the last grammar course, ENG 220). These courses are assessed summatively through classroom-based assessment and are based on the topics covered in the assigned textbook¹⁴. Classwork should include at least two in-terms and, if desired, other assessment means. The total classwork should not exceed 40% of the total mark, 100. The final examination must be out of 60 marks, which completes the classwork and adds up to a total of 100 marks (Regulation of Study and Examination for undergraduate; <https://units.imamu.edu.sa/colleges/sharia/FilesLibrary/Documents/dras.pdf>). The course teacher(s) constructs her/his own exams and quizzes during the semester, in addition to marking and reporting them. The final exam, however, is created by the course convenor¹⁵ and handed out to all the course teachers who administer the exam and correct and report the results within three days. This practice exists only at institution C and not in the other facilities mentioned in this study.

¹⁴ Grammar Sense 1, & 2 by Cheryl Pavlik / Grammar Sense 3 by Susan Bland

¹⁵ A course teacher probably with the most experience and highest degree

Institution D:

This institution functions under the umbrella of the higher-education sector. It offers a diploma degree to post-secondary students in various fields of public administration. This facility does not provide an English major pathway but hosts an English Language Centre (ELC) that offers a one-year language programme to preparatory students. The programme, which is mandatory, improves the candidates' English skills and leads them to advance the same beyond an intermediate level (however, students who score 5.5 in IELTS or 65 in TOEFL iBT can be exempted). It comprises four levels, with each level covering the subjects of reading, writing, listening, communication and grammar. Accordingly, institution D is suitable for my study context for two reasons: first, it is a public higher-educational facility; second, the grammar courses are taught and assessed independently. These courses are assessed summatively, through midterms and final exams based on the topics covered in the textbooks provided¹⁶. The course teacher(s) has to carry out the tasks of assessment (Executive Rules for training and studying at <https://www.ipa.edu.sa/ar-sa/Admission/Documents/ExecutiveRoles.pdf>). Table 2 summarizes the research context with regard to grammar assessment.

Table 3.

Grammar Courses and Assessment Means in Current Research Context

Institution A					
College	Department	Course No. & code	Credit hrs.	Level of students	Means of Assessment
Arts	DELL	ENG 106	3	Year 1	Two in-terms (40 %)
				Level 2	Quizzes (10%)
		ENG 211	3	Year 2	Assignments (10 %)

¹⁶ Basic English Grammar + Fundamentals in English Grammar by Azar and Hogen (4th Edition)

Institution A

College	Department	Course No. & code	Credit hrs.	Level of students	Means of Assessment
				Level 3	Final (40%)
COLT	DELT	ENGT 113	3	Year 1	Two in-terms (50%)
				Level 2	Quizzes (10%)
		ENGT 213	3	Year 2	Final (40 %)
				Level 3	

Institution B

Colleges	Departments	Course No. & Code	Credit hrs.	Level of students	Means of Assessment
College of	Applied	LING 131T	3	Level 2	2 In-terms (40%)
Languages	Linguistics			Year 1	Quizzes, participations
	English				(20%)
	Translation				Final exam (40%)
	English				
	Literature				

Institution C

College	Department	Course No. & Code	Credit hrs.	Level of students	Means of Assessment
COLT	DELL	ENG 118	3	Level 1	In-terms (40 %)
				Year 1	Final Exam (60%)
		ENG 119	3	Level 2	
				Year 1	
		ENG 220	2	Level 3	

Institution A

College	Department	Course No. & code	Credit hrs.	Level of students	Means of Assessment
				Year 2	

Institution D

Facility	Programme	Course No.	Credit hrs.	Level of students	Means of Assessment
ELC	English Language	Grammar 1	5	Level 1 Year1	Two In-terms (40 %) Attendance and code of
		Grammar 2	5	Level 2 Year 1	conduct (20%) Final Exam (40%)
		Grammar 3	5	Level 3 Year 1	
		Grammar 4	5	Level 4 Year 1	

As seen from the table above, there is a homogenous perspective among the four facilities with regard to grammar assessment, which, as mentioned, is conducted with a summative angle through classroom-based assessments (in-terms, quizzes and final exams).

4.3.1.3 Regulation of study and examination: Article No. 39 of the executive rules.

Shared by all higher-educational facilities, Article No. 39 is important to the present study as it constitutes students' rights to contest their exam results. If a student is not satisfied by the

outcome of the course results, she/he can make an official appeal to an internal council, which will be placed in charge of resolving disputes over exam results.

Students who wish to challenge their exam results must do so within a certain period of time after the final exam, during which the teachers must submit their exam papers to be reviewed by a designated authority, an examination board. Teachers should be able to provide evidence and justify the students' results if required.

The need of concrete evidence to justify students' overall results grounds teachers' practices within written exams since they are tangible and provide a sense of security to the teachers once their results are requested for review.

4.3.2 The participants.

This study investigated EFL teachers' beliefs and practices with regards to English grammar assessment in higher-educational facilities. Accordingly, it is essential to select an appropriate study sample (Cohen et al., 2005). Gorard (2010) distinguishes between a population and a sample as follows: 'The group you wish to study is termed the population, and the group you actually involve in your research is the sample' (p. 10). The sample in this research is a part of the actual population of EFL teachers involved in higher-educational contexts. It includes both male and female participants to obtain comprehensive and multifaceted perspectives.

For this study, I employ purposeful sampling (Creswell, 2003), a technique widely used in qualitative research for the identification and selection of information-rich cases to ensure the most effective use of limited resources. This involves identifying and selecting individuals or

groups of individuals who are especially knowledgeable about or experienced in a phenomenon of interest.

In this study, purposeful sampling indicates that the participants are intentionally selected due to their relevance to the research questions and to the present context under investigation.

Thus, the participants were EFL teachers in higher-educational facilities (initial pool) who were asked to complete and submit an online questionnaire. Afterwards, using the homogeneity strategy and based on availability and willingness, EFL teachers responsible for teaching grammar courses in English language programmes within these facilities during the academic semester of fall 2018 were contacted and asked to deeply reflect on the phenomenon under study.

Overall, in order to ensure that the collected sample represented the whole sampling population, two criteria were emphasised: first, the participants in this study must be EFL teachers in higher-educational facilities, and second, the participants must have some experience, current or previous, in teaching and assessing English grammar. Demographic information about the participants is provided in detail in later sections, along with the account of each separate instrument, since each instrument involved a different group of participants. I present a summary of the research participants in the figure below.

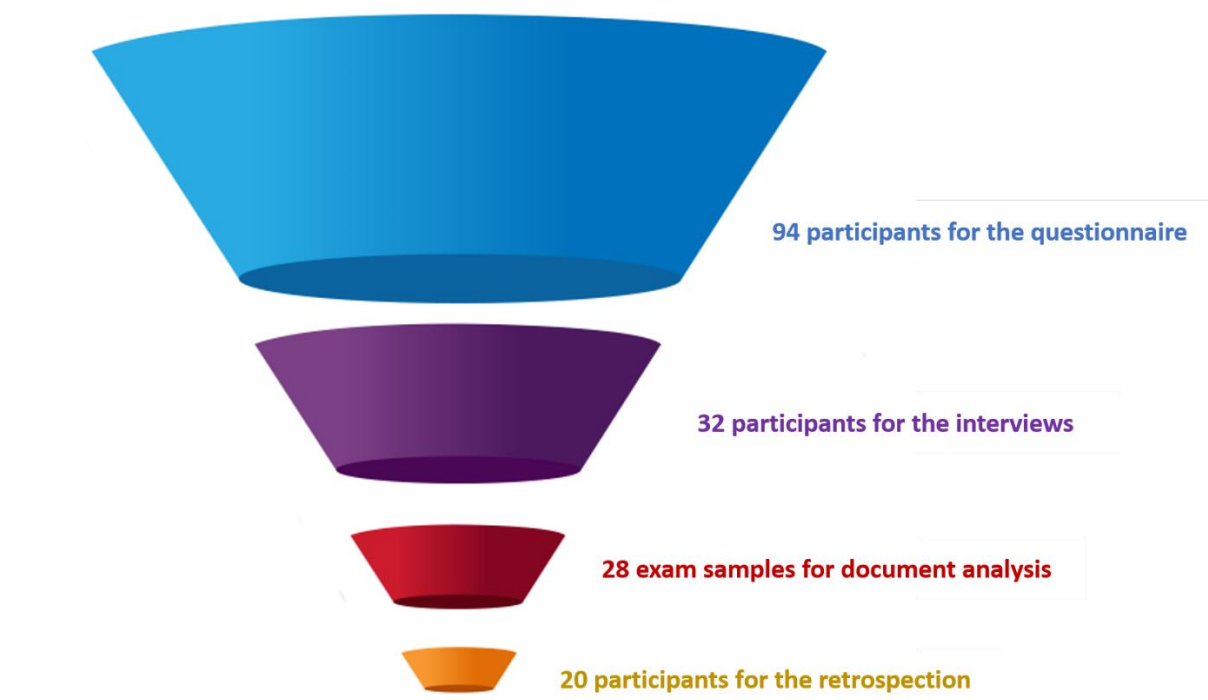


Figure 15. Overall representation of the research participants

4.4 Data Collection Instruments and Procedures for data collection and analysis

This part reviews the definition(s), rationale, advantages and drawbacks of the methods (questionnaire, semi-structured interviews, retrospective thinking and document analysis) used in this study. It also discusses how each instrument was designed, implemented and analysed in this study.

The data all comes from the September–December phase of the academic year 2018. Before the data collection phase, ethical approval and permission letters were obtained from the involved parties (University of Essex, Institutions A, B, C and D). Afterwards, the potential participants (EFL teachers in higher-educational contexts) were contacted personally (via mobile phones and emails) and invited to take part in the study. Once the participants were

recruited, information sheet about the nature of the study (see Appendix A) was provided, consent forms (see Appendix B) were signed and meeting times and places were set. Table 3 presents a detailed account of the research activities and the timeline of the data collection procedures.

Table 4.
Data Collection Timeline

Period (Fall 2018)	Date	Tasks
Week 1(Orientation)	2 Sep – 13 Sep	Recruiting the participants
Week 2	16 Sep – 20 Sep	Questionnaire administration
Weeks 3 + 4	23 Sep – 4 Oct	Interviews (Institution A)
Week 5	7 Oct – 11 Oct	Interviews (Institution B)
Weeks 6 + 7	14 Oct – 25 Oct	Interview (Institution C)
Weeks 8 + 9	28 Oct – 8 Nov	Interview (Institution D)
Weeks 10 + 11	11 Nov – 15 Nov	Document collection + Retrospective thinking (A)
Week 12	18 Nov – 29 Nov	Document collection + Retrospective thinking (B)
Weeks 13 + 14	2 Dec – 13 Dec	Document collection + Retrospective thinking (C)
Week 15	16 Dec – 20 Dec	Document collection + Retrospective thinking (D)

4.4.1 Questionnaire.

Questionnaires can be efficient and economical tools. Dornyei (2003) stated that questionnaires are used to elicit three types of data: ‘factual, behavioural, and attitudinal’ (p. 8). They can also be classified as either quantitative or qualitative based on their design.

Specifically, questionnaires that seek answers through closed-ended questions with multiple-choice scale options are analysed numerically and, thus, act as a quantitative method. On the other hand, questionnaires with open-ended questions are analysed using coding and discussions and are considered as a qualitative method.

Questionnaires are very effective as they allow researchers to collect data from a wider audience in a short time, with low or no cost requirements. However, questionnaires have certain drawbacks, e.g. participants completing a questionnaire may select answers randomly, without reading the question properly. Furthermore, sometimes the high objectivity of questionnaires may prevent respondents from expressing additional thoughts about an issue, perhaps due to the absence of a relevant question.

4.4.1.1 The questionnaire design.

To answer RQ 1, this study used an online questionnaire with closed-ended questions. Online distribution (through Google Forms) was more convenient and practical for two reasons. First, it enabled me to reach out to as many EFL teachers from various higher educational contexts in Riyadh as possible. Second, it could be completed anonymously and in private, which constituted the most conducive conditions for my participants to respond fully and honestly, since in the KSA we are usually most uncomfortable in a face-to-face setting.

The design was developed after an extensive review of the relevant research literature. The items in the designed questionnaire were either taken or adapted from previous similar questionnaires in the empirical studies reviewed in chapter 2. The questionnaire items were modified to suit my research context, i.e., the study explores issues related to EFL teachers' beliefs about English grammar assessment. Three main research studies (Barnes, Fives, &

Dacey, 2015; Elshawa et al., 2017; Muñoz et al., 2012) guided the construction of my questionnaire and helped me identify the key points which are mapped into subset themes that host the questionnaire items.

The first draft of the questionnaire (see Appendix C) was entirely written in English and consisted of two parts. The first part of the questionnaire included items on the respondents' biographical information (11 items). The second part addressed EFL teachers' beliefs and thoughts about English grammar assessment. It consisted of 46 five-point Likert scale items (plus a few open response items) where respondents would have to specify their level of agreement or disagreement with respect to a symmetric agree-disagree scale for a series of statements. These statements, or items, were grouped into subset themes (see Table 4 and Appendix D).

The questionnaire was piloted in April 2018, subsequently being administered in English to 30 EFL teachers working on a private international institution, who were not included in the main study. The comments received from the pilot-test participants were related to the length of some items, redundancy and the overall structure of the questionnaire. Three examples of modification and revision of items are given here:

- First, the questionnaire items were presented in two parts: demographical and beliefs. Items in the second part target various themes which were all mixed together. Based on the participants' feedback and my opinion, items were grouped according to their theme. This allowed for more clarity regarding what the questionnaire is about and what the participants were responding to.
- Second, some questionnaire items were made clearer, while others were rephrased because confusing structure and ambiguous items could render the questionnaire

ineffective. Accordingly, item 11 in the first draft of ‘English grammar assessment informs teaching (diagnoses strengths and weaknesses in teaching)’ was rephrased to ‘The purpose of English grammar assessment should be to inform teaching by showing the students’ strengths and weaknesses in English grammar’ (item 4 in section B).

- Third, some items were deliberately missed by most participants, resulting in low internal consistency scores. Therefore, these items were removed to allow continuity throughout the questionnaire, e.g., item 10 in the first draft ‘Assessing English grammar is a waste of time’. Other items were added, such as the approximate number of students in a class and the training received on language assessment, to the demographical part of the questionnaire (see Appendix D).

After the modifications, there remain two parts: demographic information and beliefs about English grammar assessment. There are 46 quantitative items in part two, which are grouped into five categorical themes presented in Table 5.

Table 5.
Questionnaire Categorical Themes

No.	Theme	Likert Items
1.	Teachers’ beliefs about the general nature of English grammar assessment	1-5 (5)
2.	Teachers’ beliefs about the purposes of English grammar assessment	6-13 (8)
3.	Teachers’ beliefs about English Grammar assessment methods	14-28 (15)
4.	Teachers’ Beliefs about English Grammar Assessment Formats	29-38 (10)
5.	Teachers’ beliefs about their role and sources used in constructing English grammar assessment tasks	39-46 (8)
Total		46

The questionnaire was anonymous in order to encourage the participants to respond truthfully. The questionnaire can be found on the following link:

<https://drive.google.com/open?id=1338II07eo4N6jFb-nzRFbT7SZrpaq1FU2DKR8D69ows>

4.4.1.2 Data Collection.

After official permissions were obtained to conduct the study in the four educational contexts, an email was sent to the English department chairpersons requesting the EFL teachers' participation by asking them to complete the questionnaire on the link provided and to forward the email to the these teachers. This served my purpose of collecting data from a wide range of people relevant to my research topic. Respondents were then screened based on the study needs.

4.4.1.3 The questionnaire participants demographics.

136 teachers responded to the online questionnaire. However, 36 respondents were excluded from the analysis because they were either working in schools or did not mention their workplace. Hence, they could not be regarded as genuine representatives of the targeted population in the current study.

Work place	Number
KSU	38
IMSIU	35
PNU	15
IPA	2
SEU	2
AOU	3
PSU	5
Total	100

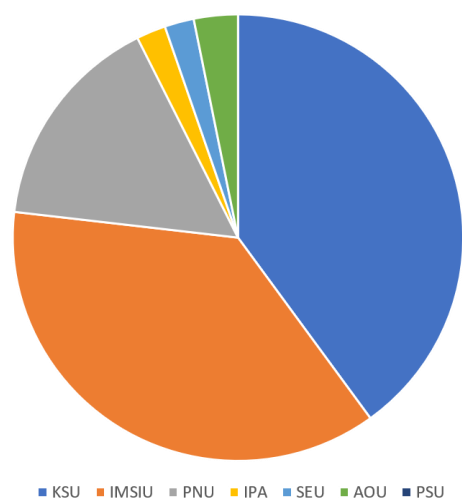


Figure 16. Participants' place of work.

A second important prerequisite for the sample was that the teachers had some experience of assessing and/or teaching English grammar. As there were six teachers who stated neither kind of grammar-related experience, those were also excluded, which left 94 EFL teachers representing the targeted population of EFL teachers.

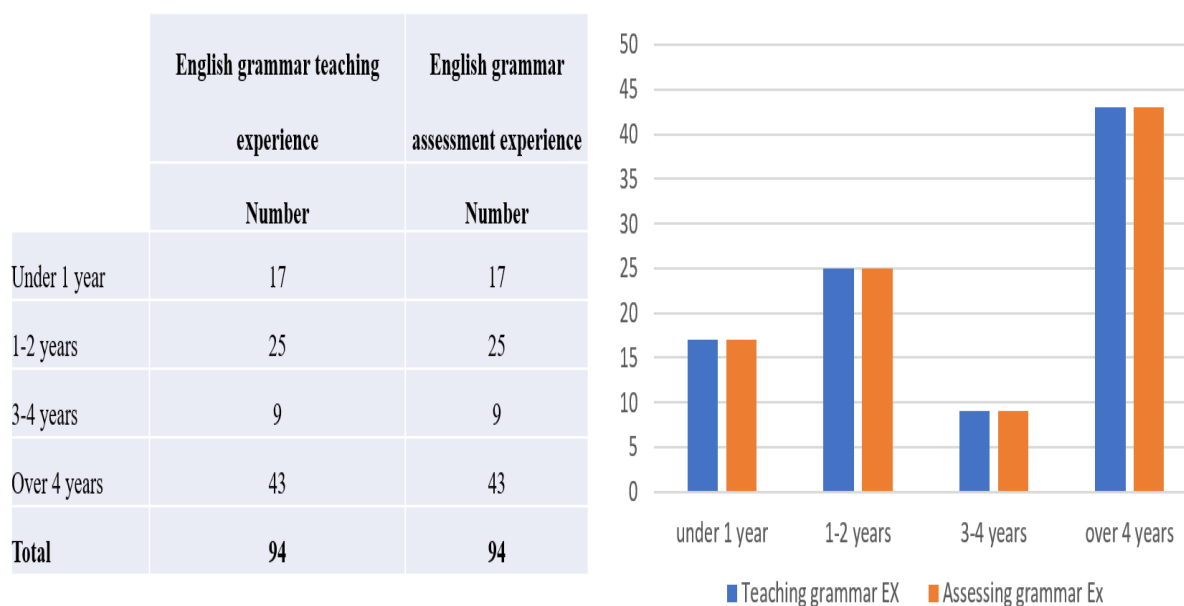


Figure 17. Participants' experience of teaching and assessing English grammar.

As a result, 94 suitable participants were identified. The cohort included, male ($N = 9$) and female ($N = 85$) participants ranging from 21 years old to over 50 years old; almost all of the participants are of Saudi origin ($N = 84$). Some of the teachers had BA qualifications ($N = 19$), while a little over half of them held MA qualifications ($N = 50$) and the rest were PhD holders ($N = 25$). The majority ($N = 74$) had received some kind of training on language assessment. The following table summarises the basic background information on the participating teachers.

Table 6.

Background Information about the Participants in the Questionnaire

Variables		Number / 94	Percent %
Training Received	None	20	21.4
	Through UG courses	36	38.4
	Through MA/ PhD courses	3	3.2
	Professional training	35	37
Gender	Male	9	9.6
	Female	85	90.4
Age	21-30	21	22.3
	31-40	43	45.7
	41-50	24	25.5
	Over 50	6	6.4
Country of Origin	Saudi Arabia	84	89.4
	Jordan	4	4.3
	Syria	1	1.1
	Egypt	2	2.1
	Algeria	1	1.1
	India	1	1.1
	USA	1	1.1
Educational level attained	BA	19	20.2
	MA	50	53.2
	PhD	25	26.6

With respect to the relevant classroom teaching matters (Table 6), almost all the participants were experienced English teachers with more than five years of experience, while a quarter had more than 15 years of experience (N = 25). In their classrooms, class sizes of 20 pupils or more were commonest, although a fifth of teachers claimed class sizes over 40 students (N = 21). The levels taught were primarily either foundation level (level 1 or 2), taken by students of almost all majors, or levels 3 or 5, where normally only English majors are in the classroom. Finally, with respect to the grammar textbook used, the most dominant textbook was found to be ‘Understanding and Using English Grammar’ (N = 34), which is consistent with what has been reported earlier in 4.3.1.2, about this textbook being used in as many as three educational contexts.

Table 7.
Participants’ Relevant Classroom Backgrounds

Variables		Number / 94	Percent
English teaching experience	1–5 years	2	2.1
	6–10 years	41	43.6
	11–15 years	26	27.7
	Over 15 years	25	26.6
Number of students in the grammar class	Under 20	9	9.6
	20–30	28	29.8
	31–40	36	38.3
	Over 40	21	22.3
	Foundation 1 + 2	30	31.9
	Level 3	26	27.7

Level of students currently taught grammar	Level 4	9	9.6
	Level 5	18	19.1
	Level 6/7	1	1.1
	Level 8/9	3	3.2
	Graduate	1	1.1
	Other	1	1.1
	Not currently teaching grammar	5	5.3
Grammar textbook being used	English Grammar in Use	25	26.6
	Understanding and Using English Grammar	34	36.1
	Interactions/Mosaic Grammar	17	18.1
	Grammar Sense	10	10.6
	Basic/Fundamentals of English Grammar	4	4.3
	Other textbooks	4	4.3

To sum up some of the essential demographics of the participants, the number of the participants was 94. All those participants were EFL teachers from public and private higher-educational facilities. The majority of them were female (90.4%). The average age of the participants is between 31–40 years. They all had undergone some kind of language assessment training at some point in their teaching careers. The duration of their grammar teaching and assessment experience varied from 1 year to over 4 years.

The participating teachers in this study do not, as Cohen et al. (2007) state, ‘represent the wider population’ (p. 104) of higher-educational facilities in Saudi Arabia, and hence the findings from this study group are not applicable in general, as they present the particular, subjective perspectives of these participating teachers.

4.4.1.4 Questionnaire data analysis procedure

4.4.1.4.1 Data handling.

The responses to the questionnaire were downloaded from Google Forms in an Excel file. All of them were then converted into numbers before the data was copied into SPSS for analysis (see Appendix O). All the belief item responses on the scale ‘strongly disagree’ to ‘strongly agree’ were represented on a scale of 0–4. In terms of analysis, first of all, internal reliability was assessed. Second, appropriate analytical tests were selected according to the normality of distribution of the data.

4.4.1.4.2 Reliability checking.

Although Cronbach’s alpha is widely used in questionnaires as well as tests to assess internal reliability (Taber, 2018), it was not deemed appropriate in the present instance as it is only applicable for sets of items in questionnaires where multiple items measure a single construct. The present questionnaire, however, largely follows the commonly found pattern of ‘one item per target construct measured’. Hence, there are no expectations of agreement between the responses within large subsets of items, and in fact the lack of agreement in responses is to be expected, rather than being misinterpreted as a sign of unreliability (Tavakol & Dennick, 2011).

Some pairs of items were, however, selected due to their logical relation in terms of meaning in some way, to check if the expected agreement in response was present. For instance, in the description of participants above, some participants reported that they had no experience of either teaching or assessing grammar. When I looked at their responses concerning the grammar textbook they used in their class, I found that they had responded 'none' or 'not teaching grammar at the moment'. Since that response was consistent with what they reported earlier about their experience, one can judge that their responses are quite reliable.

Again, among the belief items, there are some among which one might detect a logical connection, such that if a participant agrees with one belief, they must also agree with the other, unless they were not responding with care (i.e., unreliably). Two items, for example, refer in slightly different ways to the advantages of multiple-choice items in assessment instruments (Belief Items 19 and 34). Hence, if a teacher agrees with one, he/she should agree with the other, and vice versa. The Cronbach's alpha between the responses to these items was in fact .519, which is moderate as a measure of reliability (where a value of .7 or better would indicate a really high reliability). A similar check between items 19 and 31, both of which mention cloze items, yielded an alpha of .569. However, it must be borne in mind that reliability increases when all the items in a subset measure the same thing (Tavakol & Dennick, 2011); hence, a subset composed of only two consistent items cannot be expected to attain the values that a set of five or ten such items would achieve. Thus, in instances where it is sensible to assess it, there is evidence for at least a moderate reliability of the instrument.

4.4.1.4.3 Normality checking.

In order to decide what statistical tests might be appropriate, it was necessary to check the normality of the distributions of the belief ratings. Therefore, the Kolmogorov-Smirnov test

with Lilliefors correction (Lilliefors, 1967) was applied to the rating responses for each of the 46 belief items. All of them emerged as distributed highly, significantly and differently in relation to the normal distribution (bell shape), with $p < .001$. This often arises with short score scales such as five-point rating scales. Therefore, nonparametric inferential statistics were used to test significances for the results, as follows:

1. In the account of the results for all valid participants together on each belief item, the sign test was used to assess whether the teachers were expressing an overall view that definitely departed from the midpoint of the rating scale (= 2, on my scale), either higher or lower, or a view that essentially did not significantly differ from the midpoint. The sign test allowed the assessment of whether there were significantly more responses above the midpoint (i.e., 3 and 4) than below the midpoint (i.e., 0 and 1), or the reverse, or whether there was no significant difference between those above and those below (and hence no clear opinion was expressed).
2. When comparing groups such as genders, or the Saudi versus non-Saudi teacher origin, I used the nonparametric Mann-Whitney test.
3. When testing relationships between the beliefs of teachers and attributes in ordered categories, such as age groups, educational levels, or degrees of experience in grammar assessment, I used the Spearman correlation.

4.4.2 Semi-structured interviews.

In teacher cognition research, teachers should be given the opportunity to explore and reflect on their own personal beliefs (Kagan, 1990). Interviews can serve as a very useful tool to elicit in-depth and detailed information and insights about participants' beliefs, thoughts and knowledge (Bell, 2010; Cohen et al., 2011; Denscombe, 2014). However, it has been suggested that teachers may feel uncomfortable when questioned about their beliefs and

practices, and their responses may not be genuine but rather carefully structured (Kagan, 1990). Therefore, it is essential to employ strategies that would help teachers remain comfortable and encouraged to express their thoughts and beliefs. One way to accomplish this state of moderate openness can be through semi-structured interviews. In semi-structured interviews, a set of questions are prepared for the interviewee, while at the same time additional questions might be asked during interviews to clarify and/or further elaborate on certain issues.

4.4.2.1 Semi-structured interview design.

This study employed semi-structured interviews to generate the participants' own descriptive accounts of their beliefs and practices regarding English grammar assessment to help answer RQ1 and RQ2. Semi-structured interviews were also employed to illuminate the link between the beliefs of the teachers and their current practices while assessing English grammar (RQ3).

My initial plan had in fact involved a structured interview. However, as I piloted my initially structured interview (see Appendices E and F), a number of issues turned out to be problematic. Particularly, the interview questions were very restrictive in the sense that interviewees' responses were fixed and specific, almost a few words per question. This problem was highly evident with the close-ended questions. The participants in the pilot study were just satisfied with 'Yes/No' answers and showed no interest in elaborating on their responses.

The semi-structured interview therefore was deemed more suitable for my research. I felt that a fully unstructured interview could be difficult to handle well for a novice researcher such as myself. It ran the risk of wandering off target and not producing optimally relevant

information. The semi-structured format however seemed more suitable, since it consisted of flexible questions that provide a basic structure but allow the interviewer to organise a conversation and steer it properly, so it does not elicit one-dimensional answers from the interviewee as in fully structured interviews. A semi-structured interview is open and allows new ideas to be brought up in its course as a result of what an interviewee says.

The design of the semi-structured interview followed the following principles:

- Questions in the interview were adapted and based on interviews conducted in the reviewed studies (Karim, 2015; Mansory, 2016; Mussawy, 2009; Saad et al., 2013) (see Appendix G).
- The interview questions were broad and open-ended to allow the interviewees latitude in constructing answers.
- Any questions of the interview could be modified wherever required, which would allow for more relevant questions to be asked and the interviewees to clarify their responses.
- The wording, the structure and the order of some of the questions were changed based on the pilot study.
- Questions were clear, simple and short so as not to confuse the interviewees.
- Questions were designed within a time frame that suited the participants – not less than 30 minutes and no more than one hour – since long interviews might have led the respondents to experience fatigue, making them unwilling to continue (Robson, 2011).
- Interview questions were developed and grouped into four themes: 1) understanding what grammar assessment means, 2) identifying the purposes of English Grammar Assessment (EGA), 3) elaborating on the relevant factors and 4) discussing their

roles in constructing EGA. Organising the questions into themes facilitated the coding and the analysis phases (Table 8).

Table 8.

Themes for the Interview Guide and Questions

Themes	Categories	Questions
Beliefs	Type of EGA	1, 2
	Purpose	3, 7
	Role of EFL teachers in constructing EGA	8
	Factors	3, 7
Practice	Type of EGA	1, 4
	Purpose	5
	Role of EFL teachers in constructing EGA	6
	Factors	4

As the above table shows, there are 8 questions which target the core themes of the present study: beliefs and practices. Five questions were used to elicit the participants' underlying beliefs while four questions focused on the participants' practices. The questions in the semi-structured interviews were interconnected to allow the identification of the relationship between the participants' beliefs and practices.

4.4.2.2 Semi-structured interview data collection procedures.

In the present study, face-to-face semi-structured interviews with EFL English grammar teachers were conducted. English was used as the medium of communication in the interviews because all the participants spoke and understood the language very well,

including the interviewer. Moreover, using English from the start made the process of transcription easier, since there was no need for translation from Arabic to English.

All the interviews lasted for thirty minutes or more and were recorded using a digital voice recorder (see Appendix H). The use of a voice recorder assisted in the production of highly detailed and accurate transcripts, since it provided the opportunity to examine the recordings as many times as required (Silverman, 2000). Also, the use of a voice recorder proved to be easier in retrieving information and analysing the findings of the study.

Approximately 20 hours of interview data were collected. Interviews with the female participants took place in their offices (in their educational facilities) during office hours. Male participants were interviewed in the Executives Hotel¹⁷ in the hotel lobby, due to the religious and cultural aspects that regulate meeting with male strangers in public places, and with a chaperone; my husband was with me and would keep an acceptable distance, which allowed for private conversations without complete seclusion.

4.4.2.3 The interview participants' demographics.

32 EFL teachers participated in the interviews. These included both females (N = 26) and males (N = 6) and were representative of the same population which was sampled for the questionnaires (4.5.1.3). All the participants were teaching English grammar courses at the time of interviewing. They had, on average, been teaching for 12 years, the most experienced having taught for 25 years while the novices had only one year to three years of experience. With regard to participants' qualifications, most of the participants were PhD holders (N = 18) in the field of applied linguistics, theoretical linguistics, education and sociolinguistics,

¹⁷ This hotel was chosen because it is in the centre of the city of Riyadh and accessible to all areas.

while the remaining (N = 14) had a masters' degrees. The majority of the participants were Saudis. The profile of the sample is displayed in Table 8.

Table 9.
Interview Participants' Demographics

Institutions	Qualification		Average years of Experience		Origin	
	Masters	PhD	English Teaching	English Grammar teaching	Saudis	Non-Saudis
A	7	11	10	6.5	17	2
B	Ø	1	20	10	1	Ø
C	6	3	12	6	8	Ø
D	1	3	18.5	8	3	1
Total number (% of sample)	14 (44%)	18 (56%)			29 (91%)	3 (9%)
Overall Mean			12	7		

The table above sums up the interview participants' demographics. It is essential to mention here that there is no definite way to ensure that the teachers participating in the interviews took the questionnaire, since the questionnaire was anonymous. Also, personal information about age was not provided, because in Saudi Arabia most people, especially in a voluntary interview, are not comfortable talking about these aspects.

4.4.2.4 Semi-structured interview data analysis procedure.

As Merriam (1998) states, semi-structured interviews are typically analysed qualitatively. Qualitative analysis involves continual reflection and interpretation of the data obtained in order to generate sufficient information that would be tailored to answer specific research questions (Creswell, 2003). Data from the interviews were therefore subject to content

analysis. 'Content analysis is the process of organising information into categories related to the central questions of the research' (Bowen, 2009). In this respect, Cohen et al. (2007) suggest that content analysis involves not only coding and creating meaningful categories but also comparing and making links among data from different sources.

The analysis process unfolded in three phases: transcription/reading, coding and categorization. First, the audio-recorded material from each interview was transferred from the voice recorder storage unit to my personal laptop in preparations for transcription. Each interview was then imported to Dragon, the speech-to-text software (see Appendix I). Dragon allowed the transformation of voice into text within minutes and facilitated the transcription of digitalised audio-recorded files. The texts were then exported to word documents. Once the transcription of each audio material was completed, the audio along with its generated text were uploaded to oTranscribe, which is a free web app that allows one to bring both text and audio material together. Thus, one does not need to shift back and forth between the word document and the audio player. Another advantage of using oTranscribe is that it offers interactive timestamps to adapt the audio speed to one's convenience. All the transcriptions generated by Dragon were reviewed and edited in oTranscribe. Both the anonymity and confidentiality of the data collected through the interviews were guaranteed by giving each participant a number (e.g. 004) and associated initials for pseudonym (e.g. RSh) and by deleting any possible identifiable details immediately after transcription (see Appendix J). Figure 18 illustrates how the transcription process is presented in the oTranscribe template.

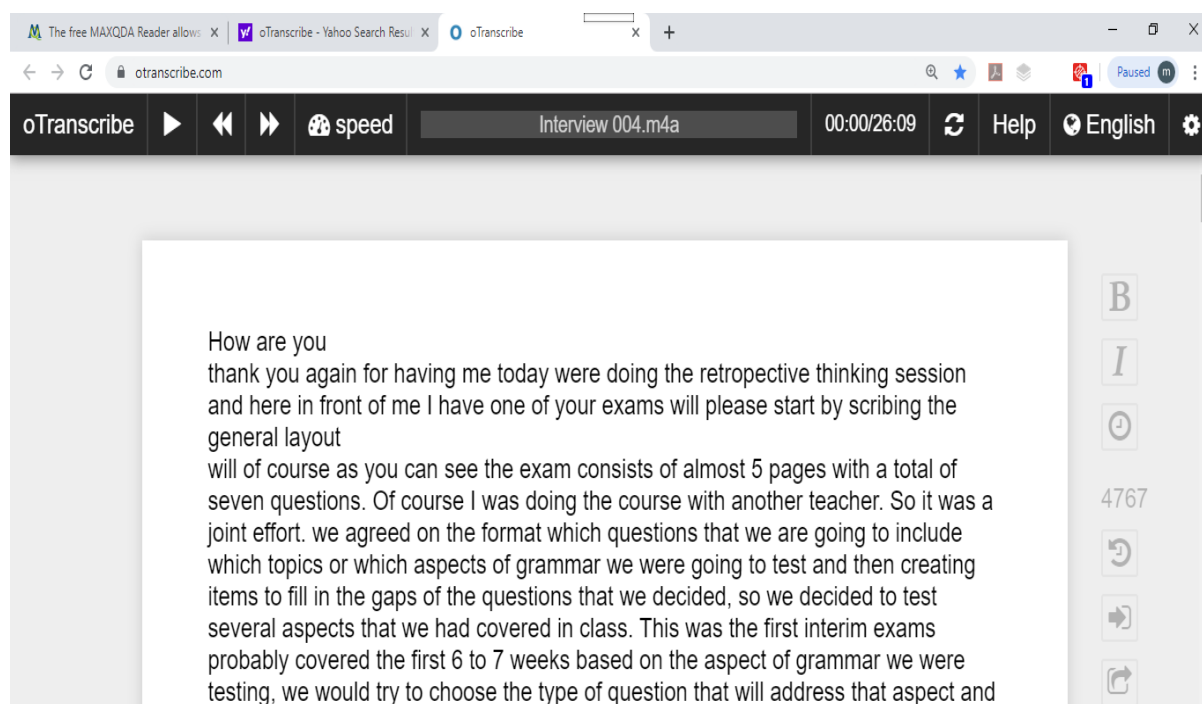


Figure 18. oTranscribe template and layout.

Soon after each transcription, the transcript was carefully checked for (verbatim) accuracy against the original digital recording. Furthermore, the accuracy of transcripts was checked and verified by one interviewee, who even requested to self-review the transcript of her interview. Checking and editing the transcripts of the interviews gave me the opportunity to familiarise myself with the data and mentally begin the coding process. As Braun and Clarke (2006) state, ‘It is vital that you immerse yourself in the data to the extent that you are familiar with the depth and breadth of the content’ (p. 87). Therefore, a systematic reading of the transcripts was conducted, and some initial thoughts and interesting points were noted before I engaged with the formal coding and the initial ideas. The process of reading was instrumental in facilitating the coding phase of the analysis.

With the first phase of interview data analysis completed, the second phase, the coding, commenced. As I mentioned earlier, the transcribed data were read multiple times;

afterwards, a preliminary list of codes based on the topics under investigation (RQs) was generated through the use of MAXQDA 2018, a software that facilitates coding frequency searches, word frequency and text searches and keyword searches. MAXQDA proved to be highly useful in allocating codes within and across the transcribed data. This programme also helped in finalising codes as well as generating specific categories and themes (Figure 19).

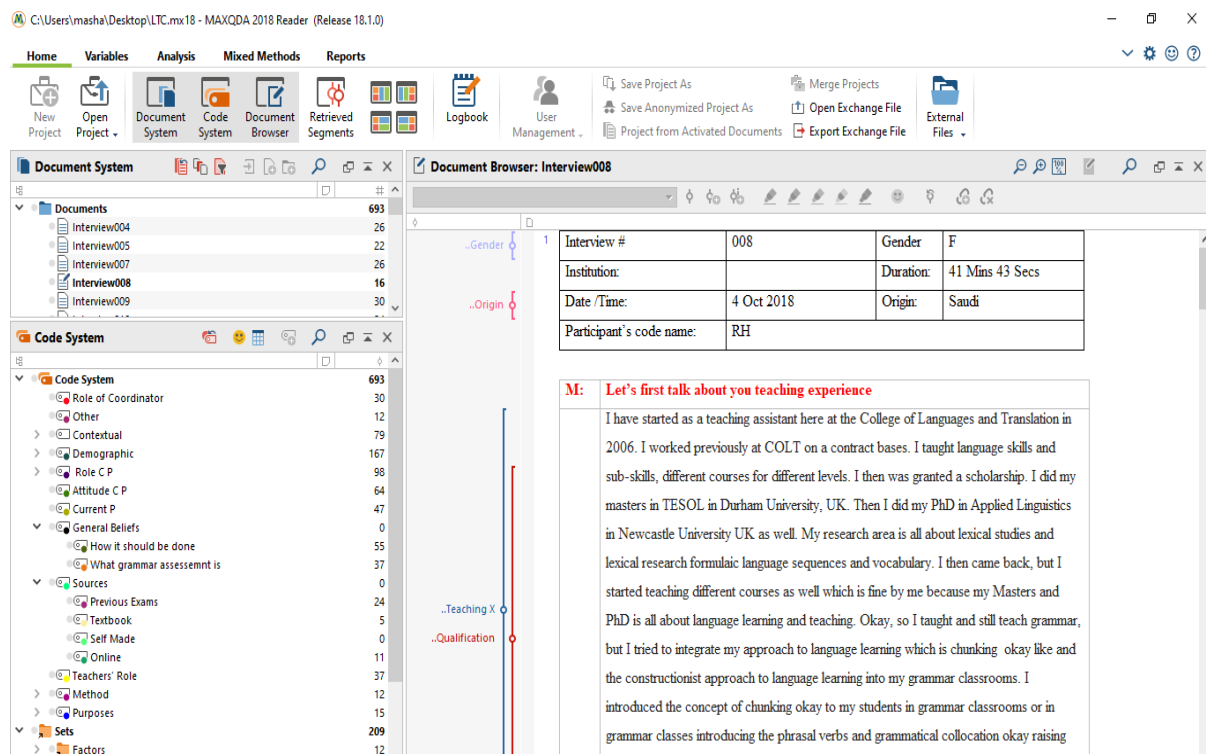


Figure 19. General overview of the coding system in MAXQDA 2018.

After quotes were coded, the third phase of data analysis began. The codes were entered under several different categories, which were grouped under three major themes: EFL teachers' beliefs, EFL teachers' practices and relevant factors (Figure 20).

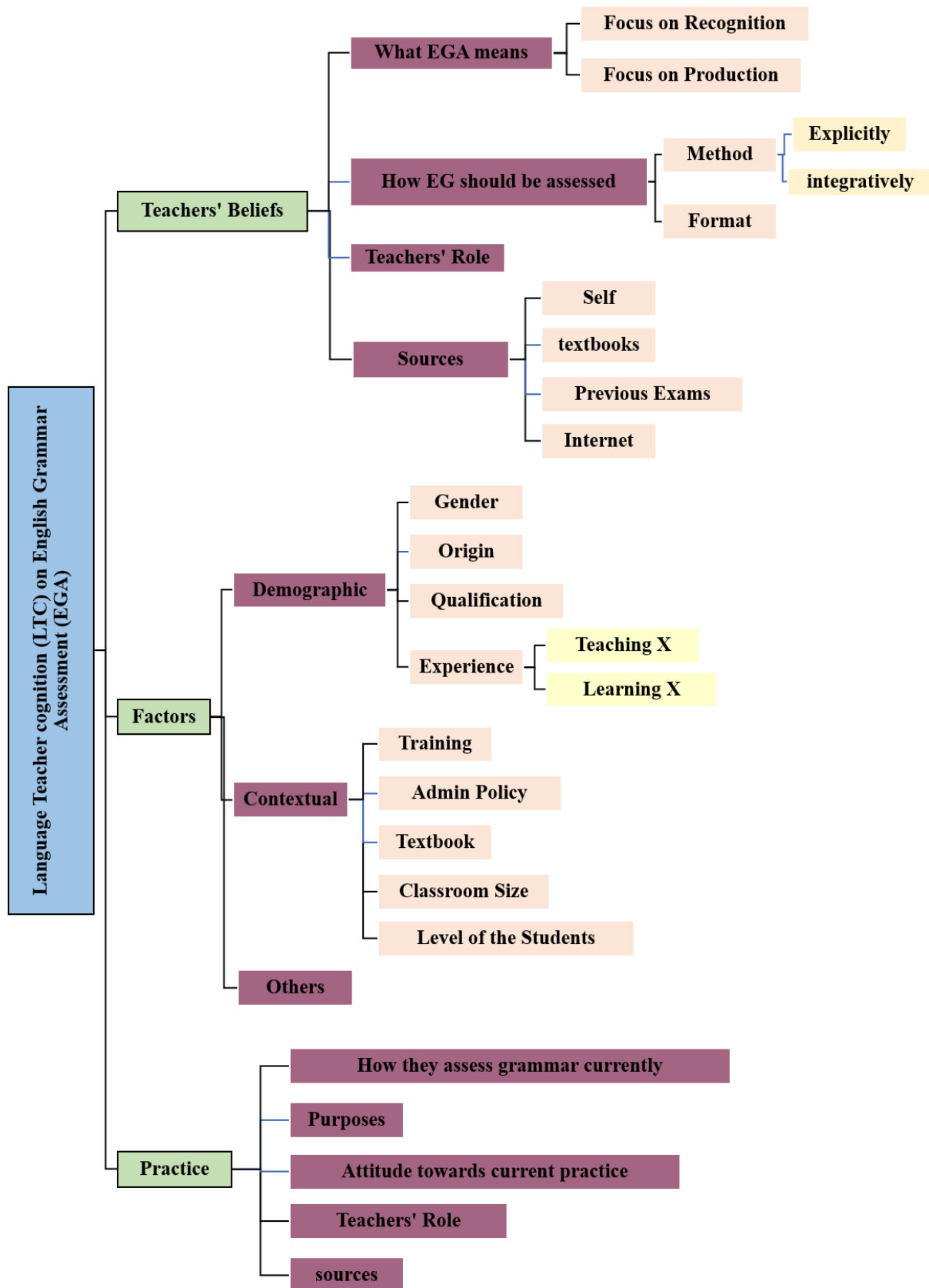


Figure 20. Themes and categories derived from the data analysis of the interviews.

The large amount of data collected in this study necessitated the recurrent reading of the transcripts, along with listening to audio recordings of the data and coding and recoding the transcripts. Braun and Clarke (2006) asserted that ‘the need for re-coding from the data set is to be expected as coding is an ongoing organic process’ (p. 21). In summary, data analysis began with the coding of the data, progressed to emerging categories and finally streamed into the themes more closely related to the research questions.

4.4.3 Retrospective thinking.

Retrospective thinking, also known as ‘stimulated recall’ or ‘aided subsequent verbal protocol’ (Van Den Haak, De Jong, & Schellens, 2003), has been widely used and is considered an effective way to study participants’ cognition and thought processes (Goo, 2010; Janssen, van Waes, & van den Bergh, 1996; Sasaki, 2008). Retrospection is based on information processing (IP) theory, which claims that information is stored in people’s short-term memory (STM) and is available for retrieval through verbal reports (Sasaki, 2008). This involves verbal reporting or commenting on people’s cognitions with reference to prior behaviours (Borg, 2006). This definition highlights two main characteristics of retrospection: first, that it is verbal, and second, that it is asynchronous, as it occurs only after a task is finished (Janssen et al., 1996).

Originally, I intended to use concurrent think-aloud to gather data. Unfortunately, during the pilot phase of the study (see below) this method to elicit a direct representation of the teachers’ cognitive processes while they were constructing their grammar assessment tasks, I encountered rejection from not only the administration party but from the teachers themselves. This was due to the fact that constructing assessment tasks, especially in the form of written exams, is subject to confidentiality and secrecy. As a result, I resorted to an

alternative, retrospective thinking. Retrospection, therefore, was employed in this study as a means to allow participants to explain/describe what went on in their minds while they construct grammar assessment tasks.

4.4.3.1 Retrospective thinking design.

As described above, retrospection is one of the main techniques used to capture teachers' interactive thoughts and decision-making processes and it was thus considered a good alternative to think aloud to gather information with regard to constructing classroom-based assessments (RQ 1 and RQ 2). Retrospective sessions, in general, are unstructured in order to encourage the participants to take the initiative and talk about any aspect of the phenomenon under study (Borg, 2006; Meijer, 1999; Woods, 1996). However, as Woods notes, the researcher should sometimes play the role of a facilitator or a prompt, because teachers vary greatly in the extent to which they take the initiative to comfortably identify episodes and comment on their own practices.

In this study, retrospection was based on the participants' classroom-based assessments. I used samples of the participants' written exams to encourage the teachers to walk me through the cognitive process, guide them in writing their exams and let them reflect on the factors that would influence their practices or prevent them from enacting their beliefs.

Conducting retrospection proved to be a very appealing method for two reasons. First, retrospective sessions were time-efficient, since the participating teachers were able to execute a task in their own manner and pace. Second, since the retrospection was carried out in a bilingual context and English was used as the means of communication, it was probably

less difficult for the participants to verbalise their thoughts in English after they had completed their written exams rather than while they worked on them.

However, there was a major concern regarding the fact that the participants might have produced biased accounts of the thoughts they had while performing the tasks, i.e., talking about how they wrote their exam. They might, for instance, have forgotten specific things that had occurred when they were writing their exams. Ericsson and Simon (1993) emphasise that vital information may be lost in the case of retrospective research, which is confirmed by several studies (e.g. Russo et al., 1989; Teague et al., 2001). Bias might have also risen as a result of the participants deciding to conceal certain thoughts they had, invent thoughts they did not have, or modify their own thoughts, perhaps due to reasons of self-presentation or social desirability.

Despite this drawback, retrospection proved to be a valuable tool to gain access to teachers' mental representations of writing exams. Furthermore, the retrospections were one of the various tools for data gathering in this study, and thus facilitated the triangulation process to ensure better validity of the research design.

4.4.3.2 Retrospective thinking data collection procedure.

Retrospective sessions took place after the semi-structured interviews were conducted (see Table 4.4). All such sessions were conducted in English. As with the interviews, the female participants had the retrospection sessions in the offices in their respective educational facilities, while the male participants had theirs consecutively with their interviews at the Executive Hotel, because it would have been inconvenient for them to make the trip twice to meet me. All the retrospective sessions were audio-recorded digitally. Session durations

varied from 10–20 minutes, depending largely on the teachers and the time available. A total of five-and-a-half hours worth of data were collected.

Each participant presented a sample of her/his written exam during the retrospection, which was used to stimulate and prompt her/him to talk about what went on inside their minds during the construction and writing of the exams.

4.4.3.3 Retrospection participants' demographics.

20 EFL teachers volunteered for the retrospection, including both females (N = 15) and males (N = 5). Half of the participants' (N = 10) were MA holders while the other half were PhD holders. Their teaching experience varied from 4 years to over 15 years. The majority of the participants were Saudis (N = 19). Table 10 summarizes the participants' information.

Table 10.
Retrospection Participants' Demographics

University/ Institution	Gender		Qualification		Origin	
	Male	Female	Masters	PhD	Saudis	Non-Saudis
A	4	8	7	5	11	1
B	Ø	1	Ø	1	1	Ø
C	1	4	3	2	5	Ø
D	Ø	2	Ø	2	2	Ø
Total	5	15	10	10	19	1
number (%) of sample)	(25%)	(75%)	(50%)	(50%)	(95%)	(5%)

It is important to emphasise here that the participants who agreed to the retrospection constitute a part of the interview participants' sample and are the one who provided a sample of written exams for the document analysis (see section 4.5.4 below).

4.4.3.4 Retrospective thinking data analysis procedure.

The transcription procedure of the retrospective audio data was identical to the interview procedure. With the aid of Dragon, I was able to transcribe all my audios to texts in a matter of a few hours. Again, I relied on oTranscribe to transcribe (verbatim) the text into MS Word documents in English. Once all the audios were transcribed and checked, I imported them to MAXQDA to begin the reading and coding processes (Figure 21).

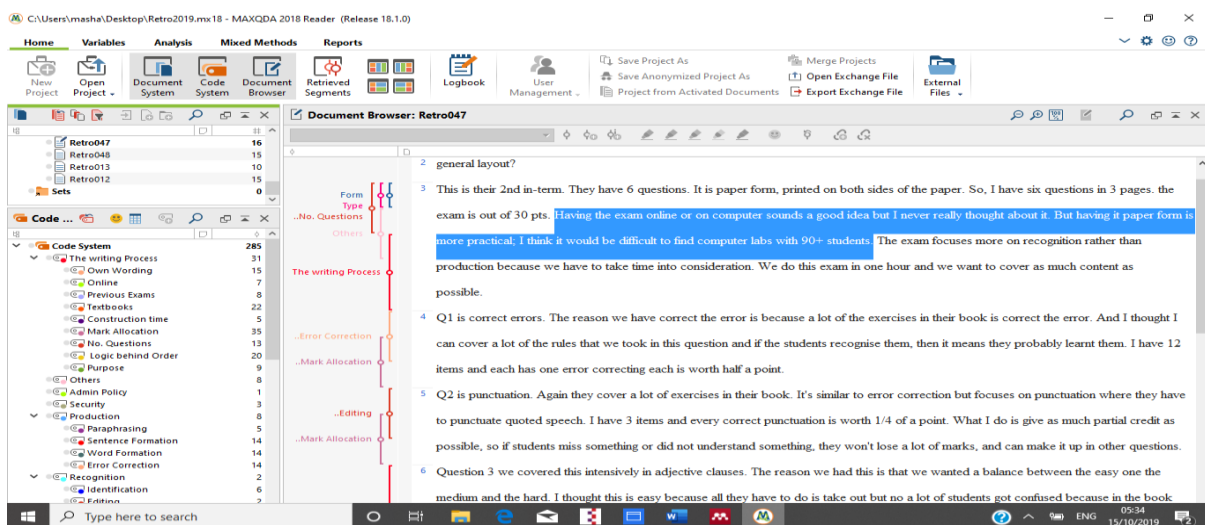


Figure 21. Coding procedure for the data gathered from the retrospection.

This coding process involved identifying the relevant segments to describe teachers' practices and the related factors. The next step was to go through the coding scheme again in order to put them together and organise them into categories. Thereafter, the categories were reorganised to generate broader themes. Figure 22 shows the categories and themes which emerged from the analysis process.

