Exploring ESL Teachers ʾAssessment Literacy, Conceptions and Practices for Assessing Oral Communication in the ESL Classroom

Amir Rezaei

A Dissertation Submitted to The Faculty of Graduate Studies in Partial Fulfillment of The Requirements for The Degree of Doctor of Philosophy

Graduate Program in Linguistics and Applied Linguistics

York University

Toronto, Ontario

September, 2026

© Amir Rezaei, 2026

Abstract

Despite growing interest in the relationship between teachers' beliefs and classroom assessment practices, little research has examined the process of classroom assessment (CA) and how teachers' assessment literacy (AL) and assessment beliefs develop in light of their classroom practices and experiences.

This study explores ESL teachers' classroom assessment practices in assessing oral communication and their relationship with teachers' beliefs and assessment literacy. It also examines the development of teachers' assessment literacy and beliefs in relation to their classroom assessment experiences and contextual factors and proposes a practical model for teachers' assessment literacy.

For this purpose, a qualitative grounded theory design was adopted. Data were collected through interviews, classroom observations, stimulated recalls, document analysis, and a questionnaire with 10 experienced college and university ESL instructors in Canada.

The analysis revealed that ESL teachers follow several classroom assessment procedures to ensure effective assessment in general and of L2 oral communication in particular. They view CA as a continuous cycle of assessment, teaching, and learning aimed at promoting student progress. They design or select tasks aligned with course objectives and student needs, apply appropriate assessment criteria, and prepare students for task performance and assessment. During assessment, they use diverse methods to diagnose learning gaps. Finally, they provide feedback, use assessment results to improve teaching and assessment practices, and reflect on their practices to identify and address shortcomings.

Furthermore, teachers' assessment practices appear to be strongly informed by their assessment beliefs or conceptions, including beliefs about teaching and assessing oral communication, assessment goals and methods, and characteristics of good assessment. Greater consistency was observed between teachers' beliefs and informal assessment practices, particularly when facing contextual constraints. Importantly, teachers' beliefs seem to start as preliminary beliefs and then evolve through real classroom assessment experiences, engagement with contextual factors and reflection.

Finally, teachers' assessment practices seem to be shaped by their assessment knowledge, education, teaching and assessment experiences, and prior learning, with classroom experience as the most dominant factor. Overall, the study has implications for classroom assessment theory and models, teacher assessment literacy, policy and curriculum development, teacher education, classroom practice, and future research.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Khaled Barkaoui, for his guidance, encouragement, patience, and support throughout my doctoral studies and the development of this dissertation. I am especially grateful for the thoughtful feedback, insightful discussions, and the time and patience devoted to reviewing and editing my work through the multiple drafts of this dissertation. This continued guidance and encouragement helped me grow as a researcher and bring this work to completion.

I would also like to thank my committee members, Professor Antonella Valeo and Professor Geoff Lawrence, for their valuable feedback, constructive comments, and support throughout this research. I particularly appreciate the different insights they brought to my work, which helped me strengthen the dissertation. I am also grateful to the examiners, Professor Shahrzad Saif and Professor John Ippolito, for the time and care they devoted to reading my dissertation and for the valuable perspectives they brought to the examination process.

I would also like to extend my sincere appreciation to Rose Frezza-Edgecombe, the Graduate Program Assistant at the DLLL, for her invaluable assistance throughout my graduate studies. Her help with administrative processes, paperwork, and the steps leading up to my defense made navigating the program requirements much easier. I truly appreciate her patience, kindness, and willingness to assist whenever needed.

I am deeply grateful to the ESL instructors who generously participated in this study and shared their experiences, practices, and perspectives with me. Their time, openness, and willingness to participate made this research possible. I would also like to acknowledge the support I received from my department and university throughout my doctoral studies. I am particularly grateful to ETS and the TOEFL Research Grant Program for providing a research grant that supported this research and contributed to the completion of the study.

Finally, I would like to thank my family and friends for their patience, encouragement, and support throughout this long journey. Their understanding and continued belief in me helped me through the many challenging stages of my doctoral studies. I am deeply grateful to everyone who, in one way or another, supported me and contributed to this journey.

TABLE OF CONTENTS

Abstract …….. ii

Acknowledgement …….. iii

Table of Contents ……... iv

List of Tables……... ix

List of Figures …… x

CHAPTER 1: INTRODUCTION …… 1

1.1 Statement of the problem …… 4

1.2 Research questions ……... 8

1.3 Overview of the dissertation …… 8

CHAPTER 2: LITERATURE REVIEW ……... 11

2.1 The importance of classroom assessment and its unique characteristics ……. 11

2.2 The purpose of assessment in the classroom: More attention to formative assessment ……. 13

2.2.1 The importance of informal assessment in the classroom ……... 15

2.3 Contextualized and dynamic nature of classroom assessment ……. 18

2.4 The validation of classroom assessment …….. 20

2.5 Teachers' conceptions of assessment ……... 23
2.5.1 The definition and importance of teachers' conceptions …… 23
2.5.2 Relationship between teachers' assessment conceptions and practices …… 26
2.5.3 Relationship between teachers' conceptions and oral communication assessment ……. 28

2.6 The knowledge base of TAL and theoretical frameworks ……... 31

2.6.1 The importance of assessment knowledge in practice: Implementation of AL in practice ... 37

2.7 Contextual factors, teachers' AL and their conceptions …….. 39

2.8 Conceptual framework ……. 43

CHAPTER 3: METHOD ……... 49

3.1 Overview of the study …….. 49

3.2 Researcher positionality and reflexivity …….. 52

3.3 Sampling and participants ……... 54

3.4 Study context and programs …… 57

3.5 Courses …… 58

3.6 Data collection procedure and tools …… 60

3.6.1 TAL Questionnaire ……... 61

3.6.2 Interviews …… 63

3.6.3 Classroom observations …… 65

3.6.4 Stimulated recalls …… 67

3.6.5 Document collection …… 69

3.7 Data analysis …… 69

CHAPTER 4: TEACHERS' CLASSROOM ASSESSMENT PRACTICES …… 74

4.1 Procedure of classroom assessment ……. 74

4.1.1 Task selection/design stage …….. 76

4.1.2 Assessment rubric selection stage …… 81

4.1.3 Preparation stage …….. 90

4.1.4 Evaluation stage: During the assessment ……. 100

4.1.5 Responding to assessment stage ……... 108

4.1.5.1 Feedback provision ……. 108

4.1.5.2 Using assessment results ……. 115

4.1.5.3 Trial and error, and reflecting on assessments …… 119

4.2 Evaluation of assessment outcome: Effectiveness criteria ……... 122

4.2.1 Teacher observation of student performance …… 123

4.2.2 Student input and feedback ……... 127

4.2.3 Grades ……... 129

CHAPTER 5: TEACHERS 'CONCEPTIONS OF CLASSROOM ASSESSMENTS ….. 131

5.1 Teachers ’general beliefs about teaching and assessing oral communication …….. 132

5.1.1 Communicative language use is central ……. 132

5.1.2 Communicative competence should be prioritized …… 134

5.1.3 Engaging students is crucial ……... 137

5.1.4 Teacher as facilitator scaffolding classroom assessment ……... 138

5.1.5 Assessment and teaching are inseparable ……... 139

5.1.6 Target language use is foundational ……... 140

5.2 Teachers’ beliefs about assessment goals …… 142

5.2.1 Assessment aims to diagnose student needs and scaffold learning …… 142

5.2.2 Assessment primary goal is improving student learning …… 143

5.3 Teachers’ beliefs about the characteristics of good classroom assessment ……. 144

5.3.1 Consistency between the assessment and the construct being assessed is foundational …. 144

5.3.2 Fair assessment and grading are vital ……. 146

5.3.3 Good assessment has positive washback ……... 154

5.4 Teachers ’beliefs about methods of classroom assessment …….. 156

5.4.1 Listening is integral to speaking assessment …….. 156

5.4.2. Use of multiple assessment methods and informal assessments …….. 160

CHAPTER 6: RELATIONSHIP BETWEEN TEACHERS’ ASSESSMENT LITERACY AND PRACTICES ……. 166

6.1 Findings from the TAL questionnaire …….. 166

6.2 The relationship between teacher past education, theoretical knowledge and classroom practices ……. 170

6.3 The relationship between teacher learning experience and classroom practices ……... 173

6.4 The relationships between teaching and assessment experiences and classroom practices ……... 175

CHAPTER 7: CONTEXTUAL FACTORS THAT INFLUENCE TEACHERS' AL, BELIEFS AND PRACTICES 180

7.1 School policies 181

7.2 Teacher's autonomy in selecting or designing assessments 183

7.3 Content-based syllabus 185

7.4 Student motivation/participation 186

7.5 Students' conceptions of assessment 187

7.6 Assessment timing 191

7.7 Graded vs. ungraded assessment 192

CHAPTER 8: DISCUSSION AND IMPLICATIONS 195

8.1 Summary and discussions of the findings 195

8.1.1 A model of the relationship between teacher AL, beliefs, context and Practices 195

8.1.2 Classroom assessment processes 198

8.1.3 Fair classroom assessment 200

8.1.4 The relationship between teacher assessment beliefs and practices 202

8.1.5 The effect of contextual factors on the relationship between beliefs and practices 204

8.1.6 Teachers' knowledge base, assessment literacy and practices 206

8.1.7 Integration of listening and speaking 207

8.2 Implications 208

8.2.1 Implications for theory and models of classroom assessment and teacher assessment literacy 208

8.2.2 Implications for policy 209

8.2.3 Implications for teacher education and training 210

8.2.4 Implications for classroom practices 212

8.2.5 Limitations and implications for future research 214

References 218

Appendices ……. 228

- *Appendix A*: Demographic Questionnaire for Participant Recruitment ……. 228
- *Appendix B*: Teacher Consent Form……229
- *Appendix C*: Assessment Literacy in Practice Questionnaire (ALiPQ)…….. 232
- *Appendix D*: Interview 1 …… 235
- *Appendix E*: Interview 2……..239
- *Appendix F*: Interview 3 ……244
- *Appendix G*: Student Informed Consent Form …….. 247
- *Appendix H*: Observation Checklist ……...250
- *Appendix I*: Stimulated Recall Questions …… ..252
- *Appendix J*: Final Assessment Form (Ryan-Doc≠ 1) ……253
- *Appendix K*: Questions for Bloom's Taxonomy (Aly- Doc≠ 1) ……254
- *Appendix L*: Debate Protocol (Hudson-Doc 1) …… .255
- *Appendix M*: Presentation Rubric (Mike-Doc≠1) …….. 258
- *Appendix N*: Communicative Task Assessment Criteria (Ryan-Doc ≠ 4) …… .259
- *Appendix O*: Debating-based-on-videos Task (Mike-DOC≠ 3) ……260
- *Appendix P*: Seminar Rubric (Sally-Doc ≠5A) …… 264
- *Appendix Q*: Target Performance Descriptors (Ryan-Doc≠ 2) …….. 266

LIST OF TABLES

Table 3.1: Overview of the Study Data Collection Tools ……. 51

Table 3.2: Participants' Demographic Data ……... 56

Table 3.3: Timeline of Classroom Observations ……... 66

Table 4.1: Procedure of Classroom Assessment …… 74

Table 4.2: Criteria for Evaluating the Effectiveness of Assessment Outcome …….. 122

Table 5.1: Teachers' Conceptions of Classroom Assessment ……... 131

Table 7.1: Contextual Factors Influencing the Implementation of Teachers' Beliefs ……... 181
and AL

LIST OF FIGURES

Figure 2.1: Conceptual Framework 46

Figure 4.1: Ryan's Progress Rubric- (Ryan- DOC≠ 3) 83

Figure 4.2: Mike's Presentation Evaluation Sheet (Mike-Doc 2)..... 84

Figure 4.3: *Ray's Oral Summary Rubric (Ray-Doc≠ 1)* 87

Figure 4.4: *Hudson's Debate Evaluation Sheet (Hudson-DOC≠2)*..... 107

Figure 4.5: Ray's Group Discussion Rubric (Ray-Doc ≠2)..... 109

Figure 4.6: Donna's Oral Group Presentation Rubric (Donna-Doc≠ 1)..... 110

Figure 4.7: Cycle of Assessment, Teaching and Learning in CAs 121

Figure 4.8: Ally's Presentation Rubric (Ally- Doc≠ 2)..... 124

Figure 5.1: Arman's Oral Presentation Rubric (Arman- Doc≠2)..... 133

Figure 5.2: Hudson's Oral Presentation Rubric (Hudson Doc≠3)..... 136

Figure 5.3: Ray's Modified Group Discussion Rubric (Ray-Doc≠ 3)..... 146

Figure 5.4: Classroom Assessment Methods and Goals 164

Figure 6.1: Teachers' Self-reported Assessment Knowledge in Different Areas..... 166

Figure 8.1: A developmental model of the relationships between Teachers' AL, Conceptions, Contextual factors and Classroom Practices 197

Figure 8.2: Classroom Assessment Processes 199

CHAPTER 1: INTRODUCTION

This study explores ESL teachers ʾassessment literacy, conceptions and practices for assessing oral communication in the ESL classroom in academic contexts. Assessment is a critical part of teaching in the ESL classroom, and it has an important impact on both teaching and learning (Black & William, 1998; Cheng et al, 2004; Leung, 2014; Stiggins, 2012). Recently, with the growing interest in formative assessment (assessment for learning), classroom-based assessment (CA) and its impact on teaching have become an important field of study, with its unique characteristics gaining more attention (Heritage, 2018; McMillan, 2012; Shepard, 2001; 2006; Stiggins, 2001; Wilson, 2018). MacMillan (2013), for example, noted that CA is becoming a substantial field of study but highlighted the paucity of research in this area and emphasized the need for establishing a comprehensive foundation for CA. On similar notes, other researchers (e.g., Campbell, 2013; Hill & McNamara, 2012; Xu & Liu, 2009) have highlighted the importance of CA and called for more research on the actual processes of CA. To understand the actual process of CA, Brookhart and McMillan (2020) emphasize the importance of the teacher's role in making assessment decisions given their knowledge of content and students' needs. Kane and Wools (2020) also emphasized the key role of teachers in validity matters in CA given their knowledge of instructional goals and the type of performance associated with them. With increasing attention to classroom-based assessment and the teacher's role in its processes, interest in ESL teachers' assessment literacy (TAL) and beliefs is gaining more impetus in the field (Basturkmen, 2012; Borg, 2003, 2011; Brown, 2004, 2008; Popham, 2009; Scarino, 2013; Taylor, 2009; Tsagari & Vogt, 2017; Zheng, 2013a).

Assessment literacy (AL) has been defined in multiple ways in the literature, reflecting knowledge-based, skills-oriented, and contextually situated perspectives (e.g., Davies, 2008; Fulcher, 2012; Scarino, 2013; Xu & Brown, 2016), but no single definition has achieved consensus.

Drawing on existing conceptualizations of teacher assessment literacy (e.g., Popham, 2009; Taylor, 2009, 2013; Xu & Brown, 2016), as well as perspectives emphasizing the dynamic, contextualized, and classroom-embedded nature of assessment (e.g., Hill & McNamara, 2012; Leung, 2004; Scarino, 2013), and aligned with the aims of the present study, I define teacher assessment literacy (TAL) as teachers' integrated knowledge, skills, and principles that guide their design, implementation, interpretation, and evaluation of formal and informal classroom assessments in contextually appropriate and pedagogically sound ways to inform instructional decision-making and support student learning. In line with this definition, teachers 'TAL in the present study is examined through data obtained from a TAL questionnaire, designed specifically for this study, complemented by information regarding their prior education, theoretical knowledge, and L2 learning, teaching, and assessment experiences. Together, these methods serve as sources of teachers 'theoretical and practical assessment literacy in L2 classroom assessment. The first source is used to understand the alignment between teachers' self-reported assessment knowledge in different areas. The second source is used to understand how TAL interacts with classroom practices and is enacted in the classroom. A more detailed discussion of these constructs is provided in Chapter 2.

Teachers are expected to be knowledgeable enough to be able to conduct quality assessment and to draw on assessment to make solid pedagogical decisions in the classroom and help students advance their learning. Unlike standardized and large-scale assessments which follow standardized criteria, classroom-based assessment is dynamic, evolving and emerging, meaning that teachers' assessment decisions are context-dependent and might differ from class to class depending on contextual factors; consequently, teachers need more knowledge of assessment to develop their own criteria and interpretations for making solid instructional decisions (Hill, 2017; Rea-Dickins & Gardner, 2000; Volante & Fazio, 2007).

Another factor that appears to be important in relation to teachers' classroom assessment practices is their conceptions of assessment. The terminology used to describe teachers' views of assessment is diverse, with researchers using terms such as preferences, perceptions, attitudes, beliefs, and conceptions. Among these terms, conceptions arguably encompass a broader scope of teachers' understandings because they capture the ways in which teachers conceptualize the nature, purposes, and functions of assessment. However, the distinction between beliefs and conceptions remains complex, and the two terms are often used with overlapping meanings in teacher cognition research (Borg, 2006; Brown, 2004; Zheng, 2013). Therefore, in this study, the terms beliefs and conceptions are used interchangeably to refer to teachers' underlying views and understandings regarding the nature and purposes of assessment, their preferences concerning assessment practices, and their understandings of effective classroom assessment. As noted by Xu and Brown (2016), teachers' conceptions act as lenses through which teachers interpret their assessment literacy (AL) and implement it. The importance of teachers' conceptions in general and assessment conceptions in particular has been supported in the literature (Basturkmen, 2012; Borg, 2003, 2011, 2018; Brown, 2004, 2008; Zheng, 2013a).

From a constructivist perspective, teachers do not seem to have a prefabricated plan for all classes. Instead, they tend to make decisions depending on the needs of the dynamic classroom context (Borg, 2003; Campbell, 2013; Willis, Adie & Klenowski, 2013; Xu & Brown, 2016; Xu & Liu 2009; Zheng, 2013a). Following this view, it seems that teachers' assessment practices and implementation of their AL and conceptions in the classroom are influenced by the teaching context. Due to the dynamic nature of classroom assessment, it is anticipated that ESL teachers might encounter a wide range of challenges in their assessment practices, but they are expected to balance these tensions and create an equilibrium between their AL, their conceptions of assessment and contextual constraints to make sound assessment decisions.

Therefore, understanding teachers' assessment practices, their contexts, and their relationship with teachers' knowledge and beliefs can give us a nuanced understanding of teachers' assessment literacy in practice and can shed light on how these elements interact with each other to shaper teachers' assessment decisions in the ESL classroom.

1.1 Statement of the problem

The relationship between teachers' assessment literacy and their classroom assessment practices is a burgeoning research area, and it has been the subject of a number of studies (e.g., Popham, 2009; Stiggins, 2012; Tsagari & Vogt, 2017). Similarly, the importance of teachers' conceptions in L2 assessment has been emphasized in the literature (e.g., Basturkmen, 2012; Borg, 2011; Brown, 2004; Xu & Brown, 2016). Despite the growing interest in teachers' AL and conceptions and their relationship with classroom practices, these two dimensions of teacher cognition have largely been examined separately. However, teachers' assessment decisions are shaped not only by their assessment-related knowledge and skills but also by how they understand and interpret the purposes, processes, and meanings of assessment. Therefore, examining teachers' AL and conceptions simultaneously can provide a more comprehensive understanding of how teachers make assessment decisions and enact assessment practices. Moreover, investigating these dimensions together allows for a deeper exploration of how teachers' assessment knowledge and understandings interact with each other and develop through classroom experiences and contextual conditions. Although interest in this area has increased, we do not know much about how ESL teachers' AL and conceptions develop in light of their actual classroom experiences and the role of different contextual factors in this development. In particular, there is a dearth of research examining how ESL teachers' AL and conceptions are related to their classroom assessment practices in assessing L2 oral communication and how contextual factors shape this relationship.

Some researchers (e.g., Fulcher, 2015; Isaacs, 2016. Luoma, 2004) have highlighted the role of L2 oral communication in enhancing academic performance, global communication, and career

prospects. In particular, in academic contexts, effective oral communication is essential for interacting with the academic community, securing a job after graduation and doing different academic tasks such as group discussions, debates and presentations. As such, the teaching and assessment of oral communication skills in addition to writing are major focuses in the ESL classroom in the academic context. Therefore, examining ESL teachers' L2 oral assessment practices and their connection with teachers' AL and beliefs can be of considerable importance. Meanwhile, the dynamic and multifaceted nature of oral communication (Luoma, 2004) due to its interactive nature makes it distinct from assessing other skills in the classroom, in the sense that it might allow teachers to practice different forms of assessment, including formal and informal assessments, balance tensions, and apply their assessment literacy and beliefs.

In the classroom context, there have been a number of studies on the assessment of speaking (e.g., Biber et al.,2004; Butler et al.,2000; Cumming et al.,2005; Fulcher,2003;) and listening as separate skills (e.g., Bejat et al., 2000; Buck, 2001; Taylor & Geranpayeh, 2011; Thompson, 1995; Vandergrift, 2011; Wagner, 2010;) but most of these studies have focused on standardized and summative tests, with little research on the actual assessment of speaking and listening (SP & LS) skills in the classroom in academic contexts where SP & LS are integrated. In classroom assessments, unlike standardized or psychometric tests, oral communication assessments are task-based and are conducted as part of the instructional process, primarily to enhance teaching and learning.