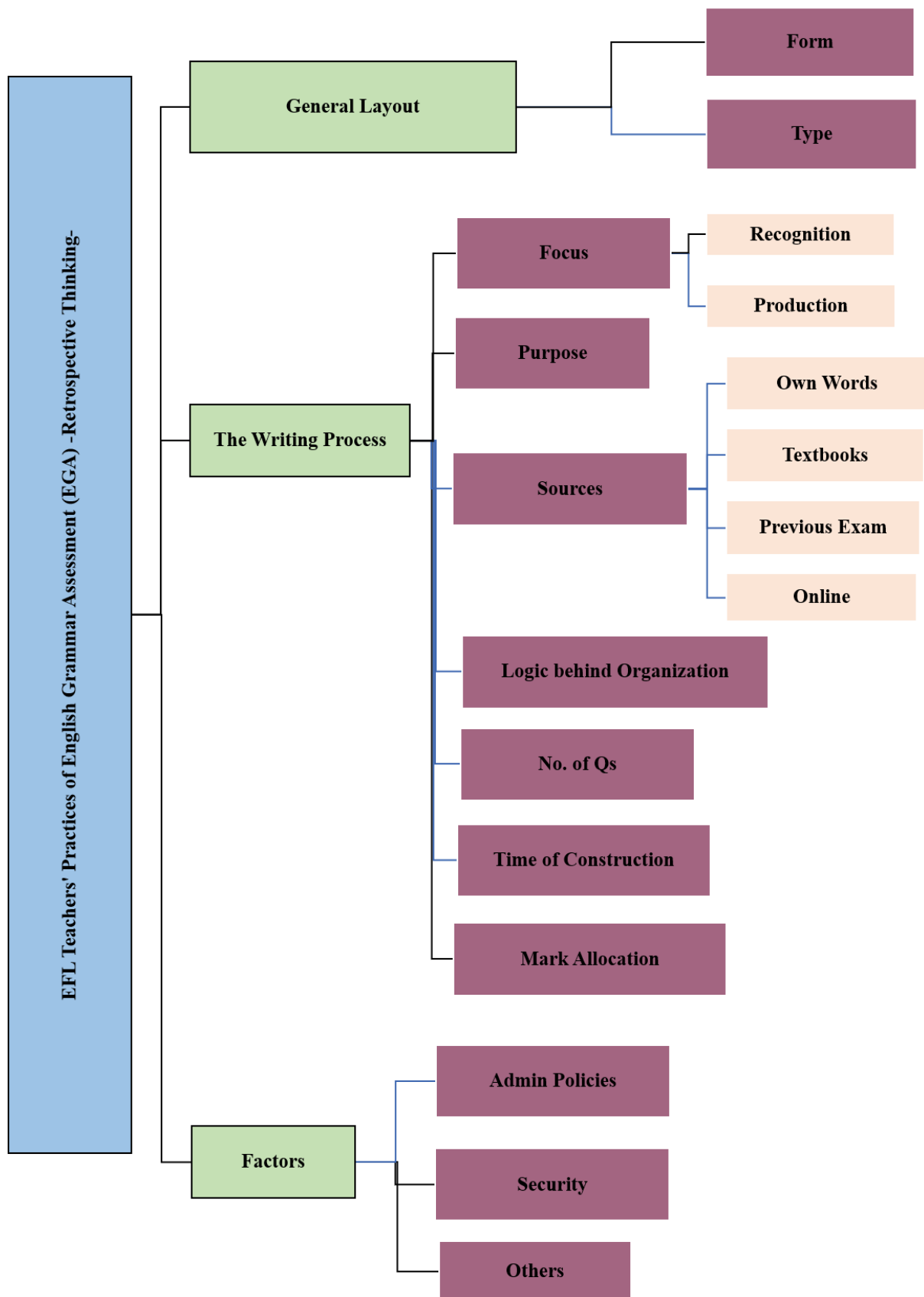


Figure 22. Categories and themes generated from retrospective thinking.

4.4.4 Document analysis.

‘Document analysis is a systematic procedure for reviewing or evaluating documents—both printed and electronic (computer-based and Internet-transmitted) material’(Bowen, 2009).

‘As a research tool, documents can produce intensive rich descriptions of a single phenomenon and represent a specific version of realities for specific purposes’ (Stake, 1995; Flick, 2014).

Mixed-methods studies sometimes include document analysis either as a standalone method or as a supplementary research tool. For example, in an attempt to explore a teacher training programme designed to address the assessment needs of LCTL¹⁸ educators working in short-term foreign language programmes in the US, Montee et al. (n.d.) conducted a mixed-method research using questionnaires and document analysis. Document reviews became the primary source of data to identify the participants’ performance while developing assessment tasks in their language programmes. Also, Sogunro (1997) used questionnaires combined with interviews, document analyses and direct observations to examine the impact of training on leadership development. The documents were additional tools which provided supplementary information about the training programme’s history, goals, objectives, enrolments and substantive content.

In this study, documentary sources refer to any written classroom-based assessment in the form of quizzes, midterms and/or final exams. These documents serve a variety of purposes as part of the research undertaken. First, the documents provide core research data regarding the participants’ actual practices of grammar assessment and the factors influencing them (RQ 2). Second, information and insights from the documents allow the detection of any

¹⁸ Less Commonly Taught Languages

mismatches between what the participants say they do and what they actually do (RQ 3).

Third, document analysis played a vital role in data triangulation.

However, document analysis is not always advantageous. Here are some of its limitations:

- **Insufficient detail:** Documents are produced here for assessment purposes and are created without a research agenda. Consequently, they usually do not provide sufficient details to answer a research question (Bowen, 2009). Therefore, in this study document analysis was used in combination with other qualitative tools, such as interviews and retrospective thinking, as the means of triangulation.
- **Low accessibility:** Due to the nature of the documents to be obtained (written exams), participants may be reluctant to show these exams because they are considered as official documents. However, in this study, once permission was given from their administrative parties, exam papers were made available based on the participants' willingness and cooperation.
- **Bias selectivity:** Because this study aims to analyse classroom-based grammar assessment in the form of written exams, only grammar teachers were requested to submit samples of these written exams. Also, the teachers got to choose which exams to deliver to the researcher, which could be modified versions of examination modes that had been administered to their students.

These are some potential flaws, but the important role document analysis plays in data collection outweighs these limitations.

4.4.4.1 Data collection for the document analysis.

Hardcopies of the EFL teachers' exams were collected according to the data collection timeline (see Table 10). Documents were submitted after the participating teachers administered them to their students. Only hardcopies were obtained because these exams were signed off by the respective departmental chairpersons. I was allowed to have these copies because I had already obtained permission letters from educational authorities, otherwise the teachers would not have shared them with me. Some of the exam copies were given to me during the retrospective thinking sessions (N = 20), while others (N = 8) were left in a sealed envelope at the department chairpersons' offices for me to collect.

These documents represented in the form of midterms, quizzes and finals are of high-stake status in the sense that their outcome is used to make important decisions about students' advancement (grade promotion or graduation for students). These assessment tasks are assigned specific marks (see section 4.3.1.2) which should all add up to 100 points. Students must score at least 60% to pass any course. Exam scores have direct consequences of students' passing or failing. Failing has major disadvantages, such as being forced to retake classes until they can be passed, not being allowed to progress to the next level or even being expelled due to low GPA. Through scores of midterms, quizzes, homework or any other course work during the semester, students are able to know if they have enough marks to pass the course after sitting for the final exams. If students do not get at least 60% of their overall course work before the final exams, they are allowed to drop the course without affecting their GPA negatively.

4.4.4.2 Participants and sample information.

28 hardcopies of EFL teachers' grammar written exams represented the sample of the document sources. The majority of the documents were midterm exam papers (N = 24), followed by final exam samples (N = 3) and, finally, one quiz. Out of the total participants who wrote these assessments, 22 were female. The majority of them were Saudi (N = 25) and PhD holders (N = 17). The table below summarizes the participants' information and the document types.

Table 11.
Participants and Documented Information

Facility	Document	Gender		Qualification		Country of Origin	
		Male	Female	MA	PhD	Saudi	Non-Saudi
A	Midterm	5	12	6	11	15	2
B	Quiz	Ø	1	Ø	1	1	Ø
C	Midterm	1	6	5	2	7	Ø
D	Final	Ø	3	Ø	3	2	1
Total (%)		6 (21%)	22 (79%)	11 (39%)	17 (61%)	25 (89%)	3 (11%)

4.4.4.3 Document data analysis.

Because all the documents obtained were hardcopies, the analysis was done manually and then the tables and figures of the results were created as MS Word documents. Document analysis involved three stages: skimming (superficial examination), reading (thorough examination) and coding. Thus, I first went through all the exam papers and familiarized myself with their content. Afterwards, I started reading each exam paper and, using different

highlighters, also began coding. The coding was a straightforward process. Since it was manual, a code was written alongside each section of the exam paper. The codes were then clustered into categories that reflected the themes (see Fig. 23).

Next, content and thematic analyses were carried out. Content analysis allowed the identification of pertinent information, which were later organized into categories related to the relevant research question (RQ 2). I constructed my own document analysis form based on one created by the National Archives and Record Administration (see Appendix K) to summarise the document contents. Information gleaned from the content analysis of teachers' written exams target the type, medium, author, audience, context, purpose, number of items, and focus (Table 12).

Table 12.
Content analysis of exam papers

No. of Docs	General Information						Focus		
	Type	Form	Author	Audience	Context	Purpose	No. of Qs	Production Items	Recognition Items
28	Quiz (4%) Midterm (85%) Finals (11%)	Computer typed paper (100%)	English grammar course teachers (100%)	Undergraduates majoring in English (89%) Post-secondary English language program (11%)	High educational facilities (100%)	Intended to explicitly measure students' abilities to use and comprehend English grammatical rules (100%)	120	80 (67%)	40 (33%)

Thematic analysis enabled pattern recognition and resulted in the generation of the salient research themes. During thematic analysis, I reread and reviewed the coding and the category construction to explore the themes pertinent to teachers' practices of EGA. Figure 23 shows the themes resulting from the document analysis.

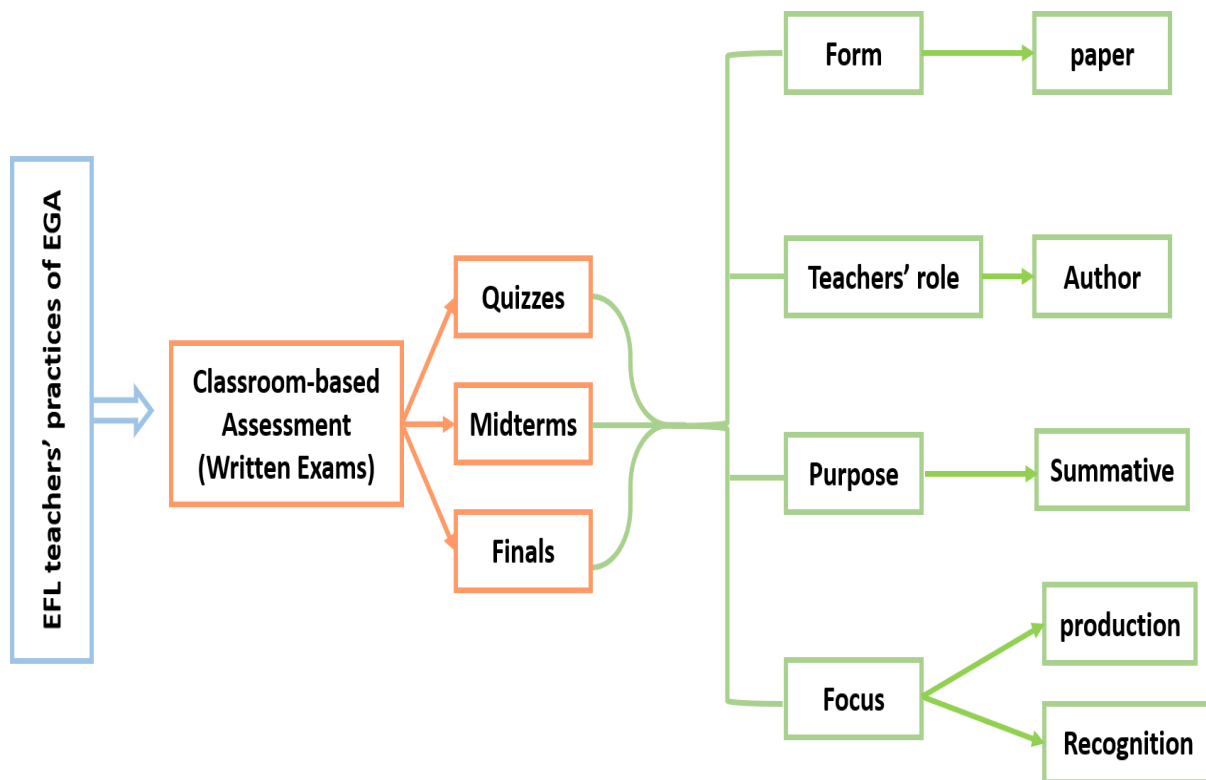


Figure 23. Resultant themes and categories from document analysis.

4.5 Positionality statement: My role as the researcher

Positionality in research embraces a number of distinct ways in which a researcher takes a stance or position within a research enterprise. It was recently summed up as of three types: One focuses on the ways in which authors engage their position through their work as a way to explore, better understand, and articulate their relationship to their work. Examples of this include how one's identity features in the work, or how one interprets data in relation to their position. A second understands positionality as a focus of the work itself, such as autoethnography or performance pieces. A third thinks through how positionality is linked to other methodological dimensions, such as validity, rigor, epistemology, etc. (Clift et al., 2018). In this case I will focus on the first of those since as the researcher I was not part of the target of the research, and I have already described my research paradigm stance (4.2.1).

Since the main goal of this research was to discover the beliefs and practices of EFL teachers, there is a need to understand the 'interpretations which [people] give of what they are doing' (Pring, 2000, p 96). To this end, it is important to understand the contexts and the participants of this study. In this section, I reflect on my role as a researcher at two levels: context experience level and the personal one. Firstly, being a teacher in one of the researched contexts and being aware of high educational facilities rules and regulations, I was able to obtain official permits to visit the facilities campuses through the right channels quickly. In addition, knowing some of the participants accelerated the process of spreading the online questionnaire and the recruitment of participants for the interviews, retrospections and document attainment. Secondly, that experience, being a faculty member, meant that I already had some idea of the sort of assessment that might be talked about and teachers' likely preferences. However, I took care to try not to assume anything when questioning the teachers or to let this experience colour my interpretations.

On the personal level, I would say that having a good previous relationship with most of the teacher participants allowed for a more friendly atmosphere, i.e., the teachers felt relaxed to speak frankly and freely about their beliefs and practices. In one key respect I was distanced from one segment of my participants, who were male, given the segregated nature of higher education in the KSA. This meant that in order to access these I had to resort to different tactics from the females. First of all, being an international student in the UK provided me with the opportunity to work with both males and females in various settings. This equipped me with the courage to contact male participants and recruit them for my study. Second, to make my male participants more comfortable, I took into consideration the shared religious and the cultural aspects. As I stated elsewhere, I met my male participants in a public place, a hotel lobby, and I had my male guardian with me which made my participants comfortable.

In all, I believe that being of the same culture and community as my participants, i.e. Saudi higher educational context and knowing some of them personally helped me make the necessary arrangements for carrying out the data gathering and at times assisted my understanding of some of what they told me. In addition, I believe that despite the differences in my positioning relative to the males compared with the females in the Saudi reality, including males added considerable value to the study, since in the Saudi context much research is single sex only.

4.6 Triangulation

Triangulation is a method used to increase the reliability and validity of research findings. By combining methods in a research study, the issues of biases that arise from the use of a single method are overcome (Noble & Heale, 2019). Denzin (1970) proposed four types of triangulation: 1) data triangulation, which refers to confirming the data obtained from different sources (e.g. teachers and students), 2) investigator triangulation, which includes the use of several researchers' works in a study, 3) theory triangulation, which encourages several theoretical schemes that can be used in the interpretation of a phenomenon and (4) methodological triangulation, which promotes the use of several data collection methods, such as interviews and retrospections.

In this study, a mixed-method research design was used where four methods of data collection were employed to investigate EFL teachers' beliefs and practices with regard to English grammar assessment:

- Questionnaire and semi-structured interviews were used to obtain a holistic view of EFL teachers' beliefs about EGA.

- Semi-structured interviews, retrospective thinking and document analyses were used to explore and explain EFL teacher's practices of assessing English grammar.

Such methodological triangulation utilising a variety of methods mentioned above aims to provide a more balanced explanation of the research findings and, in so doing, indicates both the reliability and validity of data.

However, in this study the process of triangulation has its limitations. First, I found this process to be very time-consuming. Second, as a novice analyst, combining findings from the different sources of data proved to be complex and challenging.

Nonetheless, through triangulation, I sought to overcome the weakness or intrinsic biases in the data and increase the validity of the research findings.

4.7 Research Quality

This section focuses on aspects of validity and reliability which were identified in relation to the study and describes how these were addressed to ensure research rigour and trustworthiness. As has been stated in this chapter, the study adopted a mixed-methods research design that included both quantitative and qualitative data.

4.7.1 Validity.

Validity refers to measuring what one claims to be measuring (Creswell, 2003). According to Lincoln and Guba (1985), validity is difficult to assess and has many dimensions: internal validity (credibility), external validity (transferability) and construct validity.

Internal validity is associated with the degree to which a study minimizes systemic errors or research bias, that is, the degree to which a researcher is able to say that no other variables except the ones under study have led to the results. According to Davis (1992), there are multiple ways to achieve internal validity. First, valid studies must provide evidence of lengthy engagement in a given field. In the present study, the data was collected over a period of three months. Another source of validity is the level of the richness and accuracy of the data. To strengthen validity, I provided a detailed and realistic description of how the study was conducted, including the process of data collection and how data was managed and analysed. I also stated that accuracy was accomplished via methodological triangulation in which multiple methods, i.e., questionnaire, interviews, retrospective protocols and document analysis, were brought together during the analysis and interpretation phases to ensure in-depth understanding of the phenomenon under study.

External validity involves the extent to which the research findings are replicable. According to Davis (1992), external validity is established when ‘the findings can be generalized to other contexts and/or subjects’ (p. 606). To achieve external validity, I provided a rich description of the context and the participants so that the reader can determine the degree to which the results of a study can be transferred to their contexts (McKey & Gass, 2005).

Construct validity refers to how well a research test or tool measures the construct that it is designed to measure (Creswell, 2003). In this study, construct validity indicates the extent to which a questionnaire measures EFL teachers’ beliefs about EGA. To establish this type of validity, an extensive review of previous studies to establish and justify the need of the current study was carried out. In addition, questionnaire items were adapted from studies conducted in the field of LTC and EGA, and feedback was obtained from my supervisor. The

questionnaire was also piloted with a group of volunteers whose views were as similar as possible to the target population. According to Baker (1994), the pre-testing or trying out of a particular research instrument can identify the potential practical problems with following the particular research procedure or whether proposed methods or instruments are inappropriate/excessively complicated.

4.7.2 Reliability.

Reliability refers to the degree to which a research method produces stable and consistent results (Davis, 1992). Research reliability can be divided into three categories: test-retest reliability, parallel forms reliability and inter-rater reliability (Dudovskiy, 2018). In this study, parallel forms reliability and inter-rater reliability were taken into consideration. Parallel forms reliability means that the results obtained from one assessment instrument (concerning a certain phenomenon with a group of participants) should be regenerated if a different instrument is used to measure the same phenomenon with the same participants. For example, if the results regarding EFL teachers' beliefs about EGA are obtained through a questionnaire and interviews should yield similar results that would prove the consistency of responses and allow comparison if required (triangulation). In this study, parallel forms reliability is achieved using multiple research instruments.

Another type of reliability is inter-rater and intra-rater reliability. This type of reliability is crucial to obtain, especially with qualitative research instruments (e.g. interviews, retrospections and document analysis). Inter-rater reliability asserts that the same results should be obtained by different assessors who use the same method. In this study, inter-rater reliability was established by asking a second coder to code my qualitative data using the same coding technique. The second coder was an assistant professor in Applied Linguistics

who coded one sample from each instrument (interview, retrospection and exam sample). I also provided her with the lists of codes for all datasets, their definitions and examples from the data. The final procedure required checking the similarities between the researcher and the second coder regarding the application of the codes. This was achieved by using Scholfield's (2005) formula (Figure 24). The results of the inter-rater reliability tests are presented in Table 13.

$\frac{\text{Number of items coded the same by the two raters}}{\text{Number of items (stretches of data assigned a code) coded by the researcher in the data}} \times 100$

Figure 24. Scholfield's formula for inter-rater reliability agreement (2005).

Table 13.
Inter-rater Reliability of the Coding

Data sources	Number of items coded the same by the two raters	Number of items coded by the researcher	Agreement result
Interview	12	16	75%
retrospection	7	9	78%
Document	9	9	100%

Table 13 indicates that the total percentage of agreement between the researcher and the second rater in two datasets (75%; 78%) were below 80%. Ideally a minimum of 80% agreement is recommended in the literature (Huberman, 1994; Mackey & Gass, 2005). The agreement obtained in this study for two (out of three) samples was just below 80% and was, therefore, considered acceptable; particularly given that, for one dataset, this percentage was

found to be 100%. However, it should be noted that it was not possible to discuss the differences and disagreements with the second coder due to her workload and unavailability.

As a second measure of reliability, I also conducted intra-rater reliability which refers to the degree of agreement among multiple repetitions of assessment performed by a single rater.

To accomplish this, I coded the data for the first time in January 2019 and then again in April 2019. My coding in both occasions was identical; this might be because my memory is rather strong, and I had been immersed in my data with the codes constantly present in my mind.

4.8 Ethical Considerations

The ethical considerations in this study refer to concerns about participants' rights and the sensitivity of information about personal and professional beliefs, knowledge and experiences. The ethical considerations described below are derived from the ethical guidelines of the University of Essex and the Ministry of Education in Saudi Arabia:

1) Access to and Recruitment of the participants:

First, ethical approval was obtained from the Department of Language and Linguistics ethical committee at Essex University, Colchester campus, in May 2018. Soon after, I obtained official permission from SACB to leave the UK and go to Saudi Arabia to collect the research data. Universities and institutions were contacted via email to undertake the research tasks in their facilities. Teachers from the respondent universities were approached either personally, via phone calls or messages, or via emails.

2) Information and consent forms:

Written information sheets and consent forms (see Appendices A and B) were given to the participants upon recruitment. In the process of collecting qualitative data, the participants

were also verbally reminded about their rights as they were being recorded. In all cases, the right to withdraw at any stage of the data collection process without reason was highlighted.

3) *Confidentiality and anonymity*

Steps were taken to ensure that the participants' identities, positions and institutions of work would be kept confidential. First, the participants were informed that all quotes from the raw data would be used anonymously. In addition, confidentiality was maintained throughout the analysis of the data as pseudonyms were used to identify the participants and their place of work.

4) *Data storage:*

Data collected in this study were kept in a password-protected memory stick for storage. All the data will be destroyed when the project is completed.

4.9 The Pilot Study

According to Burns (2000), the pilot study is a very important device for researchers to assess their research tools, as it not only involves acquiring data but also helps in learning how to acquire data properly and accurately, along with assisting researchers in discovering weaknesses in their methodologies. In this respect, Bell (1993) stated the following:

All data-gathering instruments should be piloted to test how long it takes recipients to complete them, to check that all questions and instructions are clear and to enable you to remove any items which do not yield usable data. (p. 84)

Therefore, all data collection instruments in this study were piloted before the initiation of the main data collection phase.

4.9.1 Context and participants.

The pilot study was carried out in spring 2018 and lasted for four weeks (March 25th – April 19th). It took place in an international private academy which offers an English language diploma (one-year programme) to post-secondary students. The programme aims to prepare candidates for majoring in English as undergraduates or to enrol in IELTS preparation course, so that they can sit for this test and score highly to pursue further studies. In this programme (female sector), there are three different grammar courses that must be completed throughout the whole year, along with the following skill courses: reading, writing, listening and speaking. In the male sector, there are only courses that cover the four main skills just mentioned. The tables below summarise the programmes in both sectors, male and female, providing details about the grammar courses.

Table 14.
English Language Diploma Programme, Female Sector.

Programme:	12-month English Language Intensive Course			
Degree:	Diploma			
Courses:	Reading	Writing	Listening & Speaking	Grammar
Accredited hrs / day:	4	4	4	4
Total hrs / week:	20 hrs			
Assessment tasks:	Midterms + Finals			

Table 15.
Grammar Courses Covered in the English Language Intensive Course, Female Sector.

Course name:	Grammar		
Levels:	1 – 4	5 – 7	8 – 12
Textbooks:	Interaction Access Middle East Diamond Edition	Interaction 1 Middle East Diamond Edition	Interaction 2 Middle East Diamond Edition

Table 16.
English Language Diploma Programme, Male Sector.

Programme:	12-month English Language Intensive Course	
Degree:	Diploma	
Courses:	Reading & Writing	Listening & Speaking
Textbooks:	Q: Skills for Success series 1 – 4	Q: Skills for Success Series 1 – 4
Accredited hrs. /day:	2	2
Total hrs. / week:	20	
Assessment Tasks:	Quizzes + Midterms + Finals	

This study included 30 EFL teachers teaching in an intensive English language programme with similar profiles in relation to the study target, that is, EFL teachers teaching in an English major programme. 54% (N = 16) of the participants were female while 46% (N = 14) of them were male. The participants had similar background features except in the aspect of country. The women were mostly Saudi, unlike the males, who hailed from various nationalities. Both genders mostly seemed to have the same educational qualifications (BA) and rather extensive experiences of teaching English (in general) and teaching and assessing English grammar (specifically) (see Table 17).

Table 17.
EFL Teachers' Demographic Information.

Variables		Gender	
		Male Count	Female Count
Age	21–30	1	3
	31–40	4	2
	41–50	4	4

Variables		Gender	
		Male Count	Female Count
	Over 50	3	1
Country of origin	Austria	1	0
	Britain	2	0
	Canada	1	1
	Egypt	6	0
	KSA	1	7
	Other	0	1
	Syria	2	0
	Yemen	0	1
Educational level attained	BA	10	9
	MA	3	1
English teaching experience	1 – 5 years	1	3
	6 – 10 years	4	1
	11 – 15 years	2	0
	Over 15 years	6	6
English grammar teaching experience	None	1	0
	1 – 2 years	2	0
	3 – 4 years	9	10
Grammar exam writing experience	None	3	1
	Less than 1 year	1	0
	1 – 2 years	1	1
	3 – 4 years	8	8

4.9.2 Instruments.

All instruments in this study were piloted (questionnaire, interviews, retrospections and document analyses). Piloting the research tools was beneficial because various flaws were found in some of the data collection tools, which later led me to modify these tools before conducting the main study.

4.9.2.1 The questionnaire.

The questionnaire was piloted in the second week (April 1st) after recruiting the participants and obtaining the signed consent forms. Hardcopies were given to the programme directors in both sectors (male and female), who were requested to hand it out to the participants. The questionnaire included 50 items addressing various issues related to EFL teachers' beliefs about English grammar assessment. The items of the questionnaire were mixed-up and written in English. For the purpose of analysis, items were grouped into their pre-set categories. Items 4, 7, 10 and 50 were excluded from the analysis because the majority of the participants refrained from answering them, which indicated they were problematic and, hence, were deleted from the final draft of the questionnaire. Due to the small sample size, the collected data were analysed based on descriptive statistics. What follows is a description of the results.

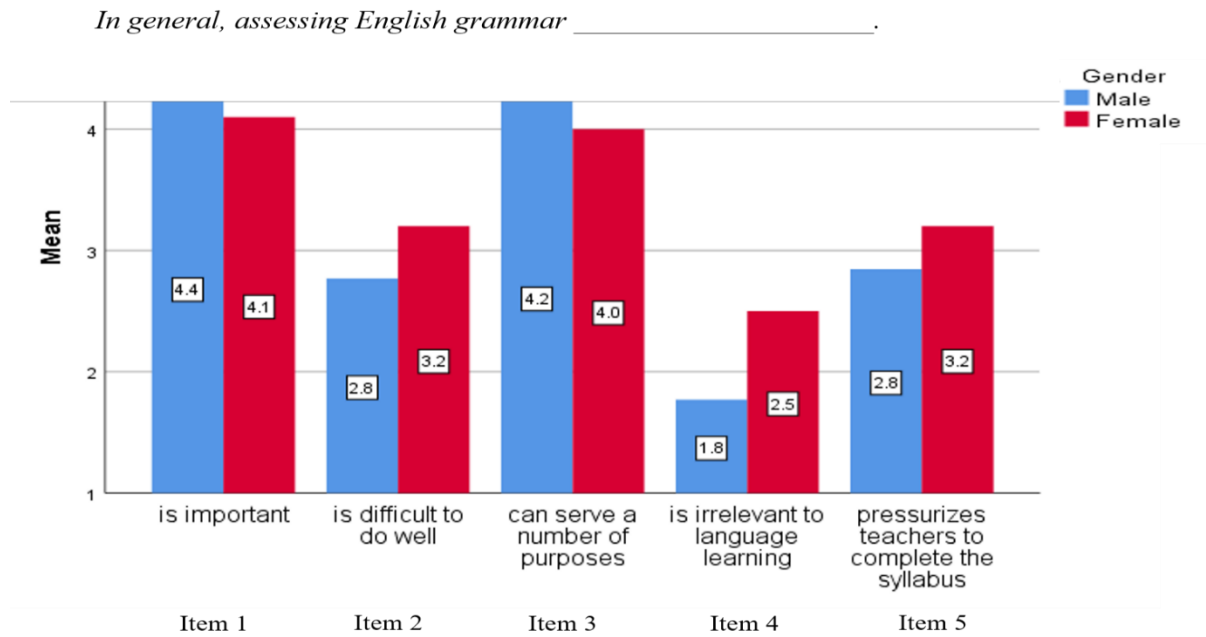


Figure 25. Teachers' general beliefs and attitudes about English grammar assessment.

As shown in the above figure, the two high agreement results (items 1 and 3) indicate positive attitudes towards English grammar assessment. These results could be attributed to the fact that the majority of participants are experienced teachers. In addition, one might expect less experienced teachers not to be sure of their ability to conduct grammar assessment appropriately, which can explain their middling on the second item. This could also be related to assessment literacy and identify whether the participants had undergone any relevant training on any sort of assessment.

As one might expect, being experienced teachers, the participants saw the value of grammar assessment for learning and did not agree with item 4. Also, since the participants themselves would construct their own assessment tasks, especially the male teachers, they did not feel pressurised by their task greatly. That pressure is more likely to arise among female respondents, since an external body – the Ministry of Education– constructs the final exams. In this context, most of the teachers can design their exams to fit whatever they have taught.

Thus, they are not highly pressurized and hence, this item witnessed scores achieved below midpoint.

In general, the purpose of English grammar assessment should be to _____.

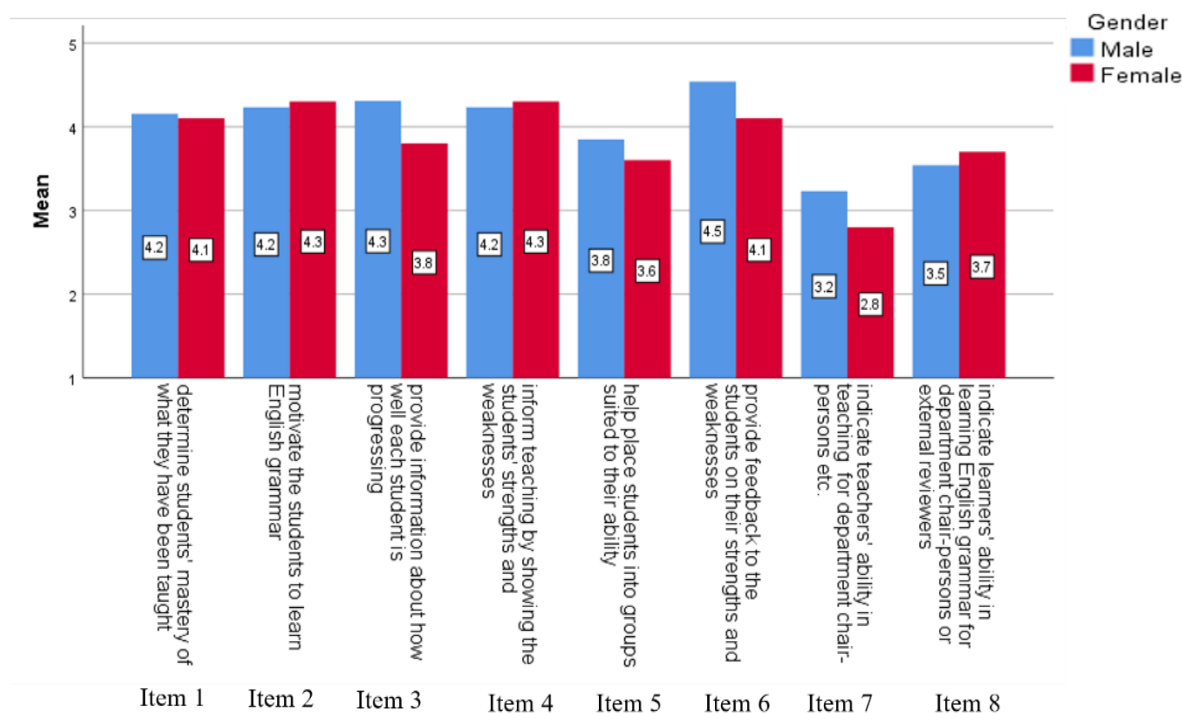


Figure 26. Teachers' beliefs about the purposes of English grammar assessment.

The responses to these items reveal that the majority of participants believe that grammar assessment fulfils different purposes: summative, formative and accountability. High scores were achieved for items 1, 3 and 6, which suggest that the participants agreed more that grammar assessment serves a more summative function than formative. The lowest scores were for items 7 and 8, related to grammar assessment used for accountability purposes. The teachers believed that grammar assessment should be used to diagnose and report students' strengths and weakness in relation to grammar learning. This could be associated with the nature of the context where assessment is used to assign scores and report learning progress.

Concerning the content and delivery of English grammar assessment, in general I believe that _____.

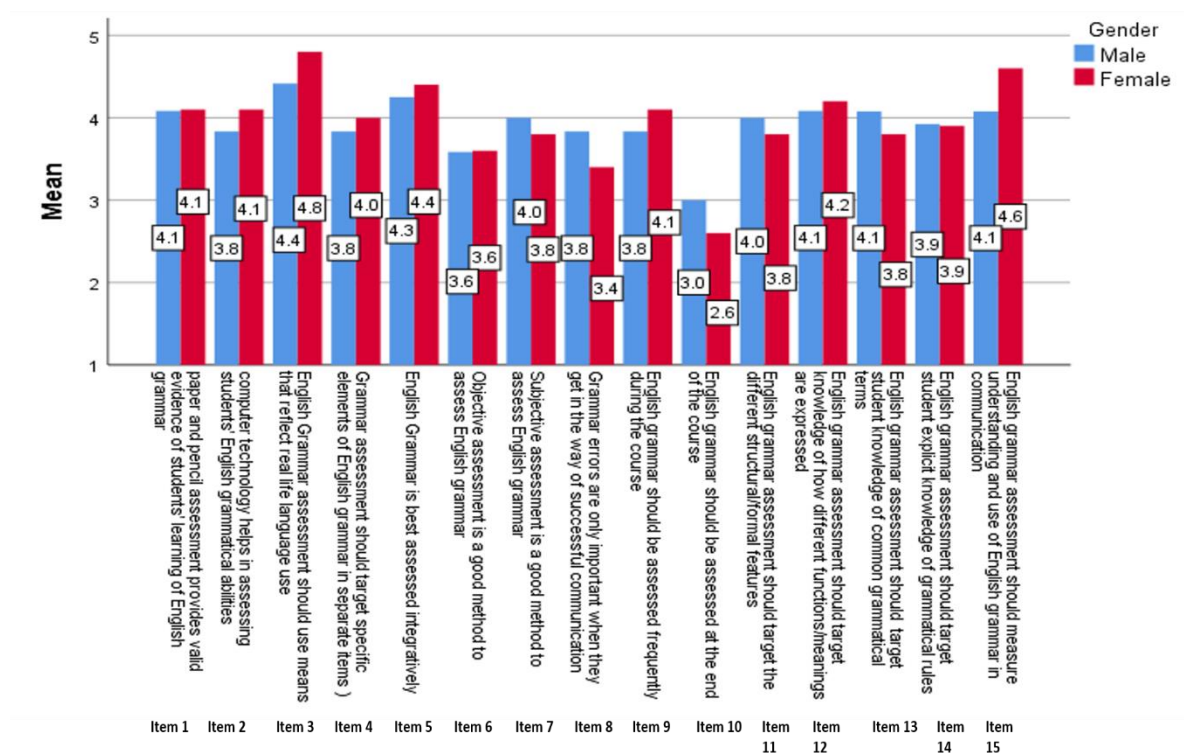


Figure 27. Teachers’ beliefs about English grammar assessment methods.

Concerning the content and delivery of English grammar assessment, participants seem to respond positively, with 14 items involving scores above midpoint. Interestingly, items 3 and 15 saw the highest scores, which indicate teachers’ agreement that the content of grammar assessment should focus more on the level of communication. Agreement on both items also suggest a level of consistency among the participants’ responses. Item 5 is ranked second highest, which means that the teachers believed that grammar is better assessed in an integrative manner. Again, this response corresponds with items 3 and 15 that focus on communication. This, however, does not deplete the value of the segmented content of grammar assessment, in which items 4, 6, 7, 8, 11, 12, 13 and 14 heralded scores of high midpoints. This indicates teachers’ positive attitude towards these items.

As for the means of delivery, teachers seem to agree positively with items 1, 2 and 9 but respond negatively to item 10. This shows that teachers believe that computer-based and paper-based assessment both have merits when it comes to grammar assessment. Participants' negative response to item 10 indicates once again a level of consistency (i.e. reliability, cf. 4.7.2) in the participants' answers. Since they agreed that grammar should be assessed continuously throughout the course (item 9) it would be logical to disagree with conducting grammar assessment only at the end of a grammar course.

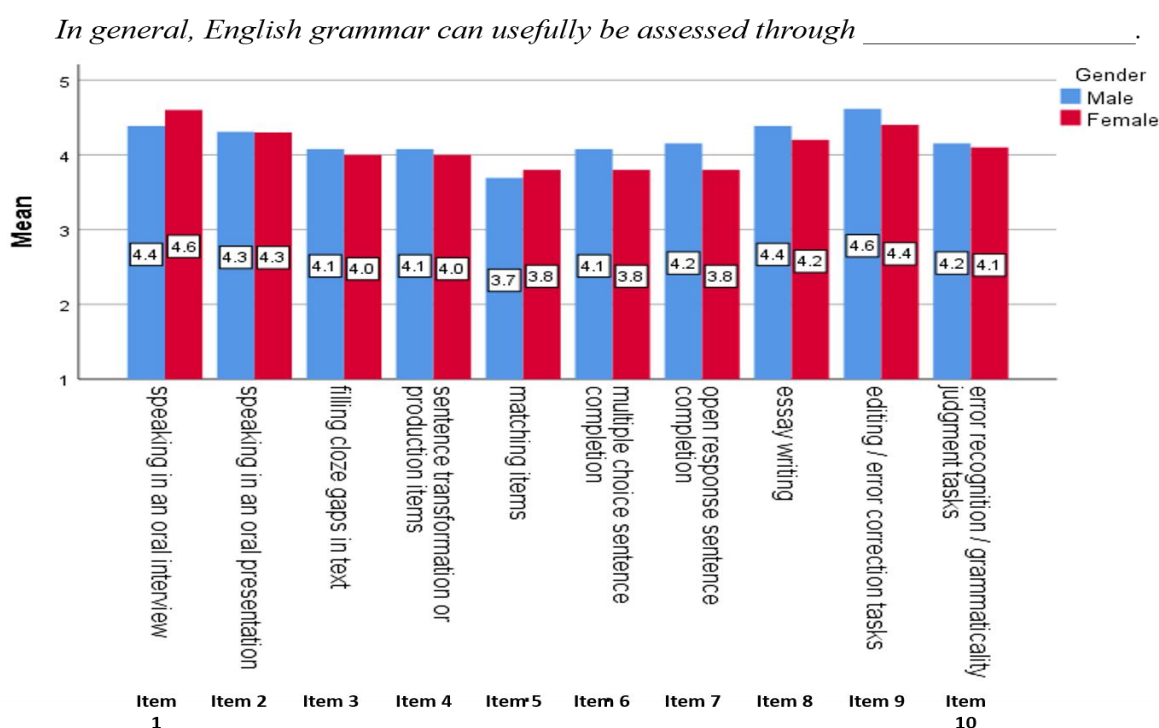


Figure 28. Teachers' beliefs about English grammar assessment formats.

The two highest scoring tasks mark an interesting contrast. Item 1 targets spontaneous spoken communicative ability and is more message-focused, while item 9 depends on awareness/metacognition of language and typically written language focus. Responses to these items would yield results once, compared to their actual assessment tasks, to identify consistency or incongruity between the teachers' beliefs and practices. To summarise, the

participating teachers seemed to agree positively with the various grammar assessment formats suggested in this section.

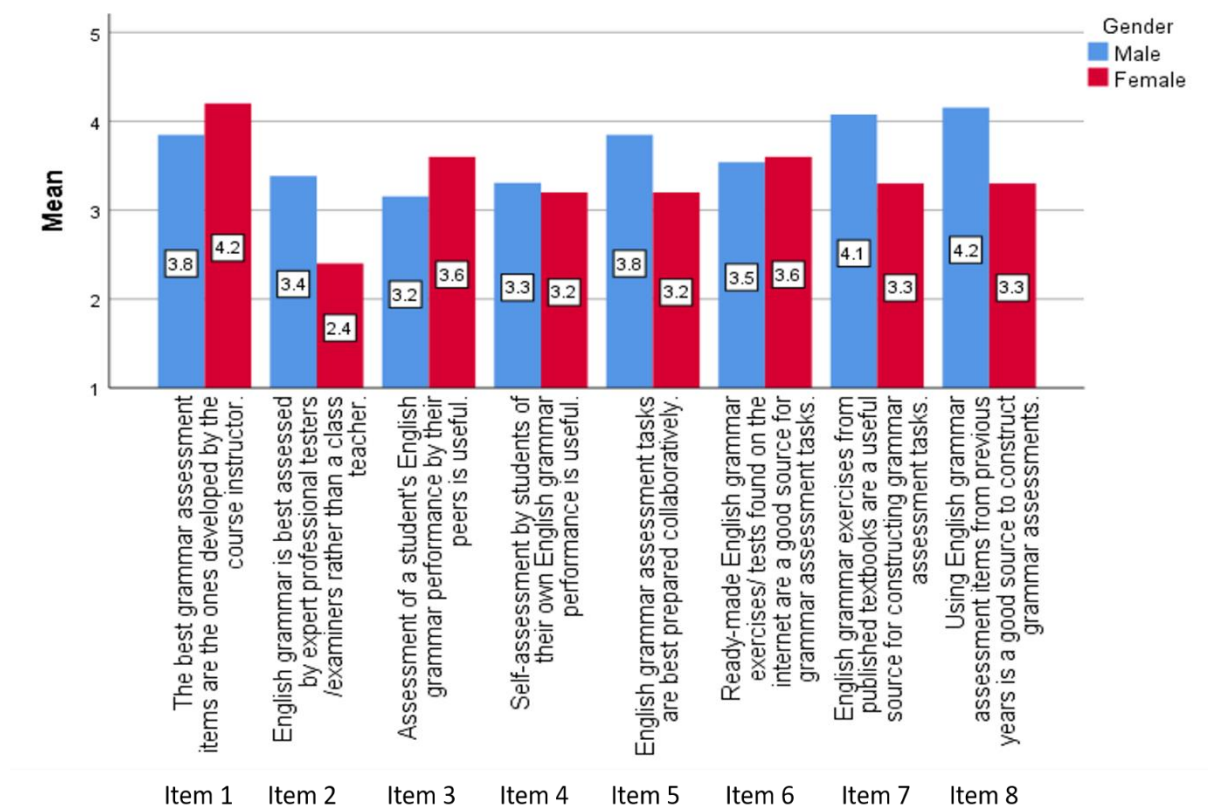


Figure 29. Teachers’ beliefs about their roles and the sources used in constructing English grammar assessment tasks.

As seen in Figure 29, the purpose of items in columns 6, 7 and 8 was to explore EFL teachers’ beliefs about their preferred source with regard to constructing grammar assessment tasks. The results indicate that most of the participants agree that ready-made grammar assessment items, either extracted from textbooks or adopted from previous assessment tasks, were appropriate sources to be used when assessing grammar.

Items 1, 2, 3, 4 and 5 refer to teachers’ beliefs about their role in constructing grammar assessment. The participating teachers believed that they should construct their own assessment tasks, preferably individually (female participants scored the highest) and

collaboratively otherwise. Moreover, the participants seemed to believe that using self-and-peer assessment could be acceptable, agreeing positively on this stance. However, item 2 shows a huge contrast between the male and female respondents. Female participants were less inclined to support the idea that exams are best written by experts, while male participants were more accepting of this notion.

4.9.2.2 Interviews.

Piloting the interview questions was crucial for two reasons: first, I needed to test myself as an interviewer, since this is the first time I have engaged in interviews. I wanted to know how I could carry myself around my interviewees and how acceptable I would be to them. Second, I had to intuit what sort of answers the interview questions would generate.

The interviews took place on various days during the pilot study. Nine teachers (F = 5; M = 4) agreed to participate in the interviews, which were about 20–30 minutes long. Interviews were audio-recorded and then transcribed into MS Word documents in preparation for analysis. Each document was given a name that would help me identify my participants and simultaneously maintain their anonymity. Once the transcription of audio recordings was complete, the documents were imported to MAXQDA to begin the coding process (see Appendix L).

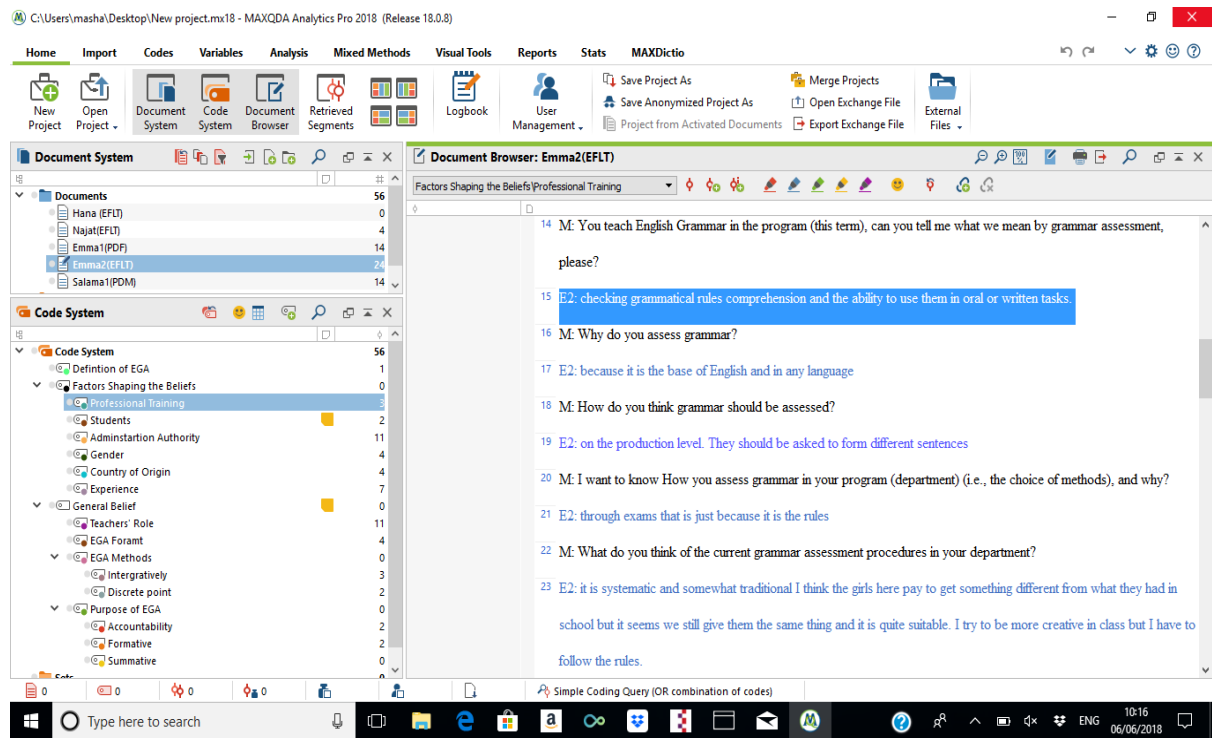


Figure 30. A screen print of how the coding system and the transcript documents are presented in MAXQDA.

The coding process involved identifying meaningful segments that were considered as relevant to teachers' beliefs and practices. I assigned each code a colour and an appropriate label. The first teacher's data that I analysed resulted in a large number of codes. However, as this process went on, the number of codes tended to decrease as the themes and categories emerged.

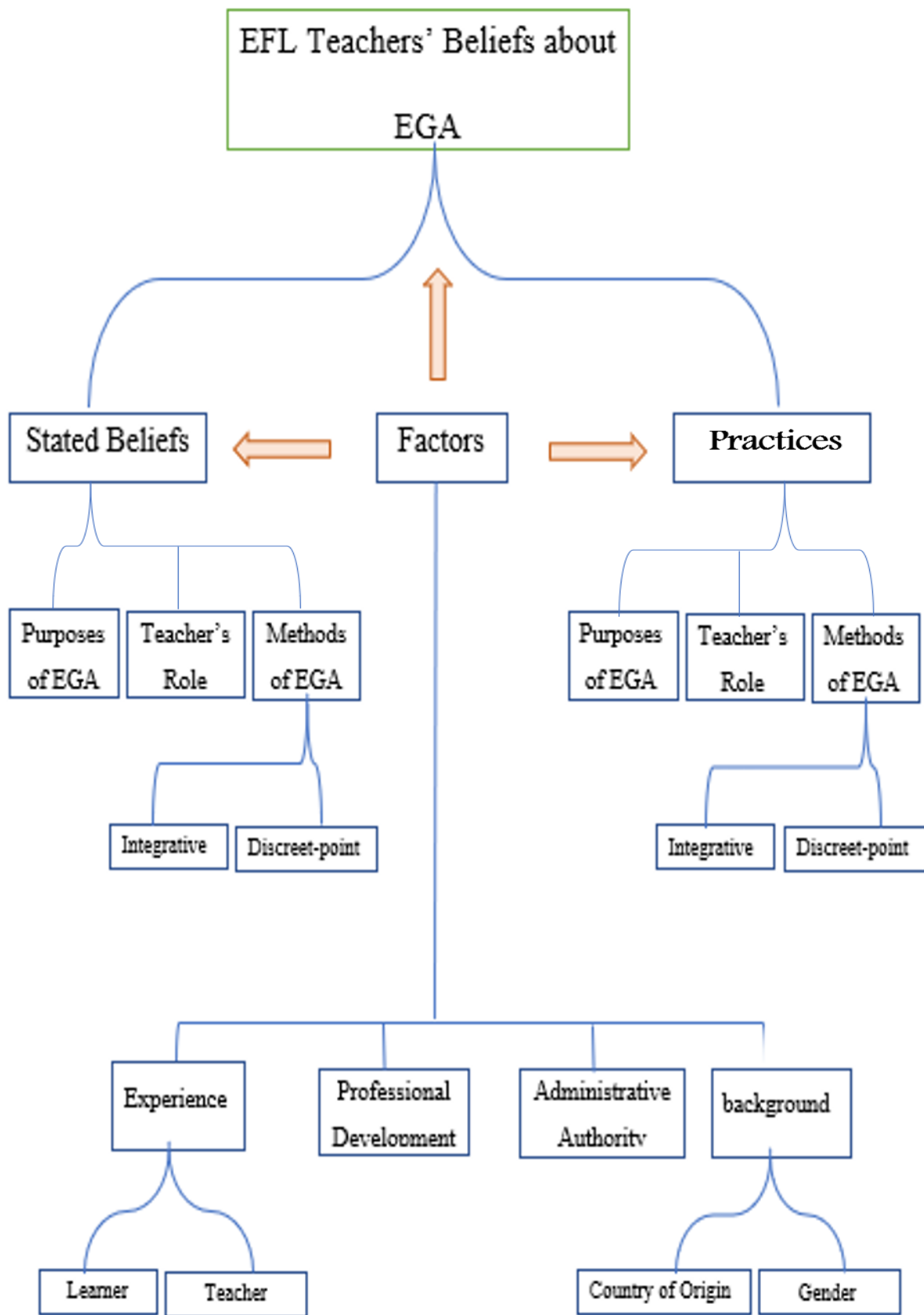


Figure 31. Themes and sub-themes developed through coding.

Preliminary data analysis showed that there are commonalities among the participating teachers' beliefs about EGA (Figure 29). For example, all the teachers stated that grammar is assessed to identify any grammatical mistakes and to report on the progress of learning. This belief indicates a summative role of grammar assessment. This result coincides with what has been found in the questionnaire responses where the majority of teachers believed that grammar assessment fulfils various functions and summative purposes witnessed the highest scores.

With regard to teachers' roles in constructing assessment tasks, the results were very interesting. Female participants strongly asserted that a teacher should construct her/his own assessment tasks or have a saying in such construction matters. This again is supported by how female participants responded to item 2 in Figure 29. The result shows that the female teachers disagreed with assigning the role of assessment construction to an expert or otherwise. The male participants, on the other hand, were more receptive to the idea of allowing a second or third party to design the assessment tasks, which they would just administer to the students. This amenable perception is also detected in their responses to item 2 in Figure 29. This can be attributed to the fact that male teachers choose their exams or quizzes (second party) from test banks. At the end of each level, students take Oxford placement tests as a final exam (third party).

As for grammar assessment methods, the teachers agreed that grammar should be assessed in an integrative manner, focusing on production level to enhance communication. Their practices, on the other hand, show some discrepancies. Female teachers assess grammar separately (discreet-point); there is a section in the midterms and the final examination where students are asked only about grammatical rules. This kind of practice is influenced by the

guidelines dictated by the academy and approved by the Ministry of Education. Males, however, assessed grammar in an integrative milieu with respect to writing and speaking during midterms, but they would resort to assessing grammar separately in their quizzes because their students would ask them to do so (because those questions would be easier to answer and guarantee high marks).

Other relevant themes, factors and comparisons between results, are addressed in more detail in the main study.

4.9.2.3 Document analysis.

Documents in the form of written exams (quizzes, midterms and finals) were collected on week 1. Surprisingly, document collection was easy. As I asked for samples of their exams, the programme directors (both male and female) offered me a large number of samples on the spot. Unfortunately, the majority of these samples were hardcopies, and so during the analysis I coded the scripts manually. Later, I generated a content analysis form in an MS Word document (see Appendix K) that could be easily imported to MAXQDA for the purposes of comparing and contrasting the themes and categories across the participating teachers.

4.9.2.4 Retrospection.

As it was not possible to conduct ‘think aloud’ protocols, I conducted retrospection instead (see 4.4.3 for details). Only one participating teacher agreed to the retrospective thinking. This teacher sat with me and gave me a sample of a quiz she had designed for her students. I asked her to walk me through the process of constructing the quiz; below is the transcript of the whole session (3 min 05 sec).

1.	Mashaël:	This is your quiz, yeah?
2.	Hana:	Yes, I just give quizzes to train my students for the midterm and the finals. No points are allocated on this.
3.	Mashaël:	So how did you decide what to put in there?
4.	Hana:	I just finished the chapter of tenses, so that is what I am testing them on.
5.	Mashaël:	How did you come up with these items?
6.	Hana:	I just took everything from the book.
7.	Mashaël:	I see. You mean you took the sentences...
8.	Hana	Yes. No. These are exercises I don't do in class and ask them to do at home. Similar to what they will have in the midterms and finals.

Evidently, the session was too short and not enough data was generated. I strived for a richer and deeper context for my main study. The retrospective thinking showed that the teacher constructed her quiz as a training task (not required by the academy): 'I just give quizzes to train my students for the midterm and the final. No points are allocated on this' (line 2). In the process of constructing this quiz, she seems to be following the academy guidelines of midterm and final exam construction: 'Similar to what they will have in midterm and final' (line 8).

4.9.3 Main findings.

As I mentioned before, the aim of the pilot study was twofold: trying out research instruments to detect any flaws or unforeseen problems and rectifying them accordingly. The other aim

was to find out if the piloted instruments would generate substantial and usable data for analyses and interpretations.

Based on the results obtained, the participating teachers acknowledged the role grammar assessment plays in learning. They believed that grammar assessment can serve various purposes, including summative, formative and accountability. They also believed that the best way to assess grammar is in an integrative means, through the main language skills (RQ 1a). These beliefs were probably shaped by the teachers' vast experiences in teaching and assessing as well as language learners (RQ 1b).

As for the teachers' actual practices, there are some mismatches between what they believe should happen and what really happens. Teachers advocate for integrative grammar assessment while the majority of the practices target grammar separately, through discrete-point items (RQ 2a). These practices are clearly influenced by administrative authority and academic policy (RQ 2b).

As for programme directors, their beliefs concur with those reported by the teachers and are probably influenced by the same factors (RQ 3a and 3b). With regard to programme directors' practices when evaluating grammar assessment tasks (RQ 4a and 4b), unfortunately the data did not generate results that could be used to answer this question, because this practice was not evaluated in their specific context.

4.10 Summary of the Chapter and Conclusion

This chapter outlined the paradigmatic and methodological frameworks underlying this study. It also highlighted the selection criterion of the research context and the participants and

provided an overview of data collection and data analysis procedures (statistical and content analyses). Research instruments were explained, justified and exemplified. Furthermore, ethical considerations and issues of validity and reliability were discussed. Finally, the pilot study was described, and some results were presented. Table 18 summarizes the key points of this chapter.

Table 18.

Summary of Research Design, Questions, Instruments and Objectives and Means of Data Analysis

Mixed-method Research Design			
Research Questions	Research Instruments	Objective	Data Analysis including software used
1. a) What are EFL teachers' beliefs about how English grammar should be assessed in their context? b) What are the factors which shape those beliefs?	Questionnaire (Quantitative tool) + Semi-structured Interviews (qualitative tool)	To obtain numerical and descriptive data about EFL teachers' beliefs on English grammar assessment and the factors which contribute to shaping such beliefs	(For quantitative tool) SPSS Descriptive statistics (means + percentages) T-test Correlation statistics (multiple regression) (For qualitative tool) MAXQDA Coding + categorization

Mixed-method Research Design			
Research Questions	Research Instruments	Objective	Data Analysis including software used
2. a) How do EFL teachers actually assess grammar in their teaching environments? b) What are the factors which influence their practices other than their beliefs?	Interviews/ Document analysis/ Retrospection (qualitative tools)	To provide insights into key participants' grammar assessment practices, roles and the relevant factors in real-life settings	(For qualitative tool) MAXQDA Coding + categorization
3. a) What is the relationship between EFL teachers' beliefs and their current practices? b) What are the factors leading to the convergence or divergence between their beliefs and practices?	Questionnaire (quantitative tool) + Semi-structured Interviews + Retrospections + document analysis	To understand the relationship between teachers' beliefs and practices and the factors that govern this relationship	Triangulation of all the processes mentioned above

The next chapter presents findings on how participating EFL teachers perceive EGA, what factors contribute to these perceptions and how these perceptions are translated into real practices (if any). Data from the questionnaire, the semi-structured interviews, the document analyses and retrospections are incorporated and triangulated to provide evidence about EFL teachers' beliefs and practices concerning EGA in public higher-educational contexts in Saudi Arabia.

5. Results

5.1 Introduction

This chapter presents the results. These are organised in three main sections, with subsections designated to answer the three research questions. First, section 5.2 presents the findings of the questionnaire and interviews to obtain a holistic understanding of EFL teachers' beliefs about EGA, Section 5.3 presents findings about the teachers' practices of EGA in their actual teaching environments, obtained from interviews, retrospection and document analyses. Finally, section 5.4. combines and integrates the data in the whole study to reveal the relationship between teachers' beliefs and practices.

5.2 EFL Teachers' Beliefs about EGA and Factors that have Shaped these Beliefs

In order to answer the first research question – What are EFL teachers' beliefs about how English grammar should be assessed? What are the factors which have helped shape those beliefs? – results from the questionnaire and the semi-structured interviews are presented concurrently wherever relevant, so that a comprehensive view of the teachers' perceptions of EGA is achieved.

5.2.1 What are EFL teachers' beliefs about how English grammar should be assessed?

This question is addressed by considering the findings of the whole dataset provided by the questionnaire from the total sample of teachers (N = 94) and by referring to teachers' responses (N = 32) to the semi-structured interviews (for question items 1, 2, 3, 7 and 8, see Appendix G).

5.2.1.1 General beliefs about English grammar assessment (EGA).

The analyses of the data show that there is strong agreement in the questionnaire, with mean ratings between 'agree' and 'strongly agree', and that EGA is important and could serve a number of purposes, which are fully discussed in the following section. No teacher disagreed with either of these general propositions and the ratings were significantly positive (Table 19).

Table 19.
General Beliefs of EGA

Item No.	In general, assessing EG....	Min	Max	Mean	Percent %	Std. Deviation	Sign test sig.
1	is important	2.00	4.00	3.6452	91.13	0.58319	< .001
3	can serve a number of purposes	2.00	4.00	3.4462	86.15	0.62320	< .001
5	pressurises teachers to complete the syllabus or textbook	0.00	4.00	2.5815	64.53	1.07998	< .001
2	is difficult to do well	0.00	4.00	2.4624	61.56	1.07904	< .001
4	is irrelevant to language learning	0.00	4.00	1.9301	48.25	1.45884	.362

This belief about how EGA is important was also endorsed by the majority of the teacher participants in the interview (63%). For example, some of the teachers stated the following:

We cannot dispense with exams. I think they are very important. They should be there, but we don't depend on them entirely; we have to think about other ways to help us get a clear idea of how well the student has mastered those skills we teach them.

In the aforementioned quote, the participant teacher clearly identifies assessment with the kind of formal assessment that is required by the institution to be used in the context of study for one of the key purposes of assessment (i.e. written exams to measure achievement).

While the teacher agrees that GA in the form of exams is very important, it is however evident that the participant's belief system goes beyond that and recognises the value of other means of assessment of whether learning has occurred. This is an implicit indication of a possible clash or contradiction (cf. AT) between teachers' beliefs, practices and institutional authorities, which would be evident from time to time in other data below, and which potentially may have led to mismatches between EGA beliefs and practices.

It's the Alpha and Omega. It's kind of upgrading and keeping track of how grammar changes through your teaching methodology using quiz or exam to assess them and you also be assessed.

This participant teacher also seems to identify assessment with the institutional version of (quiz and exams). However, unlike the preceding teacher, this participant accepts that this means of assessment are sufficient and crucial for assessment purposes (*Alpha and Omega*) without qualification. Thus, it is possible to assume that belief seems to be identical with that of the higher authorities.

There was also considerable agreement, with means between 'neutral' and 'agree', that EGA pressurises teachers to complete their syllabus. Beliefs about the interplay of EGA with the syllabus emerges in the interview data as being subtly complex, albeit thought of purely in local terms. This seemed to be an issue, as the teachers were not able to formulate a belief if it was not related to local conditions. One teacher believed that pressure would only arise,

hypothetically, if a committee would set the EG exam instead of the course teacher. In this case, the teacher stated that students would suffer if the syllabus was not completed in class:

Having a committee is a good idea, but in practice it is not realistic because not all teachers are committed to what's been required in the syllabus. Some teachers may miss lots of lectures or skip some topics that they feel trivial or common knowledge, and this will definitely affect the students. So, when the committee gives certain unified exam, students and teachers might be at a disadvantage, I think.

Another teacher, in a context where someone other than herself set the final exams, asserted the impact of exams on the teaching of the course syllabus:

I have to cover all the exercises in the book even if I don't believe they are of real value to the students because that is the problem with unseen exams. In the final, the convener writes the unified exam and I just don't want my students to be disadvantaged. You cannot lead the course the way you want.

Interestingly no teacher mentioned that in fact there is in any case pressure to complete the syllabus from another source, independent of the assessment. That is the Quality Unit who require teachers to write course reports at the end of each semester, where one of the questions is whether the syllabus (as given in the course specification document) was completed, and if not what topics were omitted and why.

There was, again, agreement significantly above the neutral midpoint of the scale for the proposition that it is in generally difficult to do EGA well. In the interviews, this was endorsed by one teacher, who commented that 'It's not easy doing grammar exams in particular'. Another teacher acknowledged the difficulty of assessing grammar well and

attributed the reason to students' English proficiency levels. For example, the teacher said the following:

I have variation among the students in my class in different classes. Some students have studied abroad and feel this course is silly, others are still learning the first steps. So, to maintain this variety or these differences...okay this is very difficult.

This teacher implicitly refers to a contradiction which would be discussed later (section 6.3). Here the participant implies that assessment should suit the level of the student. Therefore, assessment is a problem where the students are at widely differing levels, and the exams are set to assess student mastery of knowledge and skills described in the course learning outcomes (CLOs), and in a list of topics for the course, given in the course specification. That is indeed how tertiary level assessment is usually conceived around the world these days. One expects a BA student who has taken and passed a BA course in English Grammar or otherwise to have covered and know about a certain range of topics (and possess certain skills) at a certain level regardless of the prior capability of the student. However, this teacher holds beliefs that do not agree with that.

Some teachers further provided pointers to alleviate this difficulty that they perceived. One teacher said, 'I think with experience and practice I got the hang of it', indicating that in her view, experience rather than training eventually solved this problem. Another teacher was more explicit and suggested a way that she thought that difficulty can be handled by not choosing a particular assessment method:

My MA research was on communicative assessment, but still I think it's difficult to apply in our context maybe we're not used to it, and even the students they don't accept something different from what they are used to.

Again this exhibits a belief that the assessment should be suited to the students' abilities and preferences and not to any impersonal criteria such as a list of skills that the students are supposed to obtain by the end of the course (e.g. list of CLOs), even if in fact it included 'ability to use grammar in spoken and written communication'.

Finally, on the issue of EGA's relevance to language learning, there were some differences of opinion among the teachers. The mean score was close to the neutral rating of 2 and did not deviate significantly from the midpoint of the scale ($p = .362$). It was, therefore, definitely not disagreed with. As the histogram of the scores shows (Figure 32), there is some evidence of the existence of two types of teachers, ones that strongly agree and ones that disagree, with fewer teachers having views in between. This leads to a relatively high standard deviation of the ratings as well as an overall mean close to 2.

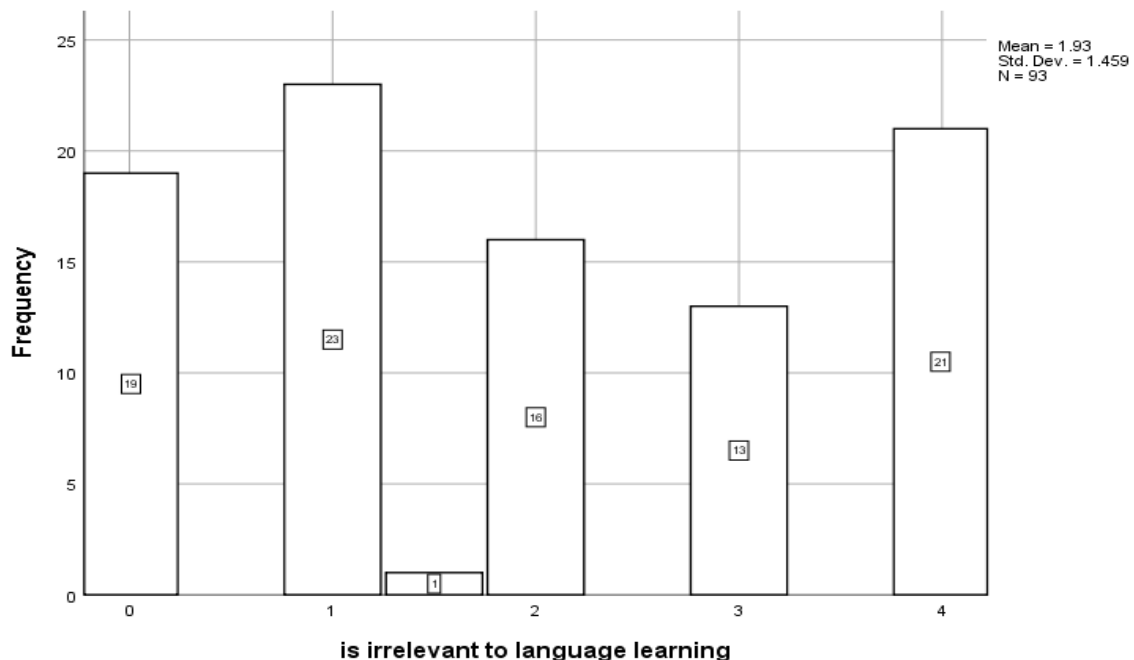


Figure 32. Participants' responses to item no.5 in the questionnaire.

Results from the interviews indicated that only few teachers (6%) believed that grammar assessment plays no role in improving the students' language learning process. Their view was that grammar assessment aims to check one's knowledge of grammatical rules explicitly, and students almost never transfer this knowledge to their implicit, spontaneous, speaking and writing ability. The statement below expresses this view:

I think and to be honest with you although I've been teaching these courses for a long time, grammar should not be assessed in isolation. This does not help the students with their language. I've seen it; we teach these rules, students memorize them, and they tend to answer the right way most of the time but when they practice when they speak or write, they don't apply these rules which is absurd.

This teacher is of course drawing attention to what is a much debated issue in applied linguistics - that of whether metalinguistic knowledge about something like grammar can ever become knowledge of grammar that is used in spontaneous utterance of language. This teacher agrees with Krashen who famously says that the former, which comes with what he calls learning, can never become the latter, which comes with what he calls acquisition (1988). There is, however, a well established view that holds the opposite, arguing that explicit knowledge of rules can become automatized into unconscious implicit knowledge (e.g. in psychology Anderson (1983), and in applied linguistics Lindseth (2016)).

Furthermore, this teacher does not consider the alternative solution of teaching and assessing implicit knowledge in the Grammar course.

5.2.1.2 Beliefs about the purposes of EGA.

The participants were asked in the questionnaire about the different purposes of EGA that are commonly recognised in the testing literature, including various summative, formative and accountability purposes. All were agreed with this in principle and were considered valid by the teachers, reaching significantly above the neutral point on the scale. Nevertheless, it was apparent that some were endorsed more strongly than others (Table 20).

Table 20.
Results related to the Purposes of EGA

Item No.	In general, the purpose of EG assessment should be to...	Min.	Max.	Mean	Percent %	Std. Deviation	Sign test sig.
6.	Determine students' mastery over what they have been taught in an English grammar course	1.00	4.00	3.4409	86.02	.63353	p < .001
9.	Inform teaching by showing the students' strengths and weaknesses in English grammar	1.00	4.00	3.4202	85.50	.57410	p < .001
11.	Provide feedback to the students on their strengths and weaknesses in English grammar as they learn	1.00	4.00	3.4202	85.50	.60154	p < .001
8.	Provide information about how well each student is progressing in her/his English grammar	1.00	4.00	3.3191	82.97	.72160	p < .001

Item No.	In general, the purpose of EG assessment should be to...	Min.	Max.	Mean	Percent %	Std. Deviation	Sign test sig.
10.	Place students into groups for English grammar instruction as suited to their ability	1.00	4.00	3.0957	77.39	.82069	p < .001
7.	Motivate students to learn English grammar	1.00	4.00	3.0914	77.28	.82405	p < .001
13.	Indicate learners' abilities in learning English grammar for department chairpersons or external reviewers	1.00	4.00	2.9202	73	.82730	p < .001
12.	Indicate teachers' abilities to teaching English grammar for department chairpersons or external reviewers	.00	4.00	2.7926	69.81	.99030	p < .001

Summative: The results from items (6, 8 and 10) show that there is a high endorsement of EGA use to simply measure students' mastery over and progress in what they have been taught, which is the classic measurement of achievement with a summative purpose, relevant to any instructed learning situation. This was recognized and reported repeatedly in the interview responses, where the participating teachers (87.5%) asserted that EGA serves to evaluate a student's knowledge of the grammatical rules that are taught during the course lessons. The following quotes are some examples of teachers' convictions about the uses of EGA:

I think it is looking at students understanding based on what I'm teaching right now.

To evaluate the student's knowledge of the grammatical lessons we're taking.

I would evaluate my students understanding of the structure and the use of the rules.

The first two teachers both indicate a focus on measuring achievement, while the last refers specifically to assessing metalinguistic/explicit knowledge. None of the three; however, indicate any ulterior pedagogical use of the assessment that is the hallmark of formative assessment.

Also, in the interviews, some general statements about the nature of EGA do seem to refer to grammar knowledge as marked by proficiency rather than achievement. For example, one teacher said, 'in general assessing grammar to me is understanding their competence of the language they are learning'. However, the teachers more often speak ambiguously, e.g., a teacher said that EGA is conducted 'to know where the students stand in relation to progressing and learning the grammatical rules'. This quote could either refer to the grammatical rules in the course syllabus or the grammatical rules of English in general. From the questionnaire, however, the teachers understandably seemed to primarily regard EGA as a measure of student learning of what they teach rather than learning of the subject matter in some course-independent sense.

Formative: Based on the average mean in table 20, questionnaire items 7, 9 and 11 have ratings between 'agree' and 'strongly agree' mark the formative and more diagnostic pedagogical purposes of EGA. This implies that the teachers, in principle, also value the use of assessment as a facilitator of the teaching/learning processes rather than just as a measurement exercise with no purpose other than to produce marks.

This attitude towards the use of EGA for formative purposes was also found among three teachers (9.3%) in the interviews. Indeed, one of them succinctly stated that ‘when it comes to assessment, I try to attend to my students’ needs’. By *needs*, it is assumed here that the teacher means what is lacking in fostering students’ leaning, as another teacher said that EGA is done ‘to assess also their <students> needs; does that student need to practice more?’ Here, there is a recognition of the diagnostic function of tests in indicating what a student does not yet know and guiding future teaching. Although not stated, of course this function could only apply to quizzes and midterm exams, not the final exam.

A corollary of this finding is that the teacher can give individual feedback to students based on information from their assessment results. This is clearly implicated by one teacher who stated that ‘the number of students should be fewer, so you can give a customised feedback’. A general belief in grammar assessment’s formative value for teachers’ own learning, however, was not clearly articulated in the interviews.

The formative purpose of promoting student motivation was also endorsed at a weaker level in the questionnaire responses (item 7). Teachers did not see this a leading reason for assessment. Thus, they do not come across as strong believers in extrinsic or ‘carrot and stick’ motivation by exam requirements. Possibly, they see other ways of motivating students as important, which do not depend on the threats of assessment (e.g. intrinsic or instrumental motivators). Some members of the interview sample, however, did seem to recognise this as an effect, if not a prime purpose, of assessment. For example, a teacher stated that ‘unfortunately, they don’t study unless there is an announced exam’.

Accountability: Finally, the two lowest rated items (although still significantly positively endorsed) are related to the assessment either of students or teachers for the benefit of external authorities (accountability). Clearly, the teachers think that the main purpose of assessment should fall within the pedagogical situation, helping the teaching/learning processes (summative/formative), rather than impressing or convincing outsiders. In the interviews, these were not referred to as definitive or ideal purposes of EGA. However, only one teacher (3.2%) referred to EGA as parallel to following departmental rules, 'I guess to me are the guidelines that we specify here in the department'.

It is possible to assume that this relative lack of recognition of the accountability function is in contrast with the view apparent in the documentation that teachers have to deal with from above. The course reports teachers have to write at the end of each semester lay great store by accountability. Teachers are required to fill in sections on what forms of assessment were used, the overall student grades obtained in bands above and below Pass, comments on those (reasons for good/poor performance), and even a breakdown in terms of average marks obtained by the student group for each of eight broad CLOs separately, with comments given on why this or that CLO was or was not mastered at an acceptable level. Clearly none of this influenced the participating teachers' beliefs about purposes of assessment. Possibly the teachers manage to compartmentalize this in their minds as irrelevant paperwork, and not really part of the teaching of a course at all.

As summed up in Figure 33, it could be said from the above analysis that there is a generalized idea of the purposes that EGA serve. Results from the questionnaire show that summative and formative purposes were highly endorsed by the participants, with more inclination towards formative (avg. = 3.3106/ 83%) than summative (avg. = 3.2852/ 82%).

This was also reflected in the participants' reports in the interviews, with the majority of teachers supporting summative function (87.5%) as compared to the formative function of assessment (9.3%). The least endorsed function by the participants in the questionnaire (avg. = 2.8564/ 71%) as well as in the interviews (3.2%) was that of accountability. The figure below shows the results in percentage from the questionnaire (N=96) and interviews (N=32).

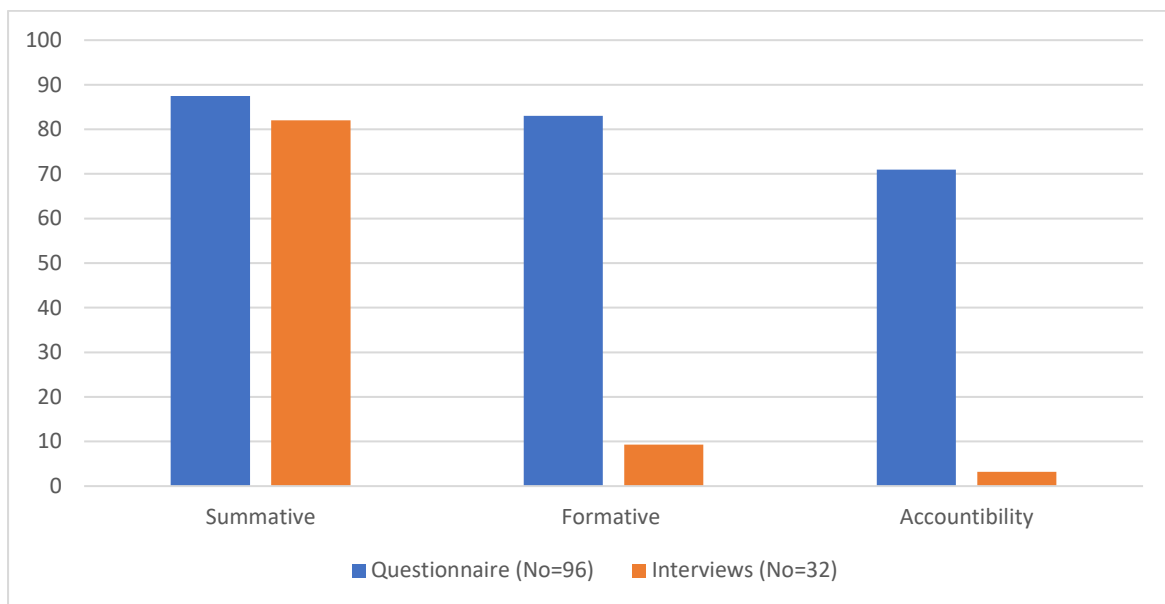


Figure 33. Summary of data results for the purposes of EGA.

5.2.1.3 Beliefs about the methods of EGA.

This section presents the results of responses on a variety of issues commonly aired in the literature concerning exactly what kind of grammatical knowledge should be assessed, when and in what mode. Table 21 shows the questionnaire results where the mean responses were significantly positive on all items, except for the last item listed. All the items in the questionnaire are clustered and discussed in detail, according to their relevance to the categories listed below.

Table 21.
Results for EGA Approaches, Means of Delivery and Assessment Frequency

Item No.	Statements	Min.	Max.	Mean	Percent %	Std. Deviation	Sign test sig.
16	English grammar assessment should use means that reflect real-life language use (not disconnected sentences or words)	2.00	4.00	3.5753	89.38	.63821	<.001
18	English grammar is best assessed in an integrative way, along with other aspects of English (e.g. through speaking or writing tasks), rather than as a separate skill	1.00	4.00	3.3656	84.14	.79105	<.001
25	English grammar assessment should systematically target students' knowledge of how different functions/meanings are expressed through English grammar (e.g. how an event in future time can be expressed, or how to make polite requests with 'Could you? Or, May I?')	1.00	4.00	3.2979	82.44	.72705	<.001
22	English grammar should be assessed frequently during the course	1.00	4.00	3.2872	82.19	.72768	<.001

Item No.	Statements	Min.	Max.	Mean	Percent %	Std. Deviation	Sign test sig.
14	Paper and pencil assessment provides valid evidence of students' learning of English grammar	1.00	4.00	3.2151	80.37	.74963	<.001
28	English grammar assessment should primarily be concerned with students' abilities to understand and use English grammar to communicate effectively and spontaneously (as a native speaker does), instead of imparting conscious knowledge about the language	1.00	4.00	3.2074	80.18	.87501	<.001
24	English grammar assessment should systematically target the different structural/formal features of English (e.g. the articles, how 'do' is used in questions and negatives, relative clause formation etc.)	1.00	4.00	3.1915	79.78	.73363	<.001

Item No.	Statements	Min.	Max.	Mean	Percent %	Std. Deviation	Sign test sig.
27	English grammar assessment should target students' explicit knowledge of grammatical rules (e.g. '-s has to be added to a verb in the simple present when the subject is third person singular', 'days of the week take 'on' while months and years take 'in')	1.00	4.00	3.1489	78.72	.78914	<.001
26	English grammar assessment should systematically target students' knowledge of common grammatical terms, such as verb, object, dependent clause etc.	1.00	4.00	3.1223	78.05	.78865	<.001
15	Computer technology helps in assessing students' English grammatical abilities	1.00	4.00	3.1183	77.95	.74600	<.001
20	Subjective assessment (e.g. rating overall grammar quality in a short essay or oral presentation) is a good method to assess English grammar	.00	4.00	3.0585	76.46	.89159	<.001

Item No.	Statements	Min.	Max.	Mean	Percent %	Std. Deviation	Sign test sig.
19	Objective assessment (e.g. through scores from sets of multiple-choice items or cloze gap filling items etc...) is a good method to assess English grammar	.00	4.00	2.9674	74.18	.95447	<.001
17	English grammar assessment should target specific elements of English grammar in separate items (discrete-point aspects)	.00	4.00	2.8333	70.83	1.02505	<.001
21	English grammar errors are only important when they get in the way of successful communication of the message being conveyed	.00	4.00	2.5213	63.03	1.18665	.004
23	English grammar should be assessed at the end of the course	.00	4.00	2.2394	55.98	1.33144	.224

Approaches to EGA – Integrative versus Discrete: The overall mean of 3.3009 (82.5%) from items 16, 18, 20, 25 and 28 indicates that the majority of the respondents seemed to have positive beliefs towards assessing English grammar in an integrative fashion. The results also show that the most strongly endorsed items fall between ‘agree’ and ‘strongly agree’, demonstrating teachers’ beliefs that grammar assessment should reflect a real-life use of

English (item 16), instead of being separated or isolated from other skills such as speech or writing (item 18).

In the interview, the integrative view was also often supported when teachers spoke about ideal assessment. The majority of the participants (64.3%) believed that grammar should be assessed in an integrative way, through writing or speaking as a potential means to increase the effectiveness of language learning. The following example illustrates this view clearly, although intriguingly the teacher does not explain why they do not assess in their preferred way. Later some possible reasons for teachers' practices not fitting their beliefs would be uncovered:

I think and to be honest with you although I've been teaching these courses for a long time, grammar should not be assessed in isolation. Grammar should be assessed in a more integrative more comprehensive ways not just in the way we do it: in isolated items and multiple-choice questions.

Another teacher expresses the same preference for integrative assessment of grammar, although the participant mistakenly contrasts it with explicit, where discrete is really the appropriate opposite. It is however true that assessment through discrete points is usually also explicit, while assessment of grammar integratively is more likely to be implicit:

I am against explicit assessment of grammar. This is old school. Nothing should be explicit; everything should be integrated all skills not just grammar, but grammar is like the basis; it's the structure and foundation. So, they have to learn how to apply the rules they learned in grammar into all the other skills.

The following teacher also makes a clear link between those two dichotomies: assessing what comes naturally to a person is implicit while assessing in context is integrative.

In my opinion the right way is not through assessing certain topics or subtopics. I think because grammar comes naturally to the person it should be assessed in context.

Another teacher also gave an eloquent defence of contextualised integrative assessment.

Again the participant teacher here implies that the belief system is not put into practice, which will be discussed later (section 5.4.3):

There are better ways of going about assessing grammar than kind of giving them blanks, and you know use this verb in this way because if you think about it in real world; we don't go around fixing verbs individually; we use language. So, I think the best way to kind of assess grammar is really through speech, I would prefer that students would be able to provide an assessment verbally and see how they use the language and how they construct it.

The third item (item 25) in the order of approval in the questionnaire endorses the role of meaning/function of the grammar. Of course in terms of the real use of language, grammar is always present in order to express some meaning or function, so the teachers often combined favouring assessment of meaning with approval of communicative assessment.

In the following quote, this combination in belief is evident, although interestingly the teacher sees explicit rule learning as relevant to be assessed even where the target is the meaning of grammar in communication:

For me okay, in assessing grammar it is knowing what rules to use to communicate which meaning.

Another teacher also endorses focus on meaning in communication, but interestingly identifies it only with speaking and writing, as if the meaning of grammatical forms is not relevant also in listening and reading, where it could equally be assessed (e.g., Widdowson, 2003):

I think that we use language to express the meanings that we want to convey either in speaking or in the written form. So, we use grammar as a tool or as mechanics to help us express the meaning that we want to say.

Another item about meaning, with a mean above 3 ('agree'), is item 28, highlighting the importance of the ability to understand and use English grammar for spontaneous communication like a native speaker, rather than possessing conscious knowledge about the language (i.e., explicit metalinguistic knowledge). This again shows that the respondents, on principle, prioritize the meaning expressed by grammar integrated within speaking or writing.

One teacher in the interviews spoke about performance ability (communication) which was associated primarily with speaking, although of course in applied linguistic scholarly discussion it applies equally to all four skills (Widdowson, 2003):

I think the best way to kind of assess grammar is really through speech, I would prefer that students would be able to communicate properly, verbally or otherwise, through correct English and I then can see how they use the language and how they construct it.

Other teachers mentioned writing as well, but not the other two skills. For the most part they did not refer to assessing explicit rules in this area, but rather spontaneous performance ability, which would be integrative, for example, 'to assess the students on how to use the

language how to use structure or form a complete sentence’, or ‘to help the student to speak correctly, to write in the right form’.

In terms of assessing English grammar as discrete-points, so typically explicitly, responses from the questionnaire (items 17, 19, 24, 26 and 27) show that teachers were fairly positive (although they did not endorse it as highly as the integrative approach) about this type of assessment being used in their classes, with an overall average of 2.9641 (74%). Indeed, later it would be evident that it is the approach they typically actually used in practice. Teachers’ responses to endorsing items 24, 27 and 26, with average scores of 3.1915 (80%), 3.1489 (78.7%) and 3.1223 (78%), respectively, were the highest in the category of assessing English grammar explicitly. Considerably lower means were obtained by items 19 (avg. 2.9674/ 74%) and 17 (avg. 2.8333/ 70.8%), both of which refer to objective assessment through MCQs and gap filling (item 19) and the assessment of EG elements separately.

Responses from the interviews, in contrast, show that there were fewer instances of teachers supporting EGA via discrete-point and decontextualized methods. Only six participants (19%) expressed positive views about assessing grammar explicitly through gap filling, MCQs, true-and-false questions etc. For example, some teachers stated the following:

You can assess them through various ways like multiple-choice questions like filling gaps, negating sentences or asking to write certain structure.

However, another was more ambivalent:

I think that assessing grammar using discrete point and independently is valid but valid doesn't mean correct. It is valid because it has been done. At some point like in proficiency test TOEFL they do ask about grammatical rules separately.

Indeed, the TOEFL does make use of multiple choice questions, and it does have a separate section about structure that assesses explicit grammatical rules. I think here the teacher's belief can be appreciated, although such tests are used by some people, that does not make them necessarily 'correct' or appropriate for the sort of GA their context is concerned with.

Four teachers (12.5%) expressed the thought that mixing both methods, both in an integrative way and explicitly, would offer more opportunity to use the language in context. As two participants stated,

I think it has to happen in an integrative way and also in isolation. I think grammar should be assessed not just through exams with multiple-choice questions and filling the gaps, and also I think they should be asked to write in a grammar class using certain context some of the activities I mentioned before like writing and speaking production specifically these two instead of just practicing them in writing and speaking classes why not in grammar class.

Within the approaches of EGA, questionnaire items 20 and 19 (avg. 3.0585/ 76.46% and 2.9674/ 74.18%, respectively) present the subjective and objective means of assessment.

These two items were ranked closely with regard to subjective assessment through writing and speaking, an integrative approach that was supported more than objective assessment, which takes place through MCQs, true-and-false questions etc. (discrete-point).

In their interviews, the participants did not talk much in terms of subjective and objective assessments but rather focused on a different dimension which was not explicitly targeted in the questionnaire set of items (Table 21): production versus recognition/reception. Here, production refers to students' ability to use the grammatical rules in writing and speaking, which is also assessed as performance ability. The majority of the teachers (62.5%) stated that assessing English grammar should target language production. One teacher saw this as isolated sentence production:

It is assessing students' abilities to produce correct sentences with correct grammar.

Another saw it more integratively and communicatively in terms of writing and speaking production.

You need to apply the rules, the grammatical rules when writing an essay or giving an oral presentation.

Assessment of recognition/reception, on the other hand, was endorsed by 37.5% of the participating teachers. Recognition or receptive grammatical knowledge can be assessed by discrete-point items. Here are some examples of their quotes:

Teachers feel more comfortable using objective questions, more economic time and effort and even students they like it as well.

I check their understanding of grammatical rules usually and I prefer multiple choice questions and error analysis. These straight to the point; students answer them quickly, and they are easy to correct.

Despite the attitude just noted in support of assessing grammar with attention to meaning, function and communicative value in realistic integrative language, item 21 (avg. 2.5213/63%) was only second from bottom. This item, unlike the others, refers to grammatical errors and shows that the teachers do not strongly believe that grammatical errors should only be negatively assessed if they damage effective communication. It therefore signals a possible contradiction in some teachers' belief systems where they on the one hand favour integrative assessment but on the other still want attention to strict grammatical correctness. In the interviews, the participating teachers did not elaborate on this issue directly but they referred to error analysis in a different sense, as a type of grammar test item given to students where they are asked to identify and maybe correct errors, for example, 'I prefer multiple choice questions and error analysis'. Such items used in exams would always require students to identify errors regardless of their communicative damage.

One teacher believed that grammatical mistakes by students majoring in English is not acceptable, which perhaps implies a belief that errors should be marked down in assessments associated with academic learning, regardless of other aspects such as success (or the lack thereof) in conveying meaning. She said the following:

They think they know how to speak, and they think they are good speakers and they can converse in English, so they don't need grammar, and when they speak, they make a lot of mistakes. So, this is okay if they are in coffee shop or in restaurants with their friends, but in an academic institution, this is not acceptable.

Another teacher shows recognition of the distinction between slips ('silly mistakes') due to performance-related factors and errors due to faulty competence (Edge, 1989) but does not indicate whether they should be scored differently: 'I see the students in class and then I see them in the test, yes most of them have full marks and they have good marks, but they make really silly mistakes and I don't think because they don't know the rule but they panic because it is a formal test.'

By contrast, a teacher indicated that she believes in a marking system that involves different levels of penalty for mistakes, including half and quarter marks, but does not say what aspects of the errors should merit greater or lesser deductions:

There is this grey area in which students argue for half a mark and that is why they need something under their hands. They need something to show them this is a mistake you lost a mark here you lost quarter here.'

Medium of assessment: As shown in Table 21, the purpose of items 14 and 15 was to address teachers' beliefs about the preferred medium to assess English grammar. The questionnaire items asked about traditional pencil-and-paper assessment versus computer-based assessment. Of those two, the teachers clearly believed more in the former (item 14, with avg. 3.2151/ 80.37%) in comparison to the latter (item 15, with avg. 3.1183/ 77.95%).

In the interviews, not all the participants talked about these choices of medium but rather spoke about written exams which, presumably, in the context, would be on paper. 12 teachers firmly supported written form of exams:

I do believe that it has to be written and not oral. It is easier for both the student and the teachers

I am a huge fan of written exams. I believe it is more academic and formal.

I prefer everything to be written. It is better when students sit, read the question and write down the answer.

Those three remarks respectively highlight three reasons for preferring written exams: ease, formality and (implied) the extra time that the students may need to write the answers to the questions compared with speaking. All would of course be equally available if exams were to be conducted on computer.

Another teacher however preferred a combination, although the participant judged oral assessment to be better done as coursework while written assessment as exams and quizzes:

I think I can assess students' performance in grammar via different tasks either orally in class through participation or presentation or written through exams, midterm, quizzes and final exam.

Frequency of Conducting EGA: Items 22 and 23 in the questionnaire ask about the beliefs of EFL teachers about how often grammar should be assessed. The results indicate a clear difference in the questionnaire responses. There was strong expression of support for conducting grammar assessment frequently during the course (item 22, with avg. 3.2872/ 82.18%). In contrast, conducting grammar assessment only at the end of course (item 23) was rated the lowest in this subset of items, and the agreement did not differ significantly from the 'neutral' response (avg. 2.2394/ 55.98%).

In the interviews, none of the participants talked about how often grammar should be assessed, because there was no question that addressed this aspect in particular. However, all the participants talked about the most appropriate sources to use when writing their quizzes and midterms, which naturally occur during the course, e.g. one teacher stated, ‘I believe previous exams to be reliable sources when writing my quizzes and exams’.

In sum, it could be said that the results from the questionnaire and the interviews indicate that there is a strong endorsement of assessing grammar in an integrative way rather than explicitly. There is also a positive belief towards using a paper-and-pen format in assessing grammar. Finally, the majority of the participants believed EGA should occur frequently during the teaching of grammar courses. Figure 34 summarizes the results from the questionnaire and the interviews.

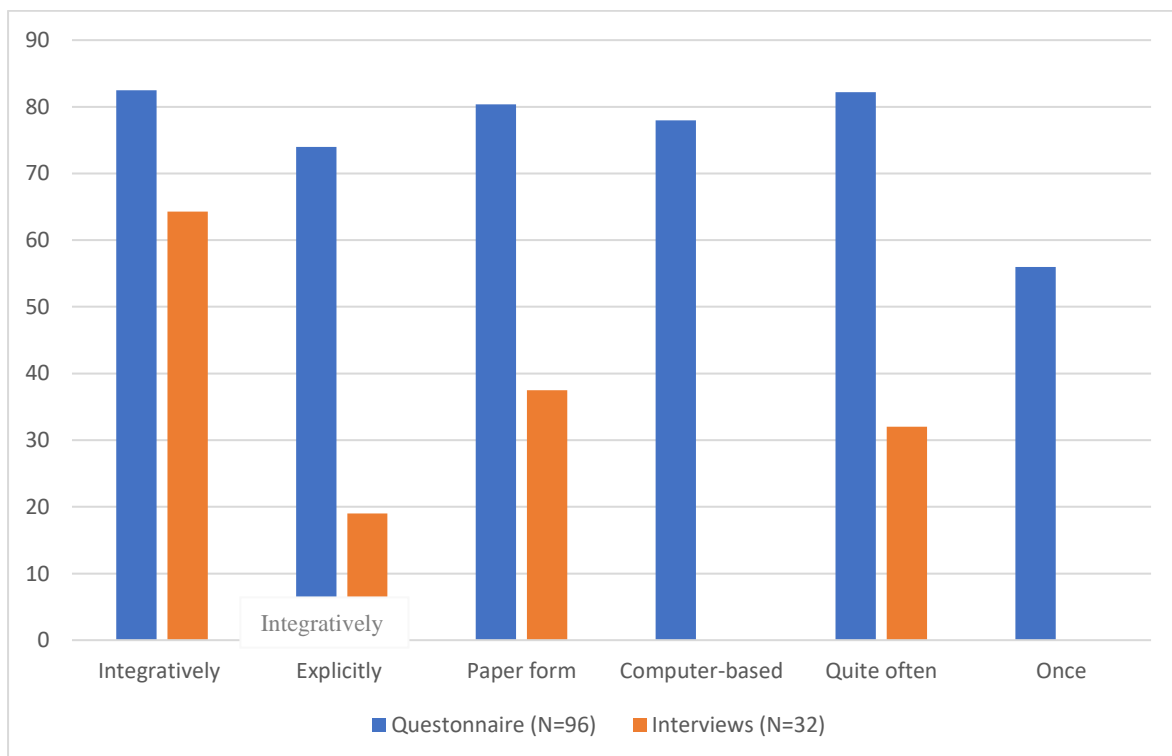


Figure 34. Summary of data results for EGA methods.

5.2.1.4. Beliefs about suitable EGA task/item formats.

Table 22 shows EFL teachers' beliefs about assessment format. All the item formats that I asked about were endorsed with ratings between 'agree' and 'strongly agree', all falling significantly above the midpoint of the agreement scale. Nevertheless, there is variation in the rank order of approval.

Table 22.
Beliefs about Suitable Format for EGA

Item No.	In general, English grammar can usefully be assessed through....	Min.	Max.	Mean	Percent %	Std. Deviation	Sign test sig.
37	editing / error correction tasks	2.00	4.00	3.4681	81.38	.59482	<.001
38	error recognition / grammaticality judgment tasks	1.00	4.00	3.4096	80.98	.67983	<.001
32	sentence transformation or production items	.00	4.00	3.2553	79.92	.74684	<.001
35	open response sentence completion	1.00	4.00	3.2553	79.65	.63392	<.001
36	essay writing	1.00	4.00	3.2394	78.32	.78879	<.001
31	filling cloze gaps in text	1.00	4.00	3.1968	76.99	.74494	<.001
34	multiple choice sentence completion	1.00	4.00	3.1862	75.13	.80317	<.001
29	speaking in an oral interview	.00	4.00	3.1330	81.38	.84617	<.001
30	speaking in an oral presentation	.00	4.00	3.0798	80.98	.79414	<.001
33	matching items	.50	4.00	3.0053	79.92	.89050	<.001

It is apparent from the table above that there is a high level of agreement on assessment by giving students error correction tasks (item 37, with avg. 3.4681/ 81.38%), followed closely by giving them error recognition tasks (item 38, with avg. 3,4096/ 80.98%). The high level of

endorsement of these items aligns with teachers' responses to item 21 in Table 21, where the participants disagreed with the statement that grammatical errors should not become a matter of concern unless they hinder communication.

In the interviews, as mentioned in the previous section, the teachers talked broadly about item formats in the English grammar exams and indicated a general approval of the use of error analysis in grammar exams. For example, one teacher said the following:

I like the questions where the students have to correct mistakes when they read and find the mistakes in the structure and write down the correction. Some students know there is a mistake, but they cannot correct it. By the time they finish analysing the sentence; they know what the problem is and correct it.

Interestingly there is a hint here that this teacher believes that students may actually learn through responding to this form of test item, rather than just provide information about their knowledge to the assessor. In fact there is a long history of study of effects of testing on learning, of which this would be one kind, although often it is seen as limited to enhancing retention of what is already learned (e.g. Larsen et al., 2013).

In addition, the results show a similar level of agreement across items that support explicit EGA (items 33, 34 and 37 with avg. 3.2198/ 80.49%), typically of discrete points, and those which are geared towards EGA of implicit knowledge (items 29, 30 and 36, with avg. 3.1507/ 78.76%), typically in integrative tasks. Teachers were probably more inclined to support more explicit item formats as they are the most common formats in the grammar exams in their context, and indeed others.

Similarly, many teachers in the interviews expressed positive perceptions towards a wide range of formats. Below are some examples of their responses:

You can assess them through various ways like multiple-choice questions, like filling gaps, negating sentences or asking to write a certain structure, or also we can do it orally by asking the students in class.

Here despite the general support for integrative assessment seen earlier, all but the last format mentioned would most likely be targeting discrete points. By contrast the following teacher suggests a balanced approach, and interestingly, as noted in an example earlier, sees the choice of item as not just benefiting assessment but also student learning.

I think students should have a variety of different questions. I think we should have a balance of both techniques. We do it in an integrative manner, so they benefit from using the structure in context and separately to help them learn the exact rules and how it is done. So, it would be a mixture of error analysis, filling in the blanks, multiple-choice questions, closed questions, so you can choose from all these types of questions.

5.2.1.5. Beliefs about sources to construct assessment items/tasks and teachers' role in the construction process.

5.2.1.5.1 Preferable sources to use when constructing EGA items.

As can be seen in table 23, all the propositions offered about the preferred sources to construct grammar assessment items were agreed upon, significantly above the midpoint.

Table 23.
Beliefs about Sources to Construct Assessment Items/tasks

No.	Statements	Min.	Max.	Mean	Percent %	Std. Deviation	Sign test sig.
45	English grammar exercises from published textbooks are a useful source for constructing grammar assessment tasks	1.00	4.00	3.2021	80.05	.63635	<.001
46	Using English grammar assessment items from previous years is a good source to construct grammar assessments	.00	4.00	2.9255	73.13	.90387	<.001
44	Ready-made English grammar exercises/ tests found on the internet are a good source for grammar assessment tasks	.00	4.00	2.7872	69.68	.91137	<.001

The results in the table above indicate that with respect to the sources that should be used for assessment items, the teachers believed that published textbooks were highly favoured as a suitable source, followed by papers from previous exams, while the least favoured was the internet.

In the interviews, the analysed data revealed three main sources that the teachers considered preferable when constructing an exam. These sources were similar to the sources presented in the questionnaire items in the above table, although they were not endorsed in the same order. Sources from the interviews were previous exams written either by the course teacher or other teachers who taught the same course, grammar textbooks other than the ones being used in teaching the current course and by resorting to ESL websites.

The teachers felt that using previous exams is justified and useful because these exams, in a sense, have been piloted and are therefore more reliable:

Yes, I do adapt and use previous exams because I benefit from the experience and I know which questions work and which are problematic.

The teachers, however, claimed that they do not necessarily use the exams exactly as they were but rather prefer to adapt and modify the questions that are suitable with respect to their content and to the level of the students:

Yes, why not if it is on the same syllabus and if it is the same topics, and she knows that this exam suits her students, I don't mind. Yes, why not.

Finally, a teacher noted that it had to be assured that the students had not seen the questions before:

I can't see a problem and if it's okay just to make a few changes but with one condition, if the students do not know these questions. If the teacher still has the copies of the previous exam, and we are sure it's not spread all over the students. I don't see why not.

15 teachers supported this finding, with similar positive statements regarding preceding exams being the most favourable source to use while writing their own exams. One teacher, however, complained that most of the previous exams she had looked at contained jargon that confused the students:

I looked at the versions of previous finals and I didn't like the kind vocabulary used; it's above the students' level and it will confuse them. They will not

look at the grammar rule because they are trying to understand the vocabulary and that is not necessary.

In technical terms, this teacher is claiming that the items found in previous exams were invalid because in fact they were testing vocab knowledge more than grammar knowledge in what was claimed to be a grammar exam. The same teacher stated that using previous exams (her own or what other teachers wrote) is an act of laziness: 'I think it is kind of lazy for me'.

Another teacher did not clearly oppose the use of previous exams but said that she never used the same exam twice, as the students change every semester and, therefore, their individual levels and their needs differ:

I have never used the exam twice ever in my 30 plus years because I have never had the same group of students. I don't need to look at other teachers' exams.

Using online sources such as ESL websites seemed to be ranked next as a reliable source to find suitable materials for exam questions and items. 11 teachers confirmed that they resort to websites to search for suitable items and that they eventually have to modify whatever they find online to suit the students' cultural and linguistic backgrounds.

I would also try to use the internet to find suitable activities I of course have to tweak some of the sentences in order to make them more acceptable to our context socially, culturally, religiously... etc.

I think the internet is full of exams and English language is well served on the internet but the problem with the internet is that the material is not tailored to

the students' need. I can get one idea here and one idea there, but I don't think you can find the perfect assessment activity or task.

That kind of adjustments identified here are, as the participants explained, would be related either to cultural aspects or students' need. By need, the teacher might be referring to the student's actual language needs, or to the needs listed in the course specification (e.g. CLOs).

Another teacher also mentioned choosing/adapting to suit the students, but was unclear if that meant adapting to student needs, as above, or to students' existing level of competence, as already commented on:

Actually, the internet is full of lots of samples in different ways to ask the students; I mean samples of tests so one chooses the best that suits their students and fits in the context.

A few teachers reported the usefulness of using grammar textbooks to find and adapt instructions and items for their exams or item type or format rather than specific items:

I would take ideas from other books the questions they're using difficulty level.

I used to look at books and take from these books the structure of the question and I take the examples sometimes from the books sometimes from the net.

Overall, a consensus was identified between teachers' perceptions in the questionnaire and their responses in the interviews. All the participants agreed, although not with equal degrees of endorsement, on three main sources which are suitable to use while constructing EGA item formats: textbooks, previous exam papers and ESL websites. Figure 35 summarises the sources endorsed by the participants on both the questionnaire and the interviews, along their level of endorsements.

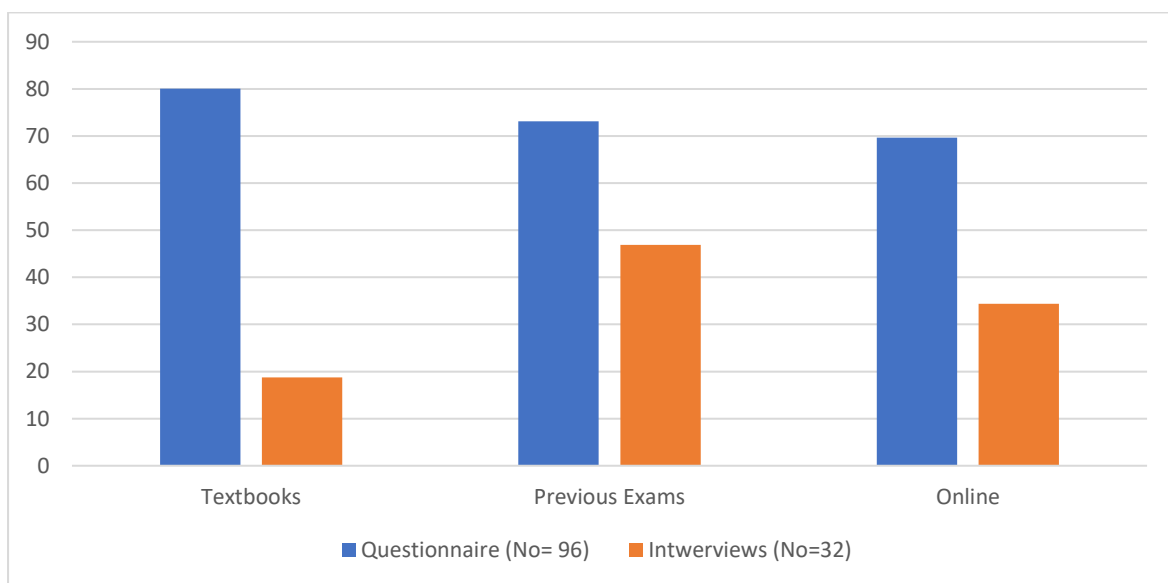


Figure 35. Summary of data results on teachers' beliefs about sources that facilitate the construction of EGA items.

5.2.1.5.2 Teachers' role in constructing EGA.

The purpose of items 39, 40, 41, 42 and 43 in Table 24 was to explore EFL teachers' beliefs about their role in EGA construction process. Interestingly, all items in the questionnaire below were perceived positively.

Table 24.
Beliefs about Teachers' Role in EGA Construction

No.	Statements	Min.	Max.	Mean	Percent %	Std. Deviation	Sign test sig.
43	English grammar assessment tasks are best prepared collaboratively	1.00	4.00	3.1223	78.05	.70209	<.001
39	The best grammar assessment items are the ones developed by the course instructor.	1.00	4.00	2.9894	74.73	.87676	<.001
42	Self-assessment by students of their own English grammar performance is useful	1.00	4.00	2.9255	73.13	.86122	<.001
41	Assessment of a student's English grammar performance by their peers is useful	.00	4.00	2.8351	70.87	.97106	<.001
40	English grammar is best assessed by expert professional testers/examiners rather than a class teacher.	.00	4.00	2.5904	64.76	1.15461	<.001

Concerning who should write the grammar assessment tasks, the most favoured option was that it should be done collaboratively (item 43, with avg. 3.1223), closely followed by item 39 (avg. 2.9894) in which teachers believed that the course instructor should be the one responsible for writing EGA tasks. Items 42 and 41 (avg. 2.9255 and 2.8351, respectively) came in close proximity to endorse peer- and self-assessments. The least favourable item, though still significantly above the midpoint in the 'neutral-agree' range, was having assessment tasks set by professional testers (third party) rather than the course teachers themselves (item 40, with avg. 2.5904).

In the interviews, the participating teachers endorsed distinct views as to whether teachers themselves should construct the exams or a committee should oversee the construction of the required exams for the grammar course.