As in the assessment of other skills, teachers bring a diverse range of personal experiences, beliefs, and expertise in assessment to their practices in evaluating L2 oral communication. Also, in the classroom context, many factors may come into play that can significantly influence teachers' assessment practices, criteria, and the overall quality of the assessment. However, despite the importance of oral communication in the academic context and the key role of teachers' AL and

beliefs in shaping classroom practices, there is little research in this area. It is not clear what assessment procedures and criteria teachers use to effectively assess oral communication in the ESL classroom and whether and how their AL and beliefs interact with their assessment practices in specific classroom contexts.

It is worth noting that despite growing interest in classroom assessment and TAL, there is little research on the actual process of CAs. The need for further research on teachers' actual assessment practices has been raised by several researchers (e.g., Campbell, 2013; Hill & McNamara, 2012; Xu & Liu, 2009). Hill and McNamara (2012) noted that most past studies focused on issues such as validity and reliability, criteria and standards, and the influence of external assessment and reporting regimes on classroom practices. By presenting these areas of research on CA, they concluded that there is a dearth of research on the "actual process" of classroom-based assessment. Campbell (2013) also emphasized the importance of studying teachers' actual practice in the classroom as a more authentic way to assess teachers' assessment literacy and the effect of teacher training on classroom assessment. Similarly, Xu and Liu (2009) stressed the importance of teachers' assessment practices and highlighted the interactive nature, uniqueness, and individuality of teachers' assessment practices, which can result in teachers using their own assessment criteria instead of using standard ones. They also pointed to our limited understanding of such practices, which highlights the need for further empirical research.

There is also evidence that the implementation of AL is dynamic and context dependent as Willis, Adie and Klenouski (2013) noted AL is a "context dependent social practice" (p. 242). Xu and Brown (2016) also highlighted the importance of studying teachers' AL in context and noted that TAL is better understood in practice. They noted the concern that measuring TAL based only on knowledge might only give an estimate of TAL, but to obtain a full picture of TAL and understand its application we need to explore it in the real classroom. Thus, it can be argued that

while a teacher's knowledge base may be stable, the implementation of that knowledge can be constrained or facilitated by different contextual factors.

In sum, past studies highlight the importance of researching TAL in context in relation to actual classroom practices to gain a better understanding of its development and its relationship with assessment practices and develop a solid model for TAL. Therefore, there is a need for further research on how teachers conduct CA, establish the effectiveness of their CA, how they enact their AL and beliefs in their practices, and how contextual factors impact the relationship between their AL and beliefs and their assessment practices. Understanding teachers' assessment practices in the classroom as well as their AL and beliefs would not only help us understand their relationships but also help us develop a better model for teachers' practical AL and knowledge base.

Given the interpretive and context-sensitive nature of this study, the researcher's positionality is also relevant to how teachers' assessment experiences and practices are understood and interpreted. My interest in this topic emerged from my experience in ESL teaching and engagement with language assessment research, which led me to become particularly interested in how teachers understand and enact assessment in their everyday classroom practices. I therefore entered the study with prior knowledge and understandings of classroom assessment and teacher assessment literacy. This background provided familiarity with the instructional context and facilitated engagement with participants, while also requiring ongoing reflexivity regarding my own assumptions and perspectives throughout the research process. The ways in which my positionality was considered during data generation, analysis, and interpretation are discussed further in chapter 3.

Therefore, to further this line of research, this study aimed to examine ESL teachers' L2 oral communication assessment practices and their relationship with their AL and assessment

conceptions. The study also aims to explore how context influences the implementation and development of ESL teachers' AL and beliefs in the classroom.

1.2 Research questions

This study addresses three specific goals. First, it aims to describe the assessment practices and criteria ESL teachers use in assessing L2 oral communication in the ESL classroom in order to propose a model for CA. Second, the study aims to examine teachers 'assessment conceptions and AL, including their theoretical knowledge and assessment/teaching experience, as well as their relationship with classroom practices when assessing L2 oral communication. Third, the study aims to identify the contextual factors that influence the implementation of teachers 'assessment literacy and conceptions in the assessment of L2 oral communication in the ESL classroom.

The following research questions are proposed:

1. What are ESL teachers' practices and criteria for assessing L2 oral communication in the classroom?
2. What is the relationship between teachers' L2 assessment conceptions and their practices in assessing L2 oral communication in the ESL classroom?
3. What is the relationship between teachers' L2 assessment literacy and their practices in assessing oral communication in the ESL classroom?
4. What contextual factors influence the implementation of teachers' assessment literacy and conceptions in the assessment of L2 oral communication in the ESL classroom?

1.3 Overview of the dissertation

The dissertation consists of eight chapters. Chapter 1 establishes the rationale of the study by highlighting the importance of classroom assessment and the need for further research on ESL teachers' AL, beliefs and practices in assessing L2 oral communication in the classroom. Chapter 2 reviews the related literature in the field. It first reviews literature on the importance of classroom

assessment and its unique characteristics. The second section discusses relevant studies on the role of teachers' beliefs in their practice in general and in assessment practices in particular. It also reviews relevant studies on the relationship between teachers' assessment beliefs and their practices for assessing different language skills in the classroom including the assessment of oral communication. The third section reviews studies pertaining to TAL and the importance of assessment knowledge in practice. The fourth section discusses contextual factors and their relationship with teachers' AL, beliefs and practices. Subsequently, the last section presents and discusses a conceptual framework for this study.

Chapter 3 describes the design of the study, providing information about the participants and sampling. It also describes the context of the study and the courses under study outlining their objectives, material, tasks and assessment methods. The next section describes the data collection procedures and tools used in the study. Finally, the chapter discusses data analysis method and ethical issues.

Chapter 4 to 7 present the results of in-depth analysis of qualitative data obtained through interviews, classroom observation, stimulated recalls, and document analysis. Chapter 4 presents results related to the classroom assessment practices, procedures, and criteria teachers use to assess oral communication in the classroom. It describes the different stages teachers go through to conduct effective assessment and evaluate the outcomes of their assessment. Chapter 5 reports findings related to teachers 'conceptions of assessments and their implementation in the classroom. It elaborates on their beliefs about L2 teaching and assessment, and their beliefs regarding the goals, characteristics, and methods of assessing oral communication in the classroom. Chapter 6 presents results concerning the relationship between TAL, and teachers 'practices in assessing L2 oral communication. Chapter 7 provides an analysis of different contextual factors that influence teachers 'AL, beliefs and assessment practices.

The last chapter synthesizes the results and presents a model of the relationship between ESL teachers' AL, beliefs, contextual factors and practices. It offers a comparative analysis and a discussion of the collected data, summarizing the key outcomes from the qualitative data analysis and relating them to existing literature. The chapter also explores the study's limitations and provides implications for teacher education, curriculum development, policy, and future research.

CHAPTER 2: LITERATURE REVIEW

2.1 The importance of classroom assessment and its unique characteristics

The importance of classroom assessment (CA) as a distinct field of study has been emphasized by several researchers (e.g., Heritage, 2018; McMillan, 2012; Shepard, 2001; 2006; Stiggins, 2001, Wilson, 2018;). MacMillan (2012) pointed to CA as a significant and distinct field of study requiring an identifiable research base. Wilson (2018) and Heritage (2018) also stressed the importance of CA as the best way to make assessment useful for education and highlighted the key role of teachers in CAs. Given that assessment forms a notable portion of classroom activities during a semester, more importance should be assigned to CAs and teachers' assessment decisions in the classroom (see Bonner, 2012).

Bonner (2012) questioned the assumption that CAs is low-stakes and that the impact of classroom assessment decisions on students (self-image, motivation, future career etc.) are low-stakes. Stakes associated with classroom assessment can have crucial effects on students and their success in education. Regarding the importance of teachers' assessment in the classroom, Rea-Dickins and Gardner (2000) also contented that sometimes teachers' decisions can be high stakes and their decisions can have important consequences on students' learning. As an example, they pointed that overrated students may get false positive judgement regarding their performance in class, which can deprive them of additional language support (p.237).

Some researchers have highlighted the differences between traditional psychometric testing and classroom assessment (e.g., Alonzo, 2020; Bailey and Duran, 2020; Black et al., 2010; Bonner, 2012; Brookhart, 2003; Heritage, 2018; Kane & Wools, 2020; McMillan, 2012; Moss, 2003). For example, Kane and Wools (2020) argued that unlike large-scale assessment, CA results are interpreted locally, and they should be used in meeting the main purpose of CA i.e., to promote effective learning and teaching. Alonzo (2020) supported Brookhart (2003) in that while most of the

criteria for validity are based on psychometric definition, CA is contextualized and has multiple purposes.

Given the differences between large-scale assessment and classroom assessment, a number of researchers (e.g., Brookhart, 2003; Heritage, 2018; Moss, 2003; Moss et al., 2005; Shepard et al, 2018; Wilson, 2018) have highlighted the need for the development of a new framework for CA, which can reflect CA procedures and validation process (primarily based on teachers' interpretations). For example, Moss (2003) emphasized the need for a new measurement theory specifically tailored to classroom assessment (CA). In the absence of such a framework, she relied on traditional models to propose principles for CA and encouraged researchers to develop methods that empower teachers to evaluate the validity of their interpretations. Similarly, Brookhart (2003) advocated for a distinct theory for CA, arguing that classroom tasks are context-specific and fundamentally different from the non-contextual nature of large-scale assessments. She pointed out that these differences necessitate unique approaches to reliability and validity in CA. Additionally, Moss et al. (2005) highlighted the importance of interdisciplinary collaboration among stakeholders to design alternative assessment practices that better address the specific needs of classroom assessments. Heritage (2018) along with Shepard et al. (2018) and Wilson (2018) believed that theories of learning are the key link between curriculum, assessment and instruction. Given the unique characteristics of CA and the drawbacks of standardized tests, Heritage (2018) suggested that the assessment community should move away from the beliefs of standardized exams to supporting teachers' assessment practices in the classroom.

In this respect, a number of researchers have pointed to the unique characteristics of CA which make it different from out of class assessment and psychometric testing. Therefore, they have called for new procedures and models to account for CA and validity in CA. Some of the unique characteristics as explained below relate to *a)* more attention to formative assessment, *b)* the

contextualized, and interpretive nature of classroom assessment and the key role of teachers, and c) validity and validation in CAs.