Seven teachers made it clear that teachers are the best candidates to write exams because of their direct involvement with students and the teaching process, which should align with the exams, as the following comments suggest:

Yes, I believe the teacher herself should write her own exam. I don't believe in unified exams and have never been part of unified exams. This is my exam; I teach them in a specific way, and I like to test them on a way that matches the way I teach them. So, each teacher should write her own exam.

This teacher clearly sees a need for assessment to match the course teacher's teaching style while the following sees a need more for it to match what was taught.

Yes, the teacher is the one in immediate contact with the students and the course material. I am the one who caters for my students' need and possible problems and when I write my exams, it is tailored to what we covered in class and how we covered it.

The following teacher also highlights the need to fit what was taught but admits that the teacher might need some training in assessment methods.

I think the best person to do that is the teacher; maybe they need to be trained with regard to assessment tools and assessment methods, but they are the best persons to do it because they know what they taught the students.

One of the teachers praised the role of writing one's exam: 'There is a pride of the teacher having her own style and writing her own exam'.

To shed more light on teachers' perception of their role in writing English grammar exams, they were also asked how they felt about having an assessment committee construct the exams on their behalf. Teachers' views represented extremes. On one hand, a high proportion of teachers (49%) expressed positive thoughts towards having an assessment committee as long as the course teachers are involved and have a say on what is to be tested and how. The teachers attributed this preference due to a couple of reasons. First, having such a committee would definitely relieve them from the burden of writing exams and the time it took to construct them. Second, having a committee may help standardise the exams and make them more valid and reliable. The following statements support the above finding:

No, I don't mind having a committee to guide me, and I can learn from them; but I want to be involved in writing my questions.

That would remove a lot of work from me. I never thought about that. Personally, I wouldn't mind. I think it would be helpful because it would contain and maintain quality; specially, if you have three, four, five sections, then all of them would receive the same questions and all of them would have to attain a certain standard that is set by the department and then, there is a clear message – this is what you need to know in terms of grammar and this is what the department expects you to know. I think I am going to lean towards that simply because it maintains standards.

Another teacher showed awareness of what was already happening in the year prior to students entering into their majors:

Excellent, it's a great idea. I think they're doing this in the prep year. Especially with the skill courses because you don't want to waste your time on designing the same test.

However, 12 teachers (41%) clearly expressed that relying on a committee to write the exams and hand them over to teachers to be administered would probably not be the best idea. The participating teachers claimed that if the exams were to be written by a committee, then all purposes of teaching would be directed to prepare and pass these exams and that would compromise and jeopardise the academic learning process. The following statements highlight the teachers' opinions and their concerns:

Having a committee makes it easier for the teacher but, then again, all the teaching process will focus on what they will have in the exam and how to make sure that the students can answer these questions. So the focus will not be on the needs of the students themselves and their level; it will be on how they can be ready for that exam.

This teacher highlights the well known phenomenon of backwash or washback of testing into teaching (e.g. Hughes, 2010). However, this is not usually seen as bad if the assessment does in fact test what the students are supposed to be learning. In the present context any committee would presumably design the assessment to test what is specified in the course specification, especially the CLOs. In that case, since the teacher is teaching to the same course specification, this should not lead students to concentrate on anything not relevant to their course.

Another teacher raised the issue that only the teacher knows the students' strengths and weaknesses:

You mean I don't get involved? No, I don't agree because I think that the teacher can collaborate with them; but I don't accept to be excluded from the group because I'm the teacher and the one who teaches the student. I'm the one who knows their weaknesses and their strengths; so, I think that it's important to construct the exam by the teacher herself or as a collaboration. So, we can share ideas and experiences; no problem with that, but I do not accept excluding the teacher from constructing the exam.

The teacher here implies (as seen in other examples above) that the assessment should be suited to the abilities of the students, which of course a committee is unlikely to really know. As it has been pointed out above, this belief however clashes with the institutional approach, which is to conduct assessment based on specified course outcomes (CLOs etc.), not students' prior abilities. Once again this teacher is assuming that what is needed is assessment, tailored to the particular group of students, rather than assessment referenced to whatever the students are supposed to be learning. A very similar view was expressed by another teacher:

I wouldn't agree with that because you know, I'm teaching the students; I know their level, so I could challenge the students more. But if I have like a committee of assessment, they wouldn't truly know the differences among the students. I think each instructor knows her students very well. I would agree if this committee would be like the coordinator's job. You know, we will put the criteria that we need to follow, but not the same test; not all of the instructors should have the same test.

As that teacher implies, the norm-referenced approach that the teacher favours might actually make the exam harder for the students than the criterion-referenced approach followed by a committee. The teacher might assess only those parts of the syllabus that he/she knows the students have problems with ('challenge'). Criterion referenced assessment, however, assesses all the part of the course or CLOs regardless of whether the students find some of them easy. In that way the teacher's belief about assessment here conflicts with that conveyed by the university in its quality documentation.

Yet another teacher drew attention first to the idea of teacher autonomy in a university context and then to a possible effect on student emotion:

I don't know. I'm not a fan of doing exactly what others ask me to do because especially I'm a teacher at the university. I should have my own space to work freely and, honestly, to say usually during the final exam sometimes things will be tense because I do not like my students to be forced to be assessed according to someone else's questions. It does not feel right that I give my students quizzes or exams or whatever that I have no opinion on. I want to be involved in every step that my students are going through as long as I am their teacher.

This teacher's belief in teacher autonomy at university has of course already been challenged by the fact that in the KSA he/she no longer has freedom to choose their own course content or textbooks or many aspects of the assessment. That is all determined by the course specification compiled by someone higher up in the university hierarchy. Possibly the teacher feels that it is 'the last straw that breaks the camel's back' if they also lose control of the setting of the actual assessment papers each semester. With respect to the impact on student

stress levels ('tense'), the teacher is probably correct in believing that students would see an exam set by a committee as more a source of anxiety than one set by the class teacher.

The above three statements not only illustrate different negative aspects that the participating teachers believe to be associated with an assessment committee but also support the belief that if a committee is to write an exam, then the course teacher should have a voice and feel involved in the questions included in the written exam.

Three teachers expressed indecisive views about whether or not to include a committee in the exam writing process. Based on their quotes, it is apparent that these teachers had concerns regarding the degree to which this committee is involved in the teaching context.

I don't think it is something I would support, but I also don't think it is not okay. If it is going to make things to take into consideration the different teachers, then why not? But I don't advocate it; I think part of the teaching, of course, is being involved in writing the exam. Unless this committee is going to involve the teachers themselves. So, the exam would be basically compilation of all the teachers' contribution or like a workshop where all the teachers write together; then why not but have a committee that wasn't involved in the teaching? I don't think I am okay with that there is something that doesn't sound right.

This kind of belief seems to represent a compromise that could become acceptable in the context, since discussed, the questionnaire item favouring collaboration in exam construction was the one that gained the highest rating (item 43). Clearly, however, the whole idea of exams being set centrally solely by people independent of class teachers is strongly resisted.

In sum, the participating teachers believed that course teachers should be the ones trusted with or heavily involved with this task. The teachers also appeared to recognise the usefulness of having a third party that can construct EGA tasks, a committee that can support them in the examination construction phase. However, they still prefer to have some autonomy and write their own questions. Figure 36 summaries the results of teachers' beliefs from the questionnaire and the interviews about who best construct EGA tasks, namely teachers, solely or collaboratively, or a third party (e.g. a committee).

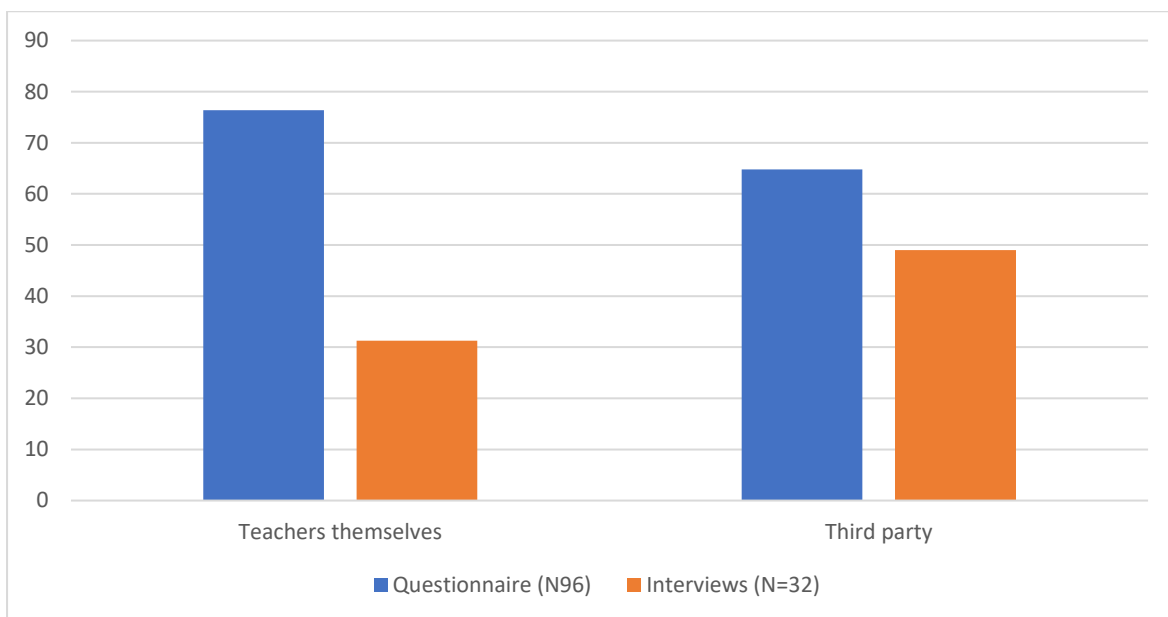


Figure 36. Results from the questionnaire and the interviews for who best construct EGA.

5.2.1.6 Summary of teachers' beliefs about EGA.

This section discussed EFL teachers' beliefs in terms of various aspects of EGA. The results from the questionnaire data analysis and the interviews were discussed separately as well as concurrently whenever and wherever relevant. The findings indicate a degree of consistency between teachers' responses in the questionnaire and their reports from the interviews. Figure 37 represents the themes and categories discussed by the participants in both the questionnaire and the interviews.

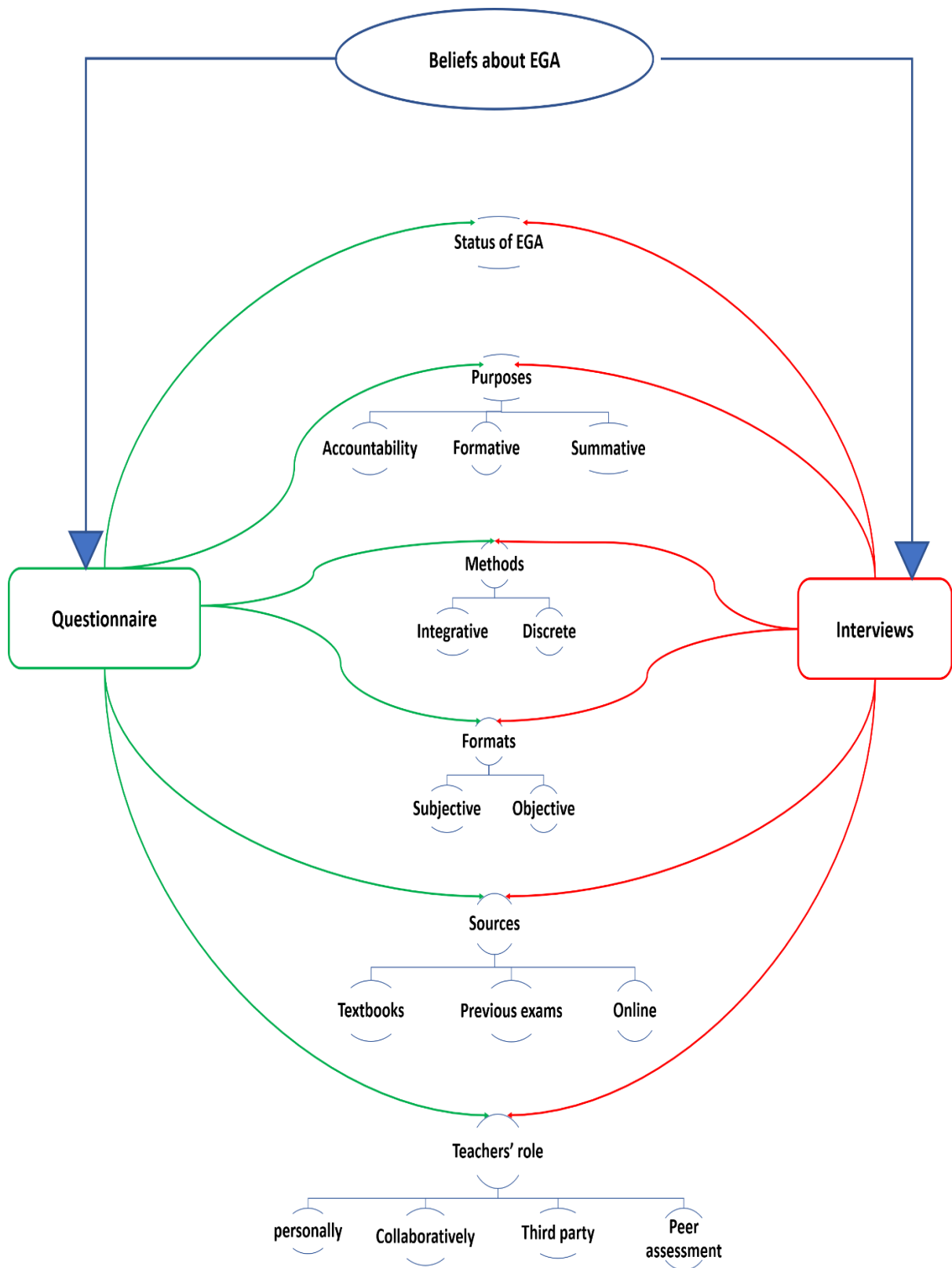


Figure 37. Teachers' beliefs about EGA: Themes and categories from the questionnaire and the interviews.

5.2.2 What are the factors which helped shape such beliefs?

This section examines some factors that may have affected teachers' beliefs about EGA. I tried to specifically locate whether EFL teachers' beliefs were affected by their gender, age, educational level, training received in assessment, length of experience of EFL teaching, or of grammar teaching or of grammar assessment. I also analysed the effect of the country of origin, although, only in the sense of Saudi versus non-Saudi teachers, since specific non-Saudi countries were only minimally represented in the data.

5.2.2.1 Gender of teacher.

Based on the questionnaire results, only two significant differences were found, and I must say that these should be seen with caution because males were not well represented in the sample (Table 6). Males believed more strongly than females that EGA could serve a number of purposes, with a mean almost on the top of the scale with strongly agree (male mean = 3.89; female mean = 3.40; $z = -2.326$; $p = .020$). Females, more than males, however, believed sentence production and transformation items to be valuable (male mean = 2.72; female mean = 3.31; $z = -2.298$; $p = .022$).

In the interviews, just like in the questionnaire, the male participants were not well represented in the sample (Section 4.4.2.3). I would say, however, that the male participants' views about EGA seemed to align either with one group of the female participants or another. Generally, it can be said that there is consistency within the male participants themselves and amongst the majority of the female participants. More conflict in perceptions seemed to exist amongst the female participants than the male ones.

5.2.2.2 Saudi versus non-Saudi origin of teacher.

This yielded seven significant differences that I have summarised in Figure 38. I must state here again that non-Saudi participants were few in number and of very mixed origins (Table 6).

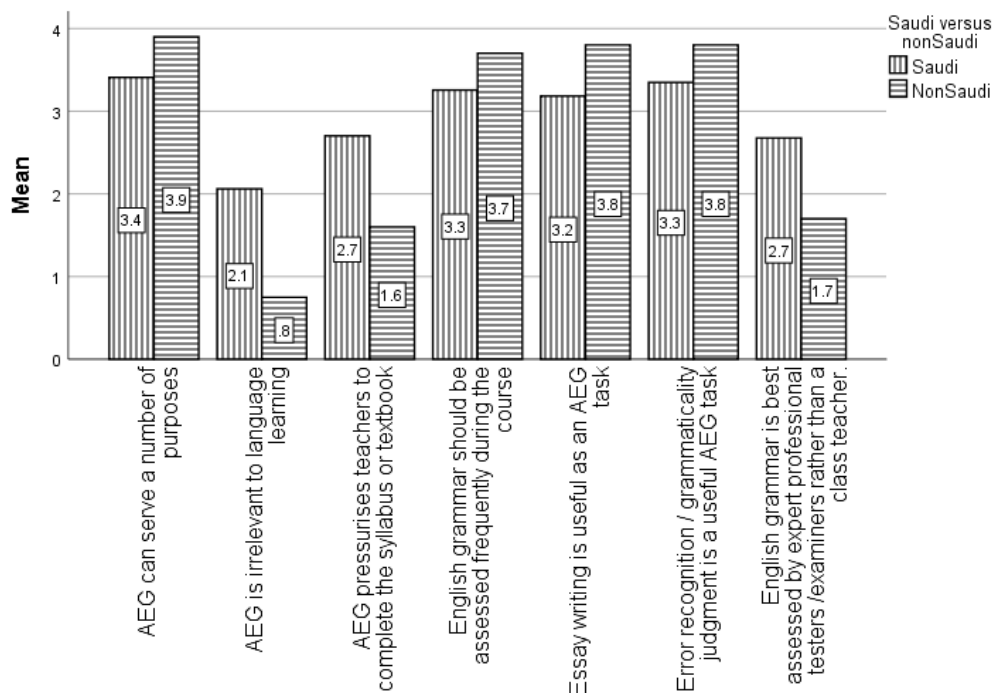


Figure 38. Differences in beliefs between Saudi and non-Saudi teachers.

Only four items were found to be regarded more strongly by non-Saudis compared to others: EGA can serve a number of purposes ($z = -2.533$; $p = .011$), grammar should be assessed often during the course ($z = -1.963$; $p = .050$) and essay writing ($z = -2.589$; $p = .009$) and error recognition/grammatical judgment tasks ($z = -2.056$; $p = .040$) are useful for EGA.

In contrast, three items that were generally found to be lowly approved, wherein the Saudi teachers on average agreed more strongly than the non-Saudis, although only in the neutral-agree range of the scale: the belief that EGA is irrelevant to language learning ($z = -2.814$; p

= .005), it pressurises the teacher ($z = -2.657$; $p = .008$) and EGA should be conducted by experts ($z = -2.380$; $p = .017$).

5.2.2.3 Age of teacher.

A negative correlation was established between age and the belief that EGA was difficult to do well by older teachers perhaps due to their greater experience. These disagreed more than younger ones ($\rho = -.284$, $p = .006$) regarding this point. Older teachers also agreed less that assessment is irrelevant to learning ($\rho = -.289$, $p = .005$) and that it pressurises the teacher ($\rho = -.238$; $p = .022$). They also agreed less than younger teachers that discrete point testing of separate aspects of grammar is a good idea ($\rho = -.246$; $p = .018$) and consistent with that, objective tests with multiple-choice and that similar items should be used ($\rho = -.239$; $p = .021$). Similarly, they favoured testing knowledge of grammar terminology less than younger teachers ($\rho = -.206$; $p = .046$). Interestingly, this did not mean that they had a more relaxed view of the importance of grammatical accuracy: they also agreed less than younger teachers that such accuracy was only important in cases where a communication confusion may occur ($\rho = -.218$; $p = .034$). Rather, age imparted a greater value of assessment subjectively and globally through overall teacher rating. This could be a sign of greater confidence in their subjective judgment rather than the need to rely on objective multiple test items.

5.2.2.4 Length of experience of EFL teaching.

Similar to age, greater general TEFL experience showed that EGA was not difficult ($\rho = -.205$; $p = .049$). However, this pressurised the teacher to complete the syllabus ($\rho = -.207$; $p = .048$). Again, like age, it lessened the value of objective multi-item tests ($\rho = -.238$; $p = .022$). Consistent with that, more experienced teachers were seen to be less convinced by the value of close multiple-choice ($\rho = -.241$; $p = .019$; $\rho = -.260$) and matching test items (p

= .011; $\rho = -.222$; $p = .031$). This variable was also found to be very nearly significant for the value of assessing knowledge of terminology ($\rho = -.202$; $p = .051$). It also showed that assessment could serve a number of purposes ($\rho = .211$; $p = .043$).

In the interviews, to explore the relationship between the participants' conceptualisation of EGA and their teaching experience, the participants were ranked into three groups. The first group included 11 teachers whose teaching experience ranged from a year to 8 years (less experience); the second group encompassed 10 teachers whose teaching experience ranged from 10 to 16 years (moderate experience) and the third group involved 6 teachers with teaching experience from 20 to 30 years (advanced experience).

The data revealed that teachers with minor experience in teaching believe that grammar assessment should be done explicitly and target the recognition of grammatical rules. They feel it is safer to mainly assess grammar recognition to maintain their productivity levels while limiting the possibility of a conflict with the student. For example, one teacher said the following:

Students argue over half a point and quarter of a point, and it is exhausting to keep justifying why you deduct this point from their written answers; but with multiple-choice and error analysis, you either know the answer or not, no room for speculation.

Written answers here presumably refers to open responses, e.g. sentence completion tasks, where the learner has to recall and produce the correct form.

Teachers with minor experience were also in favour of writing exams themselves if they wanted to be the course teachers and use previous exams to help them with the exam writing

process. The group with moderately experienced participants were found to have a more in-depth understanding of what to assess in English grammar. They believed that assessing both the receptive and the production knowledge of grammatical points should be the focus of the exams. Moreover, they embraced the idea of having an assessment committee involved in the assessing process. Participants with the most experience surprisingly agreed with the first group focusing only on the receptive/recognition knowledge of the grammatical points rather than deepening the content of the grammar assessments.

5.2.2.5 Educational level.

Teachers with higher educational qualifications agreed somewhat less than younger teachers that EGA was difficult to do well although this relationship was not found to be quite significant ($\rho = -.199$; $p = .056$), possibly because the general level of education does not necessarily involve any extra expertise in assessment. Educational level has no effect except higher-educated teachers believed that peer assessment was less valuable than less educated ones ($\rho = -.225$; $p = .029$). This implies that they value their own assessment ability over that of their peers to a greater degree.

In the interviews, educational level (qualifications) of the EFL teachers was found to also have a potential effect on their beliefs about EGA. It should be noted that all the participating teachers either had postgraduate, masters or PhD degrees (Table 9). Teachers with a PhD degree agreed that English grammar should be assessed in an integrative way. They also believed that the production of grammatical rules should be the priority of their assessment. They were found to be more inclined to accept the idea of having an assessment committee to construct and develop the exams, as long as the teachers had the final say. In addition, it can be said that the majority of PhD holders prefer to use previous exams as reliable sources to

find items for their exams. It can be assumed that PhD participating teachers are more receptive to contemporary grammar assessment and more willing to share and cooperate in comparison to MA holders. Participants with MA degrees appeared to be more conservative and tended to lean more towards assessing grammar explicitly, focusing on recognition items. These teachers preferred to write their own exams and turned down the idea of having an assessment committee in charge of writing their exams. They were also seen to rely on textbooks to find the best material for their exams, as they considered this approach safer and more traditional.

In general, one can cautiously assume that holding a PhD can influence the way teachers perceive grammar as an aspect of the language that needs to be incorporated into other skills' assessments rather than assessing it on their own.

5.2.2.6 Training received in assessment.

In the data, training in assessment showed little effect, but I necessarily had to use broad categories in the questionnaire and could not delve into the nature or quality of the training. A simple comparison between those who claimed to have had some sort of language assessment training with those who did not reveal just one significant difference. Subjective assessment of grammar by rating it in essays or speech was considered more suitable by trained teachers (mean 3.19/ 80%) than untrained ones (mean 2.61/ 65.25%) ($z = -2.316$; $p = .021$). This is consistent with them having perhaps been trained to perform such assessment reliably.

A more detailed comparison, taking into account the three different training categories that I distinguished, showed in a few instances that the MA/PhD and professional training had a stronger effect than the UG training. On the usefulness of essay writing, which is associated

with the subjective approach to assessment, those with UG training responded similarly to those with no training, stating essay-based EGA was less useful than those with the MA/PhD or professional training thought (Kruskal-Wallis Std J-T statistic = 2.503; $p = .012$). Again, those with UG training resembled those with no training in seeing EGA as pressurising the teacher more than those with MA/PhD or professional assessment training thought (Kruskal-Wallis Std J-T statistic = -2.134; $p = .033$).

5.2.2.7 Length of experience of English grammar teaching.

If beliefs are not affected by training, they may be more affected by experience if that experience is reflected upon and constitutes a source of teacher learning. Experience of grammar teaching brought with it only two significant effects: a positive relationship with the belief that EGA has multiple purposes ($\rho = .227$; $p = .029$) and that the textbook is a useful source of assessment tasks ($\rho = .251$; $p = .015$).

5.2.2.8 Length of experience of grammar assessment.

This variable correlated strongly with the preceding one ($\rho = .782$; $p < .001$) but produced a few more correlations with beliefs, as may have been expected from its closer logical connection with EGA.

The correlation was stronger with the belief that EGA serves a number of purposes ($\rho = .295$; $p = .004$). Clearly, the discovery of this comes with experience. There was also a positive correlation of assessment experience with belief in the value of paper and pencil assessment ($\rho = .222$; $p = .032$).

Experience of assessment also made possible recognising the value of textbooks ($\rho = .273$; $p = .008$) and previous papers ($\rho = .223$; $p = .030$) as sources of assessment items/task. The correlation with belief in collaborative construction of assessment papers was also very close to significant ($\rho = .202$; $p = .051$).

5.2.2.8 Own experiences as a language learner.

From the interview data, I also found evidence of another factor that affected general beliefs, which was not covered in the questionnaire: teachers' own experiences as language learners in the past.

The data related to the EFL teachers' learning experiences and their beliefs about EGA indicated a probable effect of the former on the latter, that is teachers' English learning experiences in schools and in their undergraduate English programmes had considerable impact on their beliefs about EGA. The majority of the teachers (68.75%) stated that when they were learners English grammar was taught and assessed explicitly in the traditional way (quizzes and exams), which they found to be beneficial. For example, one teacher said:

Yes, grammar was taught and assessed in the traditional way. We were like given the rules and how it is going to be formed and, then, we did some exercises in the texts and, afterwards, we sat for exams. And yes, I think for me that was useful.

Another explicitly recognised the impact of those experiences on his/her current teaching and assessing:

The assessment was based on the quizzes, midterms exams. I believe it influenced my practice as a language teacher; yes, to some extent, I see that the most accurate way to judge whether the students are able to use that certain grammatical rule in context is

through the written exam. So, what I've learned from my experience as a language learner is reflected in my teaching for that course specifically.

The remaining teachers (31.25%) could not relate their English learning experiences to how they thought grammar should be assessed because they had learned English grammar by immersing in speaking and writing contexts, which they encountered either because they spent a considerable amount of time during their schooling in an English-speaking country or because of self-learning strategies. This teacher for example relied on an approach like that recommended by Krashen (1988) - involving extensive comprehensible input:

When it comes to English learning, I went to public schools. I've never been to an English-speaking country, yet I managed to learn the language myself. My school teachers were not good enough; I used to detect mistakes and correct them. It's because I love the language, and I have a passion for it. I used to listen a lot to radio to music. This is how I acquired the grammar in an integrated manner; see, from reading, from listening; it wasn't that I would take the grammar book and learn. This is not how we do it in isolation, this is my belief.

Another had used that along with study in an English speaking country:

Actually, I was never taught or assessed in grammar explicitly only in college but, before that, in school, I used to learn grammar implicitly through reading through listening. I studied 3 years in the states, and I've never been taught grammar explicitly.

Curiously neither of those teachers seemed able to apply their experience as a basis for any belief about how EGA should be conducted. Possibly they felt that their approach to learning,

and the assessment that would suit it (e.g. communicative assessment of reading, writing, speaking and listening comprehension), were just too radically different from the methods habitually used in grammar courses in the KSA.

Overall, it can be seen that some beliefs were influenced by a range of interconnected factors, usually involving the passage of time (age and various kinds of experience), notably the beliefs that EGA served many purposes and pressurised teachers to keep up with the syllabus. Others were related just to one or two factors, such as the value of multiple-choice items and assessing English grammar in an integrative way or explicitly.

5.2.3 Further exploration of the questionnaire data: Beliefs associations/connections with each other.

As I have stated earlier, the questionnaire was not designed to contain multiple items targeting the same idea/construct. Rather, each item targeted what seemed to be a distinct background feature or belief of the participants; albeit, I did organise those in the questionnaire into logical groups such as different assessment purposes, type of task format and so on. Even in questionnaires such as this, however, it is common in applied linguistics to see respondents do not respond differently to every item (Hatch & Lazaraton, 1991, p. 496). This means they see connections between certain issues in different items and respond to them in a similar way. Thus, in the present study, I had expected to find moderate correlations between the responses to some belief items and others. Often researchers find a large number of questionnaire items to have reduced to a few sets of items, each with a broadly distinct theme and including items whose responses intercorrelate well with each other within the set, but not so much with those for items in other sets (Hatch & Lazaraton, 1991, ch 17).

To further explore the data following this line of inquiry, I used principle components' factor analysis to identify distinct subsets of intercorrelating items in my 46 belief items. On inspection of the scree plot (Figure 39), it was apparent that three underlying factors stood out from the rest in the data. Together, they accounted for 39% of the variance in the data, suggesting it is not optimally suited to the reduction in this way.

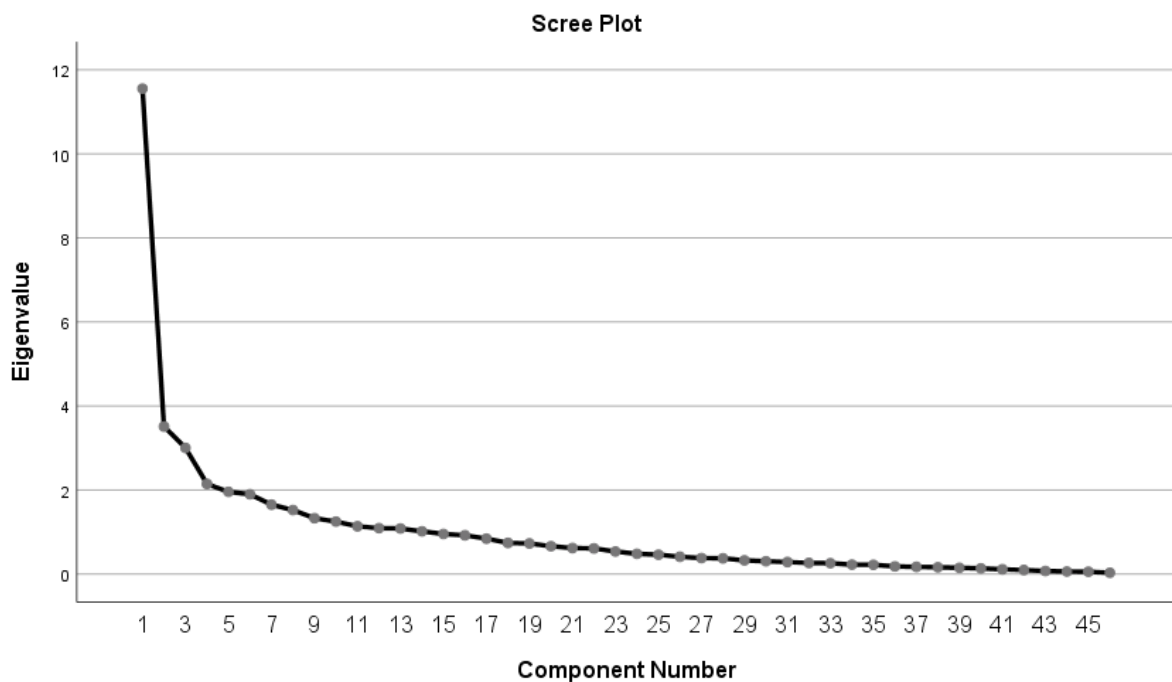


Figure 39. Screen plot of the unrotated factors underlying responses to the 46 questionnaire items.

However, it proved difficult to interpret a larger number of factors, such as the 14 with eigenvalues greater than 1. The three were, therefore, selected; the rotated loadings of the items on the three are presented in Table 25.

Table 25.
Factor Analysis Loadings of 46 Beliefs about EGA

Item	Statements	Underlying factor		
		1	2	3
11	Purpose: provide feedback to the students on their strengths and weaknesses in English grammar as they learn	.732		
6	Purpose: determine students' mastery of what they have been taught in an English grammar course	.689		
8	Purpose: provide information about how well each student is progressing in English grammar	.684		
24	English grammar assessment should systematically target the different structural/formal features of English (e.g. the articles, how 'do' is used in questions and negatives, relative clause formation)	.646		
38	Useful EGA task: error recognition/grammaticality judgment tasks	.631		
9	Purpose: inform teaching by showing the students' strengths and weaknesses in English grammar	.619		
25	English grammar assessment should systematically target student knowledge of how different functions/meanings are expressed through English grammar (e.g. how an event in future time can be expressed, or how to make polite requests with 'Could you', Or 'May I?')	.586		
7	Purpose: motivate the students to learn English grammar	.581		
22	English grammar should be assessed frequently during the course	.579		

Item	Statements	Underlying factor		
		1	2	3
26	English grammar assessment should systematically target student knowledge of common grammatical terms such as a verb, object and dependent clause	.578	.335	
3	EGA can serve a number of purposes	.567		
13	Purpose: indicate learners' ability in learning English grammar for department chairpersons or external reviewers	.556	.357	
1	EGA is important	.543		
43	English grammar assessment tasks are best prepared collaboratively	.518	.348	
10	Purpose: help place students into groups for English grammar instruction suited to their ability	.499	.363	
34	Useful EGA task: multiple-choice sentence completion	.488	.282	
27	English grammar assessment should target student explicit knowledge of grammatical rules (e.g. '-s has to be added to a verb in the simple present when the subject is third-person singular', 'days of the week take 'on' while months and years take 'in')	.487	.333	
37	Useful EGA task: editing/error correction tasks	.478		.286
12	Purpose: indicate the teacher's ability in teaching English grammar for department chairpersons or external reviewers	.447	.438	.289
15	Computer technology helps in assessing students' English grammatical abilities	.415	.319	
14	Paper and pencil assessment provides valid evidence of students' learning of English grammar	.414		

Item	Statements	Underlying factor		
		1	2	3
16	English grammar assessment use means that it should reflect real-life language use (not disconnected sentences or words)	.405		
21	English grammar errors are only important when they get in the way of successful communication of the message being conveyed		.681	
41	Assessment of a student's English grammar performance by their peers is useful		.637	
44	Ready-made English grammar exercises/tests found on the Internet are a good source for grammar assessment tasks		.601	
40	English grammar is best assessed by expert professional testers /examiners rather than a class teacher		.600	.281
42	Self-assessment by students of their own English grammar performance is useful		.583	
46	Using English grammar assessment items from previous years is a good source to construct grammar assessments		.547	
23	English grammar should be assessed at the end of the course		.530	
33	Useful EGA task: matching items	.360	.530	
17	English grammar assessment should target specific elements of English grammar in separate items (discrete-point aspects)		.529	
45	English grammar exercises from published textbooks are a useful source for constructing grammar assessment tasks	.284	.519	
39	The best grammar assessment items are the ones developed by the course instructor	.412	.485	

Item	Statements	Underlying factor		
		1	2	3
19	Objective assessment (e.g. through scores from sets of multiple-choice items or close gap-filling items etc.) is a good method to assess English grammar	.342	.480	
5	EGA pressurises teachers to complete the syllabus or textbook		.467	
4	EGA is irrelevant to language learning		.435	
2	EGA is difficult to perform well in		.431	
31	Useful EGA task: filling closed gaps in text	.313	.323	.293
29	Useful EGA task: speaking in an oral interview			.813
30	Useful EGA task: speaking in an oral presentation			.804
36	Useful EGA task: essay writing			.745
20	Subjective assessment (e.g. by rating overall grammar quality in a short essay or oral presentation) is a good method to assess English grammar		.401	.597
35	Useful EGA task: open-response sentence completion			.578
32	Useful EGA task: sentence transformation or production items	.424		.529
28	English grammar assessment should primarily be concerned with the student's ability to understand and use English grammar in communication effectively and spontaneously (as a native speaker does), not their conscious knowledge about the language		.349	.366

Item	Statements	Underlying factor		
		1	2	3
18	English grammar is best assessed in an integrative fashion along with other aspects of English (e.g. through speaking or writing tasks), rather than as a separate skill			.280

13 items stand out as uniquely related to factor 1, with 9 more having strong relationships with it (stronger than other factors at any rate). The key theme in many of the items strongly supporting factor 1 was found to be the purposes of EGA. Indeed, all the eight items 6–13 that relate to specific purposes plus item 3 about EGA having many purposes all relate to each other and factor 1. This suggests that teachers who agree with the appropriacy of any of the purposes tend to agree with all the other purposes as well, whether more pedagogical or administrative. However, the factor does not simply reinforce a grouping of items that were constructed in the design of the questionnaire, which had a separate section for assessment purpose items. It shows that teachers further associate multiple purposes with the need for frequent assessment, completed in either paper and pencil or computerised mode, since different purposes often necessitate different methods of testing, which is explicable. Perhaps, for the same reason, agreement with many purposes correlated with the agreement with a focus on assessing a range of different types of knowledge, both grammatical form and function/meaning, and knowledge of terminology and explicit rules, as well as favouring core EGA, that is, EGA tasks such as error recognition and correction and multiple-choice sentence completion. Interestingly, collaborative work to create assessment instruments was also found to be related to factor 1. Thus, a variety of aspects of EGA are seen as key ones to fulfil its purposes.

In contrast, the two other factors identified through further exploration and analysis of the data target aspects of EGA, which the teachers apparently do not see as specially related to its purposes but rather centring more on a distinction between discrete points of grammar assessed objectively versus integrative aspects judged subjectively. Factor 2 is unambiguously supported by 10 items plus substantially by 5 others. The shared theme seems to be form-based EGA supported by discrete point tasks and the objective mode of scoring, which is seen as associated with assessment at the end of the course by an expert tester/examiner or the course teacher using textbooks, or the Internet or previous papers as the source. Irrelevance to language learning and pressure on the teacher to complete the syllabus are also seen as associated with the dimension. Also, less expectedly, peer and self-assessment were found to be correlated with this factor.

Factor 3 is supported by 5 items and substantially by three others. In contrast with factor 2, the shared theme was integrative EGA, focusing on grammar as used in spontaneous communication, including the tasks that fit with that approach, such as, oral and essay tasks, and the subjective mode of scoring that is normally used with such tasks. This is, therefore, a coherent set of items contrasting with factor 2.

Thus, overall, the factor analysis suggests that at a deeper level, the participants may be conceptualising EGA as possessing three genuinely distinct belief areas, each with its own constellation of associated specific beliefs as represented in the questionnaire items. One focuses on the purposes and other aspects of EGA, such as what is tested and how which are seen to be associated with achieving those purposes. The other two concern, respectively, the discrete-point and integrative distinction and different tasks, such as teacher involvement, scoring and so on.

5.2.4 Summary of Section 5.2.

Section 5.2 presented the findings of EFL teachers' beliefs about EGA and the factors that may have shaped these beliefs. Regarding teachers' beliefs about EGA, the findings from the interviews and the questionnaire revealed that the teachers expressed similar opinions about the various aspects related to EGA:

- EGA was important and served a number of purposes
- EGA was better conducted in an integrative way through writing and oral presentations
- Various item formats were endorsed; for example, error analysis, MCQs, sentence formation etc.
- Textbooks and previous exam were the best sources teachers relied on when writing their own exams
- Constructing EGA collaboratively was favoured most by the participating teachers followed by doing it individually.

Moreover, when exploring the factors that shaped the above beliefs about EGA, it seemed apparent that educational experience, as well as educational qualification and experience as language-learners, were the most influential factors on teachers' beliefs about EGA.

The following section presents teachers' actual practices of EGA and the factors affecting these practices.

5.3 EFL Teachers' Practices of EGA and Factors that Influence these Practices

This section answers the second research question:

- a) How do EFL teachers actually assess grammar in their teaching environments?
- b) What are the factors which influenced their practices other than their beliefs?

The results obtained from the data of the semi-structured interviews (N = 32), retrospective thinking (N = 20) and document analysis (N = 28) are discussed below and related to one another when and where convenient. Moreover, examples and quotations from the actual data are presented to exemplify and support the reported findings in this section.

5.3.1 How do EFL teachers actually assess grammar in their teaching environments?

In this section, EFL teachers' grammar assessment practices are examined in-depth. The findings are discussed within the subsection according to the various themes discovered and related to the teachers' assessment practices.

5.3.1.1 Type of EGA.

The interview data revealed that the most dominant English grammar assessment method used by the participating teachers was classroom-based assessment in the form of written exams, a phrase that appeared 19 times in teachers' quotes when talking about their current practices. A couple of teachers associated 'written exams' with the word 'traditional' and 'formal'. For example, one of the teachers stated, 'So the current practice is traditional through written exams.' Another teacher explicitly said that she assesses grammar in a summative manner: 'We don't have continuous assessment; we don't have formative assessment. We only assess students through summative means.'

In their interviews, 29 teachers specified such written exams or traditional assessments are carried out through in-terms/midterms, quizzes and finals (as mentioned in Chapter 4). This indicates a holistic practice amongst the participants from different educational contexts. The following quotes illustrate what the teachers had to say about how they assess English grammar and the weighting in the context:

Well, we have to have two in-terms each 25% quizzes 10%, that's 60% classwork.

And we have one final unified exam 40%.

Recall that a quiz in the KSA is similar to a small version of a midterm exam. It is not a short low stakes test given orally by the class teacher, like in the UK. Another teacher's account was very similar:

This is the criteria of assessment. We give students two midterms two quizzes and, then, an assignment and final, and that's it.

The following teacher expanded on where the topics of the items in these assessments came from.

It is assessed in the traditional way; we use a textbook and, in that book, grammar is divided into many topics: tense, articles, positions etc ... So, the current practice is traditional through written exams and measures the ability of students on those topics. So, we give in-terms quizzes and final.

The textbook of course has been decided for the teacher as part of the course specification, so is not an aspect that he/she can easily change.

Some teachers however implied they do have some choice with respect to the assessment:

So, we teach books and give midterms, quizzes and the final. I don't ask for presentations or writing although they're good. I give them quizzes, three or four quizzes based on the time, two midterms, maybe three, and I'll choose the best two and one final. I give them a quiz at the end of each chapter, two chapters. It is a good chance for them to memorise everything.

This quote shows that the teacher had control of the number of quizzes and midterms used in the assessment procedures, and also had an option to choose an oral presentation or written essay as assessment types if desired, in place of one of the standard components.

The participating teachers' statements about their current practices with respect to the assessment components and their weight are confirmed by the assessment samples collected, which were all in the form of written exams, namely midterms, in-terms, quizzes or finals (Figure 40).

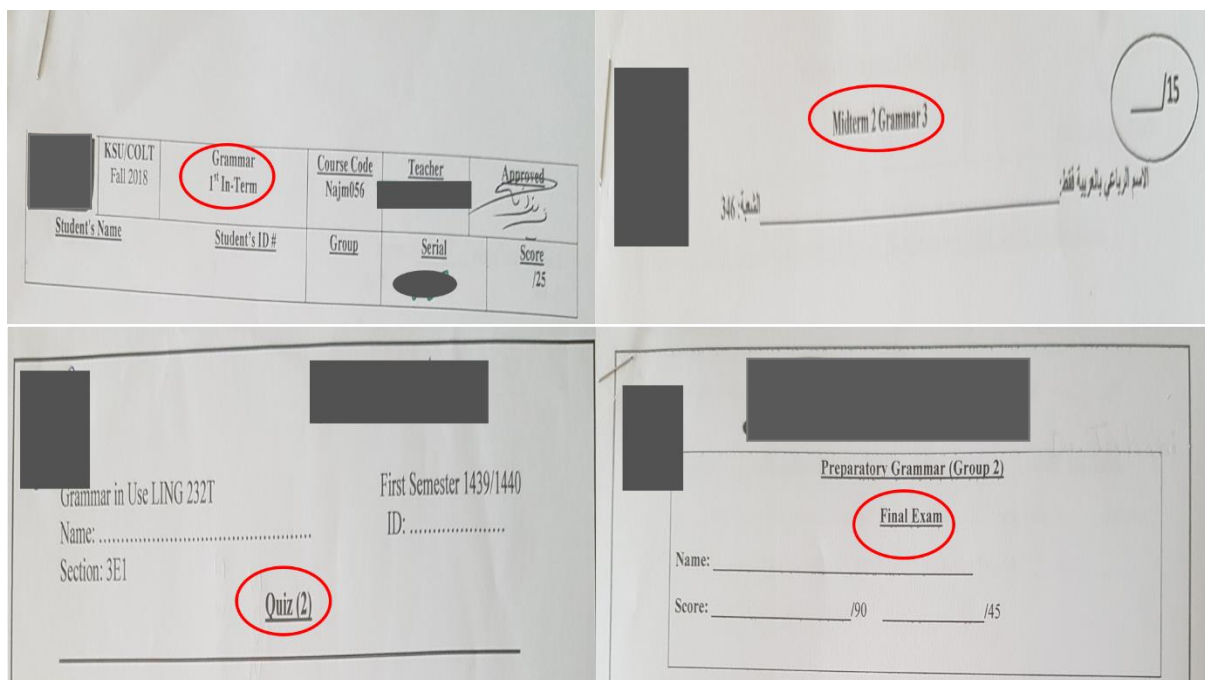


Figure 40. Written exams conducted by the EFL teachers during grammar courses.

The teachers explained they mostly used formal classroom-based assessments such as finals, midterm exams and quizzes either due to a large number of students in the class or because the educational system policy at their institution dictated this type of assessment: summative through written exams. The statements below reflect the situation of EGA, wherein the teacher participants state the division of assessment into midterm, final and quizzes are obligatory:

Through written exams, as I said before, it is the department regulation to give students two exams during the semester and one final unified exam at the end of the term.

In fact, this sort of pattern is common across courses within the department and across different departments and even higher educational facilities in the KSA. Another teacher described their practice as departing slightly from the norm:

According to the course description, okay, we supposed to be having like two in-terms; the first in-term and the second in-term each is 25% and one quiz that is 10 marks. But what I do is like I have multiple quizzes throughout the semester. So, whenever I finish a chapter, there will be a quiz and, sometimes, two quizzes if the chapter is too long, and so on. I consider that as some kind of continuous assessment and some kind of problem detection. I want to see to what extent the students grasp the grammatical rules and the application of these rules.

That teacher did not however make clear how the marks from so many quizzes were reduced to the 10 marks prescribed, or whether some quizzes were not treated as part of the official assessment. Another teacher, unusually, claimed to have control even of the % weighting of components.

Having exams is a must; I could use other means of assessment but exclude exams all together is not possible. So, I could allocate 30% for exams, maybe 15 first in-term and .15 second, and have 30% on other forms of assessment quizzes assignments or other forms of formative assessment; but, as I said, I choose not to due to the large numbers of students we have.

Indeed, amongst the participating teachers, three teachers expressed their discretion over the percentage weight of some subcomponents, the number of quizzes and the method of assessment (not necessarily all written). For example, one said:

Okay, so we have 40 marks for the midterms and 60 marks for final. So these 40 marks, I have 15 marks for the mid and 5 marks for the first quiz and five marks for the second quiz; that makes 25 marks. Then I have five marks for the handouts I told you about; then I have five marks were also each grammar book, you have a list of the irregular verbs, the past simple past participle, and, then, I have one oral activity; that's also five marks.

Another said how she included oral production under the guise of participation Thus some teachers at least seem to have found ways of 'playing the system' in such a way that they could alter some aspects of the assessment types quite markedly:

Usually I give two midterms and two quizzes and, of course, there is the final. I also dedicate 10 points to participation; so, this is oral production.

5.3.1.2 Relative weight of EGA.

An especially recurrent issue that kept arising when the teachers' spoke about EGA was the relationship of the % weighting of each component with the number of marks for that

assessment component. Nine teachers felt it is essential to explain how marks are allocated in their written exams, which should add up to 100%. Of course, each teacher talked about mark division in relation to their educational context, which seemed to largely or totally dictate the weightage: institutions A, B, C, and D (see Chapter 4). For example, participating teachers from A and B stated that 60% is allocated to assessment through midterms and quizzes while 40% is for the final: 'We have the two in-terms each 25, short quiz 10%; the total is 60 points for the classwork and, then, 40 points for the final exam.' Teachers from C said that midterms are worth 40% while 60% for the final. One of the teachers stated, 'Okay, so we have 40 marks for the midterms and 60 marks for final. So, these 40 marks, I have 15 marks for the mid ...' Lastly, teachers from D specified they have to assign 20% for two midterms, 15% for homework and 45% for the final: 'We have 20 for two midterms 15 homework and 45 for the final.'

These statements are all in accordance with the exam samples I obtained. It is the convention in these educational facilities to represent the weighting directly in the maximum score for each assessment component, out of which marks are awarded. For example, if a component is weighted 40%, then it is actually scored out of 40 rather than 100 and, then, weighed as 40% in the calculation of the final overall assessment mark along with the other components. Thus, exam papers from institution A midterms are out of 25, from C midterms out of 15 while from D final exams out of 45 (Figure 41). For an exam weighted 25% then teachers have to devise items to fit this: e.g. 5 items each scored out of 5, or 25 each scored 1, or some other combination.

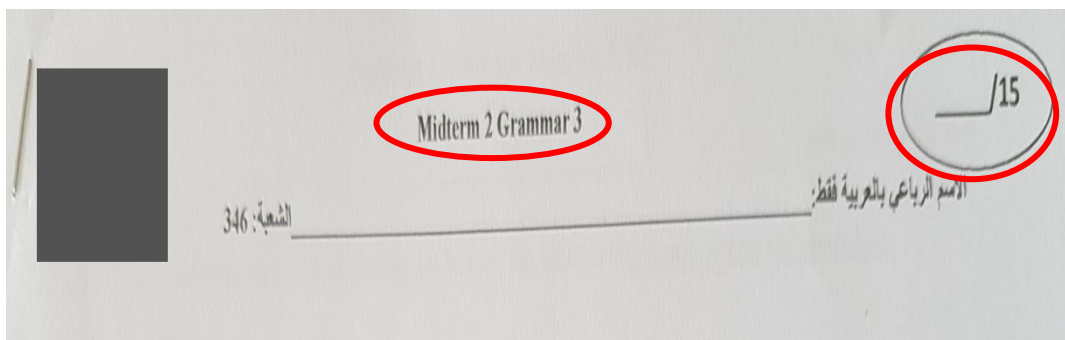
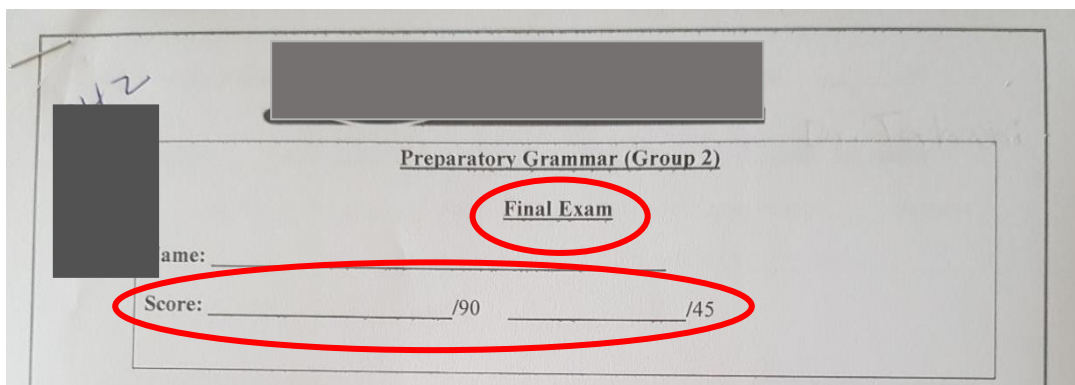
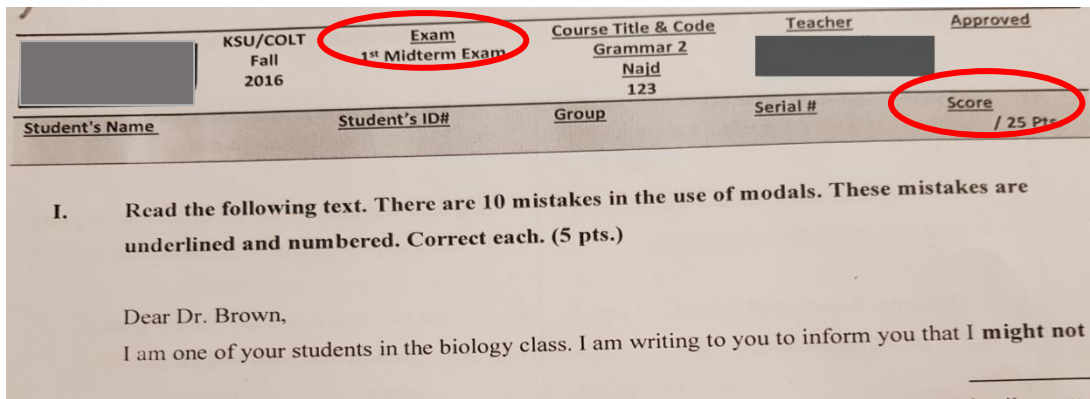


Figure 41. Section of the exam paper showing the name of the assessment component and assigned a maximum score.

All in all, the data from the interviews and document analysis show that what the teachers have said about their current practice of EGA is aligned with the document samples (written exams) they submitted.

5.3.1.3 EGA item formats.

Another important topic that the teachers shed some light on during their interviews was the kind of questions usually used in their written exams, which might be a combination of different types of items. For example, one teacher stated that she used both multiple-choice and open-response items:

Actually, through different types of questions: multiple-choice questions, fill in the blanks questions and open-ended questions with short answers.'