2.2 The purpose of assessment in the classroom: More attention to formative assessment

In terms of the goal of assessment, in traditional large-scale assessment the focus is on the product rather than the process of learning. In other words, the main goal of assessment is summative (assessment of learning); whereas in CA the focus is more on the process of learning and enhancing learning rather than merely evaluating the outcome. A number of researchers (e.g., Black & Wiliam, 2004; Crusan & Matsuda, 2018; Harlen, 2005; Heritage, 2010, 2012; Heritage et al., 2009; Kane & Wools, 2020; McMillan, 2012) argue for the compatibility of formative assessment (FA) over summative assessment (SA) with CA. They believe that CA is more compatible with FA. For example, Harlen (2005) noted that teachers' assessment in the classroom is more associated with a formative role than a summative one. On a similar note, McMillan (2012) stated that formative assessment is at the top of the most important assessment and teaching topics for CA. Parr (2012) also underlined the importance of FA in the classroom and noted that teachers can gather evidence about ongoing student learning in two ways: a) interactive formative assessment and b) planned formative assessment. Rea-Dickins and Gardner (2000) referred to formative assessment as 'silver bullet' and noted that it is more appealing to the teachers as it promotes learning and teaching in the sense that the evidence from the assessments informs the management and planning of teaching and is also used as feedback on the teaching itself.

Volante and Fazio (2007) mentioned the importance of the growth of formative assessment and referred to some past studies (e.g., Black, Harrison, Lee, Marshall, & Wiliam, 2004; Black & Wiliam, 1998) that showed that teachers get better results when they practice formative assessment in their classes instead of traditional method of assessment in which the focus is on summative assessment. Kane and Wools (2020) also underscored the formative goals of CA and noted that CA

can be used for different purposes such as diagnosing students' needs, motivating students, informing others about their performance, and evaluating their progress.

Campbell (2013) referred to the contextualized nature of learning and assessment, suggesting the priority and appropriateness of FA in classroom assessment. She explained that FA is informed by the constructivist view of learning that emphasizes discovery learning, collaborative learning, the process of learning and the dynamic nature of learning which varies in different contexts without having a predetermined format (p.79).

Despite prioritizing FA in the classroom as elaborated above, some researchers (Alonzo, 2020; Brookhart & McMillan, 2020; Scarino 2013; Stiggins, 2012) supported a pragmatic view of the goal of CA in that teachers should opt for assessment that might involve both FA and SA. In terms of the purpose of classroom assessment and approaches to assessment in the classroom, Brookhart and McMillan (2020) believed that classroom assessment can be only formative or summative (i.e., grading); any assessment beyond these is not classroom assessment. Alonzo (2020) added accountability to this list and specified three purposes for CA: Formative, summative and accountability.

Scarino (2013) emphasized that teachers should understand both traditional, psychometric-based assessment and alternative, sociocultural assessment paradigms and adapt their approach based on context to enhance assessment quality. She suggests that teachers should collect performance evidence over time and incorporate both formative and summative assessment to support learning. In a similar vein, Stiggins (2012) contended that assessment should be both the cause and measure of learning. She contended that a combination of formative and summative assessment is required in order to meet the needs of individual learners with diverse backgrounds in diverse contexts. She referred to the contextualized nature of CA and noted that no assessment method is superior to the others and that teachers need to be aware of this fact to function effectively

in the classroom (p.43). Likewise, in his definition of CA, Popham (2009) included both formal and informal assessments which can be used for grading or formative purposes. He believed that CA results can be used for improving teaching and learning as well as grading (e.g., quizzes, tests in textbooks). He noted that, in professional development programs and TAL, teachers need to gain literacy in assessment decisions related to CA and accountability.

2.2.1 The importance of informal assessment in the classroom:

As noted above, FA seems to be more compatible with classroom assessment. With the widespread use of FA in the classroom with the goal of enhancing learning and teaching, it is expected that teachers use informal assessment besides formal assessment in the classroom to meet the goals of FA. In this respect, Popham (2009, p.6) defined classroom assessment as "*those formal and informal procedures that teachers employ in an effort to make accurate inferences about what their students know and can do*." By referring to the idiosyncratic nature of CA, and variability in assessment practices, Leung (2004) also underlined the importance of teachers' informal assessment in the classroom and defined it as the "*noticing and gathering of information about student language use in ordinary (noncontrived) classroom activities, and the use of that information to make decisions about teaching (to promote learning) without necessarily quantifying it or using it for reporting purposes*" (p.20)

Some other researchers have pointed to the importance of informal assessment in CA (Bonner, 2012; Brown, 2008; Hill & McNamara, 2012; Kane & Wools, 2020; Rea-Dickins, 2001; Rea-Dickins & Gardner, 2000). Bonner (2012) noted that while informal assessment constitutes an important portion of CA, psychometric methods fail to address the needs of teachers using informal assessment (p.88). Kane and Wools (2020) posited that teachers rely more on informal assessment than formal assessment. They explained that in CA, teachers' interpretations and evaluation of inferences are not based on formal assessment and reasoning or statistical models; instead teachers

rely on less formal and quality reasoning. They emphasized patterns of inferences made by teachers, which later can be built into assessment models. They explained that if teachers' inferences are based on multiple observations of students on multiple occasions, the students' achievement evaluation can be done with higher confidence and reliability.

Brown (2008) also pointed to the importance of teachers' informal assessment and posited that in this assessment, "*standards of achievement are held in the mind of assessors, not in a marking schedule or a criteria list*". This suggests that since teachers are in close contact with their students, they understand student progress and achievement better through observation rather than formal standards and criteria. He explained that those teachers who rely on their intuition or "*headnotes*" do not consider assessment to be irrelevant. They just do not use formal assessment and instead hang to their classroom observations and intuitions for assessing their students. In Brown's opinion, the irrelevance conception of assessment stems from student centred teaching. He explained that the supporters of student-centred teaching have wider views of student learning and believe that it cannot be gauged by formal assessments.

Likewise, Rea-Dickens and Gardner (2000) argued that informal assessment is an important aspect of CA. They explained that this teacher knowledge base is not written, and it is not documented, but it exists, and teachers use it in the classroom for making decisions about students' achievement and their development. Rea-Dickins (2001) also stressed the importance of informal assessment and argued that most of classroom assessment might be informal for the purpose of deciding on how much progress a learner has made and what he/she needs to learn. Rea-Dickins (2001, p.437) suggested to plot assessment on a continuum from formal to informal. She described formal assessment (assessment 1) as formal tests or exams "*in that activity is not embedded within instruction*" and referred to informal assessment opportunities as assessment 2 and 3, which include interaction between the student and teacher, interactions among students, and teacher intervention

by asking learners for clarification/explanation to "*push learners forward in their understanding and language learning"* (p.437).

In keeping with Rea-Dickins (2001), Hill and McNamara (2012) posited that some teachers rely on their observation and intuition rather than formal assessment in assessing their students. In their framework for researching CA, they chose "*assessment opportunities*" as the unit of analysis to incorporate more intuitive forms of decision-making in CA. They took Rea-Dickins' (2001) idea of 'assessment opportunity' and extended it as follows: "*any actions, interactions or artifacts (planned or unplanned, deliberate or unconscious, explicit or embedded) which have the potential to provide information on the qualities of a learner's (or group of learners') performance.*" (Hill & McNamara, 2012, p.396).

According to Hill and McNamara (2012) classroom-based assessment incorporates both summative and formative assessment. They divided assessment activities into three categories:

1. Formal assessment (which is planned and evident)
2. Planned assessment opportunity (instruction embedded, planned)
3. Incidental assessment opportunity (instruction embedded, unplanned) (p.403)

The last two categories are embedded within instruction. The second group of activities are planned but the third one is incidental and might be practiced intuitively while teaching. However, Hill and McNamara (2012, p.396) corroborated and supported Rea-Dickins' (2006) and Leung' (2005) concern that it is hard to understand when assessment is taking place in the classroom given the intuitive and unplanned nature of informal assessment. As Leung (2005) explained,

> The difficulty here is identifying when (or whether) assessment is actually taking place, given the often-intuitive nature of teacher decision-making in classroom-based assessment (Rea-Dickins, 2006). The challenge for the researcher is thus how to understand classroom-based assessment when it takes place 'without conscious planning, as part of ordinary interaction with students, that is, when [teachers] consider themselves to be teaching' (p. 877).

Given the importance of informal assessment, some researchers (e.g., Hill, 2017; Popham, 2009) have raised the need for developing CA and TAL models that encompass both formal and

informal assessments. Many of the existing models (e.g., Inbar-Lourie, 2008; Hill & McNamara, 2012; Xu & Brown, 2016) appear to fall short of fully capturing the range of teachers' assessment decisions and practices, particularly by overlooking the informal assessments conducted by teachers. In this regard, Popham (2009) contended that both formal and informal assessment procedure should be considered in any framework of TAL and teacher professional development. In keeping with Popham, Hill (2017) also noted that the discussion of assessment literacy is not restricted to formal assessment, and it should encompass incidental forms of assessment in the classroom as well. She believed that informal assessment in the classroom has a great effect on learning, so any conceptualization of assessment should include informal assessment.

All in all, these studies indicate the importance of teachers' informal assessment in the classroom and the need for more research on it. As noted above, some researchers have underlined the importance of teachers' headnotes (e.g., Brown, 2008) or assessment opportunities (e.g., Hill & McNamara, 2012; Rea-Dickins, 2001, 2006) but there is little empirical research on these informal assessment decisions as to how teachers conduct informal assessments in the classroom, what goals they aim to achieve through informal assessment, and the relationship between their beliefs and /AL and their informal assessments. Further research in this field can help us develop CA models which encompass all types of assessment practices including informal assessments. As mentioned above, some researchers (e.g., Hill, 2017; Popham, 2009) have highlighted the need for the development of TAL models that consider both formal and informal assessments. Most of the existing TAL models (e.g., Inbra-lourie, 2008; Hill & McNamara, 2012; Xu & Brown, 2016) do not seem to capture all teachers' assessment decisions and practices i.e., they seem to disregard informal assessments.

2.3 The Contextualized and dynamic nature of classroom assessment

Another unique characteristic of CA is its dynamic and contextualized nature. A wide range of scholars (Bailey & Duran, 2020; Black & Wiliam, 2004; Bonner, 2012; Brookhart, 2003; Moss,

2003; Shepard, 2000, 2006, 2009) have emphasized that CA is closely tied to classroom interaction, instructional goals, and learners' diverse and changing needs. From this perspective, standardized and psychometric assessment models are poorly aligned with classroom assessment, as they do not account for the contextualized and process-oriented nature of learning in classrooms. Bailey and Duran (2020) argue that large-scale assessments are unable to capture the ongoing instructional processes that characterize CA. Likewise, Bonner (2012) suggests that the situated nature of classroom assessment and variation in learner needs across classrooms challenge the validity of applying standardized assessments across contexts. Building on this view, Shepard (2000, 2009) highlights the dynamic and socially mediated character of CA and proposes sociocultural theory (SCT) as a more suitable framework for understanding formative assessment practices in classroom settings.

Because classroom assessment is inherently dynamic and situated within specific instructional contexts, teachers play a pivotal role in enacting CA and in evaluating the validity of assessment decisions. Teachers' continuous interaction with students allows them to respond to emerging learning needs and to adapt assessment practices as instruction unfolds. Brookhart and McMillan (2020) emphasize that while external experts are responsible for establishing reliability, validity, and fairness in large-scale assessments, teachers are the primary experts in individual classrooms because they possess contextual knowledge of both instructional content and students' needs. Similarly, Kane and Wools (2020) highlight teachers' central role in validity matters in CA, noting that teachers are best positioned to align assessment with instructional goals and the types of performances valued in specific classroom contexts. In addition, Heritage et al. (2009) argue that teachers gather assessment evidence on an ongoing basis through everyday classroom interaction, enabling them to identify gaps between students' current learning and intended instructional outcomes. Taken together, these perspectives suggest that the dynamic and contextualized nature of

CA necessitates teacher discretion and interpretive judgment to make informed assessment decisions and to support student learning effectively.

2.4 The validation of classroom assessment

The literature reveals that while evidence in psychometric and large-scale testing is collected to assess scoring validity, in CA, the focus is on validating teachers' interpretations and decisions. Hence, classroom assessments require a distinct validation process and, accordingly, a different model to better represent teachers' interpretation in classroom assessments.

According to current validity theory, assessment validation is defined as the appropriateness of interpretations (Bachman,2005; Kane, 2006). In accord with this definition, some researchers (e.g., Black et al, 2010; Bonner, 2012; Brookhart, 1993; Kane, 2006; Moss, 2003) support interpretive approaches to the validity of CAs. According to Bonner (2012) this interpretive approach seems more compatible with CA, especially for formative classroom assessment, given its situated and qualitative nature. Kane (2006), similarly, believed that validation in CA tends to be qualitative, iterative, and informal.

Unlike psychometric assessments in which evidence is gathered to support the validity of score interpretations, in CA it is done to bolster assessment decisions in the classroom. In this respect, Brookhart (2003) noted that psychometric assessment is about gathering evidence to support interpretations of test scores, whereas CA is ongoing, and it leads to immediate changes in students. Moss (2003), also, believed that diverse information other than scores should be used by teachers to evaluate the effectiveness of their assessment practices. She believed that in addition to what students produce in response to an assignment, we should consider other factors such as how students engage through multiple drafts, observations, and the resource they had access to. On a similar note, Heritage (2018) mentioned that in the classroom it is not possible for teachers to administer tests regularly to interpret results, trace students' learning and decide on next steps.

Instead, they need to collect information regularly about each student to help them continuously make progress.

Some researchers (e.g., Alonzo, 2020; Black & William, 1998; Hill & McNamara, 2012; McNamara, 2001; Shepard, 2009) have underlined the importance of the positive consequences of CA as an important indicator of the validity and effectiveness of CA. Shepard (2009) noted that formative assessment must be beneficial to learning to substantiate validity claims. Alonzo (2020) explained that CA is an ongoing process of evidence gathering that includes everyday tasks, observation and interaction. She noted that relevant evidence is not solely what students know and can do; it should also include evidence that can be used to support student learning. She explained that in large-scale assessment, validity relates to the accuracy of interpretations of students' status; whereas, in CA it mostly concerns the diagnosis of students' needs and taking necessary actions to fill the gap in their learning.

McNamara (2001) also emphasized the importance of using assessment results (i.e., assessment use) and specified three critical dimensions of assessment: *evidence, interpretation* and *use*. Similarly, Hill and McNamara (2012) underlined the use of CA results and defined CA as follows: "*any reflection by teachers (and/or learners) on the qualities of a learner's (or group of learners') work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialization purposes*" (Hill & McNamara, 2012, p. 396).

Another concept closely related to classroom assessment validity and effective assessment is assessment fairness. The definition of fair assessment varies across different frameworks, indicating its complexity and the need for further empirical research in this field. Messick's (1989, 1994) theory of validity integrates fairness into the broader concept of validity. According to Messick, fairness is intrinsically connected to the notion that assessments must accurately measure what they are intended to measure, and the use of the assessment should have positive consequences.

Messick's theory places more emphasis on the psychometric dimensions of fairness, particularly regarding the accuracy with which assessments measure the intended construct. On the other hand, Kunnan (2004) seems to separate validity of a test from its fairness, emphasizing that "test fairness is not just about what is tested but how it is tested" (p. 160), highlighting the importance of both content and method in creating fair assessments. Kunnan's framework highlights that a fair assessment is one that provides equal opportunities for all students, removes cultural and linguistic biases, and ensures that assessment outcomes do not unfairly disadvantage certain groups. Unlike Kunnan's and Messick's frameworks, which focus on test design and validity, Black and Wiliam (1998) emphasize the processes of feedback and student involvement in assessment. Black and Wiliam advocate for the use of formative assessment as a primary tool for ensuring fairness. They propose that assessment should be an ongoing process that helps students improve rather than being just a final judgment of their ability. Kane's (2010) framework stresses the justification for assessment use in the context of learning. Fairness, therefore, depends on the extent to which educators can demonstrate that their assessment practices are grounded in evidence-based arguments about what constitutes valid and fair assessment. Tierney (2012), after reviewing the literature on educational assessment, acknowledges the complexity of defining fairness due to the diversity of classroom contexts, and notes that there are multiple and conflicting interpretations of the concept of fairness. As a general trend among relevant studies, Tierney's review underscores that fair assessment is not just about uniformity, but about providing all students with the support and conditions they need to succeed.

The above review highlights the difference between large-scale testing and classroom assessment in terms of the goal of assessment, type of assessment, and validation procedure. Given these differences, it seems that classroom assessment requires a different validation procedure. However, we know very little about what procedures teachers follow in CAs, and what

criteria/evidence they use to interpret and argue for the effectiveness, validity or fairness of their assessments. As pointed out by Bonner (2012), there is little research about the methods teachers use to improve the validity of their classroom-based assessment decisions.

The above literature also shows that CA is a distinct field of study requiring its own research base with its own assessment procedures and criteria. By underlining the key role of teachers in CA procedure and the interpretive nature of CA, it suggests that more attention should be paid to teachers' assessment decisions and interpretations, how they practice and ensure effective assessment in the classroom, what procedure they follow, what criteria they consider in their assessment and what factors affect their decisions. It also highlights the need for a new model of validation different from those proposed for psychometric testing as they fail to capture the unique characteristics of CA as outlined above.

In sum, the above review suggests that there is a need for the development of a new practice-based framework for CA considering its unique characteristics. But to develop such a framework, we need to know what assessment procedures and criteria teachers use to interpret and argue for the effectiveness of their assessment practices. Indeed, to understand CA procedures and to propose an inductive practice-based model for it, teachers' actual practices in the classroom should be the main sources of information. That is, instead of using principles of traditional measurement and psychometric testing, researchers should put more weight on teachers and their practices to come up with a sound framework for CA.

2.5 Teachers' conceptions of assessment

2.5.1 The definition and importance of teachers' conceptions

Borg (2006) refers to teachers' network of beliefs, knowledge and reflections as teacher cognition. Due to the overlap between beliefs and knowledge and their inextricably intertwined relationship in the mind of teachers, Borg used the term 'cognition' inclusively without separating

beliefs from knowledge. Borg (2018) highlighted the complexity of the relationship between beliefs and knowledge, aligning with Hoy, Davis, and Pape (2006, p. 716), who described these concepts as overlapping constructs within the literature. Similarly, Li (2017) noted that researchers often face challenges in distinguishing between belief and knowledge, prompting extensive discussions about their definitions.

Zheng (2013a) pointed out that beliefs have been defined in different ways by different researchers. She explained that part of the confusion in the definition of belief goes back to its relationship with knowledge. She stated that from an epistemological point of view they are considered separately as knowledge is closer to truth than beliefs are. But from the viewpoint of teachers' cognition, they are grouped together, and they overlap with each other. As a result, following Borg (2006, p.33), Zheng did not make a distinction between beliefs and knowledge in her study given that these constructs are not held or perceived distinctively in the mind of teachers. Likewise, Brown (2004) admitted the variance in the literature on the definition of beliefs and pointed out that there are different understandings of this term.

Thompson (1992) proposed the term "conceptions" which encompass teachers' beliefs, mental structure, concepts and preferences (p.130). According to Brown (2004), conceptions include different categories of teachers' ideas regarding educational phenomena. In his research, Brown defined conceptions as an "organizing framework by which an individual understands, responds to, and interacts with a phenomenon" (p.303). He operationalized and evaluated teachers' conceptions of assessment based on their agreement or disagreement with four purposes of assessment: a) improvement of teaching and learning, b) school accountability, c) student accountability, and d) treating assessment as irrelevant. Xu and Brown (2016) believed that "Conceptions of assessment denote the belief systems that teachers have about the nature and purposes of assessment and that encompass their cognitive and affective responses" (p.156).

In the teaching field, Cheng, Chan, Tang and Cheng (2009) identified traditional and constructivist views as two main conceptions of teaching and learning and considered teachers' views about the nature of knowledge as their beliefs. They believed that teachers' epistemological beliefs influence their conceptions, i.e., traditional or constructivist views of learning and teaching.

The above review reveals the variation in the definition of "beliefs" or "conceptions" in the field and its fuzzy relationship with knowledge. In tune with the purpose of this research, in this study, conceptions and beliefs are used interchangeably to refer to teachers' epistemological views about the nature of assessment, their preferences regarding the purposes of assessment, and their definition of effective CA. In this study, participants were asked about their beliefs regarding effective teaching and assessment, assessment purposes, and methods, the implementation of their beliefs in CA, and factors that affect the implementation of their beliefs.