Well, for me, I use different types of questions: either filling in the blanks, writing sentences, rephrases, organise, true and false. Most of them are things that are used in the book because I have to follow the style of the book. I assess based on the way I teach, right?

Interestingly the teacher above indicates that following the format used in exercises in the textbook is required to accomplish a balance between the teaching method and the assessment techniques. This was concurred with another teacher who said, 'the exam questions similar to what we do in classes: exercises of filling in the blanks, underlining mistakes, rearranging words to make a sentence. That sort of thing.'

It could be noted here that although those two teachers say they use items of the types found in the textbook for consistency, there is an implication that that will help the students by staying with what they are familiar with. What they notably do not say is that they use actual items from the textbook, which would in fact help the students even more.

In contrast to the two teachers above who said that they design their exam questions to imitate those in their textbook, one teacher stated that she composes her exam questions without relying on the course textbook:

I do everything from scratch. Every time I write a midterm or a quiz, I try to start something new. I don't repeat the exams or the quizzes; I don't follow the same typical sentences and forms in the book.

The above reports about exam questions or item format seem to concur with the teachers' exam samples collected in this study. Within the 28 exam samples, 120 questions were identified, the majority of which was open-ended questions (52.5%). The second most popular exam question was multiple-choice questions (MCQs = 23.3%) followed by error analysis questions (18.3%). The least used question formats were matching and true/false questions (3.3% and 2.5 % respectively). Table 26 shows the format questions found in their written exams, of course, a detailed discussion of these formats is presented in Section 5.3.1.5 below where the teachers have spoken in their retrospections about the exam writing process (also see appendix M for a complete account of exam questions).

Table 26.
EGA Item Formats from the Teachers' Exam Samples

Total No. of Qs	Error analysis		Open- ended Qs	MCQs	Matching	T/ F
	Recognition	Correction				
120	5	17	63	28	4	3
%	4.2	14.2	52.5	23.3	3.3	2.5

Overall, what the teachers said about their current practices in their interviews aligned with the written exam papers submitted by them. This is a welcome indication of the validity of the data on EGA practices.

5.3.1.4 Teachers' Role.

As illustrated in figures 20 and 23 in the Methodology, Chapter 4, the teachers' role in their current practices of EGA was one of the themes identified in the data analysis. All the participating teachers confirmed and emphasised they were the ones responsible for writing their exams. Notably, 90% of the respondent teachers used the sentence 'I write my own exams'. However, there was variation in how different teachers conceptualised the practice of 'writing one's exam'. For instance, one teacher stated she wrote her own exams by using questions from previous exams she had written herself in past years.

Usually, my tests are accumulation of long years, where I update them, but since we're doing this type of teaching, then grammar does not change and updating them is quite easy. We also share tests. So, some of the exercises, they are not mine, but I thought they were interesting and somewhat accurate; so, I took them from other people.

This teacher throws interesting light on the sources of her items. In particular the participant shows evidence of at least some teachers acting as a community of practice with respect to assessment, in sharing their items, although we cannot tell from this how widespread that practice is, or whether it was only within the limits of one institution. The participant also shows by implication that the items used are chosen for interest and accuracy and used repeatedly to assess something that 'does not change'. That means that the teacher is thinking in terms of assessing some fixed body of grammatical knowledge (criterion-referenced) and not inclined to vary the assessment to suit the ability of the students that happen to take the course in a given semester.

In contrast to the above view, some teachers insisted they always write their exams from scratch and never use any previous versions. That could be due to reasons such as differing students' proficiency level and, hence, different needs on different occasions. The following responses support this view:

I tend to write my own exam. I tend to write new exams every time. I never ever take previous exams because they are different students, different abilities.

Note that, associated with that practice, this teacher is implementing the norm-referenced view of assessment, mentioned earlier, which was also present in the belief systems of some teachers.

Another teacher, however, indicated that in cases where she shares the grammar course with another teacher, writing the exam is then a result of a collaboration between the course teachers:

It's a shared role between me and the other teacher putting our inputs and creating the questions of the exam. So, I would say it's a 50-50 effort to be honest, yeah. I do write my own questions and the other teacher looks at it and gives me her insight and her input regarding whether this is a valuable question to students.

During the retrospections, 20 participants brought a sample of their written exam they had previously developed and administered. These participants talked in detail about how they wrote the exam, what items they used and for what purposes and their scheme of mark allocation; the aspects have been discussed in other sections below.

In conclusion, it seems that with respect to who designs and constructs written exams, course teachers play an essential and active role in producing the assessment tasks. This statement can be confirmed by the document samples collected in this study, where 28 participating teachers submitted their own written exams signed with their names. Also, the data from the retrospections corroborated the findings reported from both the interviews and the document analysis.

5.3.1.5 The process of writing grammar exams.

To shed light on their practices of constructing EGAs, the participating teachers were asked to explain the procedure involved in writing their exams. Both interviews and retrospective thinking sessions served as the basis for the results reported in this section. The document analysis data was reported when necessary to contextualise what the participants had said in the interviews and the retrospective reports. Within the writing process, two main concepts were identified: writing guidelines and common features. The main findings for each are given below.

Aspects that were labelled by the participants as guidelines to be considered when writing exam papers include the syllabus (e.g. stated in the course description), what had been taught, the textbook, students' level and exam practicality¹⁹.

Complying with the grammar course syllabus and classroom teaching appeared 16 times in the teachers' reports, where the teachers insisted that following the syllabus and what has been taught in class were things that kept them occupied and guided them through the writing

¹⁹ Practical exam is inexpensive, within appropriate time constraints and has a scoring system that is specific and time-efficient (Brown, 2003).

process. For example, some teachers said that they strictly follow what they teach in class and the syllabus and adhere to both when writing their exams:

I look at the syllabus and what I covered in the class before I write the sentences of the exam.

This teacher implies simple implementation of a criterion referenced approach: the exam is designed to test what was supposed to be learned and was taught. Another said:

I never ask them about something not taught. So, I will never include something that was not in the course specification or the course objectives, but I do believe in varying the level of difficulty of the questions themselves.

This teacher is more ambivalent. The participant states that the exams never include something not in the course specification or objectives, but that implies that the teacher is free to leave out some things in the specification. However, the teacher's reference to also varying difficulty implies perhaps not taking the norm-referenced route of omitting items that are assumed to be too difficult for the students. Another teacher, however, referenced the textbook rather than the syllabus or classroom teaching:

I have to go through the content of each chapter and, of course, I have to be aware of the students' level. Sometimes, I have to stick to the same exercises in the book and do similar questions in the exams.

This teacher attempts to combine a norm-referenced with a criterion-referenced approach. The book, which like the syllabus and outcomes/objectives is dictated by the course specification, presumably matches those, so following the book is criterion-referenced (although it was beyond the scope of this project to actually check how far the textbook,

syllabus/course description, and CLOs did match each other). The second sentence implies a practice very close to using the textbook items directly as exam items, and clearly is designed to help weaker students, so is norm-referenced. But the two approaches cannot really be combined. If the items testing points in the syllabus or textbook are made very easy in this way, it is unlikely that getting them right demonstrates the kind of knowledge that is described in the course specification. Rather it is testing students' rote memory of the examples in the textbook, not their knowledge of whatever bit of grammar the items were illustrating.

Indeed, it is clear that quite a number of the teachers consider student level and the level of difficulty of the question when writing the exams. This was further endorsed by other teachers' responses. For example, some teachers said:

When I build up the exam, I take into consideration all the different levels of the students and their needs.

This again is ambivalent, since the needs of weaker students, whether real life needs for handling English medium teaching of subjects in their major, or needs as specified in the course specification, probably entail them learning and being tested on material that would make the exam very difficult for them.

The following teacher mentions only difficulty, which of course is relative to the ability of the exam taker. The teacher said:

The most important point in the assessment is to take into consideration the level of the questions or the difficulties of the questions.

In fact, no question is inherently easy or hard. It entirely depends on the knowledge of the person answering whether it is easy or hard. Furthermore, the teacher does not mention referencing the items to the knowledge specified by the course specification as to be learned.

Creating an exam that can be completed within the time frame and regulated by the educational system was yet another issue that needed to be considered when writing the exams. According to the teacher participants, midterms/in-terms are meant to be completed in one hour and final exams in two. Therefore, the number of items and question difficulty were governed by the time students took to complete the assessment task. For example, one of the teachers said, 'When I write my questions, I have to take time into consideration this exam is one hour'. Another teacher stated, 'I only have four questions because the students always complain that they do not have enough time to answer'.

In summary, it seems that the golden rule for some teachers when writing the exams is to construct an exam within the regulated timeframe and never deviate from the syllabus or what they have explained in class while bearing in mind students' proficiency and exam difficulty level.

As I have mentioned earlier, within the writing process, the majority of the teachers also talked about a number of shared features that can provide clear and explicit details of their practices. Some of these features identified amongst the teachers were (1) exam type and administration modality²⁰, (2) focus of the exam, (c) number of questions in the exam, (d)

²⁰ Paper and pencil or computer-based (Downing & Haladyna, 2006)

logic behind the question order, (e) mark allocation, (f) construction time and (g) potential challenges when writing the exam.

All teacher participants reported in their interviews and later restated in the retrospective sessions that their grammar assessment mostly takes the form of written exams and not oral or by research. The phrase ‘written exam’ entails the use of paper and pencil in exams, which is the case in my study since all exam samples collected were hard copies.

In their responses about what their written exams focused on, the majority of the teachers acknowledged both recognising grammatical rules and producing these rules either at the word or sentence level to follow the course syllabus and the exercises in their books. For example, some of the teachers said,

The exam covers language recognition and production because I believe both are important, and this is the focus of their syllabus.

That is a rare example of where the teacher explicitly tells us that EGA practices are in fact driven by beliefs. The statement implies that the participant was one of the minority that expressed the belief that recognition as well as production should be targeted. Another teacher also drew attention to both modes:

When I assess grammar, I don’t only ask them to produce but also how to use a particular tense to write certain sentences. I emphasise the fact they should be able to recognise well-formed sentences. I always like to include activities where they have to identify errors, locate them and correct them.

This teacher however implies that the recognition task (identifying an error) is always accompanied by a production one (correcting the error).

Sometimes the teachers indirectly indicated whether they included recognition and/or production items. For example, in the following quote, the teacher talked about writing sentences, rephrasing etc., which are production items and true/false, which is a recognition question item:

I have different types of questions, either filling in the blanks, writing sentences, rephrases, organise, true and false. Most of them are things that are used in the book because I have to follow the style of the book.

Data from the document analysis revealed that indeed all the teacher participants combine recognition and production questions in their written exams. Amongst the 120 questions analysed from these exam papers, 80 question items targeted the production of grammatical rules while 40 were assigned to assess the awareness of grammatical rules (recognition) (see Appendix M for detailed content analysis). The propensity to focus more on production is justified in the statements below:

The focus is more on production because that is what we need to stress in the exams.

By *need* here the teacher is probably referring to institutional policy, as received through the course specification. This was made clearer by other teachers, although they tended to express some confusion at the same time:

The exam focuses more on production because it's the college criteria; I don't support that. As beginners, they will do better with MCQs/true or false or matching.

This refers explicitly to a mismatch between belief and practice, and its reason, which will be revisited later. The participating teacher indicates that in assessing EG production recall items (e.g., giving a sentence with a gap and no alternatives provided) are more supported and documented in the course specification than recognition items (e.g., MCQ in which students are given a sentence with a gap and several possible words to fill that gap). One might guess that these kind of recognition items, in the mind of the teacher, are discouraged.

We focus on production because I believe this is a formal request from the coordinator and the vice-chairperson to give students in their grammar exam more production questions and less discrete point questions simply because we want to build good basis for them on how to use grammar for other future courses. And yes, I do believe it's the right way to evaluate their grammar.

This teacher interestingly mentions *a formal request* which sounds as if it is independent of the course specification, which is the default way in which institutional policies about the course syllabus, teaching and assessment are conveyed to the course teacher. Also, this teacher shows some confusion over what production and discrete point items mean. In fact, production recall items may be either discrete point (e.g. filling a sentence gap with the correct verb form with no alternatives provided) or integrative (e.g. writing/speaking a full sentence answer to a question). Hence it is unclear if the institution is actually favouring just production recall items or specifically production recall integrative items. The document analysis however revealed relatively little of the latter.

The teacher's / institution's reason (*...how to use grammar for other future courses*) is of course valid, especially if it refers not just to future grammar courses like Grammar II but to

all the courses taken through the medium of English during the BA program, since in principle the students have to do (mostly written) assignments and assessments for all those in English. When we consider future subject (rather than language improvement) courses that they might take as English majors, such as 19th century novel or Sociolinguistics, the assessment might be expected to be by writing essays, if the UK/US model were followed, and essays are of course, in assessment jargon, extended integrative and communicative versions of production recall items.

Another teacher gave more detail of what the textbook entailed:

The questions target more production items on the word level than language awareness or recognition because these what the majority of the textbook exercise cover. I believe both are important, production and recognition; but we cover more production exercises in classes.

This again supports the emphasis on production and the opposition of recognition (language awareness). However, this teacher's explicit limitation of production to the word level contradicts the previous one's mention of *less discrete point*.

In fact, teachers' written exams reflected a strong tendency to target questions that required open written responses (i.e., production) rather than selected responses, where students had to choose or circle the answers, i.e. multiple-choice (recognition). They were also in practice more a matter of supplying TL forms/words (production) rather than choosing meanings or grammatical information about TL words that were given (recognition). In addition, they were also more discrete point than integrative.

The number of questions in each exam was yet another issue discussed by the teachers. As suggested earlier, the number is in a sense controlled by the fact that the total possible marks has to add up to the same number as the % weighting of the exam, that is, if it is weighted 25% then it has to have items scored to total maximum of 25. However, within that the teacher can vary numbers of items devoted to different topics (often referred to as *questions*), and can increase or decrease the total number of items by deciding that some are scored out of $\frac{1}{2}$ rather than one, or some out of 2... etc. Based on the exam samples collected, the average number of questions written in each exam is four to five questions. Each question contains an average of 5 items targeting the same topic the question is inquiring about (Appendix M). Teachers justify the number of questions, and the number of items within questions, by saying that they have written them to reflect in due proportion the material they intend to cover in the exam. For example:

The number of items in each question is related or representative of how much time we spend in a certain topic and how much of the syllabus is covered in the semester.

I have six questions just to measure the topics that we covered during this period of time.

These statements thus suggest the teachers have criterion referencing of achievement in mind here. The numbers of questions and items reflect the syllabus/teaching, rather than what the students find hard or easy. A norm referenced approach would take students' ability into consideration and omit items that everyone would get right (Hughes, 2010), or, in the present context, from what the teachers have said elsewhere, omit items that the students would find too difficult.

Another teacher interestingly adds that he takes into consideration to ask about is also the same topic more than once in different formats so as to obtain a more reliable estimate of student knowledge.

I have six questions to make sure the students know the answers and not just by chance. For example, this question is about adverbs of frequency, and I ask about the same thing in another question but different wording.

That teacher is indeed following a well-established principle of testing that the more items target the same thing, the more reliable its measurement becomes (e.g. Hughes, 2010).

When writing the exam questions, the teachers were also asked if there were any reasons behind putting these questions in the order presented in the exam paper. At first, almost all the teachers said, 'I don't know', but after some pondering, various reasons were provided to explain why the exam questions were put in the order that they were; 40 % of the teachers explained they had arranged the exam questions starting from the easy ones to the hard ones. The rationale behind this order – easy to hard – is explained in the statements below:

I really don't know but, mentally, I thought this is the quickest way. They can do these questions and get done with them and spend more time in the production question, which is the longest. So yes, probably the easiest to the hardest.

For that teacher receptive items are easier, which indeed is the usual assumption of experts.

Another said:

I like to start with the easy questions and then move to the hard ones. Any exam taker will be nervous but, when he sees the easy one first, he will just be more relaxed; then comes the difficult part.

This indicates teacher attention to the affective state of the examinee, by adopting the order designed to reduce anxiety and so hopefully enable the student to perform better. This is a legitimate concern in testing and is quite distinct from helping the student by making the questions easy, as some teachers implied earlier. It is however long contested among experts as to whether a certain level of anxiety is in fact helpful in producing a better performance (Alpert & Haber, 1960).

In contrast to the above view, other teachers (25%) claimed they organise their exam questions from harder to easier because students at the beginning of the exam are more alert and focused and, therefore, answer the more difficult questions first and then move on to the easier ones:

Another thing that guides my order of questions is that in grammar, I like the students to start with questions that are going to require the most effort and most processing because, at the beginning of the exam, they're alert, they're aware. By the time they reach page 5, they're probably going to be a little bit exhausted.

This teacher is therefore also concerned with helping the student to do better, again not simply by making the items easier. However, the focus here is on the cognitive rather than affective side of the student, i.e. their state of attention not emotion. The following teacher supports this view as well:

For me, I know that the productive tasks are more demanding, and I don't want them to start with the easy one; I want them to be alert and in their full capacity okay and energy when they start the exam and finish from the more demanding ones and then move to the easier ones. Okay, not vice versa. Okay, to start with the easy ones; okay, and by the time they reach the more demanding ones, they are worn out.

Amongst the 20 teacher participants, just two stated that the order of the questions in their exam paper emulated the order of the chapters covered from the book.

I think the only logical explanation here is that I follow the chapters in the book and how the exercises were organised in their textbook.

I am following the structure of the book. For example, chapter one is about adverbs of frequency, then the first question covers this, and I write items to cover the aspect taught in this chapter and so on.

Although the teachers do not mention it, this of course also might in fact help the students, since one way of memorising material is by storing it in a way that relates it to what comes before and after (Buzan, 2010).

Interestingly, one teacher stated that the only reason she arranges the questions in her exam in random order was a matter of presentation, that is the teacher tried to type the whole question in one page and not break it over two pages. This was apparent in her comment below:

The only logic is I don't want to split the question into two pages. So, I don't want the student to think that this is the last item on this question, so it's more of an organisational issue than a pedagogical or academic issue.

Even though this appears to be purely a formatting issue, in fact the teacher shows that it does have a function in possibly helping the student do better, by not failing to complete a question because it was split over two pages.

Four teachers, however, asserted they do not follow any logic or reason in arranging the exam questions in a certain order on the exam paper. For example, one of the teachers said, 'I don't think I follow any logic; they just come this way'. Another teacher-supported this view by saying, 'Actually, I don't know. I have no justification for this. Question number 1 could be 2 and vice versa'.

Overall, when it comes to explaining the order of questions in the written exams, there is a mixture of conscious and unconscious decision-making and reasoning. The majority of the teachers know why they put the questions in the order they did, although they seem not to have given it conscious attention before. Most of these reasons, either implicitly or explicitly, have the effect of helping the examinees perform up to their full ability, even the methods that appeared purely organisational or a matter of presentation.

With regard to the scoring procedures, Downing and Haladyna (2006) indicate the system of scoring exams needs to be specified and communicated to exam takers. This rule seemed to be applied by all teacher participants as they designed and wrote their exams. The exam papers collected showed that the heading of each question was represented with its quantitative value (Appendix N). Furthermore, the teachers in the retrospective sessions discussed the rationale by which each exam question and item was assigned its numerical value with a maximum of one point and a minimum of a quarter of a point.

Findings drawn from the retrospections revealed a diversity of views as to how exam questions got graded. The majority of the teachers seemed to allocate marks based on how much knowledge they believed was involved in answering the question. Twelve teachers (60%) insisted on assigning one mark to production/recall questions while recognition tasks

(e.g., MCQs or true/false) got half a point. This is normal practice in a criterion referenced test where it is recognised that some targeted kinds of knowledge are more advanced than others. Below are illustrations of these teachers' reports along with their question samples:

Question Four - Reported Speech [3 marks]

Change the following sentences into reported speech. Mind your spelling and punctuation.

1. Tom said, "I had a good time at the party last night".

2. The man at the reception desk said, "The doctor sees patients every day from 4:00 to 8:00".

3. The secretary said, "Today's meeting is in the afternoon".

In question four, I got 3 items; each is one point because the students need to think and write. This question involved production and, I, as you can see in the instructions, I asked them to mind their spelling and punctuation.

These are production recall items. What the teacher says does not however make clear whether the mark would be reduced by a fraction for spelling and the like, and if so by how much, so is not fully transparent. For instance, if a student fails to change the 1st personal pronoun (I) to the third (he), $\frac{1}{4}$ or $\frac{1}{2}$. However, it is beyond the scope of this study to investigate how the teachers actually score the exam but rather explore how they allocate marks to their questions and why.

Q.1: Circle the correct word or phrase: (3. pts)

1. When the play was over, the audience arose from (**its/ their**) seats and applauded wildly.
2. In our creative writing class, we all had to write short biographies of (**yourselves /ourselves**).
3. Your child is resting comfortably, Mrs. Smith, I'm giving him some antibiotics, so the infection (**must /should**) be cleared up by next week.
4. One of the biggest problems in the world is global warming. (**Another / The other**) problem is poverty.
5. Illiteracy is still a problem in my country, but it (**would / used to**) be much worse.
6. No one knows my secret (**other/ another**) than Rosa.

Q.2 : Correct the errors in every sentence. (4.pts)

1. Pat look tired. She should gets some rest.

2. The Smiths supposed to be here at 7:00, but I think they will to be late, as usual.

3. There are many student from differents countries in this class.

4. English has borrowed quite a few of word from another languages.

Q. 3 Complete theses sentences by using the appropriate form of a modal in the appropriate tense.(4

1. It's a hospital. You _____ smoke.
2. The phone is ringing. It _____ Mary. She promised she would call.
3. He looks tired. He _____ awake all night.
4. Let's be really quite when we go into the baby's room. The baby _____, and don't want to wake her up.

So, marks are allocated based on how simple or hard the question is. The first question, because it is choosing, half a mark is enough; but question two and three because they are production and need more effort, I give them one mark each item.

This teacher illustrates the same kind of principle of differential marking. It can be seen that the teacher identifies the two types of question items: items called *production* (questions 2

and 3) and what is called *choosing* (question 1) which is in fact recognition. In addition, the teacher seems to relate difficulty to production items and easiness to recognition ones. This association of difficulty to the concepts of production and recognition seems to be widespread among the participating teachers. In fact, research shows that it is true that production is harder than recognition (Laufer and Goldstein, 2004). Another interesting feature of this quote is that the teacher refers to *effort* where it is perhaps kind or strength of knowledge of grammatical words and structures that is really involved.

1. Fill in the blanks with right form of verbs:

a. (you/ eat) your cake yet? No. Not yet.

b. I..... (think) about you a lot recently. Are you O.K?

c. Sam took off at 4. He(waiting) since 9 a.m.

d. The nurses.....(very helpful) in the clinic last week.

e. I.....(not/ be) alone in the woods before.

f. a flight? (you ever-miss).

g. Ali..... in this office since 2009. He found another job now (work)

h. Where physics? In London. (he- study).

2. Choose the right answer:

a. Thank you. You (have been- had- are) very helpful.

b. I've (decided- been deciding) to learn French this semester.

c. How long (did you have- have you had) this book?

d. We hadn't seen the accident until we (hear- heard- had heard) the ambulances

e. Who (had-have) you (been-be-being) talking to?

3. Do as shown between brackets:

a. We had been waiting outside for a long time.
.....(change to past perfect).

b. The thief escaped We call the police
(before).....(Make one sentence)

c.? No, I hadn't been drawing before. (Form a question)

4. Read the sentence, write T for true, F for false in front the statements accordingly:

A. If I had become a doctor, I would have helped sick people.

___ I became a doctor.

___ I'd have helped the sick people but I didn't become a doctor.

The true/false is half a point because just to mark true or false doesn't deserve more than 1/2 a point. But the ones that require writing should have one because I give 1/2 on correct answer and 1/2 for spelling and style.

This teacher makes the same sort of distinction but refers to one option as to *mark* something and the other as *writing*. That again describes the recognition - production dimensions. However, there is an inconsistency in that both question 4 and 2 are recognition items, yet the teacher only mentions 4 as being scored as ½ per item. In fact, item 4 is arguably more difficult than item 2 because item 4 requires interpretation of meaning which involves comprehending the language beyond words and sentences level. The other notable feature of this teacher's practice is the declaration that where an item involves *writing* (i.e. is production) it is scored out of 1, but of that ½ is for spelling etc. rather than the actual grammatical focus of the item. That means that the teacher is not in fact valuing the production of grammatical rules itself above recognition of these grammatical rules. Furthermore, the teacher does not convey any of that scoring information to the students as it is not written on the exam for each question in the way seen above.

Five teachers, on the other hand, preferred to give the most points to easy questions so that students could score well in the exam, a policy that have been already commented on as clashing with criterion-referenced principles that seem to be more embodied in the CLOs that teachers are expected to be bound by. Indeed, it is a marking system I have not found mentioned in any book on language testing. The corresponding principle was applied to hard questions, where the teacher would give fewer points to difficult questions so that students did not lose many points and score low in their exam.

V. Each sentence below has an error. Find the errors and correct them. (4 Points)

1. I would like see you tomorrow morning if you have time.
.....
2. Students would rather to have the test a week later.
.....
3. When he was a child, Ahmed can speak Spanish.
.....
4. May you give me your pen?
.....
5. Could you borrow me your car for two hours, please?
.....
6. The paper is full of stupid mistakes. You should have looked over it before you submitted it.
.....
7. He said he is going to call back us later.
.....
8. He had to ask his teacher about a three-day extension on the paper.
.....

I allocated 4 points to this question, which means 1/2 point for each item because if it is difficult, to be honest, if it's difficult, I put partial credit for this question; if it is easy, I give one mark; I want them to pass; I don't want to see them again next semester in this class.

These items are particularly demanding as they really involve first a step of recognition to find the error and then of course production to supply the correction. Hence, aside from any consideration of criterion referencing, it flies in the face of most common sense ideas of fairness to favour weaker students above those who have learned more in this way.

Another teacher however did not refer at all to the difficulty of the items in her choice to award only ½ to some items:

I have here 20 items, each is partial mark; 10 point in total. I want them to score better marks, and 20 items out of 10 will give the students the chance to get high marks.

This teacher clearly sees more items, regardless of item type, making the test easier because students have more chance to find items they can answer. Indeed, that fact is related to the point made earlier that more items make a test more reliable.

Amongst the 20 teacher participants, two teachers stressed the importance of assigning marks equally to production and recognition questions. For example, one teacher said:

I. FILL IN THE BLANKS WITH THE RIGHT FORM OF THE VERB IN PARENTHESES. [4 MARKS]

- The sun (rise) in the east. *rises*
- After Miguel (fail) his test, he left school. *failed*
- Steve (know) many people in the dormitory. *knew*
- (Do) Firas have a nice apartment? *Does*

II. FILL IN THE BLANKS WITH THE CORRECT ANSWER. [4 MARKS]

- There any bridges across the northern rivers. *aren't*
 - are
 - is
 - aren't
 - isn't
- We can't afford to visit London. We have money. *a little*
 - little
 - a little
 - few
 - a few
- How people were at the museum? *many*
 - little
 - much
 - a little
 - many
- Golden Gate Bridge is in San Francisco. *The*
 - The
 - A
 - An
 - Some

III. WRITE THE RIGHT MODAL AUXILIARY IN THE BLANKS. [4 MARKS]

- you please drive me home? (making requests – formal) *Could/Would*
- I borrow your car? (requesting permission - formal) *May/Could*
- Most people buy a car. (expressing ability) *can*
- She go to the gym regularly. (giving advice) *ought to/should*

My intention and my belief is that I should allocate one mark to the part of the skill that I aim to test. So, the definite and indefinite articles, I think it deserves one mark. I think

I allocate the marks based on the effort I made in class. That's the idea; production or recognition are both equally important and weighed similarly if taught intensively.

This is an interesting practice (and belief) since it does not refer either to the nature of the knowledge being learned, nor to the ability of the students, but to the time and effort which the teacher put into teaching the relevant knowledge. However, this practice would potentially vary semester by semester dependent not on the cohort of students but on the teacher of the course and how much effort they chose to put into teaching each aspect of the course. Hence it does not measure what a person knows or can do in a universally interpretable way.

Only one teacher expressed no concern regarding assigning marks to exam questions because she just needed to use the total exam score and cover the required content: 'I really did not think about that. I guess I have the 20, and I tried to make it work and cover the content'. The implication is that she simply writes as many items as there are marks assigned to the exam based on its weighting, with 1 mark per item.

Submitted copies of the teachers' written exams, as shown from the samples above, revealed convergence between what the teachers recalled during the retrospection tasks about the scoring procedures and the execution of these procedures on the exam questions.

Any exam writing requires some sense of timing of writing the parts (Brown, 2003). The teacher participants indicated the time required to write the exams was a minor issue because most teachers wrote their exams in one session. For example, one teacher said, 'I wrote the exam in one set'. Another teacher corroborated this view by saying,

I wrote this exam in one session because as a PhD student, I don't have time. What I did is that I looked at the chapters, and I know they have one hour. So, I write what sort of question I would include then – check the textbook or Google for items to construct the questions.

This teacher refers to various sources used (textbook and the internet), which we will review separately below. Another teacher stated that she took one hour to write the exam, but did it through several days:

The whole exam took me one hour, but I did not do it all in one session; I broke it down into several minutes simply because I am busy. Usually, I don't write the exam in one sitting; I need to think about the sentences. So, for this exam, I wrote the first question, then stopped and thought about the second question and, then, I wrote it. So, I don't write the exam in one sitting, but I revise it in one sitting.

When it comes to writing the exams, much attention and detail were paid to exam fairness, which was described by the teacher participants as a challenge. The reported responses from the interviews and the retrospections were classified into two main challenges, item appearance and student level.

According to Downing and Haladyna (2006), one way in which exam items/questions are unfair if they cause exam takers to become 'alienated, angry, distracted, fearful, offended or otherwise upset' (p 362). Amongst all these elements, the teachers in this study seemed to be mostly concerned about the questions' clarity and explicitness, which may eliminate factors such as distraction, fear and being upset. The following statements reflect this view of exam fairness through item clarity:

I don't like ambiguous questions. It is not fair for the students to be puzzled by what is required of the question. So now, most of my questions are direct, and I've been criticised for that.

Sometimes, I write a question, and I find the majority of the students did not get it right. So, I discovered that this question needs more explanation to the students, and I try to clarify what's required in class because it won't be fair for the students to mess up in the exam over ambiguous questions.

Features such as those mentioned here (question clarity and fairness) are of course standard requirements listed in books on language testing (e.g. Hughes, 2010). However, the second teacher reported of assessment practice within the classroom whether teaching session or exam administration session. Unfortunately, how assessment is brought up during regular classes or dealt with during exam administration was beyond the scope of the study. What the second teacher meant, and I reflect on that based on my personal experience, is that during the exam session if students starting asking about one particular question then it is clear that this question is problematic, thus the teacher would explain it the class (during the exam administration) so all the students would understand what is required by this question.

Another challenge brought up by the teachers is writing exam questions that can reflect and suit the students' level, a concern of our teachers that have been mentioned several times already. This particular challenge resulted from the variation amongst students' language proficiency level in the different course sections. The following statements demonstrate the teachers' views:

Major challenge, I have variation among the students in my class in different classes. Some students have studied abroad and feel this course is silly; others are still learning the first steps. So, to maintain this variety or these differences, okay, this is very difficult. Of course, you have the objectives, you can set the standards. At the end, the questions will be either very, very difficult for those intermediate learners or very easy for those advanced learners. They [course coordinator and chairperson] always tell us, okay, you can have one difficult question for those good students. So, you can discriminate between excellent students from B students or C.

This teacher recognises the problem of a mixed ability class, but notably does not take the view that have been adopted by some teachers above, that the exams should be dumbed down so as to suit the ability of the weakest students. The teacher here endorses abiding by the standards and objectives (i.e. the CLOs in the course specification), although that will inevitably mean that the exam is hard for some and easy for others. As mentioned before, that is the criterion referenced view of the exams, which is expected to assess certain fixed essential elements and not vary in its nature depending on whether the examinee happened to be a very good student or bad one. The teacher concedes to norm referencing only in including one question to challenge the really good students.

You have to differentiate between the good, the best and the weak ones. I have to give questions suitable to each level because the students, they will not answer the questions the same. So, when I give one or two challenging questions, this gives the teacher an idea to distinguish outstanding students from others.

This represents a more thoroughly norm referenced view, since it suggests suiting the items to the levels of the students and not to the objectives or topics specified for the course. In a criterion referenced approach it would not matter if the exam separates the better from the worse students if they all have the same knowledge of the core course content that is specified to be learned (CLOs).

It can be seen from the quotes above that teachers tend to consider exam fairness as a challenge they need to address when writing their exams. The teachers also suggested that writing clear instructions and reflecting students' level are major steps to accomplish exam fairness.

This section has outlined the findings on a number of key aspects of designing the written exams. In general, the findings reported above reflect both common trends and differences among teachers found in the exam writing process.

5.3.1.6 Purposes.

Identifying the purposes of assessment is crucial to teachers when planning and designing their assessment procedures (Restrepo & Aristizábal, 2003). In this study, why exams are conducted (purposes) was determined as one of the emerging themes that provided further information about the participants' practices of EGA. The data obtained through interviews (N = 32), retrospections (N = 20) and exam papers (N = 28) showed a high level of coherence and consistency among the participants' reports, which also provided information from what generalisation can be made.

The information from the teachers' interviews suggests that teachers employ assessment in the form of written exams for a variety of purposes: summative, formative and accountability.

The majority of the teachers from the interviews (53%) indicated they conduct their exams mainly to evaluate student performance and give students a grade/score (summative). The following quotes support the above statement:

The exams are intended to evaluate the mastery of the skill, whether the students have the ability to use the grammatical rules accurately.

The exam is to help us get a clear idea of how well the student has mastered those skills we teach them.

Both those are worded summatively and in criterion referenced terms, especially using the words *master...skill*. However, they also are worded in a way which suggests some agency other than the teacher decides this goal: *are intended to...* and *is to help us...* both imply that. That other agency is presumably the institution, or perhaps more specifically the course specification. Hence this confirms that some teachers at least recognise that they are expected to assess summatively in criterion referenced fashion where the scores reflect how much the student knows of what has been taught, not just whether they have achieved more than another student.

Other teachers (33%) asserted their written exams are evidence to be presented when required to prove they have done their job (accountability). For example, one teacher said:

I use exams and will always use exams because it is the easiest and safest for both the teacher and the students.

Here *safest* implies that there are possible dangers, which presumably might arise not just from inaccuracy but also from complaints either by students or the authorities (4.3.1.3.).

Another said:

Exams are tangible. They are evidence that the students and the teachers did their job.

In this case accountability is stated quite explicitly as a goal. By using the word *tangible* this teacher might be referring to the fact that the exams are written so there is a record for anyone to see which does not disappear as with an oral exam, although that overlooks the fact that oral exams can easily be recorded.

Fewer teachers, however, claimed that their exams are an opportunity to gain feedback on the students' needs and identify the gaps between the actual level of the students being assessed and the required standard of the course. One teacher, however, said, 'The exams help me assess their needs and, so, I can find ways to improve their level to reach the required standard.' She indicated that, as a consequence, she could help them better attain the required level, which is a clear sign of the formative purpose.

In summary, when describing the purposes for using written exams, many teacher participants clearly acknowledged the exams to primarily serve the summative function (identifying what students know about the taught topics and assigning grades to measure their achievement). Proving to higher authorities that teachers and students are, respectively, teaching and learning successfully (accountability) seems to come second while improving teaching and learning (formative purpose) is rarely considered by the teachers as a purpose for conducting these exams.

5.3.1.7 Sources.

This section concentrates on the discussion of sources used by the teacher participants to develop and create their exam questions and items. As mentioned earlier, the teachers in this study are the ones responsible for writing their own exams (Section 5.2.1.5.2) and, to accomplish this task, have referred to various materials they then drew upon during the exam construction process

According to the data from the interviews and the retrospection, the teacher participants seem to benefit from various resources to help them assemble their exams. The majority of the interviewees (57%) claimed they rely on previous exams written by other teachers who taught the grammar courses in past years. These exams are provided by the Quality and Assurance Unit when requested. However, during the retrospection, only three teachers (15%) asserted they used questions and items from previous exams when writing their own. For example, a teacher in the retrospective thinking said:

I prefer using prior exams, which I got from the quality unit. I got some insights as to what these exams look like. I borrowed some items and changed them in a minor way to suit the level of the students.

It is possible to note here that this teacher edits the items not because the course specification and contents may have changed (criterion referenced) but because the ability of the students may be different (norm referenced). The above quote can be supported by what some teachers said in their interviews. For example:

Yes, I do adopt and use previous exams because I benefit from the experience and I know which questions work and which are problematic.

In fact, this practice is the beginnings of what professional testers (such as the developers of the international Cambridge exams such as IELTS) do, which is to establish an 'item bank'. That contains a large number of expertly made and piloted good items from which a sample is drawn on each occasion when an exam has to be made and administered. In this case it is the Quality and Accreditation unit that holds a limited but increasing store of items in this way, although it is not clear if they do more than archive them, and actually expertly evaluate and check them as well. Clearly also there is no rule that they have to be used. They are simply available to teachers if they wish, as this teacher also described:

That's part of the quality and accreditation kind of thing. We do submit samples of our in-terms, okay, and finals to be shared, okay, and for other teachers to look at.

Okay, I know that some teachers, okay, will use these exams. I did that. For example, if I like the way this question is written, I had no problem using that question in my exam, and I know that others do not.

From the way the teachers talk it seems that the Quality and Accreditation Unit actually contributes positively to the process of exam construction through providing sample of previous exams to teachers when requested. As the second part of its title implies, a great deal of its activity is administrative and consists of assembling evidence in a multitude of categories that are required by the NCAAA in the KSA which undertakes the accreditation of Saudi universities. Part of that evidence is examples of assessment from every course together with extensive course reports written by the teachers.

Using textbooks to write exam questions and items came second to using previous exams; 13% of the teachers in the interviews referred to using textbooks, either the course textbook itself or other grammar textbooks that cover the same syllabus topics. For example, one of the

teachers said, 'sometimes, I take the same exercises from the book'. Another teacher stated, 'I would just use other books. I would go for book hunting in the library to find a grammar book that I can use'.

The former of those teachers of course illustrates a practice that our sample of teachers rarely reported explicitly, which is that of using for assessment items which are exactly items that occurred already in the course textbook (and so probably in the classroom teaching). This is reported as a widespread practice in assessment in Saudi schools (Al-Seghayer, 2015) and is of course condemned by experts for many reasons. The main one is its unwanted 'backwash' (Hughes, 2010). Students in the context will likely know and expect that the assessment will contain items chosen in that way and therefore concentrate on rote memorisation of grammatical sentences that might be tested rather than on understanding whatever grammatical feature or rule that the examples illustrate. Therefore, the exam does not actually assess the intended understanding of grammar that it is supposed or claimed to, so is invalid. Although this teacher does not say it, a typical reason for teachers using this practice is not just that it is less work than creating new items but that it helps the weaker students to pass.

These statements and others are further supported by teachers' responses during the retrospective thinking tasks. When recalling the writing process of one of their written exams, 50% of the teachers asserted they got their questions and items from the coursebook or from other textbooks.

The items I come across a few in the homework book and some of the exercises in their textbook. I modified them of course.

This teacher of course shows awareness of the dangers of using items directly from the book and says textbook items are changed. However, from the way the second sentence is added, the teacher seems to have spoken like a hasty belated correction, one wonders if the items are actually changed or only used as they are because the teacher might have realised that using material from the book is not regarded as 'good practice'. Another teacher however said

The items were mostly from textbooks. I like going through workbooks of some of the books that the students do not have. They have a lot of activities that we can use and, sometimes, I find the items from the textbook and then just change a little word or two to make it wrong to make an error in it and include this in the question.

In this case the teacher makes clear that sources used to write the exams are ones that the students do not have, so could not possibly know from classroom practice and memorise.

The last resource that teachers resort to for setting exam questions and items is the Internet; 30% of the teachers in the interviews and 15% in the retrospections indicated that ESL/EFL websites have copious materials that can be adopted, modified and used.

Actually, the Internet is full of lots of samples in different ways to ask the students; I mean, samples of tests; so, one chooses the best that suits their students and fits in the context.

The internet is of course so extensive that it is relatively unlikely that the students would find the same examples that the teacher uses. Another said:

I saw this method online on an ESL site, and I really loved the idea because it takes panic away from the student, because I'm really not testing her on how to formulate a sentence. I want to see how she uses the tenses, so I give the sentence and ask them to

formulate three-four similar sentences with different tenses. I found the exact format online.

This teacher illustrates how use of the internet can lead to discovery not just of useful items but in this case a whole item format - essentially that of sentence alteration rather than whole sentence creation in this case.

However, this resource, i.e. using online material, is not problem-free. The teachers who used questions and items on English language websites stated that online materials are sometimes above the students' level, terminology wise, and may contain many mistakes that may require time and effort to modify:

As for online, I find a lot of mistakes there, and the vocabulary sometimes are way above their level, and the context may also not be suitable to the students; so, I tend to spend a lot of time modifying and tweaking online items.

This teacher shows admirable awareness of the need to check and adjust items that are found. One has to be alert to possible errors in what is found in many sources on the internet. In grammar items there is indeed a need to ensure that the vocabulary is not difficult and problematic since that would invalidate the item as a purely grammar assessment item. The context that the teacher refers to is perhaps the cultural context. For instance, example sentences that assume knowledge of baseball or British public houses might be regarded as unfamiliar and/or inappropriate.

In general, it is apparent from the teacher participants' responses that they draw on three major sources to develop questions and items in their written exams: textbooks, previous

exams and ESL/EFL websites. The latter, however, is used cautiously due to authenticity and difficulty issues.

5.3.1.8 Attitude towards current practices.

Among the results of the data analysis, teachers' attitudes towards the current established EGA practices (written exams) in their context were identified and explored. In answering the question 'How do you feel about the current practice of grammar assessment?', teachers' answers indicated divergent views between supporters of using written exams, opposers of this practice and those who are impartial.

In this study, 61% interviewees expressed predilections for assessing grammar mainly through written exams, and thus that they did not use them only because they were forced to. On various occasions, these teachers stated that assessing grammar through written exams is a recognised and established means of assessment worldwide in most educational facilities:

I believe this way is valid. Written exams have always been part of universities' assessment systems, and this institution is no different.

They also asserted that it is an effective (easy and convenient) way to evaluate the students and the students like it:

Teachers feel more comfortable using written exams, more economic time and effort and even students, they like it as well.

I don't object on these exams. I do believe that it has to be written and not oral. It is easier for both the student and the teachers.

Another indicated their motivational value in encouraging students to study.

I do agree with it. For university students, yes, it is necessary specially since they are specialised in linguistics or translation. I do believe in written exams because it just makes it more serious for the students to study.

On the other hand, 23% of teachers expressed disinclination to use written exams as the only means to assess students' knowledge or skills on the grammar course. These teachers felt that those exams only focus on students' retention ability and ignore other skills that need to be catered for, for example, conducting research or giving presentations:

I don't agree with them. Really not 100%. I notice that students rely more on memorising the rules to write them down in the exams, but they don't apply these rules in when using the language in other courses. So, it is totally ineffective. I think there are more advanced ways of doing things, better ways to doing things, but it is not my place to say so.

This teacher seems to refer to students writing words and sentences following the explicit rules they have learned. Therefore, the teacher is perhaps really referring to the distinction (which have already been mentioned earlier) that Krashen (1988) summarises as the difference between learning and acquisition. Students may know the grammar explicitly in terms of conscious rules (by learning) but not be able to produce language following those rules in spontaneous speech or writing where unconscious grammar knowledge has to be called upon (obtained through acquisition). This teacher apparently agrees with Krashen, against some other experts, that what is explicitly learned can never really turn into spontaneous ability like that which is acquired unconsciously and is used in real life

communication. The teacher, however, expresses some subservience to the higher university authorities that decide on the assessment methods and does not propose an alternative.

By contrast another teacher does make positive suggestions:

Two words: problematic and counterproductive. I do believe that students should be given a chance to do presentations and other kinds of assessment. They are undergraduates now; they should explore and experience activities that engage higher skill thinking not just memorise the rules from the book and jotting them down on papers. Then it is pass or fail. It is not right.

This teacher suggests presentations as an assessment option. However, the teacher is not clear how presentation should be conducted in the context of grammar course. The reference to *higher thinking* means that perhaps the teacher means a presentation about some grammatical point (e.g. when to use *will* vs *shall*) based on some research in more than one grammar book or website, so applying critical thinking beyond just repeating the rule in the course textbook. If so, that would of course not meet the objection of the previous teacher since it would still be in the realm of talking about grammar explicitly. On the other hand, maybe the teacher is thinking of a presentation just on any general topic of interest, where the purpose is to develop oral fluency in grammar in general through practising speaking (integratively and communicatively). That would accord with the previous teacher's idea.

Interestingly, the remaining five teachers (16%) seemed to acknowledge the shortcomings of these written exams, and some expressed such exams should be complemented by other assessment means, such as oral presentation and research. For example:

It is not the best, and it is not the worst. We are doing it for a reason. We are teaching here academically, catering for the need of the students, not for the learning, I would say, but for their sense of security.

This teacher clearly has reservations but does not suggest any specific improvement. The teacher seems to see the current system as suiting academic needs of the students but not their actual language *learning*, possibly has in mind again Krashen's opposition (i.e., *academic needs* = Krashen's learning and this teacher's *learning* = Krashen's acquisition). The reference to security implies perhaps that students feel safer where academic knowledge about grammar is assessed explicitly rather than performance ability in using grammar integratively in speech or writing. Therefore, students would score high and get the grades need to pass the course (section 4.4.4.1).

I think it's a bit inaccurate way of doing it. If it is going to be through written exams and presentations or research that would be, I think, more valid.

This teacher clearly favours assessments though means where grammar has to be used integratively and perhaps more spontaneously, although it remains unclear whether the topics of the texts, presentations or research are in fact to be on grammar itself or general topics.

In general, the majority of the teacher participants were satisfied with assessing grammar through written exams and justified this practice as being practical and academically endorsed by most educational facilities worldwide. Other teachers felt using written exams is rather an outdated practice and targets students' superficial learning skills, which is not the goal of undergraduate studies. A few teachers, however, indicated that written exams could

be useful if combined with other means that do involve grammar as a skill in real extended language use.

5.3.2 What are the factors which influence their practices other than their beliefs?

This section presents the factors that may have influenced the EFL teacher participants' current practices of EGA. EFL teachers in Saudi's high-educational contexts tried to apply what works best for their students when it comes to EGA. However, constraints may occur relating to 'the contextual factors that may have facilitated or hindered the kinds of decisions teachers were able to make' (Borg, 2003, p. 98).

In this study, a number of factors were identified that constrained or geared the participants' EGA practices towards a fixed assessment method, written exams. These factors include class size, learners' readiness, teachers' training and development and educational culture.

5.3.2.1 Educational culture.

One of the most significant contextual issues in the present study was educational culture, which, with respect to assessment, could be defined as the norms and guidelines set by the educational facility to govern and guide teachers' assessment practices (Inbar-Lourie, 2008). According to the teachers, assessing English grammar through written exams is a must, dictated by the Regulation of Study and Examination equally in institutions A,B, C and D (Section 4.3.1.2) and required for Quality and Assurance Unit documentation. In addition, the role of the departmental policies of the high educational facilities usually also dictated the division of the assessment into quizzes, midterms and final exams at different times, the percentage weight of marks from each of those and even the dominance of production

grammar items over recognition ones in the papers. That was confirmed by teachers' statements such as these two:

I assess grammar through written exams; as I said before, it is the department regulation to give students two exams during the semester and one final unified exam at the end of the term.

We do follow what is normal or the norms in assessing the course of grammar, which is the exams. You know, at the beginning of each semester, I receive the course specification, and there it is written for assessment two in-terms out of 25 assignments and quizzes 10 marks and final 40. Everything comes detailed, so I have no say in it. The exam is paper form because I believe this the way is done here in the college. Also, the college wants this paper exam as evidence for quality.

As seen elsewhere, some teachers such as the latter regard the university practices that are imposed on them as the *norm*, and so were disinclined to challenge them.

5.3.2.2 Class size.

Another important issue raised by the teachers was class size, which many of the teachers felt could prohibit any other assessment activities due to the difficulty of grading many students and providing them with appropriate feedback. One teacher used this to argue against computer-based assessment:

Assessing grammar through online activity on computers or apps sounds a good idea, but I cannot really do that. But, having it in paper form is more practical; I think it would be difficult to find computer labs with 90+ students.

Another used it to show why assessment via oral presentations would be impracticable:

Even if I want to incorporate oral presentation, it would be near impossible to grade 50 students effectively, and the students, you know, they argue over half a mark. So, exam papers work well with my large classes.

5.3.2.3 Learners' readiness.

In this particular study, learners' readiness refers to how likely students are to accept, comprehend and participate in any mental or behaviour change in the assessment process.

Learners' readiness is the first step to check when planning to initiate any particular concept or skill at a given time, for example, introducing new means of assessment (Tomlinson & Moon, 2013).

Teachers who wish to employ any new method of assessment should consider a number of factors that might influence learners' readiness. Firstly, anything that pushes the students out of their comfort zone or causes anxiety or fear can affect students' ability and motivation to adopt the new assessment method. Secondly, teachers should provide assessment tasks within student's zone of proximal development. According to Tomlinson & Moon (2013) when assessment tasks are at or below students' current state of proficiency, no progress will be detected. The same situation would present itself if assessment tasks are well above students' level of knowledge- frustration and confusion will result, but no growth will occur.

In this study, a number of teachers reported that their students are more accustomed to having written exams and securing marks, which might hinder the use of more innovative assessment methods. Another teacher, for example, gave an extended account of an attempt to introduce corpus-based data driven learning into a class. She said:

So anyway, I wanted to give them the chance to do something different. So, I introduced the corpus to them. Okay, I showed them two different types of corpora, okay, I show them the BNC, and I showed them COCA, okay, and I showed them another application that has parallel corpus, okay, English, Arabic corpus, and we talked about the phrasal verbs okay as being grammatical collocations. Okay, these types of collocations are grammatical first, and they make abundance of mistakes when using those. So, I asked them to go look for the phrasal verbs okay in the monolingual corpus okay, which, is the BNC. I even make some kind of a training session where I introduced the notion of collocations, the notion of corpus. I showed them like how it's done the search and everything, okay, and I asked them to go and look for 50 grammatical collocations and to notice the way it is used in real language, okay, because corpus is all about the real deal. Okay, it's not ready-made sentences like those they have in their textbook. I want them to see how these grammatical collocations occur in real language. I asked them to provide me with one example from the monolingual, okay, and one example from the bilingual, okay, and like my aim was for them to see the translation because they are going to be translators and interpreters. They did that I'd rather say clumsily, okay; they kept coming back and forth with their assignments to show me the layout, the whatever, and I'm like this is not how it should be done, okay, and it's not about searching for the information. I thought the exposure to the language in this wealth of data, there must be something that would stick in their head, but on the other, as I told you, I sense that it was just a task to be done and over with rather than a process for language learning or grammar learning. The students did the activity just for the sake to get marks and get high score for classwork, not for learning

Possibly the teacher introduced too many different novelties at once, such as collocations, several corpora, and the idea of working up from examples to rules (data driven learning) rather than the usual pattern from rules to examples. However, the last two sentences also suggest that there is a student culture to combat, possibly ingrained originally in the school system. This takes the form that grades themselves have become the goal of their tertiary level activity, and not the learning that is supposed to be the primary activity, with grades simply used to measure progress and success of that learning. This focus on grades further takes the harmful form that students strategically choose courses and options within courses that will likely get them better grades in place of those that involve learning what they are most interested in or really need for their future life. This however is not a problem limited to the KSA. It is widely recognised worldwide in tertiary education (Holtgreive, 2016).

5.3.2.4 Teachers' training and development.

In the previous example, the teacher could have asked rather too much at once from her students and, indeed, Carless (1999, p. 23) argued that 'teachers need to acquire the skills and knowledge to implement something, particularly if it is slightly different to their existing methods'. Accordingly, the lack of teacher training and development in various assessment methods could lead the teachers in this study to feel unsuccessful if they innovate and fall back on traditional means of assessment (written exams). This is consistent with Carless' (1999, p 23) argument that 'if teachers are not equipped to deal with the implications of a new approach, they are likely to revert to the security of their previous behaviour and the desired change may not take place.' The following statements support the argument above:

I guess I could use other ways of assessment, but what is out there and how to use it. Even for online activities, it sounds easy and trendy nowadays, but simply I am not good at this technology and I am sorry to say that. If I have the courage to use the

Internet in assessing my students, that would be more practical, less consuming paper, and time for correcting.

Now I am doing a workshop on how to use portfolio in assessing students' progress. It is very interesting. I am not using it now but maybe next semester, but not in grammar. I think the idea is more applicable in writing and reading courses.

The former teacher is clearly in need of some support in the form of training or professional development. The other teacher, on the other hand, seems to be progressing steadily through attending a workshop on how to use portfolios which would not be applicable in grammar courses since it diverges from the written examination model.

The findings of this study strongly therefore indicate not only the need for additional training to foster teachers' knowledge of various assessment methods and how they can be conducted, but also some need for the institution to become more accepting of variations from the standard written exam model. If that is done, teachers' confidence to apply contemporary and innovative assessment methods may be raised.

In sum, the results drawn from the interviews, retrospective reports and document analysis indicate all the teacher participants assessed grammar through written exams, which were designed and developed by the teachers themselves. This practice, however, seems to be governed by a number of factors, namely class size, learners' readiness, teachers' training and development and educational culture, which are presented in Figure 42.