The importance of teachers' beliefs has been emphasized in the literature. According to Borg (2003, 2011) and Basturkmen (2012) beliefs are the bases of actions, as they inform teachers' practices in the classroom. In this respect, Borg explained that "beliefs are propositions individuals consider to be true and which are often tacit, have a strong evaluative and affective component, provide a basis for action, and are resistant to change. In the context of language teacher education, beliefs are seen to be a key element in teacher learning and have become an important focus for research" (Borg, 2011, p.371). Fives and Buehl (2012) emphasized that teachers' beliefs serve as "filters for interpretation, frames for defining problems, and guides or standards for action" (p. 79). Li (2017) emphasized that examining teacher cognition provides researchers with the opportunity to analyze teaching and learning through the lens of teachers, offering valuable insights into their perceptions of effective instructional practices. It, also, allows for a detailed investigation of how various components of the teaching process such as the choice of instructional materials, the design of learning tasks, the dynamics of classroom interactions, and assessment decisions interconnect and

influence one another. Exploring beliefs provides valuable insights into teachers' perspectives, enhances teaching quality, drives educational reform (Skott, 2014), and plays a critical role in supporting teachers' professional development (Borg, 2018).

Brown (2004, 2008) also stressed the importance of teachers' beliefs and noted that they exert influence on the implementation of policies mandated by the school or curriculum. Teachers' decisions in the classroom are guided more by their conceptions rather than the rules or regulations set by the school and policy makers. He added that the success of any policy depends on the conceptions of teachers and the consideration of this factor as an important element. He concluded that studying teachers' conceptions is important because they inform teachers' teaching methodology and influence students' learning. Brown pointed to the individualistic nature of teachers' conceptions and noted that teachers might follow their idiosyncratic values and beliefs in assessment practices.

2.5.2 Relationship between teachers' assessment conceptions and practices

Concerning the relationship between beliefs and practice, Basturkmen (2012) explained that this relationship is interactive, in that they affect each other. After reviewing the literature on the relationship between teachers' beliefs and their classroom practice, he noted that most of the studies that focused on the planned aspects of teachers' practice such as instructional approach and selection of activities for teaching in the classroom reported alignment between teachers' reported beliefs and their classroom practices.

Davison (2004) stated that teachers' beliefs affect the way assessment criteria are interpreted and utilized in the classroom and most of teachers' decisions in the classroom go back to their beliefs. Likewise, Zheng (2013a) pointed out that teachers' conceptions inform their judgment in teaching and shape their classroom practices. Xu and Brown (2016) also emphasized the importance

of teachers' beliefs and argued that teachers' conceptions of assessment can act as lenses through which their knowledge is interpreted. They explained that "teachers' beliefs are an interpretive and guiding framework by which they mediate their uptake of theoretical knowledge and its implementation" (p. 156). Scarino (2013) postulated that if the purpose of AL is to benefit classroom practice, it is necessary to reconsider the knowledge base to develop AL that includes teachers' personal knowledge, beliefs and values. She considered dialogue and reflection on experience and conceptions as a key to the development of AL.

Recently, with the growth of interest in teachers' cognition, there have been a number of empirical studies examining the relationship between teachers' beliefs and their classroom practices in teaching or assessing different skills such as reading (e.g., Boubris & Haddam, 2020; Farrell & Ives, 2014; Kuzborska, 2011), grammar (e.g., Phipps & Borg, 2009), and writing (Guadu & Boersma, 2018; Lam, 2019; Montgomery & Baker, 2007; Wang et al., 2020).

Farrell and Ives (2014), for instance, investigated the connection between a novice teacher's beliefs about teaching L2 reading and his classroom practices in an academic context in Canada. Their findings revealed alignment between the teacher's stated beliefs and his instructional practices for L2 reading. Nonetheless, they also identified some inconsistencies between the teacher's professed beliefs and his actual classroom practices. In another study, Boubris and Haddam (2020) examined the relationship between EFL teachers' beliefs about effective reading skills development and their assessment practices. Their findings indicated a strong connection, revealing that teachers' interactive understanding of reading significantly influenced their evaluative practices.

Some researchers explored the relationship between teachers' beliefs and their practices in assessing writing. Research on teachers' writing assessment beliefs and practices (e.g., Guadu & Boersma, 2018; Montgomery & Baker, 2007, Wang et al., 2020) highlights inconsistencies between beliefs and actions. Wang et al. (2020) studied the alignment between teachers' beliefs about writing assessment and their self-reported practices, as well as factors influencing these aspects. They found

that teachers favored assessment for learning (AfL) over assessment of learning (AoL) for supporting writing development. However, their practices did not consistently reflect these beliefs. Similarly, Guadu and Boersma (2018) found that while EFL teachers valued formative assessment for improving learning, this was not reflected in practices like progress monitoring or scaffolding. Lam (2019) employed a mixed-method approach to explore the knowledge, beliefs, and practices of teachers regarding classroom-based writing assessment. The study revealed that while teachers could differentiate between AoL and AfL, they lacked understanding of assessment as learning (AaL). Despite this, participants made efforts to implement alternative writing assessments, such as peer evaluation and post-writing activities, but faced challenges due to institutional constraints. Montgomery and Baker (2007), also, observed mismatches in teachers' use of written feedback, suggesting that training may shape beliefs but not necessarily practices.

2.5.3 Relationship between teachers' conceptions and oral communication assessment practices

Oral communication is widely acknowledged as a crucial skill for achieving success in various domains, including academic, professional, and social settings. In particular, within ESL and EAP programs, effective oral communication is essential for interacting with the academic community and completing different academic tasks. The importance of oral communication for improving academic performance, gaining access to higher education, enabling global communication, and enhancing employment opportunities has been highlighted by several researchers (e.g., Fulcher, 2015; Luoma, 2004; Isaacs, 2016). According to Luoma (2004), speaking skills play a crucial role in language teaching curricula, making them a key focus of assessment. Luoma emphasized that speaking skills are essential components of language proficiency and should be closely integrated into both the curriculum and assessment processes. She discussed the importance of evaluating speaking as it reflects real-world communication abilities.

However, most of the published studies in the field have focused on psychometric testing and the assessment of oral communication in standardized tests (Fulcher, 2015, Ginther, 2013). Despite

the centrality of oral production in general and in academic context, it has received relatively less attention in classroom assessment research especially when it comes to the relationship between teachers' beliefs and their assessment practices. As described below, to the best of my knowledge, there are only three studies specifically on the relationship between teachers' beliefs and their assessment of oral communication in the classroom.

Özdemir-Yılmazer and Özkan (2017) conducted a study examining the perceptions and practices of speaking assessment among English teachers at the tertiary level in Turkey. The research focused on two ESL teachers instructing graduate university students in a preparatory course, culminating in a proficiency exam at the end of the semester. The findings revealed that both teachers emphasized student interaction, prioritized fluency over accuracy, and considered feedback provision essential in their stated beliefs and speaking assessment practices. The study also highlighted individual differences in the teachers' beliefs regarding the components of speaking assessment, which led to variations in their judgment. These differences were attributed to the teachers' distinct learning experiences and styles.

Ahmed (2021) conducted a study using semi-structured interviews to investigate the beliefs and practices of internationally educated teachers (IETs) regarding speaking skills assessment in ESL/EAP classrooms. The study examined three teachers' perspectives on the purposes and forms of assessment, feedback, and their classroom practices. His results showed that most participants believed that assessment should promote learning and provide opportunities for feedback. All participants valued formative assessment for its ability to generate feedback and help teachers identify student needs and challenges. The participant teachers also supported learner-centered pedagogy, using formative feedback to evaluate performance and suggest improvements.
In practice, the teachers implemented formative assessments and feedback strategies that aligned with their belief in supporting student learning. Their approaches were influenced by task-based and communicative language teaching, aiming to enhance communicative competence. Ahmed

concluded that while internationally educated teachers' beliefs and practices were interconnected, the alignment varied depending on individual experiences and teaching contexts, reflecting the complexity of the relationship between beliefs and practices.

In another study, Baker and Burri (2016) explored five experienced teachers' beliefs in relation to providing feedback on pronunciation in an EAP program in North America. Their results showed that the teachers shared similar beliefs regarding the importance of feedback in learner development and enabling students gain comprehensible pronunciation. In tune with this belief, in their practices, they used approaches such as peer feedback on comprehensibility, whole -class feedback, and recasting.

The above studies seem to revolve around teachers' general beliefs regarding formative and. summative assessment and their views about feedback provision in assessing oral communication, without examining the procedures and criteria they use for assessing oral communication and ensuring effective assessment. Also, it is not clear whether and how informal and formal assessment practices are related to teachers' beliefs regarding the assessment of oral communication, and what contextual factors influence their implementation. Also, past studies in the field do not seem to explore whether and how teachers' beliefs are shaped by teachers' past education, theoretical knowledge and assessment experience in the classroom. Furthermore, the cognitive and affective aspects of teachers' beliefs, which have been mentioned by some researchers (e.g., Borg, 2011; Brown 2004, 2008; Golombek & Doran, 2014; Xu & Brown, 2016), have not been adequately addressed in past research. In particular, it is not clear whether and how teachers enact emotional aspects of their beliefs, which, unlike the cognitive aspects, seem to be less grounded in their knowledge base and less aligned with assessment principles, in their assessment practices to ensure the fair assessment of oral communication.

It is worth noting also that oral communication in the classroom is highly interactive, making its assessment distinct from the assessment of other skills. Due to its nature, teachers tend to use informal assessments alongside formal ones, primarily for formative purposes. This flexibility might give teachers more opportunities to apply their own beliefs and ensure effective assessment. Therefore, further research on oral communication can reveal the nuances of the relationship between teachers' beliefs and assessment practices as well as the development of their beliefs and /AL in the classroom.

2.6 The knowledge base of TAL and theoretical frameworks

With the growing emphasis on formative assessment and assessment for learning, teachers' assessment literacy has become increasingly important and a key focus of educational research. As formative and alternative assessment practices expand, teachers face greater assessment responsibilities, making a solid understanding of assessment principles essential for supporting student learning and guiding instructional decisions (Fulcher, 2012; Stiggins, 2010; Taylor, 2009; Tsagari & Vogt, 2017). Research shows that teachers devote substantial time to assessment activities, highlighting the need for assessment literacy to design and use appropriate assessments, distinguish high-quality assessments from poor ones, and interpret assessment results effectively (Cheng, 2001; Popham, 2009; Stiggins, 2012). Despite this, teachers' assessment literacy remains differently conceptualized and underexplored, highlighting the need for further research (Rea-Dickins & Gardner, 2000; Vogt & Tsagari, 2014; Sheppard, 2000; Tsagari & Vogt, 2017). Building on this, it is essential to review how assessment literacy has been defined, the knowledge base it comprises, and the factors influencing it in order to develop a conceptual framework. Hence, in this section different definitions of TAL and knowledge base as well as different influential factors in TAL are presented and then a preliminary conceptual framework of TAL is proposed.