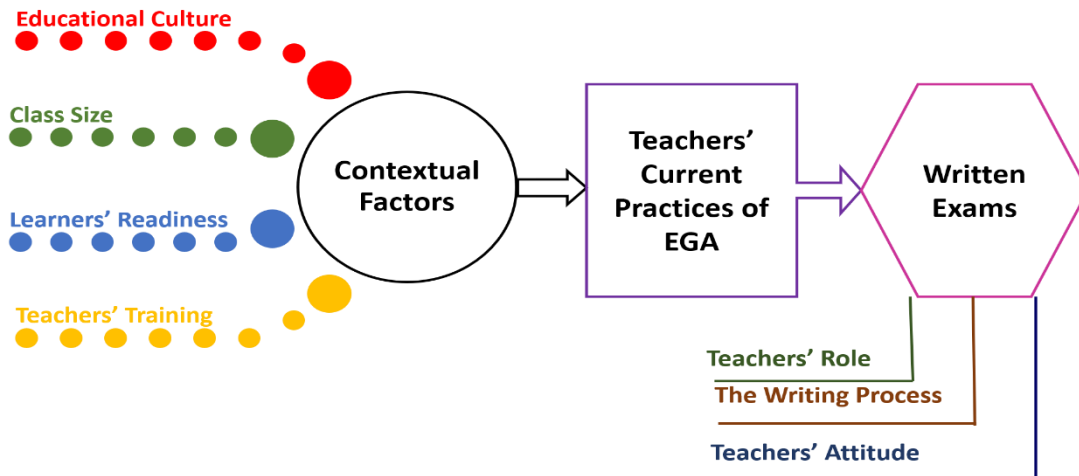


Figure 42. The interrelationship between EFL teachers' current practices of EGA and contextual factors.

The figure above summarises the interrelationship between teachers' current practices of EGA and the contextual factors that have potentially hindered the development of other contemporary and possibly more beneficial assessment tasks.

5.3.3 Summary of section 5.3.

In this section, I shed some light on EFL teachers' actual practices of EGA and factors that have influenced such practices. The findings from the interviews, retrospection and document analysis revealed all the teacher participants use classroom-based assessment in the form of written exams:

- Essential and served a number of purposes
- Conducted explicitly
- Included both production and recognition item formats
- Constructed by the course teachers themselves, drawing on sources such as textbooks and previous exams.

The findings have also provided evidence that teachers' practices of EGA are affected mainly by class size, learners' readiness, teachers' training and development and educational culture.

The following section compares teachers' beliefs and practices to find out the extent to which their beliefs match/ mismatch their practices. Whereas in 5.3.1.8 we gave some idea of the teachers' own view on that relationship, the account in 5.4 is provided by the researcher based on comparison of much of the information in 5.2 with that in 5.3, wherever the same themes occurred in both.

5.4 The Relationship between EFL Teachers' Beliefs and Practices regarding EGA

According to Borg (2003), teachers' cognition is frequently cited as exerting a powerful influence on their current practices. However, study of the mainstream literature on teachers' beliefs and their practices revealed that the relationship between what teachers say and what they do is almost always incongruent. This section focuses on investigating whether or not there is any relationship between teachers' stated beliefs and their current practices with regard to EGA.

In order to answer RQ3 – a) What is the relationship between EFL teachers' beliefs and their current practices? b) Which factors lead to the convergence or divergence between their beliefs and practices? – findings from the research various sources were accumulated, compared and contrasted to identify significant themes and provide in-depth analysis of the relationship between EFL teachers' beliefs and practices of EGA. Three topics are discussed in this section: the congruity between teachers' beliefs and practices regarding EGA (section

5.4.1), the incongruity between teachers' beliefs and practices regarding EGA (section 5.4.2), and the factors impacting teachers' beliefs and practices (section 6.4.3).

The analytic process began with teachers' beliefs about how English grammar should be assessed (RQ1) and then proceeded to find out what they actually did in assessing English grammar (RQ2). The final stage sought to identify and explore the extent of the convergence and divergence between the stated beliefs expressed by the teachers and their actual EGA practices to provide reasonable interpretations of the relationships between these variables (RQ3). Figure 43 presents the shared themes between teachers' beliefs and practices.

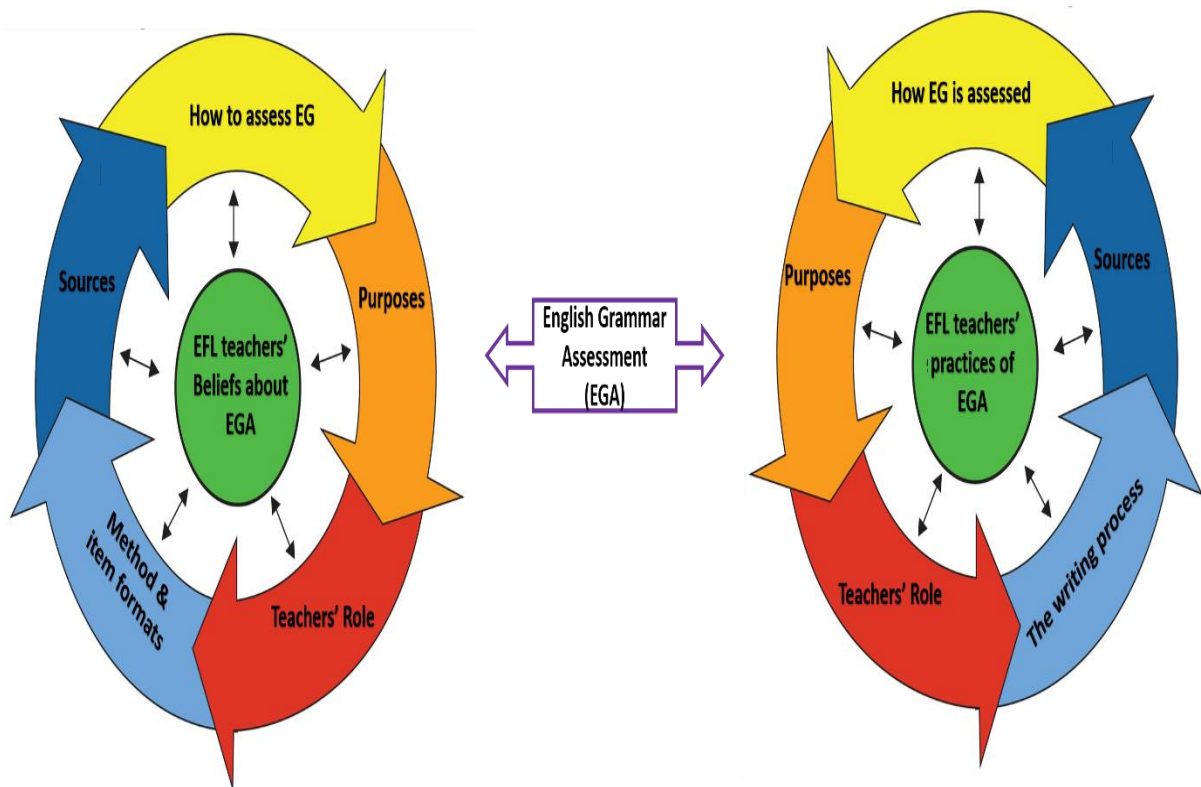


Figure 43. Themes identified to understand the relationship between EFL teachers' beliefs and practices of EGA.

As the figure shows, five themes were identified in both EFL teachers' beliefs and their assessment practices, which formed the base of comparing and contrasting EFL teachers' beliefs with their practices. The findings are presented according to EGA issues rather than

per teacher for two reasons. First, the number of the participants was large and different in each data collection method and presenting the findings about each teacher required more space. Secondly, presenting data in a holistic view highlights more clearly the similarities and differences between teachers' beliefs and practices, which helps answering the third research question.

5.4.1 Congruity between teachers' beliefs and practices in relation to EGA.

In this section I briefly consider the extent to which teachers' practices were congruent with their stated beliefs with regards to the themes represented in figure 43. In doing so, I summarised and categorized the participants' responses from their interviews, retrospections and documents individually and later presented it as a whole. Questionnaire responses were excluded here because the participants were anonymous, and I could not match the participants' beliefs in the questionnaire with those in the interviews, retrospections and exam samples.

The participants' responses from the qualitative data were divided into three groups: group A has 4 participants who did only the interview, group B with 8 participants who did the interviews and submitted samples of their written exams, and group C includes 20 participants who did the interviews, retrospections and submitted samples of their written exams (see Appendix O for detailed analysis).

- *How to assess EG.* Only one teacher from group A seems to be consistent in stated beliefs and stated practices regarding assessing EG Explicitly through discrete-point items like MCQs and filling in blanks... etc. From group B, only 3 teachers were found to be consistent with regards to their stated beliefs, stated practices and their samples of their written exams. as with the teacher in group A, teachers in group B

believed that EG should be assessed explicitly through discrete-point items and they acknowledged doing so in their practices which was mirrored in the sample exams. In group C, two teachers believed that EGA was best done explicitly and stated that in their practices which was further supported by what they said in their retros and their submitted written exams. In total, 6 teachers (19%) were found to be congruent in the beliefs about and practices of EGA.

- *Purpose of EGA.* Looking at teachers' responses (group B & C). The findings reveal that the beliefs of the EFL teachers were congruent with their practices regarding the purposes of EGA. As shown in section (5.3.1.6), the teachers first and foremost believed that the function of EGA is to report about students' progress and achievement usually in numerical terms (summative purpose), which was evident in their practices as well.
- *Item formats.* Congruity between teachers' stated beliefs and practices was also evident in relation to EGA item formats, that is, assessing EG subjectively through speaking and writing or objectively through discrete-point question items). Interestingly, teachers who were congruent in their beliefs and practices with regards to assessing EG explicitly were also found among the teacher participants who showed congruity in assessing EG through objective item formats.
- *Teachers' role in constructing EG exams.* 22 teachers (68.75%) believed that the course teachers are the best ones to write EG exams. Their beliefs were in line with their practices since their exam samples submitted during data collection phase were written by the teachers themselves. (Section 5.3.1.4).
- *Sources.* According to the findings reported in sections 5.2.1.5.1 and 5.3.1.7, teachers' beliefs and practices are in alignment when it comes to the sources used when constructing EGA tasks. Teachers believe that textbooks and written exams

administered sometime in the past were the best sources to rely on when constructing EGA tasks. In their practices, the teachers indicated that books and previous exams are among the frequent sources they rely on to write their exams.

In conclusion, the findings of the data indicated a level of congruence between teachers' beliefs and practices in five different aspects related to EGA, each of which has its own value. More explicitly, teachers' beliefs about how to assess EG, purposes, teachers' role, item formats and sources match their practices within these aspects.

5.4.2 Incongruity between teachers' beliefs and practices in relation to EGA.

Tensions between teachers' stated beliefs and practices were apparent in three main aspects of EGA: how to assess EG and item formats, purpose of EGA and teachers' role. These are presented as follows:

- *How to assess EG.* According to the findings reported in section 5.2.1.3, teachers' beliefs were incongruent with their practices in section 5.3.1.1 regarding the method of assessing English grammar. 22 Teachers (64%) believed that EGA should be assessed in an integrative manner through writing essays or research and giving oral presentations whereas, in practice, they all assessed grammar explicitly through discrete-point items.
- *Purpose of EGA.* Among the participating teachers only three teachers believed that EGA should be geared to help students' progress in learning the English language through identifying students' weaknesses and working towards amending these weaknesses (section 5.2.1.1). However, those teachers in their practices used EGA to report students' results and determine if those students have passed or failed the course and accordingly progress to the next level or not (section 5.3.1.2).

- *Teachers' role.* 49% of the teacher participants believed that having a committee to construct the EGA tasks is accepted and preferred (section 5.2.1.5.2). in their practices, all the teachers stated that they write their own exams (5.3.1.4).

Despite the difference discussed above, these tensions between teachers' beliefs and practices can be seen as a gate way to provide an opportunity to inspect and improve the fundamental issues related to EGA that cause conflict between teachers' beliefs and practices (see section 6.3).

5.4.3 Factors that influence the relationship between teachers' beliefs and practices in relation to EGA.

Congruity and incongruence found between teachers' beliefs and practices regarding EGA could be attributed to a number of demographic and contextual factors (sections 5.2.2 and 5.3.2). It seems that teachers' learning experience has had a strong impact on how they assess grammar in their teaching environment. The majority of the teachers (Section 5.2.2.8) confirmed they assess grammar explicitly because that is how they were assessed when they were language learners themselves; thus, there is congruity between teachers' beliefs and practices.

However, their beliefs about assessing grammar in an integrative manner have been challenged by various contextual factors, therefore, resulting in incongruity. For example, since the regulation of study and exam (Section 4.3.1.2) mandate that all theoretical courses (e.g. grammar course) are assessed via two midterms and final exams, teachers have no other choice but to comply with these demands and feel constrained to employ innovative assessment tasks. Class size, teachers' lack of training and development as well as learners'

readiness are also evident factors that cause teachers' practices of EGA to deviate from their beliefs.

5.4.4 Summary of section 5.4.

It was generally found that the relationship between teachers' beliefs and practices was highly congruent regarding teachers' role in constructing EGA tasks and using textbooks and previous exams as sources to find exam questions, partially congruent in preferable item formats and purposes of EGA, and vastly incongruent regarding the method of assessing English grammar.

5.5 Summary of the Chapter and Conclusion.

This chapter has addressed the research questions of this study in relation to teachers' beliefs and practices regarding EGA in high educational facilities in Saudi Arabia. The results presented above were obtained from the analyses of the data from a questionnaire, interviews, retrospections and document analysis. The results obtained from the questionnaire and the interviews allow the development of a more comprehensive picture of the teachers' beliefs about EGA. Results from the interviews, retrospections and document analysis provided accurate description of the teachers' current practices of EGA.

The findings show that the participating teachers held various beliefs about different aspects of EGA. For example, most teachers believed that grammar should be assessed in an integrative way through other English language skills, namely writing and speaking. Moreover, all the teachers were consistent in their practices, where they assessed grammar explicitly through written exams.

This study also offers a window to inspect and understand the relationship between EFL teachers' beliefs and practices. This kind of relationship found may involve congruence or incongruence between the teachers' beliefs and practices with regard to English grammar assessment. The findings show more congruent than incongruent relationships between teachers' beliefs and practices in this study.

Table 27.
Summary of Data Findings

English Grammar Assessment (EGA)			
Themes	EFL teachers' beliefs	EFL teachers' practices	Relationship between EFL teachers' beliefs and practices
Assessing EG	In an integrative manner	Explicitly	Incongruity
Purposes	Summative, formative and accountability	Summative, accountability and formative	Partial congruity
Teachers' role	Collaboratively and individually	Individually	Congruity
Item formats	Error analysis, sentence completion, MCQs etc.	Errors analysis, sentence completion, MCQ etc.	Congruity
Sources	Textbooks, previous exams and ESL websites	Previous exams, textbooks and ESL websites	Congruity
Factors	Teaching and learning experience, qualifications and teacher training	Class size, learners' readiness, teacher training and educational culture	∅

The table above summarises the findings of the study about EGA within EFL teachers' beliefs and practices. It also sums up the factors that affected these beliefs and practices and

defines the relationship between them. The following chapter discusses the main findings of the study in light of the existing literature and within the theoretical framework of Activity Theory.

6. Discussion

6.1 Introduction

This chapter discusses and presents the findings of the study within the theoretical framework of Activity Theory (AT) and in relation to existing literature. The key findings drawn from the analysis of the qualitative and quantitative data focused on aspects related to EFL teachers' beliefs and practices regarding English grammar assessment (EGA) in higher educational facilities in Saudi Arabia. AT is used as a framework for the discussion of these findings given its explanatory potential, i.e. to account and further explore the findings by viewing them as principles of an activity network which combines and links the teachers' beliefs system to their practices system. Throughout this chapter and based on AT theory, it would be evident that beliefs and practices are interactive and are interdependent (Jonassen, 2002).

6.2 The Activity Systems and The Activity Network

It has been already discussed elsewhere in this thesis (Chapter 3), that the theoretical framework of Activity Theory and particularly the third generation of AT would be used to interpret the findings of the data in this study. As pointed out by Engeström (1999), one activity system is often related to other activity systems. In this study, I consider EFL teachers' EGA beliefs as representing one activity system and their EGA practices another. These systems interact within a larger construct called the activity network. Figure 44 presents the themes of the study as AT components along with the contradictions which were uncovered (represented by the wiggly lines).

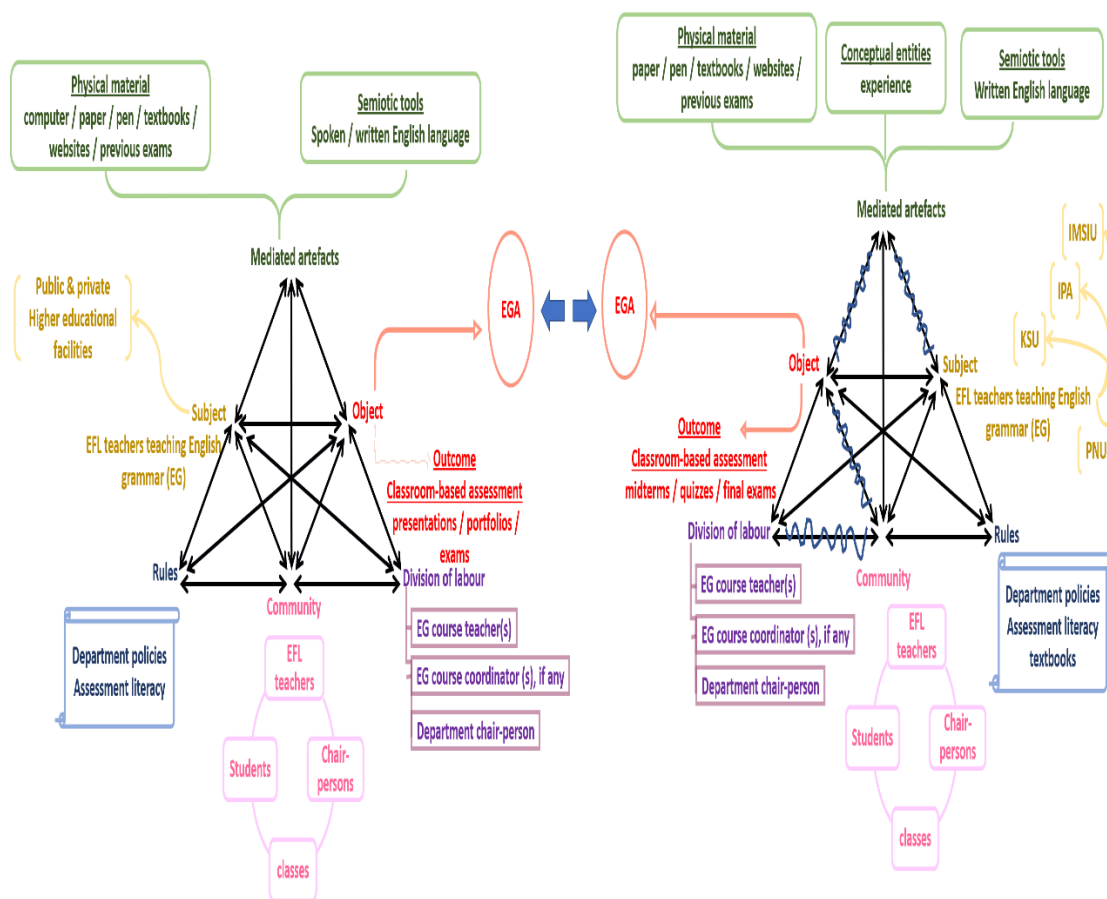


Figure 44: Representation of the activity network and themes of the study as AT components

In Figure 44, the activity network hosts two systems: the system of belief on the left, and the system of actual practice on the right. Thus, whenever a teacher engages in the activity of EGA she/he is modelled as doing so at two levels: the teacher works at a conceptual level of beliefs where ideal EGA would be performed and at the level of actual practices which yields the actual EGA. The teacher who is constructing EGA is in both the subject, and the object - outcome is in both the EGA product itself. At the apexes of the triangles are the 'tools' or mediating artefacts used by the subject in producing the EGA. These are essentially the same in both the belief and practice system except during the actual practices, the teachers rely on their experiences from writing, administering and grading previous exams as a tool to construct their current exams. The bases of the triangles are occupied by contextual social factors which affect the activity at the time when it is executed. These are divided by AT into

three types - rules, community and division of labour. Along with AT's six main elements and of great interest in LTC research field, and in the present study, are the historical and background factors that affect how teacher beliefs and practices come to be, how they are at the time of doing any activity, such as teacher experiences in the past as learners/testees, and any assessment training they have received. Through AT rationalization, these do not have a separate place to be entered but are regarded as part of the makeup of the subject.

Within the activity network as a whole, the subjects (EFL teachers) participating in both activity systems belong to the same community, that is, all the participants are EFL teachers currently working in higher educational facilities in Riyadh, Saudi Arabia. They share the same object: assessing English grammar. However, teachers' beliefs and practices are expressed by two systems because the subjects in each system approach the object differently. This difference contributes to the divergence between the beliefs and practices of the EFL teachers in relation to EGA which leads to the resulting contradictions and tension (see further discussion in section 6.3). In the following subsections, I discussed the findings under each of the six AT framework components: the high-level categories and themes which were presented as key findings in chapter 5 (Table 28).

Table 28

Summary of themes as AT components

Domain Components	Beliefs	Practices
Subject	Personal background sources of current teacher beliefs	Personal background sources of current teacher practices
Object → Outcome	Belief-based ideal EGA produced: integratively → oral presentations + lengthy written tasks (e.g., research or essay)	Actual EGA produced: explicitly → written exams with discrete-point items
Mediating Artefacts	Ideal belief-based use of MA:	Actual use of MA:

Domain Components	Beliefs	Practices
	Physical: textbooks, previous exams and ESL websites Semiotic: spoken or written English language Conceptual: assessment literacy	Physical: previous exams, textbooks and ESL websites Semiotic: written English language Conceptual: assessment literacy
Division of labour	Ideal belief-based division among course teachers, coordinators and/or exam committees	Actual division among course teachers, coordinators and chairpersons
Community	Ideal belief-based community: EFL teachers, chairpersons, coordinators, students	Actual Community of practice: Teachers + coordinators + chairpersons Discourse community: teachers + coordinators + chairpersons + students
Rules	Educational policy + Assessment literacy	

6.2.1 Subject

The subject in each of the two activity systems under consideration, i.e., beliefs and practices, was the sample of EFL teachers in higher education in the KSA. The subject in the beliefs system includes EFL teachers from public or private higher educational institutions all with some experiences of EGA, while the subject in the practices system included EFL teachers from four public higher educational institutions, with a majority coming from KSA.

6.2.1.1 Teachers' age

This study showed that teachers' age, instructional and learning experiences and qualifications all contributed to forming their current beliefs about EGA, such as that EGA should be done integratively through subjective item formats, and that it is done mostly for summative purposes. With regards to the effect of teachers' age on their beliefs about assessment, the results in this study were, therefore, not in line with those found in other studies which reported that age has no significant effect on teachers' beliefs about assessment (Chan, 2006; Mehrgan, Hayati, & Alavi, 2017).

6.2.1.2 Teachers' qualifications

As for teachers' qualifications, it was evident that teachers with higher educational degrees had stronger beliefs about the various aspects of EGA mentioned in section 5.2.2.5 compared to teachers with lower degrees. This was consistent with findings reported in studies like Mehrgan et al. (2017) and McMullen (1999, 2003).

6.2.1.3 Teachers' length of teaching experience

The influence of teachers' teaching experience on their assessment beliefs was also evident in studies by Chan (2006), Sahinkarakas (2012) and Mehrgan et al. (2017) in which results displayed a significant relationship between teachers' beliefs and years of ESL teaching experience.

It was evident that in the present study the EFL teachers' assessment experiences as language learners centered around traditional written exams, and that they viewed these experiences as positive. A similar result was also found in Karp & Woods (2008) in which teachers perceived their assessment experiences as learners to be practical and useful.

The study also revealed that teachers' professional development and assessment training, or lack of them, affected their EGA practices, causing teachers to rely mainly on traditional written exams. In line with this previous statement, Wang (2004) mentioned that 'it is incumbent upon science educators and teacher training program to provide the teachers with more opportunities to examine their beliefs and make their implicit beliefs explicit' (p. 110).

The discussion presented in this section thus is related to RQ1 b), factors which have helped shape beliefs as well as RQ2 b) factors, other than the teachers' beliefs, which have influenced their practices. The results on the subject of the EGA belief activity system

revealed that teachers' age, qualification and teaching experience were distinct factors that affected teachers' beliefs about EGA. In addition, teachers' experiences of assessment as language learners and their professional training on assessment seem to influence their EGA practices.

6.2.2 Object and Outcome

The Subjects of both systems were oriented towards assessing English grammar (the Object). Three themes from the data analysis related to the attainment of the object: How to assess English grammar, for what purposes, and what methods are best to assess English grammar. It is useful to note here that the themes discussed below in relation to the object component in both activity systems are linked to RQ1 a and RQ2 a which concern EFL teachers' beliefs and practices with regards to EGA.

6.2.2.1 Assessing English grammar

The participant teachers' conceptualization of EGA was found to entail assessing English grammar integratively, with the intention of tapping into the total communicative abilities of language learners. This belief is in alignment with the findings reported by Hodgson (2017) and Ahmadi and Shafiee (2015). In Hodgson's study, it was reported that assessing English grammar explicitly was unrealistic and impractical and deviates greatly from what language is used for (communication). I interpret this view as a call for assessing grammar integratively within the language through reading, writing, listening and speaking. As for Ahmadi and Shafiee's study, it was found that teachers believe that assessing grammar at discourse level is more beneficial than assessing it separately through discrete-point items. The outcome of the belief activity system has to be imagined, but beliefs favouring

integrative assessment such as those just described imply that it would involve more oral presentations and essay or research writing.

In their practices, however, teachers assessed grammar explicitly, which contradicts their beliefs stated above. However, and according to Purpura (2004), the teachers' EGA practices are in line with their teaching practices since they assess the way they teach. The consequence of this is that the outcome of their practices (i.e., how the EFL teachers assess EG) is classroom-based grammar assessment tasks in the form of written exams: as we saw in the actual midterms, quizzes and finals that they set.

6.2.2.2 Purposes

In relation to the second object related theme, purposes of EGA, the research evidence revealed commonalities in teachers' overall beliefs about the purposes of EGA and their practices. Teachers believed that EGA is first and foremost used summatively to obtain valuable information about students' overall performance at a specific point in their English grammar learning. This was directly translated into their practices of assessing English grammar through exams which reported on students' achievement usually in grade-related or numerical terms. In published studies, teachers often talk about the important values underpinning assessment in education (Ağçam & Kaya, 2017; Barnes et al., 2015; Chew & Lee, 2008; Elshawa et al., 2017; Mussawy, 2009; Opre, 2015; Sikka et al., 2007). Although most of the studies report on formative purposes or accountability, the findings in this study observed them not to the same extent as summative purposes. One study, however, by Dayal & Lingam (2015) stated that the participants, pre-service teachers, focused on the summative function of assessment in which assessment is used to provide students with scores on how

much they knew about the subject area. Our study matches this although our participants were not pre-service teachers.

6.2.2.3 Method and item formats

The third theme related to the AT object is the method and item formats by which the outcome is achieved. Teachers believed that executing grammar assessment through paper-and-pencil exams is preferable and more reliable. In addition, they seem to believe in various question types, such as error analysis, sentence production and multiple-choice questions (MCQs). Their beliefs were again in line with their practices since all the EGA samples collected in this study were hard copies of written exams had various item formats endorsed in their beliefs were evident in their exam papers. Other empirical studies reported similar attitudes towards paper-based assessment and item formats (e.g., MCQs and true and false). For example, in Karim's study (2015) it was found that all the participants acknowledged the use of paper-and-pencil assessment methods to grade their students while in Elshawa et al. (2017) teachers perceived essay writing, MCQs, and true/false items to be suitable question formats for assessing various language skills.

6.2.3 Mediating Artefacts (MA)

In AT, artefacts (tools) mediate the subject's thought and behaviour during the engagement in the activity, in this case EGA. It is important to acknowledge that the role mediated artefacts have in EGA is incorporated within the system of EFL teachers' beliefs about and practices of EGA (RQ1 a and RQ2 a). Furthermore, the concept of mediating artefacts enables me to foreground a number of aspects highlighted in the findings about teachers' EGA beliefs and practices. EFL teachers seem to use and refer to various instruments to support their beliefs

and practices ranging from abstract to fully contextualized concepts. The sections below elaborate on these themes.

6.2.3.1 Physical materials: Sources

The first artefact to be discussed here is the sources which the participating EFL teachers think are suitable to adopt exam questions from (beliefs), and those they currently used when writing the exam questions (practices). Sources can be divided into two types: complementary materials and teaching materials. The former refers to textbook exercises/tests, previous exams and online materials which could be used in the construction of exams. The latter refers to the grammar course textbooks used by the teachers and how they influence teachers' decision-making when writing their exams.

Teachers believed in, and actually used, textbooks and previous exams as preferable reliable sources when constructing EGA tasks. Teachers also referred marginally to using ESL websites when writing their exams which was in alignment with their beliefs. In addition to using the previously mentioned sources, the teachers in their practices also referred to using the teaching material assigned by the department to guide their exam construction.

Grounding the exams within the course textbooks or the teaching material also reflected underlying tensions in the activity systems in the sense that teachers wanted to be more innovative and assertive in their question items, but they were constrained by being required to follow the course textbook exercises and the teaching method from which they cannot deviate (see below section 6.2.6).

6.2.3.2 Semiotic tool: Language

In AT, language is construed as a psychological tool. Language allows the teachers to control their behaviour, i.e. their practices of EGA, through social interaction. It also enables them to regulate their minds, i.e. their beliefs (Vygotsky, 1978). Whether talking about their beliefs or reporting their practices, English language (written or spoken) is the means to achieve EGA.

Although the teachers did not talk about this directly in the interviews, it was clear that English language was a valuable tool to accomplish the object (EGA). In the context of this study, it is impossible for EGA activities to be carried out without spoken or written form of English. The need to assess students' grammatical knowledge through language production was reflected in their beliefs that grammar should be assessed integratively and communicated through the language as a whole. In their practices, however, teachers sought to measure explicit language use of particular grammatical topics and restricted language production to its minimal (i.e., word or sentence level). This was evident in the exam samples collected which focused more on assessing grammar explicitly through discrete-point items (see Appendix M).

6.2.3.3 Conceptual tool: Assessment literacy (AL)

In the context of this study, knowledge about how to assess EG was obtained from two sources: teachers' instructional experiences (a posteriori knowledge) either as learners or teachers, and /or teachers' qualifications/training.

With regards to teaching experience, it was apparent that teachers with greater experience believed more in assessing EG integratively through extended oral and written tasks. This

was due to the fact that those teachers have assessed grammar repeatedly through discrete-point items and the results were not satisfying, that is, students put the rules in the paper exam but were unable to apply those rules in a real-life situation which requires the use of correct language for successful communication.

My interpretation is that experienced teachers rate their own language skills too highly and thus they are confident in using oral presentations and essay writings as means to assess EG. A similar finding was reported in Sahinkarakas (2012) where experienced teachers were found to be more courageous in conducting assessment effectively. In the present study, however, those teachers did not put their beliefs/knowledge into practice because it would not be allowed (section 6.2.6).

Teachers' instructional experiences along with their qualifications substantially contributed to teachers' knowledge about assessment (see section 5.2.24). When assessing EG (practices), teachers seemed to attend to vital aspects related to writing exams, such as, exam fairness, validity and practicality all of which indicate that the participating teachers were to some degree literate in assessment. Such issues were discussed and referred to in various instructional assessment handbooks, for example, Downing & Haladyna (2006) and Brown (2003). With regards to exam fairness, the teachers often reported that their exams have to be fair to all students taking into consideration the different students' proficiency levels. In line with this concept, exam fairness, other studies (Deneen & Brown, 2016; Saad, Sardareh, & Ambarwati, 2013; Munoz, Palacio, & Escobar, 2012) showed that the teacher participants asserted that exams should be fair to all students.

When it comes to exam practicality, the teachers adhere to aspects of time constraints and scoring systems. Teachers design their exams to accommodate one hour if it is a midterm and two hours if final. All their exams are weighted properly according to a specific scoring system that meets the course objectives. In addition, the teachers seem to always bear in mind class parameters, that is, how many students there are in each section and how long each exam session is. Knowing that they have over 30 to 50 students in each class, teachers realize and acknowledge that written exams with discrete-point items are more practical than doing oral presentations or asking students to write essays. The former might require multiple sessions to complete while the latter might take longer time in correcting and scoring.

In all, the number of students and the time and timing of exam sessions in addition to exam modality—were significant in the case of teachers' assessment practices. Teachers on various occasions showed how awareness of classroom parameters played a role when teachers assessed students in classroom-based assessment.

6.2.4 Division of Labour

The AT concept of division of labour here refers to the explicit role of teachers in constructing EG exams in contrast with the roles of course coordinators when represented institutionally. Another consideration with regards to division of labour relates to the 'division of authority and status' (Roth & Tobin, 2002, p 114). A useful example in this regard is that of the power relationship between the authority of the department chair-persons, the coordinators and the teachers. This relationship among the members of the community could result in internal tensions and contradictions with regards to division of labour when assessing EG as discussed in the following section 6.3.

6.2.4.1 Teacher vs. coordinator role

According to Sahinkarakas (2012), the ‘teachers’ role in assessment is inevitable as they are the core of this process’ (p 1787). Concerning teachers’ roles in the development of EGA, one of the major findings of this study was that the majority of the teachers were in favour of being in charge of writing their own exams either collaboratively or individually (see section 5.2.1.5.2). This was evident in their EGA practices where all the participating teachers submitted samples of grammar exams which they wrote themselves. These findings were also reported in another study (Elshawa, Abdullah & Rashid, 2017) where teachers exercised a degree of freedom in constructing their assessment tasks and they did so by themselves.

As for course coordinators and department chair-persons, findings reveal complex and various roles based on the teachers’ reports. In KSU, teachers mentioned that the coordinators’ job does not exceed proofreading the exams and making sure they are of the required level of difficulty, for example, ‘maybe you can say like proofreading sometimes’. In IMSIU, coordinators have a more vital role since the course coordinators are the ones to write the final exams which are unified between male and female sections and which might result in tensions between the course coordinators and the course teachers within the practices system (section 6.3), for example:

The coordinator writes the final; it’s unified. I write the midterm and quizzes. So, he writes the final and if we all agree that's it. When presented with the exams, we give our opinion and suggest some changes, changes like the wording of the question, like grammatical mistakes. Some questions may even be hard on the other students because he taught his students like this and it's unfair. So, we have to change to an easier one.

In summary, teachers' role(s) in EGA is complex and multifaceted as shown in section 5.3.1.4. In addition, teachers' roles in EGA in most cases surpass that of course coordinators.

6.2.4.2 Authority & status

Institutional regulations (see sections 4.3.1.3 and 6.2.6 below) play a crucial part in assessment because they represent authoritative instructions that dictate teachers' role in assessment. These regulations are executed by the department chair-persons, monitored by course coordinators, if any, and followed by the teachers. According to these regulations, course teachers are the ones responsible for developing, designing, implementing and correcting any assessment tasks, within a set framework (as referred to in section 4.3).

However, if extenuating circumstances exist, the department then has the right to assign the role of constructing an assessment to whom they see fit, for example, the course coordinator.

The findings of this project clearly showed the source of power and responsibility for EGA and whether or not that power was being shared within specific institutions. The teachers developed and produced classroom-based assessments within strict guidelines and schedules imposed by the administration. In the design phase, the teachers determine what is to be assessed, chapters and topics, and then items and tasks were produced in accordance with the purpose of the assessment and the type of language being assessed (e.g., production or recognition). The development phase includes deciding on the number of items to be included in the assessment, taking into consideration teachers' knowledge of the students and the time available. This phase concentrates on ensuring that the items and types of tasks are suitable in order to produce a valid EG exam. For the operational phase, coordinators, if the institution had such role, would review the exams in order to make sure they have easy and clear instructions and fall within the course objectives. Finally, the administration phase

commences after the department chair-persons sign off the exams. As stated by one teacher, ‘okay well a coordinator to be honest just sees the exams before the teachers administer them to the students, and also the head of the department has to read and sign the exam.’

In sum, the discussion of division of labour suggests demarcation of roles exists among the various members in the activity system. This leads then to questions about how the work is shared out within a community, which is discussed next. The discussion in section 6.2.5 which follows, like that in the above section, is related to RQ1 a and RQ2 a in which roles of teachers, coordinators, chair-persons and students are part of teachers’ beliefs about and their practices of EGA.

6.2.5 Community

The AT notion of community in my study includes all the agents in the activity systems: the teachers, the coordinators, the chair-persons and the students. The term community here does not imply necessarily co-presence of all the members; however, it does imply their participation in an activity system whose members share understanding concerning EGA and exercise some influence on how EGA is done. For example, the coordinators, the chair-persons and the students are not really physically present in this study, however, these members do have an influence over the teachers when it comes to EGA decision-making. The teachers in my study for example keep talking about how they need to take into consideration students’ level and how that is a major factor in designing their exams and selecting question formats (section 5.3.1.5).

The teachers also refer to the coordinators’ role in reviewing their exams and giving feedback. The chair-persons’ participation exists only through the physical presence of their

signatures on exam papers which indicates that the exams are up to the required standards and follow the faculty regulations of assessment.

Participants in the community could in fact be placed into two categories: community of practice and discourse community. The former refers to members who share the same interests, make diverse contributions to an activity, and hold varied viewpoints, while the latter refers to a group that shares a set of values and goals and means to communicate these values and achieve these goals (Swales, 1990). The common features between the community of practice and discourse community are that both emphasize engagement with a shared goal or, in AT terms, object, in a local context and both involve language and activity. However, language here is a tool (section 6.2.3.2) rather than the focus for members of the community of practice.

6.2.5.1 Community of Practice (CoP)

A community of practice within higher educational institutions is an organized group of people which is primarily composed of teachers and students, with the support of other members of the organization such as department chair-persons and coordinators. Originally community of practice was a term related to a theory of learning (Wenger 1998), and it is significant in this study of teacher cognition as a means to explore how teachers' knowledge - assessment literacy- and beliefs about EGA are evolving with their participation in the EGA activities in the higher educational community, and through interaction with other members. In this sense, experience in performing assessment practices and possibly the feedback from course coordinators might be regarded as a way of teacher learning or professional development.

In this study, a community of practice model can be particularly suited to EFL teachers who acquire knowledge about EGA in one formal context and transfer this knowledge into their practices of assessing EG. Some of the participant teachers mentioned that they had course coordinators who supervise and advise the teachers during the writing of the EG exams while others stated that their exams have to be signed off by the department chair-person. During the process of finalizing the exam and prior to exam administration, those teachers were in the process of learning about the norms governing EGA in their community.

The research found that EFL teachers in the higher educational community follow two types of norms: the first is explicit rules (see section 6.2.6 below) and visible behaviours of academic practices of EGA. It also contains explicitly expressed statements of beliefs and values of the role, methods and standards of EGA all of which were represented in this study. The second is an implicitly accepted culture of assessment which involves usually unspoken assumptions about its purposes and appropriate practices, the understanding of academic requirements, relationships and the routine operations of practice.

There is a further evidence in this research suggesting the existence of two CoPs: 1) that of the teachers making the assessments, 2) that of the authorities higher up (deans, quality assurance unit... etc.) who lay down the rules and make the course specifications where key aspects of assessment are determined. The coordinators and heads of Dept have an awkward position between the two as they are usually teachers themselves who have to implement policies from higher up, and they seem to identify themselves more with 1 than 2 though.

Furthermore, some EFL teachers in CoP 1 tend to follow a culture or world view that focuses on helping the students (often weak) in doing as well as possible, by teaching what they can

cope with and making the assessment easy enough so they can pass, while COP 2 is a bit out of touch with reality and lives in a world of ideal targets that look good to outside accreditation bodies but are way beyond what the students can manage, so their ideas on assessment do not get fully put into practice by CoP 1. There is then a colossal 'fudge' in communication between CoP 1 and 2 which then causes much tension in the EFL teachers' practices of EGA.

6.2.5.2 Discourse community (DC)

Swales (1990) lists six features of a discourse community which are here reduced to four features to illuminate the interpretation of my study:

1. A discourse community has a threshold level of members with a suitable degree of relevant content and discursal expertise. This includes the teachers, the coordinators and the department chair-persons. All these members (subject in Activity Theory terms, see section 6.2.1) are EFL teachers with high degrees in various fields of English. They also have some experience of EGA and share an understanding of how and why English grammar should be (beliefs) or is (practices) assessed (section 6.2.1).
2. A discourse community possesses one or more genres which are used for communication involved in the activity. This refers to language used in relation to the object in the activity systems and within the activity network, in our case EGA (section 6.2.2).
3. A discourse community has mechanisms of intercommunication among its members. These are represented in the artefacts including physical materials, semiotic (especially linguistic) and conceptual. Within this discourse community, the members have developed and acquired some specific lexis which is the terminology commonly

a part of assessment literacy and referred to by the teachers, for example, exam validity, fairness and practicality (section 6.2.3)

4. A discourse community has a broadly agreed set of common public norms. These are the rules and regulations of assessment dictated by the educational organisations (section 6.2.6).

Swales (1998) further divides discourse communities into place discourse communities and focus discourse communities both of which are applicable in this study. The former emphasizes local participation for a mutual project. All the participants in this study are faculty members of high educational facilities who design, administer and mark their own grammar exams. The latter stresses the common interest among the members of a discourse community which is represented by the activity object (EGA).

Some might argue that the participants of this study would not form a DC because the EFL teachers would not have common channels of interaction or a common project. Responding to that I would say that the participating teachers in this study are likely to be a DC because of cultural discursive practices, that is, the EFL teachers are engaged in similar EGA activities (writing exams, administration of exams, scoring ...etc.). The EFL teachers not necessarily repeat their own practices; instead, a teacher may perform EGA for the first time but, through direct or indirect observation, the teacher has knowledge of the norms of EGA in his/her community. In addition, the participating EFL teachers might, through intense and long-term engagement of EGA, develop 'habits of mind' that would shape many areas of the EGA practices (section 5.4).

In sum, the community in this study, whether conceived of as a community of practice or discourse, has a common object (assessing English grammar), shared rules (regulations), specified roles (division of labour), and tools (artefacts) to mediate and to achieve the object. Both the tools that teachers use, including their own beliefs and knowledge about EGA (which may not be shared), and the context within which teachers work, influence their assessment practices. Individual teachers are regarded as part of the community of assessment practices (subject) rather than isolated individuals. Teachers are the subject who conduct actions according to the division of labour in the community to achieve the expected goal of assessing English grammar (object).

Furthermore, the community participates in an activity system in which all members of the community are engaged. The community is both durable and dynamic. With regard to the former, means of assessment and regulations are developed from years of practice and are not easily subject to change. This is evident in the Regulation of Study and Examination (RSE) of each educational context (section 4.3.1.3) and supported by teachers' reports in their interviews and retrospections (section 5.3.2.1). On the other hand, the community is dynamic in that it engages in a process of continual negotiation and co-construction through the joint effort of its members.

The community of both activity systems, conceived of in any of the ways mentioned above, is considered a factor that shaped teachers' beliefs and influenced their practices (RQ1b and RQ2b).

6.2.6 Rules

The AT framework draws our attention to the fact that rules include both explicit and implicit norms, conventions, and regulations governing the performance of the activity and enabling the subject to interact with other members and fit into the community. I will now discuss both types of rules in relation to EGA in the present study.

6.2.6.1 *Explicit: RSE*

As can be seen from section 4.3.1.3, public higher educational facilities follow strict, rigorous and clear regulations when it comes to assessment practices. These rules understandably yield a structured and controlled assessment environment in which the teachers are compelled to follow these regulations, for example, teachers said:

So that's why I do this type of testing because I am required to do these written exams.

I don't know if there's another way of doing this, a different way of assessing the students, if they [the department administrators] would let me change it. I don't know what's my limitation, What I could do or could not do, but yeah, I'm sure it's mandated because it was given to us

How educational policies dictate assessment practices was echoed in other studies as well. For example, in Wales, Leung & Rea-Dickins (2007) reported that educational institutions interfere with assessment processes and manipulate data to promote themselves in a favourable light. Along these lines, in the USA, Pellegrino (2004) indicated that the need for high educational attainment and public demand for accountability are main factors to shape assessment standards which are then executed by the teachers. Similarly, Bigg (1999) reported that in Hong Kong there is a strong tradition of rigorous norm-referenced summative

assessment which teachers have to follow to accommodate national or local requirements for certification and accountability.

However, this standardization of how assessment is conducted limits the teachers' decision-making power which then may lead to tension between the rules and the subject within the activity system (further discussion in section 6.3 below).

6.2.6.2 Implicit: fundamentals of test development

Designing classroom-based assessment tasks, specially written exams, is governed by well-defined basic principles of educational measurement (e.g. Hughes, 1989). Although it would be oversimplistic to say that all aspects of this are agreed upon by experts, it can be said that if the teachers are literate about these basics, they would typically make better decisions when it comes to writing their exams, if they have the freedom to do it within the rules. As described in section 2.4.3, knowledge of how to conduct assessment and why is known as assessment literacy.

As shown in the result chapter (section 5.3.1.5), the teachers in this study on several occasions referred to common themes related to assessment, for example, exam validity, fairness and practicality. These aspects indicate that teachers seem to possess knowledge in this area which ground their practices although the study did not go so far as to test their actual assessment literacy.

The influence of implicit rules when referring to the subject of an activity system was also reported in a study by Chu et al. (2016) where implicit rules in the system had a pertinent impact on students' collaboration during classroom marking. Similarly, Binjumah (2019)

reported in his study that there were no official rules to guide the teaching practices of the participants, however, the teachers seemed to have acquired standardized habits of science teaching from previous experience in schools which governed their teaching practices.

6.2.7 Summary of section 6.2

In this section, findings of the study were presented and discussed within the theoretical framework of Activity Theory. Themes and categories generated from the data were aligned with AT components and interpreted within this theoretical and analytical framework.

The following section moves on to discuss the findings related to the contradictions within the current activity systems, and between the teachers' beliefs activity system and the neighbouring system: teachers' practices.

6.3 Contradictions

An underlying principle of activity theory and related to RQ3a (What is the relation between EFL teachers' beliefs and their current practices?) is the notion of contradictions. However, despite it being one of the most commonly employed concepts of activity theory, there does not seem to be a universally agreed meaning for the terms contradictions or tensions in current literature (Engeström and Sannino, 2011). Therefore, the interpretation of these terms depends on the broader understanding of the particular activity system or network in the research context.

In section 5.4.2 and as discussed in some of the previous sections, three types of contradictions within and between components of the activity systems have been identified. These contradictions and tensions were identified at three different levels: primary, secondary

and quaternary (section 3.4.3). At each level, contradictions and tensions were grouped into one of four types:

- Dilemmas are expressions of, or exchanges about, incompatible evaluations, either between people or within the discourse of a single person.
- Conflicts take the form of resistance, disagreement, argument and criticism.
- Critical conflicts are situations in which people face inner doubts that paralyse them when faced by contradictory motives unsolvable by the subject alone.
- Double binds are processes in which actors repeatedly face pressing and equally unacceptable alternatives in their activity system, with seemingly no way out (Engeström and Sannino, 2011).

The first type of contradiction is a primary one (see section 3.4.3) which exists locally within the component of division of labour (section 6.2.4) in the activity system of teachers' EGA practices. Manifestation of this level of contradiction fall into two categories: double binds and conflicts. The former exists where the teachers seem to find themselves in a continuous process of checking and rechecking the exams which the teachers find redundant but unavoidable, for example:

And then again we pass the exams to a coordinator of grammar and she has the final word. Let's say in terms of whether it is suitable for the level of students that we are teaching and then she presents it to the vice chair-person, where she approves it or not. Again, we've been going through a cycle of returning the exams to us just because we need to change this word or rephrase sometimes change this question. It is a bit tiresome and time consuming.

The second category of contradiction, conflict, appeared when there was resistance, disagreement, argument and/ or criticism between the teachers and the coordinators with regard to EG written exams. For example, teachers in IMSIU kept emphasizing that final exams are written by the grammar course coordinator which causes pressure on the teachers and sometimes a high level of disagreement over the exam questions. In the quotes below, some teachers expressed frustration about how final exams are constructed by the course coordinator:

So, the coordinator, he writes the final, and he sets a meeting. We meet and when presented with the exams, we try to give our opinion and suggest some changes. Some questions may even be hard on the other students because he taught his students like this and it's unfair.

We have that in the finals. They [course coordinators] constructed the written exam, and I'm not so happy about it because it's kind of different from my style and the way I teach my students.

But in the finals, I have no role in doing anything; I can suggest, but I do not have a role in actually saying what I can and cannot add, and it is frustrating.

The situation described above could also be treated as double binds since the teachers seem to find themselves in a loop where absence of role in writing final exams repeats itself every semester.

Secondary contradictions (section 3.4.3) also existed between a number of elements of the teachers' EGA practice system. The first occurred between the mediating artefacts and the

subject in the teachers' EGA practice activity system. Within the mediated artefacts, the teaching materials seem to cause tension to the teachers because the course textbooks restrict the teachers' choice of question formats when developing their assessment formats. In the following examples, the teachers explain that they cannot introduce new question formats to target more language production, namely speaking and lengthy writing, because the focus of the textbooks used is purely explicit rule recognition and decontextualized item production. The tensions created between the teaching materials and the subject could be described as both conflicts and double binds. The former manifests through teachers' disagreement with following the textbooks. The double bind manifestation arises because the teachers always find themselves bound by the textbooks and cannot escape this fact.

The textbook we use, *Understanding and Using English Grammar*, endorses discrete-point items throughout the exercises, So I just focus on them in assessment; I wish I could include more integrative sort of questions, but I cannot.

Within mediating artefacts, class parameters also seem to hinder teachers' desire to employ assessment tasks other than written exams to assess EG (the object in AT terms). The contradictions are presented here as double binds since the teachers find themselves in an inescapable situation where the class size interferes with the outcome of the activity object (EGA frequency in this example).

The number of students enrolled in the classroom doesn't allow me to ask for weekly nonetheless daily assignments. I mean let's be honest if you have 40 students in each section, and you are teaching three to four sections; it is not going to be humanly possible for us to correct all of this and give them [students] feedback.

Because of time constraint also because of the size of the class that is why I feel limited in my options [ways to assess EG].

Contradictions between the mediated artefacts and the object manifests once again in the form of dilemmas as class parameters seem to cause teachers to struggle between choosing to employ more integrative means to assess grammar (oral presentation), which would be difficult due to time constraints and class size, and traditional assessment (written exams), which are practical but not entirely effective. For example, a teacher explained:

I think I have some freedom in choosing how to assess my students. I want to use other ways, oral presentation, for example. But can you really imagine how much time that would take? The alternative is just exam and quizzes, not really that effective; I'm struggling here.

Another secondary contradiction was identified within the EGA practices activity system between the subject and the object. The teachers seem to face a critical conflict as they acknowledge that the lack of training on other assessment means or the absence of assessment training might have been a factor influencing their choice of traditional assessment. For example:

I have not gone through any professional training in assessment, but during my MA I have taken Language Assessment [name of a course]. But in language assessment I haven't taken any workshops. So this is maybe what we're missing, the way or the skill to employ other assessments, and that's why we end up doing it the traditional way.

The last example of secondary contradictions diagnosed in this study took place between the community and the object. Within the community, students' readiness affects the object hence the outcome. The teachers on various occasions reported that they cannot employ new assessment methods due to the students' resistance to change the traditional methods which they have become accustomed to throughout the years of learning. For example, one teacher refrained from conducting oral presentations, although she had the freedom to do so, because the students would probably shy off and that would possibly affect the assessment process. She said,

I believe in giving presentations, for example, in Europe and in the United States, they give everyone a chance to present something from the beginning of the course. But I think our students shy to present something, and they fear to speak in public, and I don't want to make this assessment one to one. It should in front of the other students.

Another teacher mentioned the same problem in which she wished to conduct oral presentations as part of the assessment procedure in EG classes; however, the students complained that they do these presentations in speaking classes, and they do not want to be burdened with that during grammar assessment tasks as well.

Quaternary contradictions also exist between the two systems within the activity network: teachers' beliefs and their practices concerning EGA. This occurs because teachers' beliefs about how grammar should be assessed and how they actually assess EG are incongruent. The presence of this belief-practice gap in EGA is attributed to various reasons, namely educational policies of assessment and absence of professional development – assessment training.

Findings (section 5.2.1.1) indicate the participating teachers believe that grammar should be assessed integratively, preferably through spoken and written tasks which require lengthy and comprehensive language production. In their EGA practices, however, teachers tend to rely entirely on assessing EG explicitly through discrete-point items. Divergence between teachers' beliefs and practices resulted first from the rules mandated by the educational institutions which steer teachers' practices towards formal summative classroom-based assessment. Another factor contributing to beliefs-practices divergence is teachers' lack of training. In cases where teachers have some degree of freedom in assessing EG by different means, for example creative portfolios or projects, teachers argue that they did not receive proper training on how to use these assessment methods, hence they could not employ them even if they want to and are allowed to do so.

Reports about incongruity between teachers' beliefs and their practices has been presented in several other studies. Büyükkarcı (2014), for example, found that teachers had positive perceptions of formative assessment. However, in their practices they mostly used exam papers (summative methods) because of large class size, teaching loads and time constraints. A similar mismatch between teachers' beliefs and practices with regards to assessment was reported in Davis & Neitzel's (2011) study in which the research indicated that divergence between beliefs and practices resulted from teachers' need to comply with school regulations and to attend to accountability requirements.

In sum, this section discussed the main contradictions identified within the activity network. In this study it was revealed that various types of contradictions occurred within and between the activity systems. The first one was identified within the division of labour (primary). The second type of contradiction existed between the mediated artefacts and the subject and the

mediated artefact and the object (secondary). The last type of contradiction was related to teachers' beliefs in the first activity system and their practices in the second activity system (quaternary). These contradictions were categorized either as dilemmas, conflicts, critical conflicts or double binds. On a heartening note, it is crucial to point out here that these contradictions are not seen as obstacles but rather as a source for change and development, which may open up opportunities and call for novel solutions that can lead to transformations in teachers' assessment practices (Engeström, 1987).

6.4 Summary of the Chapter

This chapter focused on the theoretical framework adopted for the discussion of this research. AT was found to be an effective interpretive framework to conceptualize the findings of this study. First, the themes and categories from the results were presented and discussed in relation to AT components and in relation to previous studies. Second, AT shed light on factors which create contradictions and tensions within and between the activity systems. The next chapter presents the conclusion in which main findings are reported, the implications of this study, recommendations and suggestions for further research.

7. Conclusion, Implications and Suggestions for Future Research

7.1 Introduction

This study set out to explore EFL teachers' beliefs and practices regarding English grammar assessment (EGA) in higher educational facilities in Saudi Arabia using the underpinning research questions:

RQ1:

- a) What are EFL teachers' beliefs about how English grammar should be assessed?
- b) What are the factors which have helped shape these beliefs?

RQ2:

- a) How do EFL teachers actually assess grammar in their teaching environments?
- b) What are the factors, other than their beliefs, that have influenced their practices?

RQ3:

- a) What is the relation between EFL teachers' beliefs and their current practices?
- b) What are the factors which have led to a convergence or divergence between such beliefs and practices?

This chapter first outlines the key findings of the research which have arisen from the results which were presented in the earlier chapters and discussed in the preceding chapter. The contributions of the research include its significance to knowledge within its context and its contribution to theory. The implications of this study in relation to the wider research community as well as to educational practices are presented as well. Recommendations arising from the research then follow, divided into three parts: recommendations relating to the higher educational sector in general, recommendations relating to the specific higher educational contexts studied, and recommendations relating to higher educational faculty

members. The limitations of the study are next discussed followed by suggestions for further research. Finally, a summary of the chapter is provided.

7.2 Key Findings and Contributions to knowledge

This section reports on the most interesting findings obtained. These findings are presented according to the sequence of the research questions. The first question explored what EFL teachers' beliefs about EGA are and what factors shaped these beliefs. The second research question aimed to examine how EFL teachers in Saudi Arabia currently assess EG in their classrooms and factors other than their beliefs that might have affected these practices.

Finally, the third research question investigated the relationship between what EFL teachers state about AEG and what they actually do in AEG. The main answers to these research questions are presented below.

7.2.1 EFL teachers' beliefs about EGA and factors shaping these beliefs

In order to explore teachers' beliefs and understand the factors which contributed to shaping these beliefs, questionnaire and semi-structured interviews were used as data collection methods. Broader themes and sub-themes, both pre-determined and emergent, were presented based on all the teachers' questionnaire and interviews responses.

The findings of the study suggest that the collective subject in the activity system (EFL teachers) desire to assess English grammar (the object) integratively using spoken and written language production tasks. The desired outcome would have been assessment in the form of oral presentation, essay writing and probably research. It has been found that subject's prior instructional experiences and their educational qualifications have an impact on the activity, and they are considered as sources of their current value beliefs. Therefore, the subject of the

activity shaped and refined their beliefs through knowledge they acquired throughout their learning and teaching years.

This way of looking at teachers' beliefs within a conceptualized theoretical framework portrays EFL teachers' beliefs about EGA as a dynamic activity system in which all components interact between each other and amongst one another causing contradictions and tensions within the activity system. This approach on cognition and assessment processes enriches the research cycle of LTC and contributes to the study of EGA which seems to be lacking in the field of educational research.

7.2.2 EFL teachers' practices of EGA and factors influencing these practices

In RQ2, the focus was on teachers' professional practices of EGA through the accounts of EFL teachers from four public higher educational contexts. Semi-structured interviews, retrospections and document were used to collect data about teachers' EGA practices.

Based on the findings, teachers use classroom-based assessment, namely written exams, simply to provide grades to students. It is noteworthy that in an exam-oriented country where this study took place, emphasis is placed more on the scores as outcomes and measures of abilities; therefore, it would be natural to expect the teachers to employ EGA as a summative tool to grade the students' performance. In addition, teachers differed in the focus of EGA question items, some questions targeted language production which concentrated on the sentence level while others looked at rules recognition without referring to any particular context. Although commonly practiced in higher educational contexts in Saudi Arabia, some faculty members doubted the effectiveness of traditional forms of assessment. They showed

how traditional ways of assessing EG hardly allowed the students to progress in their language learning.

The findings also revealed a number of interesting factors which influenced the current EGA status. In particular, rules and regulation dictated by the higher educational sector, teachers' instructional experiences, teaching methods, teachers' professional development (assessment training) and class parameters including students' readiness were amongst the most significant factors affecting teachers' practices of EGA. These factors, sometimes, prevented the subject (i.e., EFL teachers) from enacting their beliefs towards the attainment of the object (i.e., EGA).

It was apparent that extensive explicit rules and regulations created tensions within teachers who desire to assess EG through means other than written exams. This implies that relying on the traditional forms stems from following rules and not from their own conviction of the effectiveness of assessing grammar through written exams. Also, some faculty members complained that they lacked the authority in terms of changing assessment methods or incorporating other assessment means in their classroom-based assessment practices.

In addition, teachers' lack of knowledge on how to implement other means of assessment (e.g. portfolios) and students' reluctance to be assessed by means other than written exams hinder the teachers from conducting various means of assessment. However, these contradictions within and between system elements could act as motivation for improving teachers' assessment training programs and reforming educational policy of assessment practices.

Moreover, because part of the study focuses on assessment practices, and assessment is carried out by all teachers, findings from the study provide relevant insights into the thinking which happens when teachers write their grammar exams. It is possible to assume that exam writing thinking resonates with all teachers irrespective of subject specialism. Considering the findings of this study, its relevance contributes to the subfield of teacher cognition, thinking.

7.2.3 The relationship between EFL teachers' beliefs and practices with regards to EGA

As far as the third RQ was concerned, both congruence and tensions between teachers' beliefs and practices were found. Teachers' beliefs were greatly congruent with practices regarding the purposes of EGA, preferred item formats, sources from which items are best adopted and teachers' role in constructing the written exams. Conversely, teachers' beliefs were incongruent concerning the method by which EG is assessed. From an Activity Theory perspective, therefore, the subject within the system appears to be oriented towards different objects and thus sources of conflict appear and result in different outcome. These conflicts may result in either discouragement among the teachers to seek employing assessment means other than written exams, or they became the driving forces of teachers' practices to expand (Engeström 2001; Engeström and Sannino 2010), as they caused the teachers to question and try to develop the current practices (i.e., written exams) by adopting other assessment means (e.g., research, oral presentations ...etc.).

In addition, the study found several contextual factors related to the overall educational context which caused the mismatches between teachers' beliefs about and practice of EGA, namely educational institutions (e.g. rules and class size), teachers (e.g. teachers' assessment training), and students (e.g. learners' readiness).

The exploration of the factors affecting EFL teachers' beliefs and practices with regards to EGA (RQ3) provides a more in-depth understanding of the extent to which teachers' views are encapsulated in their practice. As the findings of the present study revealed (section 5.3.2), teacher EGA beliefs and practices are very complex systems in which inconsistency existed due to the various contextual factors that prevented teachers from enacting their beliefs into practices. However, the exploration of the contradictions between teachers' beliefs and practices could be used as a springboard for devising meaningful teachers' professional development programs (Golombek and Johnson, 2004).

7.2.4 Main contributions of the study

At the start of this study (chapter 1) I was at pains to make clear the distinction between the present study and research in contiguous but distinct fields. I believe the findings of the study have fully justified my premise that there exists an under-researched area of EFL teachers' beliefs and practices with regards to EGA. In addition, the research sought to highlight the existence of contradictions within the activity network and contributed to the work of activity theory in the field of LTC and EGA.

7.2.4.1 Practical contribution to the research context

The study highlights the importance of LTC research in EGA which is quite lacking in the context of Saudi Arabia, in particular, and other educational contexts in general. The readers can clearly see how this research contributed to both the body of knowledge in the field and to higher educational assessment practices in Saudi Arabia. Since the second research question dealt with contextual practices of EGA in higher educational facilities in Saudi Arabia, in section 4.3 of the research, I provided readers with basic understanding of Saudi Arabia's higher educational context. A final point to make is that involvement in the study

brought practical benefits to the participants by helping them notice aspects of their assessment and stimulating critical reflection. This reminds us that teacher cognition researchers, through working closely with teachers, can also contribute to their development.

7.2.4.2 The contribution to LTC research

The present study provides several contributions related to the field LTC research, confirming Borg's claim that understanding teachers' cognition is essential to the process of understanding their practices (2006). Only by exploring EGA from teachers' perspective can we begin to comprehend the complexities of teachers' cognitions and how they make sense of their assessment practices. This study, by seeking to understand this complexity, shows the multi-faceted nature of EGA when seen from teachers' perspectives. Moreover, in this study, practices were broadly conceptualized to include not only the teachers' processes of writing exams but also the factors which governed teachers' behaviour and derived their decision-making. In addition, the present study has not only examined the mis/match that exists between teachers' beliefs and practices, but it has also highlighted the potential tension(s) that existed in the teachers' beliefs and practice systems (see the section below).

7.2.4.3 The contribution to theory

The use of third generation activity theory (see Chapter 6) as a theoretical framework has contributed to the exploration of the beliefs and practices activity systems of EFL teachers in EGA, in particular the EFL teachers in four public higher educational facilities in the Saudi context. It has revealed important results on a number of issues in the relationship between the activity systems that have an impact on how EFL teachers perceive EGA and how they actually do it. One of the most important of these is the recognition of the contradictions and

tensions which illustrate that the activity of EGA is a multidimensional developmental process, in which teachers mediate and negotiate their views about EGA in wider institutional contexts. Through the lens of AT, it was also possible to view contradictions and tensions as an opportunity for developing EFL teachers' skills in EGA (see section 7.3).

7.3 Pedagogical Implications and Recommendations

Several pedagogical/educational implications of the study can be identified. They concern many of the stakeholders in any university assessment situation from the highest university officials down to the ordinary lecturers. While many of these are practical suggestions suiting local conditions, several implications can also be drawn from this study which provide practical recommendations for EGA teachers and educational administrators more widely.

The study was of what teachers believe with respect to EGA, and their EGA practices, and reasons for those. It was therefore descriptive and explanatory, not evaluative. I have aimed to illuminate those areas but not rule on which are good or bad. Therefore, the research does not immediately yield recommendations about how EGA should be changed in the contexts studied so as to become 'better'.

What can be instead suggested is some measures that I believe would ease the process of thought, communication and negotiation about EGA between the key agencies involved, and indeed within the main agent that the study was concerned with, the teacher. These suggestions, prompted by our findings and AT analysis, I imagine would be relevant in almost any context, not just my own.

First, it is strongly recommended that the relevant authorities encourage teachers' collaboration with other relevant agencies within the university, especially the Deanship of Students' Affairs and the University council work on and finally sanction the regulation of study and examination (RSE), to provide an opportunity for them to work together to explore ways in which they can act on reforming ideas, as well as find solutions to tensions created at the level of the classroom-based assessments in the educational context. At present it is apparent that, for some teachers at any rate, there were clashes between their own beliefs and the practices forced on them which ultimately came not from the coordinators or head of department, but the rules, as described in section 4.3.1.3, which at present operates very much in top-down mode, with no dialog with teacher (or coordinator) behind their decisions about how the teaching process should be conducted, including the assessment types, frequency, medium etc. Without some dialog between the university sectors with the power, and the relatively powerless teachers, bitterness and frustration may ensue and fruitful changes in assessment may not take place. In AT terms, I am in a sense therefore calling for less division of labour, and for the university top management to redefine the concepts of authority and power in relation to how assessment is conducted and improved.

Second, the study also suggests that institutional support is needed to foster better understanding of alternative assessment approaches in higher educational institutions. It is the responsibility of certain university sectors, as explained in the previous paragraph, to constantly review, and when necessary reform, their rules and regulations of exams. However, for this to occur, relevant knowledge needs to be established and updated, in our case specifically of EGA. It was apparent that some of our participants lacked such expertise and few had had any relevant pre-service training specifically in testing or assessment. This therefore implies the need for in-service training / continuing professional development

(CPD) to be provided of a type that exposes teachers and all relevant agencies to the latest ideas in assessment and testing, at a practical level. In that way a dialog between the key agencies such as I suggested earlier can be conducted that is properly informed.

On light of what has been stated above, it would be prudent to initiate some sort of council which includes countries that carry out the same principle of assessment in higher educational facilities, for example, UAE, Egypt and China. Similar to the Bologna process, ministerial meetings have to be conducted to evaluate the assessment practices and suggest possible means to reform and develop the procedures and standards of assessment approved among the countries of this council to the end of improving learning outcomes.

Third, I would suggest that promoting teachers' reflective ability may allow implicit thoughts and beliefs to become explicit and subsequently to influence teachers' practices or enable them to voice their true opinions where the practice is imposed. Teachers' in-service training and continuing professional development (CPD) should not; therefore, be directed only at providing the latest information and wisdom about assessment but also at developing teacher's competence in reflecting about the educational assessment experiences that they have in their day to day work. This would then enable teachers to select sensibly among assessment ideas that they might hear about based on a proper understanding of what would work in their class. In this way they could engage in more effective assessment practices in order to provide high-quality educational outcomes.

7.4 Limitations of the study

Any research is bound to have limitations. Best and Khan (1989) state that 'limitations are those conditions beyond the control of the researcher that may place restrictions on the

conclusions of the study and their application to other situations' (p. 37). In this section, I indicate my awareness of the boundaries caused by time, place, the sensitive issue of assessment and other uncontrolled circumstances.

First, this study is limited by its subject specificity, EFL teachers in higher educational institutions in Riyadh, Saudi Arabia. I did draw my research sample from more than one institution; however, EFL participants from different regions in SA could have allowed the claim of representativeness of public universities across Saudi Arabia. In addition, I would say that there are institutions outside of Saudi Arabia around the world where some similar conditions to my research context might apply. Hence the findings might resonate in other contexts. This includes not only nearby EFL countries such as UAE but more distant ones like China and Malaysia where teachers are also heavily involved in the assessment practices.

Another limitation related to the participants of the study is the gender and the origin of the participants. The majority of the participants were female and Saudi. The first of those was of course influenced by the cultural norms in Saudi Arabia which makes it hard to conduct cross-gender studies where the researcher is of a different gender from the participants. This again makes it difficult in the strict sense to generalize the findings of this study beyond the populations I actually sampled. However, it is not a common finding of belief studies that genders differ substantially, so I may cautiously suggest that the findings might apply more widely.

Furthermore, the research study had to be conducted within a maximum period of three months. The short period did not allow for further investigation of other contexts which

would have doubtless yielded more insights into teachers' beliefs and practices of assessment in higher educational institutions more widely.

Another constraint that I experienced was imposed by the university authorities, who prohibited the use of think aloud data gathering. I had planned to gather think aloud data while teachers actually sat and constructed quizzes or exam papers. This might have yielded even better information than what I obtained from the retrospective reports on exam construction obtained some time afterwards, since it would have been obtained right at the time of performing the activity that was being researched. Later reports about what they did might have suffered from forgetting some of the details, post rationalisation of what was done, and halo effect²¹. However, because of the strict requirements concerning keeping the contents of exam papers secret, the university authorities did not allow this.

7.5 Potential for further research

Given the contributions and implications of this study as presented above, it is clear that there is a need for further research in this area. Thus, this study has opened up various areas worthy of future research:

First, there is a need to conduct longitudinal research to obtain in-depth understanding of EFL teachers' beliefs and practices towards educational assessment in general and course related assessment specifically as they change and evolve over time. For instance, do teachers early in their careers tend to exhibit more conflicts between beliefs and practices which disappear over time as they adjust their beliefs to practices that are imposed upon them? While the data

²¹ A cognitive bias which allows the transfer of feelings about one aspect of something to another (Thorndike, 1920)

in places hinted at this, it can only be established properly by longitudinal research where the same teachers are followed over a period of time. An AT perspective here would again be useful to reveal for instance the role of the community of practice in such changes/acclimatisation, in contrast with the internal reflective activity of the teacher subject herself.

Second, similar studies on EGA could be conducted in different contexts, in order to explore the extent to which the issues and conflicts in EFL teachers' beliefs and practices differ or coincide with the present context. These could be contexts both within and outside of the KSA and at different levels. For instance investigating English assessment in schools in the KSA may be a large contributor to illuminating teacher beliefs and practices in an AT framework which could be a valuable contribution to moving school assessment to become more effective and so have a backwash effect on English teaching and learning in school to help raise standards.

Third, further research should also go beyond teachers' beliefs, thoughts and subjective knowledge to investigate other elements of cognition, for example, teachers' identity, emotions and motivation. There is also work to be done on teachers' knowledge in the objective sense which I referred to in this study but did not measure systematically. If indeed experts can agree on a body of component pieces of information about assessment that every teacher should know - i.e. assessment literacy - then the way forward is open to create a test of such knowledge (in the objective sense, opposed to belief).

Fourth, the study looked only at the creation/development of assessment instruments. It did not consider the administration of the assessment, nor its scoring/correction. These deserve

attention also in future in order to fully understand the role of teacher cognition in assessment.

Finally, training in educational assessment is not a well-researched area and there is a need for further research into teachers' beliefs, practices and transformation with EFL teachers who attend training courses or workshops in assessment which aim to develop teachers' knowledge about various aspects of assessment. The impact on teacher beliefs of pre-service teacher education/training in general has often been found to be weak (Peacock, 2001 & Song 2014). However rather less is known about the impact of in-service training and particularly of training in relation to assessment. This then could also be a fruitful area for future attention, to measure any change in teachers' beliefs and practices after completing the training sessions.

7.6 Summary

In conclusion, the study has made a contribution to the state of knowledge about what EFL teachers do and what they state as knowing (i.e. believe) about assessing English grammar. It has also provided a clearer picture of the relationship between teachers' beliefs and their practices regarding EGA. It is hoped that further studies such as this can expand our knowledge of foreign language EGA. In addition, the innovative use of AT in this research proved to be informative by placing EFL teachers' beliefs and practices, and conflicts between them, into their broader interactive contexts and exploring the factors that support or hinder teachers' beliefs about, and practices of, EGA.

References

- Ağçam, R., & Kaya, E. (2017). Analysing prospective teachers' perceptions on assessment and evaluation practices in higher education, (2007), 2007–2008.
- Alavi, S.M., Kaivanpanah, S. & Danesh, F. (2019). A comparative study of writing assessment using Activity Theory-based assessment model (ATBAM) and a traditional approach. *Journal of English language Teaching and Learning University of Tabriz*, 11(23), 1-25.
- Anderson, J. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Armstrong, L. E., & Kraemer, W. J. (2015). *Introduction to Research Methods. ACSM's Research Methods*. Pearson Education. Retrieved from <https://doi.org/10.4324/9781315212791-1>
- Alpert, R., & Haber, R. (1960). Anxiety in academic achievement situations. *The Journal of Abnormal and Social Psychology*, 61(2), 207-215.
- Al-Seghayer, K. (2015). Salient key features of actual English instructional practices in Saudi Arabia. *English Language Teaching*, 8(6), 89-99.
- Baker, T. L. (1994). *Doing social research*. New York: McGraw-Hill.
- Bandura, A. (1982). The assessment and predictive generality of self-percepts of efficacy. *Journal of Behavior Therapy and Experimental Psychiatry*, 13, 195-199.
- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In *International handbook of research on teacher beliefs* (pp. 284–300).
- Basharina, O. K. (2007). An activity theory perspective on student-reported contradictions in international telecollaboration. *Language Learning & Technology*, 11(2), 82–103.
- Basturkmen, H., Loewen, S., & Ellis, R. (2004). Teachers' stated beliefs about incidental focus on form and their classroom practices. *Applied Linguistics*, 25(2), 243–272.

Retrieved from <https://doi.org/10.1093/applin/25.2.243>

Bell, J. (1993). *Doing your research project : A guide for first-time researchers in education and social science*. Open University Press.

Berge, O., & Fjuk, A. (2006). Understanding the roles of online meetings in a net-based course. *Journal of Computer Assisted Learning*, 22(1), 13–23.

Biggs, J. (1998) Assessment and Classroom Learning: a role for summative assessment?, *Assessment in Education: Principles, Policy & Practice*, 5:1, 103-110, DOI: 10.1080/0969595980050106

Binjumah, S. (2017). Using Activity Theory to Explore the Perspectives of Participants on an Initial Teacher Education Programme for Science Teachers in the Kingdom of Saudi Arabia. Retrieved Nov 20, 2019 form <https://ore.exeter.ac.uk/repository/bitstream/handle/10871/31032/BinjumahS.pdf?sequence=1&isAllowed=y>

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy, *Assessment in Education: Principles, Policy & Practice*, 25:6, 551-575, DOI: 10.1080/0969594X.2018.1441807

Borg, S. (1999). Studying teacher cognition in second language grammar teaching. *System*, 27(1), 19–31.

Borg, S. (1999). Teachers' theories in grammar teaching. *ELT Journal*, 53(3), 157–167.

Borg, S. (2003a). Teacher cognition in grammar teaching: A literature review. *Language Awareness*, 12(2), 96–108.

Borg, S. (2003b). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), S0261444803001903.

Borg, S. (2008). Teacher cognition and language education. *Continuum*, 183–186.

- Borg, S. (2009). Introducing language teacher cognition. *The Cambridge Guide to Second Language Teacher Education*, 163–170.
- Borg, S. (2015). A longitudinal study of the impact of the TKT on teachers' classroom practices. Final report to Cambridge English Language Assessment.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Quality Research Journal*, 9(2), 27–40.
- Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education*, 11(3), 301–318.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Brown, H. D. (2003). *Language assessment: Principles and classroom practices*. Pearson/Longman.
- Burns, R. B. (2000). *Introduction to research methods*. Pearson Education.
- Büyükkarcı, K. (2014). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction*, 7(1), pp.107–121.
- Buzan, T. (2010). *Memory Book: How to Remember Anything You Want*. Pearson Education.
- Chew, A., & Lee, I. (2008). Teachers' Beliefs and Practices of Classroom Assessment in Republic Polytechnic, Singapore.
- Clift, B., Hatchard, J., Gore, J. & Jordan, A. (2018). Introduction: How do we belong? Researcher positionality within qualitative inquiry. The Fourth Annual Qualitative Research Symposium, University of Bath.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. London: Routledge.
- Cole, M. (1996). *Cultural psychology: A once and future discipline*, Cambridge: Harvard

University Press.

Creswell, J. W. (2003). *Research design : Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.

Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Thousand Oaks, CA: SAGE Publications, Inc.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.

Davis, D. S., & Neitzel, C. (2011). A self-regulated learning perspective on middle grades classroom assessment. *Journal of Educational Research, 104*(3), 202–215.

Dayal, H., & Lingam, G. (2015). Teachers' Conceptions of Assessment. *Procedia - Social and Behavioral Sciences, 209*(8), 229–233.

<https://doi.org/10.1016/j.sbspro.2015.11.222>

Deluca, C., & Klinger, D. A. (2010). Assessment in education: Principles, policy & practice assessment literacy development : Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice, 17*(4), 37–41.

Deneen, C. C., & Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education, 3*(1), 1–14.

Denzin, N. K. (2009). *The research act : A theoretical introduction to sociological methods*. AldineTransaction.

Dippe, G. (2006). The missing teacher: Contradictions and conflicts in the experience of online learners. In *The fifth international conference on networked learning*. Lancaster: United Kingdom.

Dörnyei, Z. (2003). *Questionnaires in second language research : Construction, administration, and processing*. New York: Lawrence Erlbaum Associates.

Downing, S., & Haladyna, T. (Eds.). (2006). *Handbook of test development*. NY: Routledge.

- Edge, J. (1989). *Mistakes and correction*. UK: Longman
- Elshawa, N., Abdullah, A., & Rashid, S. (2017). Malaysian instructors' assessment beliefs in tertiary ESL classroom. *International Journal of Education & Literacy Studies*, 5(2), 29–46.
- Engeström, Y. (1999). Introduction. In *Perspectives on activity theory* (pp. 1–16). Cambridge: Cambridge University Press.
- Engeström, Y. (2001) Expansive learning at work. Toward an activity theoretical reconceptualization n. *Journal of Education and work*, 14: 133–156.
- Engestrom, Y., & Miettinen, R. (2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics*, 43(7), 960–974.
- Engeström, Y., & Sannino, A. (2011). Discursive manifestations of contradictions in organizational change efforts: A methodological framework. *Journal of Organizational Change Management*, 24(3), 368-387. doi:10.1108/09534811111132758
- Ericsson, A., & Simon, H. A. (1993). *Protocol analysis : Verbal reports as data*. MIT Press.
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—principles and practices. *Health Services Research*, 48(6 Pt 2), 2134.
- Figueiredo, S., Alves, M., & Silva, C. (2016). How teachers' perceptions affect the academic and language assessment of immigrant children. In *Proceedings of the ICCSBS 2016 - the annual international conference on cognitive - social, and behavioural sciences* (pp. 90–108).
- Fleer, M. (2016). The Vygotsky Project in education- the theoretical foundations for analysing the relations between the personal, institutional and societal conditions for studying development. In D. S. P. Gedera & P. J. Williams, eds. 230 *Activity theory in education: Research and practice*. Rotterdam: Sense Publishers, pp. 1–15.

- Fujioka, M. (2014). L2 student–U.S. professor interactions through disciplinary writing assignments: An activity theory perspective. *Journal of Second Language Writing* (25), 40-58
- Gorard, S. (2010). *Quantitative methods in educational research : The role of numbers made easy*. London: Bloomsbury Publishing.
- Gutbrie, L. I. (2005). *Teachers' beliefs about adopted formative assessment strategies in teaching miming in the primary school. A case study*. University of Newcastle Upon Tyne.
- Hatch, E., & Lazaraton, A. (1991). *The research manual*. Boston, MS: Heinle and Heinle.
- Herazo, J., Davin, K. & Sagre, A. (2019). L2 Dynamic Assessment: An Activity Theory perspective. *The Modern Language Journal*, 103(2), 443-458.
- Holtgreive, J. (2016). Too Smart to Fail? Inside Higher Ed. Retrieved June 26th, 2020 from <https://www.insidehighered.com/views/2016/08/16/students-focus-too-much-grades-detriment-learning-essay>
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge University Press.
- Hughes, A. (2010). *Testing for Language Teachers* (2nd ed). Cambridge University Press
- Inbar-Lourie O. (2008). Language assessment culture. In Hornberger N.H. (eds) *Encyclopedia of Language and Education*. Springer, Boston, MA
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Johnson, K.R. & Golombek, P.R. (2018). Making L2 teacher education matter. In Lantolf, J, et al *The Routledge Handbook of SCT and Second language development* (ch. 28). UK: Routledge.
- Jonassen, D. H. (2002). Learning as activity. *Educational technology*, 42(2), 45-51.
- Jones, C. (2005). *Assessment for learning*. Learning and Skills Development Agency.

- Kagan, D. M. (1990). Ways of evaluating teacher cognition: Inferences concerning the Goldilocks principle. *Review of Educational Research*, 60(3), 419–469.
- Kaptelinin, V. (2013). Activity Theory. In *The Encyclopedia of Human-Computer Interaction* (2nd Ed.). Interaction Design Foundation.
- Karim, B. H. H. (2015). The impact of teachers' beliefs and perceptions about formative assessment in the university ESL class. *International Journal of Humanities, Social Sciences and Education*, 2(3), 108–115.
- Karp, G. G., & Woods, M. L. (2008). Preservice teachers' perceptions about assessment and its implementation. *Journal of Teaching in Physical Education*, 27, 327–346.
- Kim, A. (2014). Examining How Teachers' Beliefs about Communicative Language Teaching Affect Their Instructional and Assessment Practices: A Qualitative Study of EFL University Instructors in Colombia. *RELC Journal Journal*, 45(453), 337–354.
<https://doi.org/10.1177/0033688214555396>
- Krashen, S. (1988). *Second Language Acquisition and Second Language Learning*. Prentice-Hall International.
- Kuutti, K. (1995). Activity theory as a potential framework for human-computer interaction research. In *Context and Consciousness:: Activity theory and human-computer interaction*. Cambridge, MA: The MIT Press.
- Larsen, D., & Butler, A. (2013). *Test-enhanced learning*. Walsh, K. (ed.). In *Oxford Textbook of Medical Education*. pp. 443–452.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength and computer adaptiveness. *Language Learning* 54(3), 399-436.
- Lenn, M. (1992). Global Trends in Quality Assurance in Higher Education. WENR.
Retrieved June 16, 2020 from
http://www.wes.org/ewenr/wenrarchive/QA_HighEdQualAssTrendsSpr92.pdf

- Leont'ev, A. (1978). *Activity, consciousness, and personality*. Englewood Cliffs: Prentice-Hall.
- Lilliefors, H. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318), 399.
- Lindseth, M. (2016). The Effects of Form-Focused Instruction on the Acquisition of Subject-Verb Inversion in German. *Foreign Language Annals*, 49(1), 10-22.
- Mackey, A., & Gass, S. (2005). *Second language research : Methodology and design*. Lawrence Erlbaum.
- Mansory, M. M. (2016). EFL teachers' beliefs and attitudes towards English language assessment in a Saudi university' s English language institute. *PhD Dissertation*, 1–220.
- Mehrgan, K., Candidate, P. D., & Language, E. (2017). Investigating the impacts of EFL teachers' age, educational background, instructional experience and gender on their beliefs about formative assessment, 5(18).
- Merriam, S. B. (1998). *Qualitative research and case study applications in education: revised and expanded from case study research in education*. San Francisco: Wiley.
- Montee, M., Bach, A., Donovan, A., & Thompson, L. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of the National Council of Less Commonly Taught Languages*, 13, 1–31.
- Moseholm, E., & Fetters, M. D. (2017). Conceptual models to guide integration during analysis in convergent mixed methods studies. *Methodological Innovations*, 10(2), 205979911770311.
- Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *PROFILE*, 14(1), 143–158.

- Mussawy, S. (2009). Assessment practices : Student's and teachers' perceptions of classroom assessment. *Center for International Education*, 1–115.
- Noble, H., & Heale, R. (2019). Triangulation in research, with examples. *Evidence-Based Nursing*, 22(3), 67–68.
- Önalın, O. (2018). JOURNAL OF LANGUAGE AND LINGUISTIC STUDIES A study on Turkish EFL teachers ' beliefs about assessment and its different uses in teaching English, 14(3), 190–201.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332.
- Peacock, M. (2001). Pre-service ESL teachers' beliefs about second language learning: A longitudinal study. *System*, 29, 177-95.
- Pereira, D., & Flores, M. (2017). University teachers' perceptions about assessment practices: A study in five Portuguese universities. 494–502.
- Popham, W. J. (2009). Assessment literacy for teachers : Faddish or fundamental ? 4–11.
- Pring, R. (2000). The false dualism of educational research. *Journal of Philosophy of Education*, 34(2), 247-260.
- Purpura, J. E. (2004). *Assessing grammar*. J. C. Alderson & L. F. Bachman, L, (Eds.). Cambridge University Press.
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(S1), 190–208.
- Qian, D. D., & Cumming, A. (2017). Researching English language assessment in China: Focusing on high-stakes testing. *Language Assessment Quarterly*, 14(2), 97–100.
- Rahim, S., Venville, G., & Chapman, A. (2009). Classroom assessment: Juxtaposing teachers' beliefs with classroom practices. Australian Association for Research in

- Education: International Education Research Conference. Retrieved June 14, 2020 from <http://aare.edu.au/09pap/abd091051.pdf>.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472–482.
- Richards, J., Gallo, P., & Renandya, W. (2001). Exploring teachers' beliefs and the processes of change. *The PAC Journal*, (1986), 47–48.
- Roth, W., & Tobin, K. (2002). Redesigning an 'urban' teacher education program: An activity theory perspective. *Mind, Culture and Activity*, 9(2), 108–131.
- Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, 17(6), 759–769.
- Saad, M., Sardareh, S., & Ambarwati, E. (2013). Iranian secondary school EFL teachers' assessment beliefs and roles. *Life Science Journal*, 10(3), 1638–1647.
- Sahinkarakas, S. (2012). The role of teaching experience on teachers' perceptions of language assessment. *Procedia - Social and Behavioral Sciences*, 47, 1787–1792.
- Sasaki, T. (2008). Concurrent think-aloud protocol as a socially situated construct. *International Review of Applied Linguistics in Language Teaching*, 46(4), 349-374.
- Saydee, F. (2016). Foreign language teaching: A study of teachers' beliefs about effective teaching and learning methodologies. *Journal of the National Council of Less Commonly Taught Languages*, 18, 63–91.
- Scholfield, P. (2005). Simple statistical approaches to reliability and item analysis. *Department of Language and Linguistics, University of Essex. Unpublished Manuscript.*
- Shannon-Baker, P. (2016). Making paradigms meaningful in mixed methods research. *Journal of Mixed Methods Research*, 10(4), 319–334.

- Sheehan, S., & Munro, S. (2017). *Assessment: Attitudes, practices and needs*. London: British Council.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Sikka, A., Nath, J., & Cohen, M. D. (2007). Practicing teachers beliefs and uses of assessment. *International Journal of Case Method Research & Application*, XIX(3), 239–253.
- Sogunro, O. A. (1997). Impact of training on leadership development. *Evaluation Review*, 21(6), 713–737.
- Song, S. (2014). Teachers' beliefs about language learning and teaching. In M. Bigelow, & J. Ennser-Kananen (Eds.). *The Routledge Handbook of Educational Linguistics*, (pp. 263-275). London: Routledge.
- Suwannasom, T. (2010). *Teacher cognition about technology mediated EFL instruction in the Thai tertiary context*.
- Swales, J. (1998). *Other floors, other voices: A textography of a small university building*. NJ: Lawrence Erlbaum.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Taras, M. (2005). Assessment - summative and formative - some theoretical reflections. *British Journal of Educational Studies*, 53(4), 466–478.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.
- Teague, R., De Jesus, K., & Ueno, M. N. (2001). Concurrent vs. post-task usability test ratings. In *CHI'01 extended abstracts on Human factors in computing systems - CHI'01* (p. 289). New York, New York, USA: ACM Press.

- Terrell, S. (2017). Mixed-methods research methodologies. *The Qualitative Report*, 17(1), 254–280.
- Thomas, M. (2012). Teachers' beliefs about classroom assessment and their selection of classroom assessment strategies. *Journal Research and Reflections in Education*, 6(2), 103–112.
- Tomlinson, C., & Moon, T. (2013). *Assessment and Student Success in a Differentiated Classroom*. Association for Supervision & Curriculum Development
- Van Den Haak, M. J., De Jong, M. D. T., & Schellens, P. J. (2003). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour and Information Technology*, 22(5), 339–351.
- Vygotsky, L. S. (1978): *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Wang, P. (2004). CHINESE SCIENCE TEACHERS' BELIEFS ABOUT AND PRACTICES OF ASSESSMENT by PING WANG (Under the Direction of Lynn A. Bryan).
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.
- Widdowson, H. (2003). *Defining issues in English language teaching*. Oxford: Oxford University Press.
- William, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.). *The future of assessment: Shaping teaching and learning*. (pp. 53-82). New York: Lawrence Erlbaum Associates.
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualizing teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241–256.

Wolf, D., Bixby, J., Glenn, J., III, & Gardner, H. (1991). To use their minds well:

Investigating new forms of student assessment. *Review of Research in Education*, 17, 31–74.

Woods, D. (1996). *Teacher Cognition in Language Teaching*. Cambridge: Cambridge University Press.

Woods, D., & Cakir, H. (2011). Two dimensions of teacher knowledge: The case of communicate language teaching. *System*, 39, 381-390.

Yin, M. (2017). Understanding Classroom Language Assessment Through Teacher Thinking Research. *Language Assessment Quarterly*, 7(2), 175–194.

<https://doi.org/10.1080/15434300903447736>

Appendices

Appendix A: Project Information Sheet



Department of Language and Linguistics

Researcher

Masahel AlSalem

Email: ma16709@essex.ac.uk

Supervisor

Dr. Adela Gánem-Gutiérrez

Email: aganem@essex.ac.uk

Project Title: Language Teacher Cognition: Exploring EFL Teachers' Beliefs and Practice
in relation to English Grammar Assessment

Purpose of the Study

The study aims at investigating what EFL teachers know, believe and think about how English grammar should be assessed and the exploring the factors that helped shape these beliefs. It also aims at comparing these beliefs to actual practice of English grammar Assessment.

Project Procedure

Participating in the study will involve the following: completing a questionnaire about beliefs regarding assessing English grammar, participate in interviews both on an individual basis with the researcher as well as in a group to discuss aspects of English grammar assessment

procedures. The participants will be observed during their teaching sessions, and these classes will be audio recorded. The participants will also be asked to take part in retrospective thinking sessions in which they share their thoughts about how they constructed their English grammar assessment tasks, and these will be audio recorded. Finally, a sample of their assessment tasks will be collected for analysis purposes only. All data will be anonymised and will only be used for research purposes.

Participants' Right

You have the right to:

1. Decline participation
2. Decline to answer any particular question without giving reasons
3. Withdraw from the study (at any stage) without giving any reasons or suffering any consequences
4. Ask for the audio recorder to be turned off at any time during the interview
5. Decline attending any group meetings
6. Decline giving a sample of your assessment tasks
7. Deny the research from attending all or any of your teaching sessions

Means of Contact

If you have any questions or comments regarding this research project, please feel free to email the researcher and/or the supervisor.

Thank you

27th April, 2018

Appendix B: Participant Consent Form



Department of Language and Linguistics

Researcher

Masahel AlSalem

Email: ma16709@essex.ac.uk

Supervisor

Dr. Adela Gánem-Gutiérrez

Email: aganem@essex.ac.uk

Project Title: Language Teacher Cognition: Exploring EFL Teachers' Beliefs and Practice
in relation to English Grammar Assessment

Tick the Box

- | | |
|---|--------------------------|
| 1. I confirm that I have read and understood the Information Sheet dated 27 th April 2018 for the above study. I have had the opportunity to consider the information, ask questions and have had these questions answered satisfactorily. | <input type="checkbox"/> |
| 2. I understand that my participation is voluntary and that I am free to withdraw from the project at any time without giving any reason and without penalty. | <input type="checkbox"/> |
| 3. I understand that the identifiable data provided will be securely stored and accessible only to the members of the research team directly involved in the project, and that confidentiality will be maintained. | <input type="checkbox"/> |
| 4. I understand that data collected in this project might be shared as appropriate and for publication of findings, in which case data will remain completely anonymous. | <input type="checkbox"/> |
| 5. I agree to take part in the above study. | <input type="checkbox"/> |

Participant's name:

Date

Signature

Researcher's name:

Date:

Signature

Appendix C: First Draft of the Questionnaire

Project Title: Language Teacher Cognition and English Grammar Assessment in the Educational Context

Dear teachers,

This questionnaire aims to explore your grammar assessment beliefs as an English language instructor. The responses you provide will strictly be used for research purposes only and remain at all time anonymous.

Part I: Demographic Information

1. **Name (optional):** _____
2. **Age:** 21-30 31-40 41-50 above 50
3. **Gender:** Male Female
4. **Country of origin:** Saudi Arabia Other: _____
5. **Working place:** KSU IMSIU PNU Other: _____
6. **Educational background:** BA MA PhD Other: _____
7. **English teaching experience:** 1-5 years 6-10 years 11-15 years more than 16 years
8. **English grammar teaching experience:** none less than a year 1-2 years 3-4 years
9. **Grammar exam writing experience:** none less than a year 1-2 years 3-4 years
10. **Grammar textbook used:** English Grammar in Use Understanding and Using English Grammar Interactions / Mosaic Grammar Other: _____
11. **Level of students:** Foundation Year levels 1+2 level 3 Level 4 Level 5

Part II: Please tick (✓) your response to the following statements by using the scale below.

SA= Strongly Agree A= Agree N= Neutral D= Disagree SD= Strongly Disagree

		SA	A	N	D	SD
1.	Assessing English grammar is important.					
2.	English grammar can be assessed through oral presentation.					
3.	Assessing English grammar determines students' mastery of learning.					
4.	English grammar assessment facilitates language learning.					
5.	English grammar can be assessed through cloze items.					
6.	Paper and pencil exam format provides valid evaluation of students' learning.					
7.	English grammar assessment provides evidence of pedagogical achievements to external reviewers.					
8.	English grammar should be assessed frequently during the term.					
9.	Assessing English grammar motivates students to learn.					
10.	Assessing English grammar is a waste of time.					
11.	English grammar assessment informs teaching (diagnoses strengths and weakness in teaching).					
12.	English grammar can be assessed through sentence transformation or production items.					
13.	English grammar can be assessed through editing tasks.					
14.	Assessing English grammar provides information about student progress.					
15.	English grammar assessment pressurizes teachers to complete their syllabi.					
16.	Computer technology helps in assessing students' grammatical abilities.					
17.	Objective testing (e.g., multiple choice items, matching items, cloze items) is a good method to assess English grammar.					

		SA	A	N	D	SD
18.	English grammar can be assessed through matching items.					
19.	English grammar is best assessed by expert professional testers /examiners rather than a class teacher.					
20.	English grammar assessment serves a number of purposes.					
21.	English grammar can be assessed through multiple choice items.					
22.	English grammar assessment should systematically target student knowledge of how different functions/meanings are expressed through English grammar (e.g. how an event in future time can be expressed, or how to make polite requests with <i>could you?</i> , <i>may I?</i> <i>I wonder if...?</i> etc...)					
23.	English grammar can be assessed through essay writing.					
24.	English grammar assessment helps to group students for instructional purposes.					
25.	English grammar assessment is irrelevant to language learning.					
26.	Best grammar assessment items are the ones developed by the course instructor.					
27.	Exam questions should reflect real life language use.					
28.	grammar exams are best prepared collaboratively.					
29.	English grammar assessment holds teachers accountable for their teaching.					
30.	Assessing English grammar is difficult to do well.					
31.	Ready-made grammar exercisers found on the internet are a good source for constructing grammar exams.					
32.	English grammar is best tested integratively with other aspects of the language.					
33.	English grammar assessment holds students accountable for their learning.					
34.	Grammar can be assessed through open response sentence completion.					

		SA	A	N	D	SD
35.	Grammar exercises from published textbooks are a better source for constructing grammar exams than those found on the internet.					
36.	Grammar exams should target specific elements of English (discrete-point aspects).					
37.	English grammar can be assessed through error recognition.					
38.	English grammar exams target the structural-functional aspect of the English language.					
39.	English grammar can be assessed through oral interview.					
40.	English grammar assessment should systematically target student knowledge of common grammatical terms such as <i>verb, object, dependent clause</i> .					
41.	English grammar assessment should target student explicit knowledge of grammatical rules (e.g. '-s has to be added to a verb in the simple present when the subject is third person singular', 'days of the week take <i>on</i> while months and years take <i>in</i> ')					
42.	English grammar should be assessed at the end of term.					
43.	English grammar can be assessed through self-assessment tasks.					
44.	Grammar errors are only important when they get in the way of successful communication of the message being conveyed					
45.	Using previous grammar exam items is a good source to construct grammar exams.					
46.	English grammar can be assessed through peer-assessment tasks.					
47.	Subjective testing (e.g., short essay, sentence completion) is a good method to assess English grammar.					
48.	English grammar assessment should primarily be concerned with student ability to understand and use English grammar in communication effectively and spontaneously (as a native					

		SA	A	N	D	SD
	speaker does), not their conscious knowledge about the language.					
49.	English grammar assessment provides feedback to the students as they learn.					
50.	English grammar can be assessed through _____ (please specify and respond accordingly)					

Thank you 😊

Appendix D: Final Draft of the Questionnaire

Project Title: Language Teacher Cognition and English Grammar Assessment in the Educational Context

Dear teachers,

This questionnaire aims to explore your grammar assessment beliefs as an English language instructor. The responses you provide will strictly be used for research purposes only and remain at all times confidential.

Part I: Demographic Information

1. **Age:** 21-30 31-40 41-50 above 50
2. **Gender:** Male Female
3. **Country of origin:** Saudi Arabia Other: _____
4. **Current Working place:** KSU IMSIU PNU Other: _____
5. **Educational background:** BA MA PhD Other: _____
6. **English teaching experience:** 1-5 years 6-10 years 11-15 years more than 16 years
7. **English grammar teaching experience:** none less than a Year 1-2 years 3-4 years
8. **Experience in grammar assessment** none less than a Year 1-2 years 3-4 years
9. **Training received on language assessment:** none undergraduate courses professional training Other: _____

10. **Approximate no. of students in a class:** less than 20 21- 30 31- 40 above 40
11. **Grammar textbook used currently:** English Grammar in Use Understanding and Using English Grammar Interactions / Mosaic Grammar Other: _____
12. **Level of students currently taught:** levels 1+2 level 3 Level 4 Level 5

Part II: Please tick (√) your response to the following statements by using the scale below.

SA= Strongly Agree A= Agree N= Neutral D= Disagree SD= Strongly Disagree

Section A: Teachers’ Beliefs about the General Nature of English Grammar Assessment

In general, assessing English grammar _____.		SA	A	N	D	SD
		5	4	3	2	1
1.	is important					
2.	is difficult to do well					
3.	can serve a number of purposes					
4.	is irrelevant to language learning					
5.	pressurizes teachers to complete the syllabus or textbook assigned to the course					

Section B: Teachers’ Beliefs about the Purposes of English Grammar Assessment

In general, the purpose of English grammar assessment should be to _____.		SA	A	N	D	SD
		5	4	3	2	1
1.	determine students’ mastery of what they have been taught in an English grammar course					
2.	motivate the students to learn English grammar					

In general, the purpose of English grammar assessment should be to _____.		SA	A	N	D	SD
		5	4	3	2	1
3.	provide information about how well each student is progressing in English grammar					
4.	inform teaching by showing the students' strengths and weaknesses in English grammar					
5.	help place students into groups for English grammar instruction, suited to their ability					
6.	provide feedback to the students on their strengths and weaknesses in English grammar as they learn					
7.	indicate teachers' ability in teaching English grammar for department chair-persons or external reviewers					
8.	indicate learners' ability in learning English grammar for department chair-persons or external reviewers					

Section C: Teachers' Beliefs about English Grammar Assessment Methods

Concerning the content and delivery of English grammar assessment, in general I believe that _____.		SA	A	N	D	SD
		5	4	3	2	1
1.	paper and pencil assessment provides valid evidence of students' learning of English grammar					
2.	computer technology helps in assessing students' English grammatical abilities					
3.	English Grammar assessment should use means that reflect real life language use (not disconnected sentences or words)					
4.	Grammar assessment should target specific elements of English grammar in separate items (discrete-point aspects)					
5.	English Grammar is best assessed integratively along with other aspects of English (e.g. through speaking or writing tasks), rather than as a separate skill					
6.	objective assessment (e.g., through scores from sets of multiple choice items or cloze gap filling items etc...) is a good method to assess English grammar					

Concerning the content and delivery of English grammar assessment, in general I believe that _____.		SA	A	N	D	SD
		5	4	3	2	1
7.	subjective assessment (e.g., by rating overall grammar quality in a short essay or oral presentation) is a good method to assess English grammar					
8.	Grammar errors are only important when they get in the way of successful communication of the message being conveyed					
9.	English grammar should be assessed frequently during the course					
10.	English grammar should be assessed at the end of the course					
11.	English grammar assessment should systematically target the different structural/formal features of English (e.g. the articles, how <i>do</i> is used in questions and negatives, relative clause formation)					
12.	English grammar assessment should systematically target student knowledge of how different functions/meanings are expressed through English grammar (e.g. how an event in future time can be expressed, or how to make polite requests with <i>could you?</i> , <i>may I?</i> <i>I wonder if...?</i> etc...)					
13.	English grammar assessment should systematically target student knowledge of common grammatical terms such as <i>verb</i> , <i>object</i> , <i>dependent clause</i>					
14.	English grammar assessment should target student explicit knowledge of grammatical rules (e.g. '-s has to be added to a verb in the simple present when the subject is third person singular', 'days of the week take <i>on</i> while months and years take <i>in</i> ')					
15.	English grammar assessment should primarily be concerned with student ability to understand and use English grammar in communication effectively and spontaneously (as a native speaker does), not their conscious knowledge about the language					

Section D: Teachers' Beliefs about English Grammar Assessment Formats

In general, English grammar can usefully be assessed through _____.		SA	A	N	D	SD
		5	4	3	2	1
1.	speaking in an oral interview					
2.	speaking in an oral presentation					
3.	filling cloze gaps in text					
4.	sentence transformation or production items					
5.	matching items					
6.	multiple choice sentence completion					
7.	open response sentence completion					
8.	essay writing					
9.	editing / error correction tasks					
10.	error recognition / grammaticality judgment tasks					
11.	Others... please identify _____					

Section E: Teachers' Beliefs about Their Role and the Sources Used in Constructing

English Grammar Assessment Tasks

		SA	A	N	D	SD
		5	4	3	2	1
1.	The best grammar assessment items are the ones developed by the course instructor.					
2.	English grammar is best assessed by expert professional testers /examiners rather than a class teacher.					
3.	Assessment of a student's English grammar performance by their peers is useful.					
4.	Self-assessment by students of their own English grammar performance is useful.					
5.	English grammar assessment tasks are best prepared collaboratively.					
6.	Ready-made English grammar exercises/ tests found on the internet are a good source for grammar assessment tasks.					
7.	English grammar exercises from published textbooks are a useful source for constructing grammar assessment tasks.					

8.	Using English grammar assessment items from previous years is a good source to construct grammar assessments.					
----	---	--	--	--	--	--

Appendix E: Interview Questions for EFL teachers who currently teach Grammar (First Draft)

This study will investigate LTC about Grammar assessment and classroom assessment practices.

1. First, I want to know how many foreign languages you learned.
2. Tell me about your experience in learning this language, please.
3. Tell me about English grammar, how were you taught and assessed when you were a student?
4. What kind of grammar assessment procedures do you remember as a student?
5. Do you think the way you used to be assessed as a student affected your current grammar assessment practices? How?
6. You teach English Grammar this term, can you tell me what we mean by grammar assessment, please?
7. Why do you assess grammar?
8. How do you think grammar should be assessed?
9. I want to know How you assess grammar in your department (i.e., the choice of methods), and why?
10. What do you think of the current grammar assessment procedures in your department?
11. What would you like to change in the current grammar assessment practices? Why?
12. In your opinion, do you think there are better assessment strategies to assess grammar other than exams? What are they, and why are they better?
13. Do you think that your department may allow you to induce any new grammar assessment methods (e.g., self-evaluation, interviews...etc)?

14. What is your role in the construction of grammar exams?
15. What do you do when you start writing your grammar exams? Please walk me through your thinking process as best as you can?
16. Are there any specific decisions you have to make with regards to your grammar exams (e.g., date, duration, marks distribution)?
17. Do you have any freedom when writing your grammar exams?
18. Do you explain to your students how the exam will be like? Do you give them any mock exams ahead? Why or why not?
19. How do you make sure your exams are fair and valid?
20. What power do the course coordinator and the vice chair-person of the department have over your written grammar exams?
21. Could any other teacher write the grammar exams and hand it to you to administer?
22. Do you think teachers should concentrate on teaching only and leave the exam preparation to an assessment committee?
23. Would you please tell me about any challenges you may face when writing the grammar exams?
24. What strategies do you use to overcome these challenges?
25. Have you ever attended any professional development program about assessment? If yes what was it about, and did it change the way you believe grammar should be assessed? How?
26. Do you think these current grammar assessment practices in the form of written exams lack the accuracy to assess the students' grammatical abilities? Why or why not?
27. Would like to add anything before we finish this interview?

Appendix F: Interview Questions for EFL teachers who currently hold the post as coordinators for the English Grammar Course in their departments (First Draft)

This study will investigate LTC about Grammar assessment and classroom assessment practices.

1. I understand you are the coordinator of the English grammar course in this department, how long have you been the coordinator of the Grammar course?
2. What does being a coordinator involve? What is your job as a course coordinator?
3. Tell me about your experience as Grammar course coordinator?
4. What challenges have you faced as a course coordinator? Why?
5. How do you overcome these challenges? (e.g., if a teacher does not accept your proposed changes to their exams)
6. Let us talk a bit about your experience as a language learner, tell me about English grammar, how were taught and assessed when you were a student?
7. What kind of grammar assessment procedures do you remember as a student?
8. Do you think the way you used to be assessed as a student affected the way you evaluate the teachers' current grammar written exams? Why or why not?
9. You check and evaluate the teachers' written exams on grammar, can you tell me what do you assess in these exams, please?
10. Why do you assess grammar in the forms of written exams?
11. In your opinion, how should grammar be assessed?
12. What do you think of the current grammar assessment procedures in your department?
13. What would you like to change in the current grammar assessment practices? Why?

14. In your opinion, do you think there are better assessment strategies to assess grammar other than exams? What are they, and why are they better?
15. Do you think that the DELT may allow you to induce any new grammar assessment methods (e.g., self-evaluation, interviews...etc)?
16. What do you do when you evaluate the grammar exams submitted by the course teachers? Please walk me through your thinking process as best as you can.
17. Do you explain to your grammar teachers why you want them to make any changes to their written exams or do you have the authority to make changes on these exams without going back to the teachers?
18. How do you make sure these exams are fair and valid?
19. Could any other teacher write the grammar exams and hand it to the other teachers to administer?
20. Do you think teachers should concentrate on teaching only and leave the exam preparation to an assessment committee?
21. Have you ever attended any professional development program about assessment? If yes what was it about, and did it influence the way you evaluate the grammar exams submitted by the course teachers? How?
22. Do you think these current grammar assessment practices in the form of written exams lack the accuracy to assess the students' grammatical abilities? Why or why not?
23. Would like to add anything before we finish this interview?

**Appendix G: Interview Questions for a) EFL teachers who
currently teach English grammar course and b) current course
coordinators, if any (Final Draft)**

Q1a/b: What do you think grammar assessment means?

Q2a/b: In your opinion, how should grammar be assessed? (Types)

Q3a/b: Why should grammar be assessed in that way? (Purposes)

- Were you assessed this way when you were students?
- Does the administration have a saying in that?
- Have you taken any training courses with regards language assessment?

Q4a/b: How is grammar assessed here in your context? Why? (Type & Factors)

Q5a/b: What abilities of students do you intend to assess?

Q6a: What is your role as a course teacher when it comes to constructing assessment tasks?

Please walk me through the process.

- What decisions do you have to make?
- What influenced these decisions?
- What challenges do you frequently face?
- How do you overcome these challenges?