Davies (2008) in his review of the textbooks in the field of second language testing/assessment from Lado (1961) till 2008 identified two trends in the field of language testing. The first one related to the expansion of teaching material within the discipline to meet the needs of teacher training programs. The second trend, according to Davies, concerns the expansion of a skills+ knowledge approach and the inclusion of principles. He reported that principles regarding ethics, validity and professionalism have become important besides the skill+ knowledge approach. Skills refers to item writing and analysis while knowledge pertains to background in measurement, familiarity with language learning, language teaching and language testing models and theories. Principles concern the appropriate use of language tests and their impact on different stakeholders. Davis (2008, p. 335-336) described the relationship between these three elements as follows:

> Skills, meaning techniques and methodologies, on their own are no longer enough, skills + knowledge are inadequate without the addition of principles. For teaching, as for learning, there is a need for careful balancing of the practical (the skills) with the descriptive (the knowledge) and the theoretical (the principles). All are necessary but one without the other(s) is likely to be misunderstood and/or trivialized.

Inbra-lourie (2008, p.389) focused on five modules introduced by Brindley (2001) for programs for professional development in language assessment: the social context of assessment, defining and describing proficiency, constructing and evaluating language tests, assessment in the languages curriculum and putting assessment into practice. Inbra-Lourie argued that teachers should ask and answer critical questions about assessment matters. Being informed by a social-constructivist view, she stressed a dynamic understanding of assessment in context. Taylor (2009) supported Inbra-Lourie's perspective on assessment on the grounds that it is a broader view of assessment, so it is more tangible to a larger audience. Following Inbra-Lourie, Taylor concluded that current frameworks/trends should not only consider the three components proposed by Davies (2008), i.e., skills, knowledge and principle, but also the social and educational roles of assessment.

In the field of general education, Stiggins (2012) seemed to adopt a knowledge-based constructivist approach towards assessment literacy. He put forward the idea of "quality

assessment" and believed that assessment is both a cause and a measure of learning. He pointed out that there are five key features that should be taken into consideration for doing quality assessment: 1) the purpose of assessment, 2) a clear and appropriate vision of learning targets, 3) the target and information need of the user, 4) communicating the results to the intended user, and 5) student involvement in the process while they are learning. He believed that every successful teacher should know these five keys. Following a constructivist approach, Stiggins noted that every teacher should know different assessment methods and should understand what assessment method is appropriate to sample student performance in a specific context and minimize bias that can distort the results. Stiggins suggested that teachers should understand that no method is superior to the other one and that they are not interchangeable.

The American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA) (1990) jointly developed seven standards for teacher competence in educational assessment. The standards are as follows:

> 1. Choosing assessment methods appropriate for instructional decisions.
> 2. Developing assessment methods appropriate for instructional decisions.
> 3. Administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods.
> 4. Using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
> 5. Developing valid pupil grading procedures that use pupil assessments.
> 6. Communicating assessment results to students, parents, other lay audiences, and other educators.
> 7. Recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information (Campbell, 2013, p.77).

But as emphasized by Campbell (2013), these standards exclude some important competencies such as formative assessment and accountability assessment that a teacher typically needs. Campbell believed that assessment for learning guided by a constructivist perspective of learning differs from context to context; consequently, determining a clear-cut set of standards for all the contexts does not seem warranted.

Fulcher (2012) in his model of TAL addressed the importance of context besides skills, knowledge and principles. He noted that most assessment techniques are based on large-scale assessment and are not applicable to the classroom. Fulcher suggested that teachers should have a grasp of classroom-based approaches besides large-scale testing. He considered three dimensions for TAL: Practice, concept and context. He proposed the following working definition of AL based on the findings of his study:

> 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. (p.125)

Based on these three dimensions, in their TAL model proposed for CA research, Hill and McNamara (2012) addressed four questions. 1.What do language teachers do in the classroom? 2. What do teachers look for? 3. What theory or standard do they use? and 4. Do learners share the same understanding? In her new model, Hill (2017) added a new question on the effect of context on CA i.e., 5. How does the context for teaching shape assessment practices? Hill noted that this framework is grounded in CA practices and has been elaborated with reference to the principles of TAL which can be used by teachers to understand their practices. Hill (2017) criticized Fulcher's (2012) definition of TAL in which he addressed merely planned and formal types of assessment and focused on tests and test processes (p. 2). She underlined the importance of informal assessment in the classroom and noted that the first question above is divided into two subcategories and encompasses both formal and informal assessment. She noted that informal assessment can have a strong impact on learning and suggested that any conceptualization of assessment should consider informal and incidental assessment.

In their TAL model, Xu and Brown (2016) tried to connect the skill component with the knowledge component plus details of the sociocultural context. Since the previous TAL models did not account for the needs of pre-service and in-service teachers, Xu and Brown proposed a new model which covers all stages of teacher education and development. They underlined the importance of the knowledge base and stated that, without a knowledge base and/or a set of standards, appropriateness of classroom practices cannot be evaluated. Based on past research and models proposed for the knowledge base and the findings of recent research, Xu and Brown proposed a new knowledge base for TAL that includes the following components:

1-Disciplinary knowledge and pedagogical content knowledge (PCK): Knowledge of discipline and how to teach content

2. Knowledge of assessment purposes, content, and methods: why assess (formative vs summative), connection between assessment methods and content, different assessment strategies

3. Knowledge of grading: using rubrics, objectively marked testing, subjective marking and consistency

4. Knowledge of feedback: different type of feedback and their strengths and weaknesses

5. Knowledge of assessment interpretation and communication: interpreting evidence and assuring the warrants

6. Knowledge of student involvement in assessment: benefits of peer and self-assessment, strategies for using them in different tasks

7. Knowledge of assessment ethics: use, storage and dissemination of results; equity and social justice (Xu and Brown, 2016, p.156)

They noted that a knowledge base is necessary but not sufficient for TAL in practice (TALiP) because it is like a “decontextualized guideline” for dynamic and complex classroom practices. They cautioned that these forms of knowledge show general principles without considering the dynamic context which can cause the knowledge base to change. They argued that following the contextualized view requires consideration of many stakeholders such as students, administrators and policy makers in AL. They pointed out that teachers’ decision-making is under the influence of external factors and that teachers’ beliefs, values, and decisions are balanced with

new tensions and contextual needs. Xu and Brown (2016, p.157) concluded that AL is better understood in practice where teachers experience contradictions and balance their classroom practices with contextual factors and constraints. In their framework proposed for TALiP, Xu and Brown considered teachers' knowledge base, their conceptions and their classroom practices as the main components and noted that these elements impact each other under the influence of micro and macro contextual variables i.e., the dynamics at work within and outside the classroom.

For language assessment literacy, Scarino (2013) believed that the knowledge base and the process for its development should be taken into account. She stated that the knowledge base should consist of not only knowledge of assessment but also knowledge of theories and practices, evolving theories of language and culture, and different epistemological views about the nature of knowledge. She believed that, in addition to theoretical knowledge, teachers' interpretive frameworks which are developed as a result of teachers' experience and beliefs should be considered.

Some researchers (e.g., McNamara, 2001; Willis, Adie & Klenowski, 2013; Xu & Liu, 2009;) define assessment as a social practice, arguing that teachers' real experiences in the classroom should be considered to gain a deeper understanding of their AL. In this regard, Xu and Liu (2009) emphasized the practical aspect of TAL and argued that some aspects of each teacher's knowledge are personal and dependent on the teacher as each teacher's experience is unique. They explained that since assessment is a social practice, research on assessment should be done in the real classroom setting to fully grasp teachers' real experiences. To illustrate the difference between teachers' knowledge in the classroom and outside the classroom, Xu and Liu used the metaphor of *secret* stories and *sacred* stories. Sacred stories refer to out of class visions which are influenced by the researchers, policy makers and administrators, according to them. Secret stories occur in the classroom, and are the lived stories of practice. This suggests that teachers might report following

principles set by policy makers, administrators and researchers but what they actually do in the classroom can be different. Xu and Liu reached the following conclusions about teachers' knowledge, practice, and development in assessment. First, the tension between teachers' implicit knowledge of assessment and prescribed standards calls for new methodology to address the uniqueness and individuality of teachers' assessment practice. Second, the salient issue of teachers' actual practices in assessment and the limited understanding of such practices require further empirical research. The interactive nature of teachers' practice also highlights the importance of situating teachers' assessment practices in larger sociocultural contexts.

All in all, the above studies highlight the importance of TAL and the need for further research in this area. They call for further empirical research on teachers' assessment practices in the classroom to gain a better picture of their AL and propose a solid comprehensive model for it (e.g., Hill, 2017, Xu & Liu, 2009). In the dynamic classroom context, under the influence of different factors, teachers use their own interpretations to make effective assessment decisions in the CAs. Therefore, in any model for teachers' AL and knowledge base, teachers' actual experience, interpretations and arguments should be considered.

Regarding the different components of the knowledge base, the models proposed for TAL and knowledge base seem to be more theoretical, without considering teachers' real practices and the different forms of knowledge they actually use in the classroom. Further research on the assessment procedures and criteria teachers use to argue for the effectiveness of their CAs can help us better understand the different forms of knowledge that teachers use and, accordingly, develop a better model of their knowledge base.

2.6.1 The Importance of assessment knowledge in practice: Implementation of AL in practice

As for the implementation of AL in assessment practices, a distinction should be made between teachers' theoretical knowledge and their practical knowledge. Shulman (1987) made a

distinction between subject matter knowledge (i.e., discipline knowledge) and pedagogical content knowledge. He noted that discipline knowledge by itself is not enough to be transmitted and used for teaching purposes; it should be modified to be used. Following Shulman, Scarino (2013) argued that subject matter knowledge by itself is not adequate, and a distinctive knowledge is required to enable teachers to implement discipline knowledge. She maintained that "knowledge is contextualized in the realities of teachers' context of practice as pedagogical or practical and experiential knowledge" (p316). Scarino explained that teachers' theoretical knowledge base is the main source of input i.e., it is the main component of each teacher's assessment literacy. Nonetheless, if the goal of AL is to bring change to teachers' practice and understanding, in addition to theoretical knowledge, their interpretive frameworks, which are developed as a result of their personal experience, understanding and beliefs, should be considered. As such, she posited that teachers need more training on the construct of assessment on a theoretical, practical and institutional level. At the end, she noted the need to consider the interplay between theoretical knowledge and teachers' personal understanding of this knowledge and how it is reshaped for the purposes of assessment and teaching.