Q6b: What is your role as a course coordinator when it comes to the constructed assessment tasks?

Q7a/b: How do you feel about the current practice of grammar assessment?

- Would you suggest any changes? Why or why not?

Q8a/b: Do you believe that course teachers should be the ones to construct assessment task?

- Can the assessment tasks be constructed by one teacher and passed on to the others to administer?

Would be with or against having an assessment committee in charge of constructing the assessment tasks and handing them over to the teachers to administer?

Appendix H: Information about the Digital Voice Recorder

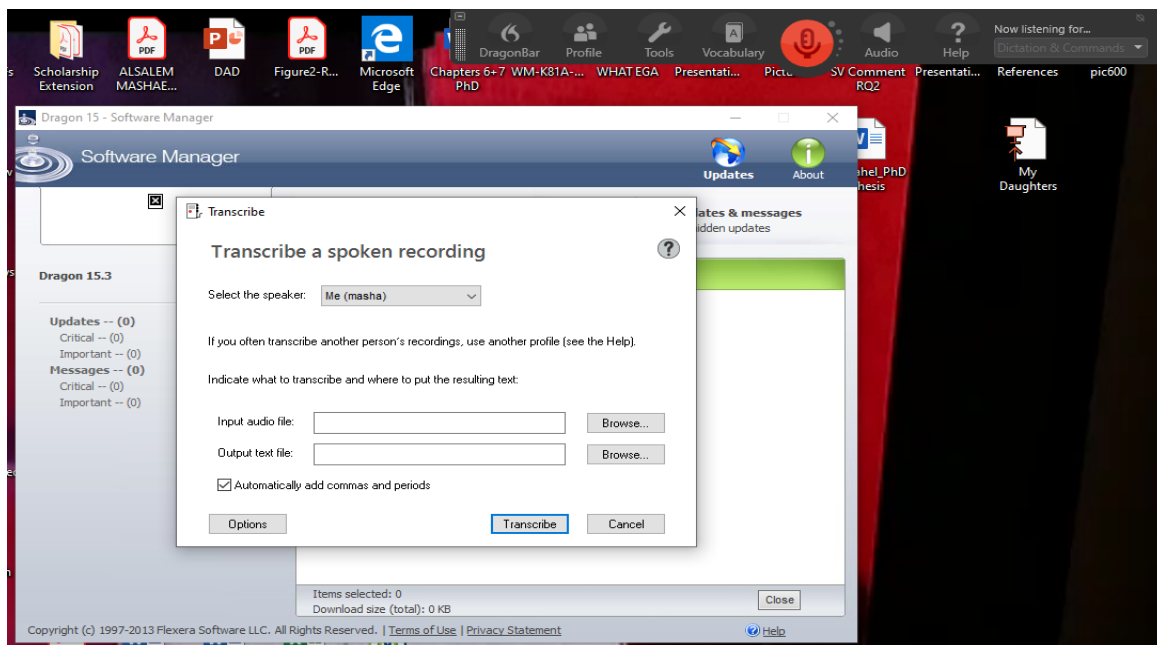
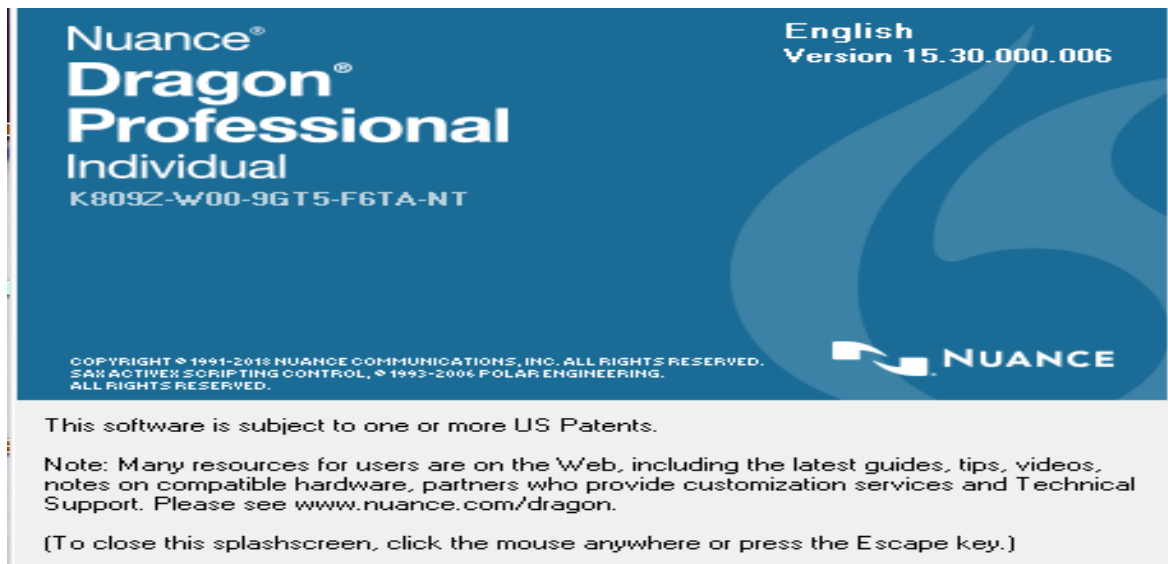


- **【Powerful Memory & Speed Adjustment】** - This small voice recorder with 8GB internal memory let you store up to 576 hours of recordings (record at 256kbps/ 32KB/s) or 1500 songs. It is 16 playback speed design and support fast forward and rewind during playback, which makes you more comfortable and convenient to adapt different voice speed, to meets different needs to listen to the sound recording or music;
- **【Excellent Warranty】** - Just connect the recorder to your computer using the supplied USB cable and drag the files over to your PC/MAC to make files transferred. Please read carefully the precautions in the manual before use. 1-year warranty, if you have any further questions about SKEY dictation machine, please feel free to contact us.
- **【Ambient Noise Reduction & MP3 Player】** - Upgrade Integrated intelligent professional audio noise reduction IC provide provides the rate of up to 1536kbps, SKEY digital voice recorder efficiently helps reduces ambient noise for clearer recordings with 2 built-in microphones, can record clearly even long distances(maximum 50 M). Upgrade high quality aluminum metal case let you enjoy light and ultra-portable experience, supports MP3,WMA,OGG,APE,FLAC,WAV music formats, can be use as MP3 player;
- **【High-efficiency & One Button Recording】** - Go out of the old style Dictaphone needing complicated operation, we bring you convenience to avoid troublesome while you can set recording automatically at a certain time. It is easy to stop and save by switching the middle button to avoid recording with a mistake of operation. It is easy to operate for those who are not good at machines, even beginners, elderly people and so on. Name files by date and time automatically, easily found and managed;
- **【A-B Repeat Playback & Wide Use】** - Automatic voice reduction function, the dictaphone voice recorder starts recording when sound is detected or turn to standby state. The A-B Repeat and Variable Speed Play are ideal helper for listening and singing. And the monitoring function, the segmented recording, and the timer recording, suitable for Students, Doctors, Lawyers, Writer, Journalists, businessmen ect. They use dictaphone for lectures, class, conferences, evidences, music, interviews ect;

Information retrieved @ https://www.amazon.co.uk/SKEY-Dictaphone-Microphone-Professional-Rechargeable/dp/B07HBTHVMV/ref=pd_sbs_23_t_0/259-2230062-2056401?encoding=UTF8&pd_rd_i=B07HBTHVMV&pd_rd_r=cd421461-43db-4cf4-b700-7790f7a7be1f&pd_rd_w=xnFld&pd_rd_wg=t2JOz&pf_rd_p=e44592b5-e56d-44c2-a4f9-dbd09b29395&pf_rd_r=MWT31BXB4J9AQTD AEXJ3&psc=1&refRID=MWT31BXB4J9AQTD AEXJ3

Appendix I: Dragon- speech to text software @

<https://www.nuance.com/dragon/business-solutions/dragon-professional-individual.html>



Appendix J: Sample of a Complete Interview Transcription

Interview #	004	Gender	F
Institution:	KSU (COLT)	Duration:	26 Mins 09 Secs
Date /Time:	26 Sep 2018	Position:	Teacher
Participant's code name:	RSh		

M: Let's first talk about you educational and teaching experience

RSh: I have BA in Applied Linguistics and masters in TESOL. I've been teaching English for 16 years. 5 years in schools and 11 years here in the College of Languages & Translation

M: How long have you been teaching Grammar?

RSh: Grammar 5 years. Let's say I've taught grammar in different approaches. Like when I used to teach in other schools never taught grammar explicitly. All indirectly through reading and other language skills. But I taught like 4 no six semesters so three years here. I teach explicit grammar courses.

M: What are the grammar courses you taught or teaching?

RSh: I taught grammar 2 and 3. I now teach grammar 3 which is all about clauses. Oh right I taught another grammar in Eng. 101 where we taught basic grammar rules for students in other department

M: Since you've been teaching grammar for so long that must have involved the process of assessment as well Let's focus a bit on grammar assessment, what do you think grammar assessment means?

RSh: Assessment as we do it here in the department or in general my thought?

M: Your thoughts please.

RSh: So in my head not the head of the department. It is to check how advanced our learners are in the English language or their competence in this language. Of course English is not their native language so when I assess grammar I am assessing their language proficiency and their competence of the language and how they deal with it because some of the students actually not all of the time but 90% of them in the exam are able to answer the exam because their language proficiency is high even though they don't study. We want them to understand that certain rules are set and they need to study them as science and not just rely on their language proficiency.

So here the whole assessment process changed later on. Even though their language proficiency is high they were unable to answer all the questions because they need to understand how to explain these rules they understand in their head. So in general assessing grammar to me is understanding their competence of the language they are learning.

M: In your opinion, how should grammar be assessed?

RSh: I think it depends first on the objective of the institution. I believe that assessing grammar should be on how to use the language accurately. However, with the specification of the courses we are teaching lately it is not about using the language accurately but explaining how and why certain structures are used in a certain way and to know the terms and how to form complex structure and understand how to label them. So, it focuses on more linguistic details rather than just using the language itself which makes more sense to me. I mean it is a grammar course not just a language course. They should be able to understand all these issues. We also assess them depending on the objective of the course and on what I taught them in class. I assess them on the content but also I believe there should be essay questions where they explain their answers.

M: Why should grammar be assessed in that way?

RSh: So students can demonstrate their abilities to explain why we choose to use this rule over the other

M: Were you assessed this way when you were students?

RSh: Actually I was never taught or assessed in grammar explicitly only in college but before that in school I used to learn grammar implicitly through reading through listening. I studied 3 years in the states and I've never been taught grammar explicitly but later in college maybe because my major was in Linguistics, we did have a lot of linguistic courses starting with grammar and then we moved on to more detailed and more specialized courses in language. This is when I learned why this structure is correct and explain grammar rules and all of that.

M: Does the administration have a saying in that?

RSh: Yes, it is an administration thing. We have to have two in-terms and one final but with the quizzes and what goes on during the semester we have more saying in what we are doing. I do believe in quizzes because it just makes it more serious for the students to study

M: Have you taken any training courses with regards language assessment?

RSh: Yes, during my masters, I had a course on language testing and also during the BA studies, I did have a course on language assessment. And during the years I taught in schools, we had weekly training sessions. It was a two-hour session every week where we trained in all different skills. It wasn't grammar. It was assessing grammar through reading and writing because we didn't teach grammar explicitly. So we were trained on how to construct exams and everything but not on grammar specifically but assessing grammar through writing through reading and that's how we did it then. The training also covered the general educational dos and don'ts, but I have never got training on grammar

M: How is grammar assessed here in your context? Why?

RSh: Through written exams as I said before it is the department regulation to give students two exams during the semester and one final unified exam at the end of the term.

M: What abilities of students do you intend to assess?

RSh: To check student competence and knowledge about the rules they study

M: What is your role as a course teacher when it comes to constructing assessment tasks? Please walk me through the process?

RSh I write my exams usually from scratch. When it comes to writing the exam it is my job to cover everything. The most important points in the assessment and to take into consideration the level of the questions or the difficulties of the questions. The level of difficulty, it should be challenging but also include simple questions that everyone can answer and have some challenging questions that only excellent students can answer but never on something not taught. So I will never include something that was not in the course specification or the course objectives but I do believe in varying the level of difficulty of the questions themselves. I do use more objective question than essay ones because we've got a lot of rules to do. I use a lot of multiple-choice questions. They are a bit challenging specially if we have a lot number of students so we choose questions teacher-friendly but challenging as well.

M: What sort of decisions do you have to make? Exam dates, mark allocation and exam format, and What influenced these decisions?

RSh: Well there are general rules and there are things that are flexible. General rules include the grading of the exam itself. So we do have 60 marks for classwork and

40 for the final but the 60 is divided between 25 on the first in-term and 25 on the second in-term and 10 points that is flexible for quizzes, participation or whatever. But even the exams they are different from one semester to the other according to the coordinator and the department. Sometimes they do allow us to allocate a mark on each item but sometimes they say if it is an objective exam, we should have half a mark on each one so we should have more items. So, there are some rules that are fixed but also there is some flexibility in choosing the items. Now they actually highlight that you have to have a question for each learning objective, but these objectives are general.

As for exam dates, the final is fixed but the midterms are just an agreement between the different groups

M: What challenges do you frequently face? And How do you overcome these challenges?

RSh: You know most of my questions are direct and I've been criticized on that. So I try to make my questions more challenging although I believe in the different level of difficulty of the questions I don't like ambiguous questions. So, the struggle I have how to make it challenging and not ambiguous. So, this a struggle I always face in grammar specially. When I write the objective questions specially the multiple-choice items sometimes two answers are possible. The students think differently from what I am thinking. So some of these questions happen and I realize them after conducting the exam. So even if the exams are checked and double checked and all is correct. I notice two answers when I am correcting the papers and I accept both answers. So piloting the exam is very helpful that is why I depend on my former questions for the exam and adjust them.

M: How do you feel about the current practice of grammar assessment? Would you suggest any changes? Why or why not?

RSh: I do agree with it. For university students yes it is necessary specially since they are specialized in linguistics or translation. I don't object on these exams. I do believe that it has to be written and not oral. It is easier for both the student and the teachers. I do believe that if there is a little bit more flexibility during the semester maybe not in the final; I know we have to have 40 for the final but with 60 marks divided in a certain way I think we could have more flexibility like asking students to write a research paper, for example, and one exam why two. There are lots of

ways to make sure the students understood what they have been taught but we don't have this flexibility. Another issue is that every teacher has her own way of thinking so if they give us such freedom there will be difference in groups which has caused lots of problems, and I understand why they have these fixed rules because of complaints from the students; this teachers is easier than the other and perhaps if a teacher asks for a research paper she may give the students high grades but the other will not. So I think they are trying to avoid any complaint and make sure all the groups are being assessed fairly.

M: If you have the freedom to assess grammar however you like, how would you do that?

RSh: At the beginning I would the same thing I do now because I am more comfortable with it. I mean I never thought about it before but it could be done in different way. Research I believe in research I strongly believe that you learn a lot when you do research than just taking exams and memorizing rules. So we could for example ask the to search a certain grammatical issue and talk about it. So even if it is a mini research and the giving presentation will be very beneficial

M: Do you believe that course teachers should be the ones to construct assessment task? Or should there be an assessment committee in charge of constructing the assessment tasks and handing them over to the teachers to administer?

RSh: Having a committee makes it easier for the teacher but then again all the teaching process will focus on what they will have in the exam and how to make sure that the students can answer these questions. So the focus will not be on the needs of the students themselves and their level; it will be on how they can be ready for that exam. But if the course teacher herself is responsible for exams I think she can do needs analysis she understands the level of the students. It will be more beneficial.

M: What if you are sharing a course with other teachers, how does that work?

RSh: I usually have a saying in all exams but sometimes it is helpful to have someone help you with that when I am overwhelmed with work. When we share a course, we usually divide the work. If she does the first exam, I will do the second but again we do need to agree on the questions included and make sure we are on the same page. This happened to me once, I was so busy with work and did not go through the exam written by my colleague and that caused a lot of problems. The way she was teaching was different from my way. She focused on certain issues that I didn't. So,

her students, I don't know, outperformed mine. So, I learned a listen to always read the questions meet with the teachers and agree with them on the type of questions.

M: What about adapting exams or questions from previous exams?

RSh: Yes, I do adopt and use previous exams because I benefit from the experience and I know which questions work and which are problematic. And specially if I am teaching the course for the first. You benefit a lot from the test bank and look at exams written by teachers who taught the course before you. You know to see what the other teachers focus on. I never used it as it is, but I do change it a little bit depending on my students and if I am comfortable with the questions or not.

Appendix K: Document Analysis

Content Analysis Form for English Grammar Assessment Tasks

To be filled or ticked as appropriate

Document no.	_____		
Document type:	<input type="checkbox"/> quiz	<input type="checkbox"/> midterm	<input type="checkbox"/> final exam Other: _____
Purpose why the document was produced:	_____ _____ _____		
Audience/ target (level of students):	_____		
Author:	<input type="checkbox"/> teacher her/himself	<input type="checkbox"/> co-authored	<input type="checkbox"/> assessment committee
Source of material:	<input type="checkbox"/> internet	<input type="checkbox"/> textbooks	<input type="checkbox"/> Previously made material
Language:	_____		
Method of assessing grammar:	<input type="checkbox"/> Integratively		<input type="checkbox"/> Discreet- point
Format:	_____ _____ _____ _____		
Others:	<input type="checkbox"/> Solicited		<input type="checkbox"/> Unsolicited
	<input type="checkbox"/> Edited		<input type="checkbox"/> Unedited
	<input type="checkbox"/> signed		<input type="checkbox"/> anonymous

Appendix L: Sample of MAXQDA Coding Outcome

Color	Document name	Code	Segment	Author	Creation date	Area	Coverage %
●	Interview 004	Demographic\Qualification	I have BA in Applied Linguistics and masters in TESOL.	masha	13/02/2019 20:04:03	55	0.43
●	Interview 004	Demographic\Experience\Teaching X	I've been teaching English for 16 years. 5 years in schools and 11 years here in the College of Languages & Translation	masha	13/02/2019 20:04:24	11 9	0.93
●	Interview 004	Demographic\Experience\Teaching X	Grammar 5 years	masha	13/02/2019 20:04:40	15	0.12
●	Interview 004	General Beliefs\What grammar assesemnt is	It is to check how advanced our learners are in the English language or their competence in this language.	masha	13/02/2019 20:05:31	10 6	0.83
●	Interview 004	General Beliefs\What grammar assesemnt is	they need to understand how to explain these rules they understand in their head. So in general assessing grammar to me is understanding their competence of the language they are learning.	masha	13/02/2019 20:06:38	18 8	1.48
●	Interview 004	General Beliefs\How it should be done	I believe that assessing grammar should be on how to use the language accurately.	masha	13/02/2019 20:17:08	81	0.64
●	Interview 004	Contextual\Admin Policy	However, with the specification of the courses we are teaching lately it is not about using the language accurately but explaining how and why certain structures are used in a certain way and to know the terms and how to form complex structure and understand how to label them. So, it focuses on more linguistic details rather than just using the language itself which makes more sense to me.	masha	13/02/2019 20:21:19	39 2	3.08
●	Interview 004	Demographic\Experience\Learning X	Actually I was never taught or assessed in grammar explicitly only in college but before that in school I used to learn grammar implicitly through reading through listening. I studied	masha	13/02/2019 20:33:36	25 1	1.97

			3 years in the states and I've never been taught grammar explicitly				
●	Interview 004	Contextual\Admin Policy	Yes, it is an administration thing. We have to have two in-terms and one final but with the quizzes	masha	13/02/2019 20:39:07	99	0.78
●	Interview 004	Attitude C P	I do believe in quizzes because it just makes it more serious for the students to study	masha	13/02/2019 20:39:26	87	0.68
●	Interview 004	Contextual\Training	Yes, during my masters, I had a course on language testing and also during the BA studies, I did have a course on language assessment. And during the years I taught in schools, we had weekly training sessions. It was a two-hour session every week where we trained in all different skills.	masha	13/02/2019 20:40:28	28 8	2.26
●	Interview 004	Current P	Through written exams as I said before it is the department regulation to give students two exams during the semester and one final unified exam at the end of the term.	masha	13/02/2019 20:44:04	16 8	1.32
●	Interview 004	Role C P	I write my exams usually from scratch. When it comes to writing the exam it is my job to cover everything. The most important points in the assessment and to take into consideration the level of the questions or the difficulties of the questions. The level of difficulty, it should be challenging but also include simple questions that everyone can answer and have some challenging questions that only excellent students can answer but never on something not taught. So I will never include something that was not in the course specification or the course objectives but I	masha	13/02/2019 20:45:53	91 6	7.19

			do believe in varying the level of difficulty of the questions themselves. I do use more objective question than essay ones because we've got a lot of rules to do. I use a lot of multiple-choice questions. They are a bit challenging specially if we have a lot number of students so we choose questions teacher-friendly but challenging as well.				
●	Interview 004	Role C P\Decisions	General rules include the grading of the exam itself. So we do have 60 marks for classwork and 40 for the final but the 60 is divided between 25 on the first in-term and 25 on the second in-term and 10 points that is flexible for quizzes, participation or whatever. But even the exams they are different from one semester to the other according to the coordinator and the department. Sometimes the do allow us to allocate a mark on each item but sometimes they say if it is an objective exam, we should have half a mark on each one so we should have more items. So, there are some rules that are fixed but also there is some flexibility in choosing the items. Now they actually highlight that you have to have a question for each learning objective, but these objectives are general. As for exam dates, the final is fixed but the midterms are just an agreement between the different groups	masha	13/02/2019 20:57:08	88 9	6.98

●	Interview 004	Role C P\Challenges	You now most of my questions are direct and I've been coitized on that. So I try to make my questions more challenging although I believe in the different level of difficulty of the questions I don't like ambiguous questions. So, the struggle I have how to make it challenging and not ambiguous. So, this a struggle I always face in grammar specially. When I write the objective questions specially the multiple-choice items sometimes two answers are possible. The students think differently from what I am thinking. So some of these questions happen and I realize them after conducting the exam. So even if the exams are checked and double checked and all is correct. I notice two answers when I am correcting the papers and I accept both answers. So piloting the exam is very helpful that is why I depend on my former questions for the exam and adjust them.	masha	13/02/201 9 21:02:25	85 9	6.75
●	Interview 004	Attitude C P	I do agree with it. For university students yes it is necessary specially since they are specialized in linguistics or translation. I don't object on these exams. I do believe that it has to be written and not oral. It is easier for both the student and the teachers.	masha	13/02/201 9 21:03:11	26 7	2.10

●	Interview 004	Attitude C P	I know we have to have 40 for the final but with 60 marks divided in a certain way I think we could have more flexibility like asking students to write a research paper, for example, and one exam why two. There are lots of ways to make sure the students understood what they have been taught but we don't have this flexibility. Anther issue is that every teacher has her own way of thinking so if they give us such freedom there will be difference in groups which has caused lots of problems, and I understand why they have these fixed rules because of complaints from the students; this teachers is easier than the other and perhaps if a teacher asks for a research paper she may give the students high grades but the other will not. So I think they are trying to avoid any complaint and make sure all the groups are being assessed fairly.	masha	13/02/201 9 21:12:19	84 1	6.60
●	Interview 004	General Beliefs\How it should be done	Research I believe in research I strongly believe that you learn a lot when you do research than just taking exams and memorizing rules. So we could for example ask the to search a certain grammatical issue and talk about it. So even if it is a mini research and the giving presentation will be very beneficial	masha	13/02/201 9 21:18:25	31 0	2.43
●	Interview 004	Method\Integrative	Research I believe in research I strongly believe that you learn a lot when you do research than just taking exams and memorizing rules. So we could for example ask the to search a certain grammatical issue and talk about it. So even if it is a	masha	13/02/201 9 21:18:30	31 0	2.43

			mini research and the giving presentation will be very beneficial				
●	Interview 004	Teachers' Role	Having a committee makes it easier for the teacher but then again all the teaching process will focus on what they will have in the exam and how to make sure that the students can answer these questions. So the focus will not be on the needs of the students themselves and their level; it will be on how they can be ready for that exam. But if the course teacher herself is responsible for exams I think she can do needs analysis she understands the level of the students. It will be more beneficial.	masha	13/02/2019 21:29:08	50 0	3.93
●	Interview 004	Role C P	I usually have a saying in all exams but sometimes it is helpful to have someone help you with that when I am overwhelmed with work. When we share a course, we usually divide the work. If she does the first exam, I will do the second but again we do need to agree on the questions included and make sure we are on the same page. This happened to me once, I was so busy with work and did not go through the exam written by my colleague and that caused a lot of problems. The way she was teaching was different from my way. She focused on certain issues that I didn't. So, her students, I don't know, outperformed mine. So, I learned a listen to always read the questions meet with the teachers and agree with them on the type of questions.	masha	13/02/2019 21:29:49	73 8	5.80

●	Interview 004	Sources\Previous Exams	Yes, I do adopt and use previous exams because I benefit from the experience and I know which questions work and which are problematic.	masha	13/02/2019 21:30:27	13 5	1.06
●	Interview 004	Sources\Previous Exams	You benefit a lot from the test bank and look at exams written by teachers who taught the course before you.	masha	13/02/2019 21:48:48	10 8	0.85
●	Interview 004	Teachers' Role	I never used it as it is, but I do change it a little bit depending on my students and if I am comfortable with the questions or not.	masha	13/02/2019 21:49:14	13 3	1.04
●	Interview 004	Demographic\ Gender	F	masha	13/02/2019 22:12:23	1	0.01
●	Interview 004	Demographic\ Origin	Saudi	masha	13/02/2019 22:16:44	5	0.04

Appendix M: Document Analysis (English Grammar Written Exams)

Format	P #	Instruction	Type	#Items	Score Weight
Production	004	Read the following text. There are 10 mistakes in the use of modals. These mistakes are underlined and numbered. Correct them.	Error correction	10	5
		Form sentences using models to accommodate the function in each item.	Sentence formation	5	5
		Change the percentages in the paragraph to modals or modal like expressions and make any necessary changes.	Word formation	5	5
	047	Read the following texts. Underline and correct each error. Some texts have more than one error.	Error correction	10	6
		Reduce the following sentences.	Rewriting sentence	4	4
		Change the active to passive if possible	Paraphrasing	5	5
		Make sentences to accommodate the function in each item.	Sentence formation	6	6
	009	Change the following sentences into passive if possible.	Paraphrasing	6	6
		Write information questions about the underlined words in each of the sentences below.	Sentence formation	2	2
		Each of the sentences below has an error. Underline and correct them.	Error correction	10	5
		Change the following sentences into reported speech.	Paraphrasing	3	3
		Fill in the blanks with the correct forms of verbs.	Word formation	12	6
		Complete the following sentences with your own words.	Word formation	4	4
	050	Complete the sentences by changing the quoted speech to reported speech.	Paraphrasing	3	3
		Correct the underlined error in each of the following sentences	Error correction	2	2
		Use 'Would you mind...' to ask someone to do something.	Sentence formation	1	1

Format	P #	Instruction	Type	#Items	Score Weight	
	010	Answer the following questions.	Sentence formation	3	3	
		Correct the underlined error in each of the following sentences	Error correction	8	10	
		Make sentences based on the situation below.	Sentence formation	4	4	
		Write sentences to accommodate the instructions between brackets.	Sentence formation	5	5	
			Decide if the sentences below are correct (C) or incorrect (I). Write (C) or (I) on the lines and make necessary changes to correct the incorrect sentences.	Error correction	8	10
	008	Read the following text. There are 10 mistakes. Underline and correct each one of them.	Error correction	10	5	
		Complete the sentences by adding your own verbs.	Word formation	6	3	
		Combine each pair of sentences with the bolded words.	Rewriting sentences	5	5	
		Change the sentences into conditional sentences using if.	Sentence formation	5	5	
		Use Since in two adverbs clauses, once to show time and once to show cause and effect.	Sentence formation	2	2	
	056	Each of the sentences below has an error. Underline the errors and correct them.	Error correction	12	6	
		Write the correct form of the given nouns.	Word formation	12	6	
		Write meaningful sentences using the following expressions.	Sentence formation	5	5	
		Complete the sentences. Use the possessive form of the nouns in parentheses.	Word formation	6	3	
	019	Complete the following sentences using the verbs between parentheses.	Word formation	6	6	
		Rewrite the sentences using since and for	Rewriting sentence	2	1	
		Read the following then answer the questions.	Sentence formation	2	2	
	011	Fill in the blank with the appropriate tensed form of the verb provided.	Word formation	13	6.5	
Form sentences with the following words given the tenses provided.		Sentence formation	4	4		

Format	P #	Instruction	Type	#Items	Score Weight
		Mark the mistake in the following sentences. Explain why it is incorrect.	Error correction	2	2
	023	Use the correct form of the given verbs to complete the sentences.	Word formation	5	5
		Change the following sentences to questions.	Sentence formation	5	5
		Correct the mistakes. Each sentence has one mistake.	Error correction	5	5
	022	Correct the errors in every sentence.	Error correction	4	4
		Complete these sentences by using the appropriate form of a modal in the appropriate tense.	Word formation	8	4
		Fill in the blanks with the correct answer.	Word formation	8	4
	032	Complete the sentences with the words in parentheses using the correct tense.	Word formation	16	8
		Each of the sentence below has an error. Underline the errors and correct them.	Error correction	12	6
		Write complete sentences using the verb tenses mentioned in each point.	Sentence formation	4	4
	017	Fill in the blanks with the right form of the verb in parentheses.	Word formation	4	4
		Fill in the blanks with the right form of the verb in parentheses.	Word formation	4	4
	060	Combine both pair of sentences into 1 parallel structure.	Sentence formation	5	5
		Use paired conjunctions to combine the following pairs into a compound sentence.	Rewriting sentence	5	5
	033	Check the sentences below and correct any tense mistakes.	Error correction	8	4
	041	Combine each pair of sentences with a suitable word of your own.	Sentences formation	6	6
	035	Reorganize the following sentence and modify the sentences if needed	Rewriting sentence	6	3
		Using the following sentence, form 5 sentences in the given tenses.	Sentence formation	5	5
	054	Fill in the blanks with the correct form of the verb in parentheses.	Word formation	3	3
		Define the following terms. Provide one example for each.	Sentence formation	2	2

Format	P #	Instruction	Type	#Items	Score Weight
		Provide one example for the following.	Sentence formation	3	3
	034	Fill in the blanks using the correct form of the verb	Word formation	12	12
		Find the errors in the sentences and correct them.	Error correction	6	6
	048	Change these statements into negative/ question.	Sentence formation	5	5
		Give the past tense of these following verbs.	Word formation	4	2
	027	Fill in the blanks with the right form of verbs.	Word formation	8	8
		Do as shown between brackets.	Sentence formation	3	3
	061	Change the following to passive	Sentence formation	6	6
		Do as shown between brackets.	Sentence formation	6	6
		Correct the mistakes between brackets.	Error correction	5	5
	044	Circle the mistake and write the correct form.	Error correction	10	10
		From the words below, make correct sentences.	Sentence formation	10	10
		Write the plural of the following.	Word formation	5	5
		Complete using the correct Wh-word.	Word formation	5	5
	042	Write the correct pronoun in the following sentences.	Word formation	9	9
		Write the sentences in the negative form - as Yes/No questions.	Sentence formation	10	10
	043	Write the correct form of the following verbs.	Word formation	14	14
		Write yes/no questions or information question to the answers given below.	Sentence formation	5	10
		Correct the mistakes in the following sentences.	Error correction	7	7
		Complete the sentences using the correct form of the words in parentheses.	Word formation	13	13
	045	Fill in the blanks using the correct prepositions.	Word formation	10	10

Format	P #	Instruction	Type	#Items	Score Weight
		Combine the two sentences in each pair. Use the second sentence as an adjective clause.	Sentence formation	4	4
		Change the quoted speech to reported speech.	Paraphrasing	4	4
		Complete the sentences with the correct conditional.	Word formation	10	10
	057	Do as shown between brackets.	Word formation	10	10

Format	P #	Instruction	Type	#Items	Score Weight
Recognition	004	Choose the stronger, the more formal, or the more polite sentence in each pair.	MCQ	10	5
		Complete the sentences below by choosing the correct answer.	MCQ	10	5
	047	Add any needed punctuation and capitalization.	Editing	3	4
		Choose the best completion for each sentence.	MCQ	10	5
	009	Choose the answer that best completes each of the sentences below.	MCQ	8	4
	050	Read each statement. Then, circle the correct answer.	MCQ	6	3
		What do the underlined modals express in each sentence?	MCQ	3	3
		Underline the adjective clause in each sentence. Then, draw an arrow to the word it modifies.	Identification	3	3
	010	What is true of the following examples?	MCQ	6	3
	008	Complete the sentences below by choosing the best answer.	MCQ	10	5
	056	Complete the following sentences with a, an, the or Ø.	Matching	10	5
	019	Circle the correct verb in the following sentences.	MCQ	8	4
		Decide which of the following phrases best describes	Matching	4	2
	011	Answer the questions after the sentences that follow.	MCQ	5	2.5
023	Choose the correct answer.	MCQ	20	10	

Format	P #	Instruction	Type	#Items	Score Weight
	022	Circle the correct word or phrase.	MCQ	6	3
	032	Choose the correct answer that best completes each of the following sentences.	MCQ	14	7
	017	Fill in the blanks with the correct answer.	MCQ	4	4
		Write the right modal auxiliary in the blanks.	MCQ	4	4
		Write (S) if the sentence is simple, (C) if the sentence is compound and (CX) if the sentence is complex.	Identification	4	4
	060	Choose the correct completion.	MCQ	6	3
		Select (✓) the following sentences that are grammatically correct.	T/F	4	2
	033	Is it the simple present (SP) or is it the present progressive (PP).	Identification	10	10
		Choose the correct word to complete each sentence.	MCQ	8	4
	041	Choose the correct answer.	MCQ	14	7
		Use one of the conjunctions provided to fill in the gap.	Matching	12	6
	038	Identify each sentence as 'complex' or 'compound' or 'compound-complex'.	Identification	4	4
		Do as shown between brackets.	MCQ	2	2
	035	Read the following sentences and choose the best verb option provided below.	MCQ	20	5
	054	Circle the letter of the best answer	MCQ	12	12
		Circle (T) for true sentences and (F) for false ones.	T/F	3	3
	048	Choose the correct answer.	MCQ	12	6
	027	Choose the right answer.	MCQ	5	5
		Read the sentences. Write T for true F for false.	T/F	2	2
	061	Choose the right answer.	MCQ	6	3
	044	Choose the correct answer.	MCQ	10	10
	042	Circle the correct answer.	MCQ	5	5
	043	Choose the correct answer.	MCQ	8	4
	045	Fill in the blanks with one verb from the box.	Matching	12	6
	057	Choose the correct form.	MCQ	10	10

Appendix N: Sample of English Grammar Exam with Marks

Allocation

Question 1: Fill in the blank with the appropriate tensed form of the verb provided. (6.5 Marks)

1. He hurt his head while he (play) _____.
2. Mary (like) _____ to travel. She (go) _____ abroad almost every summer. Next year, she (plan) _____ to go to Peru.
3. Jack (climb) _____ the ladder when he (slip) _____ and (fall) _____.
4. Before I (go) _____ to the supermarket, I (check) _____ what was missing and made a list.
5. We were late because we had some car problems. By the time we (get) _____ to the train station, Susan (wait) _____ for us for more than two hours.
6. The students (be/ usually) _____ taught by Mrs. Monty. However, this week Mr. Tanzer will be giving them their lessons.
7. This is a problem. When Maha (come) _____ we will talk to her.

Question 2: Form sentences with the following words given the tenses provided. You may add whatever cues you believe necessary to aid in constructing your tense. (4 Marks)

He/ write/ a letter

1. **Present Perfect**

2. **Simple Present**

3. **Future Progressive**

4. **Past Perfect**

Question 3: Answer the questions after the sentences that follow. (2.5 Marks)

- Jane talks on the phone.
Bob has been talking on the phone for an hour.
Mary is talking on the phone.

Who is not on the phone now? _____

- I'm going to make dinner for Frank.
I'm making dinner for Judy.
I'll make dinner for Mary.
I make dinner for Ted.

Who are you willing to make dinner for? _____

- Jane left when Tim arrived.
Bob left when Tim had arrived.
When Tim arrived, Mary was leaving.
John had left when Tim arrived..

Who did not run into Tim? _____

- Jane is talking in class.
Bob always talks in class.
Mary is always talking in class.

Whose action bothers you? _____

- Jane never left Jamestown.
Bob has never left Jamestown.

Who is still alive? _____

Question 4: Mark the mistake in the following sentences. Explain why it is incorrect. (2 Marks)

- I will paint my room tomorrow.

.....

- The soup is tasting salty.

.....

Appendix O: Examples of Quantitative Analysis of the Questionnaire Data in the form of SPSS Output

Part of the one-sample Kolmogorov-Smirnov test

Showing that the data is not normally distributed

	is important	is difficult to do well	can serve a number of purposes	is irrelevant to language learning	pressurises teachers to complete the syllabus or textbook	determine students' mastery of what they have been taught in an English grammar course	motivate students to learn English grammar
N	93	93	93	93	92	93	
Normal Parameters ^{a,b}	Mean	3.6452	2.4624	3.4462	1.9301	2.5815	3.4409
	Std. Deviation	.58319	1.07904	.62320	1.45884	1.07998	.63353
Most Extreme Differences	Absolute	.427	.261	.329	.190	.194	.317
	Positive	.271	.159	.247	.190	.151	.251
	Negative	-.427	-.261	-.329	-.148	-.194	-.317
Test Statistic	.427	.261	.329	.190	.194	.317	
Asymp. Sig. (2-tailed)	.000 ^c	.000 ^c	.000 ^c	.000 ^c	.000 ^c	.000 ^c	

- a. Test distribution is Normal.
- b. Calculated from data.
- c. Lilliefors Significance Correction.

Example of the binomial test (aka the sign test) performed for all participants on each item

item

This tests whether there is a significant preference for or against the statement. Note that N is reduced because those that chose the midpoint of the scale (2) as their response are omitted: they express no preference.

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
determine students' mastery of what they have been taught in an English grammar course	Group 1	<= 2	1	.01	.50	.000
	Group 2	> 2	88	.99		
	Total		89	1.00		
motivate the students to learn English grammar	Group 1	<= 2	4	.05	.50	.000
	Group 2	> 2	74	.95		
	Total		78	1.00		
provide information about how well each	Group 1	<= 2	3	.03	.50	.000
	Group 2	> 2	86	.97		

student is progressing in English grammar	Total		89	1.00		
inform teaching by showing the students' strengths and weaknesses in English grammar	Group 1	<= 2	1	.01	.50	.000
	Group 2	> 2	92	.99		
	Total		93	1.00		
help place students into groups for English grammar instruction suited to their ability	Group 1	<= 2	6	.07	.50	.000
	Group 2	> 2	78	.93		
	Total		84	1.00		
provide feedback to the students on their strengths and weaknesses in English grammar as they learn	Group 1	<= 2	1	.01	.50	.000
	Group 2	> 2	91	.99		
	Total		92	1.00		
indicate the teacher's ability in teaching English grammar for department chair-persons or external reviewers	Group 1	<= 2	10	.14	.50	.000
	Group 2	> 2	60	.86		
	Total		70	1.00		
indicate learners' ability in learning English grammar for department chair-persons or external reviewers	Group 1	<= 2	4	.06	.50	.000
	Group 2	> 2	65	.94		
	Total		69	1.00		

Example of correlation

Using nonparametric correlation

			English grammar teaching experience	is important	is difficult to do well	can serve a number of purposes	is irrelevant to language learning	pressurises teachers to complete the syllabus or textbook	In the ass sho
Spearman's rho	English grammar teaching experience	Correlation Coefficient	1.000	-.009	-.057	.227*	.009	-.113	
		Sig. (2-tailed)		.930	.585	.029	.934	.284	
		N	94	93	93	93	93	92	
	is important	Correlation Coefficient	-.009	1.000	.184	.383**	-.084	.108	
		Sig. (2-tailed)	.930		.078	.000	.421	.305	
		N	93	93	93	93	93	92	
	is difficult to do well	Correlation Coefficient	-.057	.184	1.000	.016	.386**	.367**	
		Sig. (2-tailed)	.585	.078		.880	.000	.000	
		N	93	93	93	93	93	92	
	can serve a number of purposes	Correlation Coefficient	.227*	.383**	.016	1.000	-.047	.128	
		Sig. (2-tailed)	.029	.000	.880		.652	.225	
		N	93	93	93	93	93	92	

Example of group differences

Using nonparametric Mann Whitney test

	is important	is difficult to do well	can serve a number of purposes	is irrelevant to language learning	pressurises teachers to complete the syllabus or textbook
Mann-Whitney U	319.000	266.500	233.000	193.000	205.500
Wilcoxon W	3805.000	321.500	3719.000	248.000	260.500
Z	-1.484	-1.933	-2.533	-2.814	-2.657
Asymp. Sig. (2-tailed)	.138	.053	.011	.005	.008

a. Grouping Variable: Saudi versus nonSaudi

Factor analysis

Material additional to that in the text

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11.553	25.115	25.115	7.959	17.303	17.303
2	3.511	7.633	32.748	6.032	13.113	30.416
3	3.002	6.526	39.274	4.075	8.858	39.274
4	2.144	4.662	43.936			
5	1.957	4.255	48.191			
6	1.898	4.127	52.318			
7	1.652	3.592	55.910			
8	1.523	3.310	59.220			
9	1.330	2.891	62.111			
10	1.249	2.715	64.826			
11	1.138	2.474	67.300			
12	1.092	2.374	69.673			
13	1.085	2.359	72.032			
14	1.018	2.213	74.245			
15	.955	2.075	76.320			
16	.922	2.004	78.325			
17	.843	1.833	80.157			

18	.742	1.613	81.770		
19	.728	1.583	83.353		
20	.664	1.444	84.797		
21	.619	1.345	86.143		
22	.611	1.328	87.471		
23	.536	1.165	88.636		
24	.481	1.046	89.682		
25	.461	1.001	90.684		
26	.412	.896	91.580		
27	.382	.830	92.410		
28	.374	.813	93.223		
29	.327	.710	93.933		
30	.305	.664	94.598		
31	.286	.622	95.219		
32	.264	.575	95.794		
33	.258	.562	96.356		
34	.225	.490	96.846		
35	.222	.483	97.329		
36	.183	.398	97.727		
37	.172	.373	98.100		
38	.163	.353	98.453		
39	.149	.323	98.776		
40	.135	.294	99.071		
41	.113	.245	99.315		
42	.096	.209	99.524		
43	.072	.156	99.680		
44	.060	.130	99.810		
45	.055	.119	99.930		
46	.032	.070	100.000		

Extraction Method: Principal Component Analysis.

Rotated Component Matrix^a

	Component		
	1	2	3
is important	.543	-.083	.153
is difficult to do well	-.155	.431	.250
can serve a number of purposes	.567	.025	.002
is irrelevant to language learning	-.070	.435	-.089
pressurises teachers to complete the syllabus or textbook	.114	.467	-.076

determine students' mastery of what they have been taught in an English grammar course	.689	.266	.030
motivate the students to learn English grammar	.581	.137	.108
provide information about how well each student is progressing in English grammar	.684	.055	.235
inform teaching by showing the students' strengths and weaknesses in English grammar	.619	.117	.119
help place students into groups for English grammar instruction suited to their ability	.499	.363	.023
provide feedback to the students on their strengths and weaknesses in English grammar as they learn	.732	-.010	.152
indicate the teacher's ability in teaching English grammar for department chair-persons or external reviewers	.447	.438	.289
indicate learners' ability in learning English grammar for department chair-persons or external reviewers	.556	.357	.277
paper and pencil assessment provides valid evidence of students' learning of English grammar	.414	.112	-.110
computer technology helps in assessing students' English grammatical abilities	.415	.319	-.052
English grammar assessment should use means that reflect real life language use (not disconnected sentences or words)	.405	.048	-.112
English grammar assessment should target specific elements of English grammar in separate items (discrete-point aspects)	.225	.529	.165
English grammar is best assessed integratively along with other aspects of English (e.g. through speaking or writing tasks), rather than as a separate skill	.238	.195	.280
objective assessment (e.g., through scores from sets of multiple choice items or cloze gap filling items etc...) is a good method to assess English grammar	.342	.480	-.058
subjective assessment (e.g., by rating overall grammar quality in a short essay or oral presentation) is a good method to assess English grammar	.037	.401	.597
English grammar errors are only important when they get in the way of successful communication of the message being conveyed	-.059	.681	.182
English grammar should be assessed frequently during the course	.579	-.004	.129
English grammar should be assessed at the end of the course	.083	.530	.122
English grammar assessment should systematically target the different structural/formal features of English (e.g. the articles, how 'do' is used in questions and negatives, relative clause formation)	.646	.194	.111
English grammar assessment should systematically target student knowledge of how different functions/meanings are expressed through English grammar (e.g. how an event in future time can be expressed, or how to make polite requests with 'could you?', 'may I?')	.586	.111	.153
English grammar assessment should systematically target student knowledge of common grammatical terms such as verb, object, dependent clause	.578	.335	-.007

English grammar assessment should target student explicit knowledge of grammatical rules (e.g. '-s has to be added to a verb in the simple present when the subject is third person singular', 'days of the week take 'on' while months and years take 'in')	.487	.333	.029
English grammar assessment should primarily be concerned with student ability to understand and use English grammar in communication effectively and spontaneously (as a native speaker does), Not their conscious knowledge about the language	.135	.349	.366
speaking in an oral interview	.025	.008	.813
speaking in an oral presentation	.064	-.039	.804
filling cloze gaps in text	.313	.323	.293
sentence transformation or production items	.424	.122	.529
matching items	.360	.530	.222
multiple choice sentence completion	.488	.282	.111
open response sentence completion	.128	.188	.578
essay writing	.124	-.010	.745
editing / error correction tasks	.478	.003	.286
error recognition / grammaticality judgment tasks	.631	-.041	.242
The best grammar assessment items are the ones developed by the course instructor.	.412	.485	-.013
English grammar is best assessed by expert professional testers /examiners rather than a class teacher.	.029	.600	.281
Assessment of a student's English grammar performance by their peers is useful	.121	.637	-.073
Self-assessment by students of their own English grammar performance is useful	.177	.583	-.080
English grammar assessment tasks are best prepared collaboratively	.518	.348	.076
Ready-made English grammar exercises/ tests found on the internet are a good source for grammar assessment tasks	.168	.601	.181
English grammar exercises from published textbooks are a useful source for constructing grammar assessment tasks	.284	.519	.163
Using English grammar assessment items from previous years is a good source to construct grammar assessments	.125	.547	.245

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Component Transformation Matrix

Component	1	2	3
1	.744	.568	.350
2	-.654	.725	.214
3	-.132	-.389	.912

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Appendix P: Summary of Participants' Results from the Qualitative data

Group A: Participants who only did the interviews

Ps	Stated Beliefs					Stated Practice				
	Means	Purpose	Format	Sources	Author	Means	Purpose	Format	Sources	Author
005	Integrative	Formative	Subjective	PE	Teacher	Explicit	Summative	Objective	PE	Teacher
007	Integrative	∅	Subjective	∅	Teacher	Explicit	Summative	Objective	Questions made up by teacher	Teacher
045	Integrative	Formative	Subjective + Objective	∅	Committee	Explicit	Summative	Objective	PE	Teacher
057	Integrative	Formative	Subjective	∅	Committee	Explicit	Summative	Objective	Books	Teacher

Group B: participants who did the interviews and submitted exam samples

1) Interviews

Ps	Stated Beliefs					Stated Practice				
	Means	Purpose	Format	Sources	Author	Means	Purpose	Format	Sources	Author
010	Explicit	Summative	Objective	PE	Committee	Explicit	Summative	Objective	PE	Teacher
019	Integrative	Summative	Subjective	∅	Teacher	Explicit	Summative	Objective	online	Teacher
042	Integrative	Summative	Subjective	∅	Teacher	Explicit	Summative	Objective	PE	Teacher
043	Explicit	Summative	Objective	∅	Teacher	Explicit	Summative	Objective	PE	Teacher
044	Integrative	Summative	Subjective	∅	Committee + Teacher	Explicit	Summative	Objective	PE	Teacher
054	Explicit	Summative	Objective	∅	Teacher	Explicit	Summative	Objective	Books + online	Teacher
056	Integrative	Summative	Subjective	∅	Teacher	Explicit	Summative	Objective	PE	Teacher
060	Integrative	Formative	Subjective	∅	Committee	Explicit	Summative	Objective	Textbook	Teacher

2) Document Analyses

Ps	Exam Samples				
	Means	Purpose	Format	Sources	Author
010	Explicit	Summative	Objective	∅	Teacher
019	Explicit	Summative	Objective	∅	Teacher
042	Explicit	Summative	Objective	∅	Teacher
043	Explicit	Summative	Objective	∅	Teacher
044	Explicit	Summative	Objective	∅	Teacher
054	Explicit	Summative	Objective	∅	Teacher
056	Explicit	Summative	Objective	∅	Teacher
060	Explicit	Summative	Objective	∅	Teacher

Group C: participants who did the interviews, retros and submitted sample of their exams

1) Interviews

Ps	Stated Beliefs					Stated Practice				
	Means	Purpose	Format	Sources	Author	Means	Purpose	Format	Sources	Author
004	Explicit	Formative	Objective	PE	Committee + Teacher	Explicit	Summative	Objective	PE	Teacher
008	Integrative	Summative	Subjective	Teacher made	Committee	Explicit	Summative	Objective	PE + Textbook	Teacher
009	Integrative	Formative	Subjective	∅	Teacher	Explicit	Summative	Objective	Books + Online	Teacher

011	Integrative	Summative	Subjective	∅	Committee	Explicit	Summative	Objective	Textbook + ESL website	Teacher
016	Integrative	Summative	Subjective	∅	Committee	Explicit	Summative	Objective	PE + self made up items	Teacher
017	Integrative	Summative	Subjective	∅	Teacher	Explicit	Summative	Objective	PE	Teacher
022	Integrative	Summative	Subjective	∅	Teacher	Explicit	Summative	Objective	Textbook	Teacher
023	Explicit	Summative	Subjective	PE	Teacher	Explicit	Summative	Objective	Textbook	Teacher
027	Explicit	Summative	Subjective	PE + self made up items	Teacher	Explicit	Summative	Objective	PE	Teacher
028	Integrative	Summative	Subjective	Textbook	Teacher	Explicit	Summative	Objective	Textbooks	Teacher
032	Integrative	Summative	Subjective	∅	Committee	Explicit	Summative	Objective	Books	Shared effort
033	Integrative	Formative	Subjective	∅	Teacher	Explicit	Summative	Objective	Textbooks	Teacher
034	Integrative	Formative	Subjective	∅	Teacher	Explicit	Summative	Objective	Self made up items	Teacher
037	Integrative	Formative	Subjective	∅	Committee	Explicit	Summative	Objective	Textbook + Made up items	Teacher
039	Integrative	Summative	Subjective	∅	Teacher	Explicit	Summative	Objective	PE	Teacher
041	Integrative	Summative	Subjective	∅	Committee + Teacher	Explicit	Summative	Objective	Books	Teacher
047	Integrative	Summative	Subjective	∅	Committee	Explicit	Summative	Objective	Books	Teacher
048	Integrative	Summative	Subjective	∅	Committee	Explicit	Summative	Objective	Textbook + made up items	Teacher
050	Explicit	Formative	Objective	∅	Committee	Explicit	Summative	Objective	Textbook	Teacher
061	Integrative	Summative	Subjective	∅	Teacher	Explicit	Summative	Objective	Made up items	Teacher

2) Retros + Document Analyses

Ps	Retros					Exam Samples				
	Means	Purpose	Format	Sources	Author	Means	Purpose	Format	Sources	Author
004	Explicit	Summative	Objective	Books	Teacher	Explicit	Summative	Objective	∅	Teacher
008	Explicit	Summative	Objective	Online	Teacher	Explicit	Summative	Objective	∅	Teacher
009	Explicit	Summative	Objective	Made up items	Teacher	Explicit	Summative	Objective	∅	Teacher
011	Explicit	Summative	Objective	Textbook + online + PE	Teacher	Explicit	Summative	Objective	∅	Teacher
016	Explicit	Summative	Objective	Textbook + made up items	Teacher	Explicit	Summative	Objective	∅	Teacher
017	Explicit	Summative	Objective	Online + textbook	Teacher	Explicit	Summative	Objective	∅	Teacher
022	Explicit	Summative	Objective	Textbook + ESL website	Teacher	Explicit	Summative	Objective	∅	Teacher
023	Explicit	Summative	Objective	Made up items	Teacher	Explicit	Summative	Objective	∅	Teacher
027	Explicit	Summative	Objective	PE	Teacher	Explicit	Summative	Objective	∅	Teacher
028	Explicit	Summative	Objective	Textbook	Teacher	Explicit	Summative	Objective	∅	Teacher
032	Explicit	Summative	Objective	Textbook + Books	Teacher	Explicit	Summative	Objective	∅	Shared effort
033	Explicit	Summative	Objective	PE	Teacher	Explicit	Summative	Objective	∅	Teacher
034	Explicit	Summative	Objective	Made up items	Teacher	Explicit	Summative	Objective	∅	Teacher
037	Explicit	Summative	Objective	Textbook + made up items	Teacher	Explicit	Summative	Objective	∅	Teacher
039	Explicit	Summative	Objective	Made up items	Teacher	Explicit	Summative	Objective	∅	Teacher
041	Explicit	Summative	Objective	Books + made up items	Teacher	Explicit	Summative	Objective	∅	Teacher
047	Explicit	Summative	Objective	Books + PE	Teacher	Explicit	Summative	Objective	∅	Teacher
048	Explicit	Summative	Objective	Textbook + made up items	Teacher	Explicit	Summative	Objective	∅	Teacher
050	Explicit	Summative	Objective	Textbook	Teacher	Explicit	Summative	Objective	∅	Teacher
061	Explicit	Summative	Objective	Made up items	Teacher	Explicit	Summative	Objective	∅	Teacher