On a similar note, Xu and Brown (2016) expressed their concern over the transferability of assessment knowledge to practice. They pointed out that TAL can be estimated by different measures used in the field, but the implementation of AL knowledge in practice is not guaranteed. In interviews with ESL teachers, Vogt and Tsagari (2014) found that although the teachers reported that they were familiar with different language testing and assessment methods, they were not able to clarify and provide concrete examples of how they used those methods in their classes. Similarly, in Graham's (2005) study some teacher candidates were not able to enact their theoretical assessment knowledge. This suggests that assessment literacy does not merely consist of theoretical or conceptual knowledge of assessment; it encompasses teachers' knowledge of how to implement

different assessment methods as well. That is, the scope of assessment literacy should be wider than theoretical knowledge about assessment.

Despite acknowledging the importance of theoretical assessment knowledge, the above studies suggest that it is not enough by itself to ensure effective CA. The theoretical knowledge is decontextualized, so its implementation in assessment practices is not guaranteed. Classroom dynamics and different contextual factors might give teachers a deeper understanding of context needs and lead them to modify their earlier assessment knowledge.

2.7 Contextual factors, teachers' AL and their conceptions

A number of researchers (e.g., Borg, 2003; 2006; Willis et al., 2013; Xu & Liu, 2009; Xu & Brown, 2016) have posited that teachers' AL is dynamic and their practices in the classroom are influenced by contextual factors. Due to the interactive and dynamic nature of CA, teachers' decisions might differ from class to class depending on various contextual factors including students' needs, course objectives, school policies, and classroom events. Cheng, Rogers and Hu (2004, p.378) reported that different contextual factors such as needs, level of students, teaching/learning environment and standardized exams contribute to the complexity of CA practices and teachers' decisions. In their study, Cheng et al. found that in Beijing teachers use assessment to plan and improve their instruction (instructional purposes) while in Canada and Hong Kong, they use it to provide feedback to their students and motivate them to learn (student-centred purposes).

Xu and Liu (2009, p.496) believed that teachers' knowledge is dynamic, and it is an ongoing process. They stressed the importance of context and power relations in assessment practices in the classroom and argued that practicing formative assessment is complex due to the "interactive nature and multiplicity of the contexts" in which it is practiced. They postulated that teachers' assessment practices are influenced by external factors such as researchers, policy makers, and administrators

and internal class related factors as CA is a social practice and teachers use their discretion in assessment practices.

Willis, Adie and Klenowski (2013) stated that assessment literacy is dynamic, context-dependent social practice and that assessment literacy implementation depends on many intervening factors. On a similar note, Xu and Brown (2016) argued that the classroom context is dynamic and that teachers' assessment decisions are influenced by contextual factors. They posited that teachers experience a lot of tensions in classroom practices, which might lead to the modification of their AL and their conceptions. They speculated that policies, power relations in the workplace and teachers' awareness of their identity as assessors can be possible influential factors that might impact the implementation of TAL in the classroom.

Concerning the relationship between teachers' cognition and practices, some researchers believe that teachers' cognition and practice are interrelated, and that contextual factors affect the extent to which teachers can practice activities in line with their cognition (e.g., Borg, 2003, 2006, 2018; Brown et.al., 2019; Fives & Buehl, 2012; Wang et.al., 2020; Lam, 2019; Zheng, 2013a). Borg (2003) noted that there are many factors influencing teachers' practices in the classroom and these factors might interact or even conflict with each other in shaping classroom practice. After reviewing research findings on teachers' beliefs and practices, he concluded that there is little reference in past studies to the contextual factors which constrain or facilitate teachers' assessment decisions. Specifically, as an example, he pointed to teachers' freedom and set curriculum that can play a crucial role in understanding language teachers' decisions.

Borg (2006) pointed out that classroom environments both influence and are influenced by teachers' cognition and practices. This interaction can lead to adjustments in teachers' beliefs or create conflicts between their beliefs and actions. He emphasized that teachers' personal, social, and historical experiences shape their cognition, which then impacts their teaching methods. Borg

(2018) further highlighted that the connection between beliefs and practices is complex and non-linear, as multiple beliefs interact and are shaped by external factors like policies and curricula. In a similar way, Brown et al. (2019) emphasized that teacher conceptions of assessment are not universally uniform, as they are shaped by the unique historical, cultural, social, and policy contexts in which teachers operate. As a result, these beliefs are always influenced by local factors.

Similarly, Zheng (2013a) stressed the importance of context and contextual factors in shaping a teacher's behaviors in the classroom and explained that teachers' practices are influenced by the needs of a context and classroom events. Zheng explained that the organic nature of interaction between teachers' beliefs, practice and classroom events can lead to inconsistencies. Zheng supported situated views of teachers' beliefs and noted that teachers' beliefs are context based and situational in the sense that teachers use their interpretation when applying their beliefs to make pedagogical decisions in different contexts. "*Teachers' beliefs are a complex and interrelated system, which are tacitly held and grounded in the experiences teachers pass through in the socio-cultural context.*" (Zheng, 2013a, p.399).

Fives and Buehl (2012) mentioned several internal factors that influence how teachers align their actions with their beliefs, including personal beliefs, knowledge, self-efficacy, and identity. Additionally, external contextual factors, such as students' attitudes, educational policies, and cultural norms, may also impact their practices.

Wang et al. (2020) explored whether teachers' beliefs aligned with their self-reported writing practices and found that inconsistencies between teachers' reported beliefs and their classroom practices can be attributed to contextual factors at the micro, meso, and macro levels. At the micro-level, inadequate training in language assessment and concerns about students' English proficiency and motivation hindered the application of AfL. While teachers supported AfL, they struggled to implement it effectively due to limited training. At the meso-level, institutional factors such as syllabus requirements, assessment policies, workload, and large class sizes created

challenges. For instance, schools prioritized the quantity of writing and summative scores over formative feedback, forcing teachers to balance AfL principles with administrative demands. At the macro-level, China's high-stakes testing culture emphasized scores, prompting teachers to overuse numerical evaluations despite preferring comment-only feedback.

Lam (2019) found that while teachers in her study were knowledgeable about the differences between AoL and AfL, they lacked understanding of AaL. Lam also recognized that AfL was more effective in enhancing learning than AoL. However, the study revealed that while teachers tried to implement alternative writing assessments, such as peer assessment and post-writing activities, they faced institutional barriers. Lam concluded that teachers' assessment practices were influenced by meso-level constraints, including a lack of collaborative culture, school support, autonomy, and opportunities for professional development.

Similar to Xu and Brown (2016), Lam (2019) and Harris and Brown (2009) noted a significant tension between what teachers believe is best for students and what is required for school accountability. They suggested that teachers should choose assessments that strike a balance between societal and school assessment expectations and the needs of students, in order to reduce this tension. In this respect, Phipps and Borg (2009) made a distinction between core and peripheral beliefs and defined core beliefs as broad, overarching principles about learning that exert a powerful influence, while peripheral beliefs are narrower in scope. They posited that inconsistencies between beliefs and practices can stem from the "tension" between core and peripheral beliefs. They suggested that understanding these discrepancies requires examining the distinctions between these two belief subsystems.

These studies suggest that contextual factors can influence teachers' practices and their application of AL or beliefs in their assessment decisions in the classroom. In their practices, teachers might act against their AL or beliefs about assessment due to contextual factors i.e., they might withhold their real assessment beliefs to align themselves with the needs of the context and

practice strategies and techniques in the classroom that are inconsistent with their AL or beliefs. There is a need for further empirical research in this area concerning which contextual factors negatively or positively impact the implementation of teachers' AL or beliefs in assessment practices in general and assessing oral communication in particular. That is, which contextual factors impede the application of teachers' AL or beliefs or, alternatively, cause their revision/reconstruction, and how these factors contribute to the process of AL or beliefs development. As noted by Campbell (2013), there is a need for further research in different contexts given the possibility that "*contextually unique classrooms may reveal barriers not previously identified*" (p.80).

All in all, the review above reveals some of the limitations of existing TAL models for the classroom context. Most of them are theoretical; not based on the results of empirical research. As noted above, research on teachers' practices and their interpretations in the classroom can provide useful insights to inform the development of a model for TAL. Additionally, because an important part of teachers' AL and, accordingly their beliefs, seems to develop in the classroom, we need to examine their real assessment practices, their lived experiences, the criteria they consider for effective assessment, their interpretations regarding solid CA, the contextual factors and challenges that might influence their assessment decisions, and the strategies they use to deal with these challenges in order to better understand their AL and knowledge base as well as the changes, or developments that their AL and beliefs undergo in the classroom.

2.8 Conceptual framework

Among the assessment literacy frameworks reviewed in the preceding sections, Xu and Brown's (2016) framework of Teacher Assessment Literacy in Practice (TALiP) was considered particularly appropriate for the present study because it provides a more comprehensive representation of assessment literacy as enacted in classroom assessment. While some earlier

frameworks primarily conceptualize teacher assessment literacy in terms of assessment knowledge, skills, competencies, or individual teacher characteristics, Xu and Brown's framework connects teachers' knowledge base, conceptions of assessment, assessment practices, and institutional and sociocultural contexts within a broader practice-oriented conceptualization of teacher assessment literacy. Their framework was developed from a synthesis of 100 studies and was intended to move the conceptualization of teacher assessment literacy beyond a primary focus on the knowledge base toward a situated, dynamic, and evolving system in which teachers negotiate assessment in relation to contextual conditions.

This orientation was particularly relevant to the present study because the study sought to examine not only what ESL teachers know about assessment, but also how their assessment knowledge and conceptions of assessment relate to their classroom assessment practices and how contextual conditions influence the enactment of assessment. In particular, the framework provided a conceptual basis for examining two related dimensions of teacher cognition–assessment knowledge and conceptions of assessment–together, rather than focusing on either dimension in isolation. This was important for developing a more comprehensive understanding of classroom assessment and teachers' assessment decision-making. The framework's emphasis on assessment literacy in practice was also consistent with the study's focus on teachers' actual classroom assessment rather than on assessment knowledge as a decontextualized body of theoretical knowledge. In addition, its explicit recognition of institutional and sociocultural contexts provided a useful basis for examining how contextual conditions may facilitate, constrain, or shape the implementation of teachers' assessment knowledge and conceptions.

Xu and Brown's (2016) framework consists of six interrelated components: the knowledge base, teacher conceptions of assessment, institutional and sociocultural contexts, teacher assessment literacy in practice, teacher learning, and teacher identity (re)construction as assessors. The authors

emphasize that these components are interconnected rather than independent, and their model uses multidirectional relationships to represent the dynamic nature of TALiP. In their account, the knowledge base is a necessary but insufficient condition for assessment literacy in practice; teachers' conceptions, contextual conditions, teacher learning, and identity development all contribute to how assessment literacy is enacted. They conceptualize TALiP as emerging through compromises among tensions involving teachers' conceptions, the expected knowledge base, and contextual conditions.

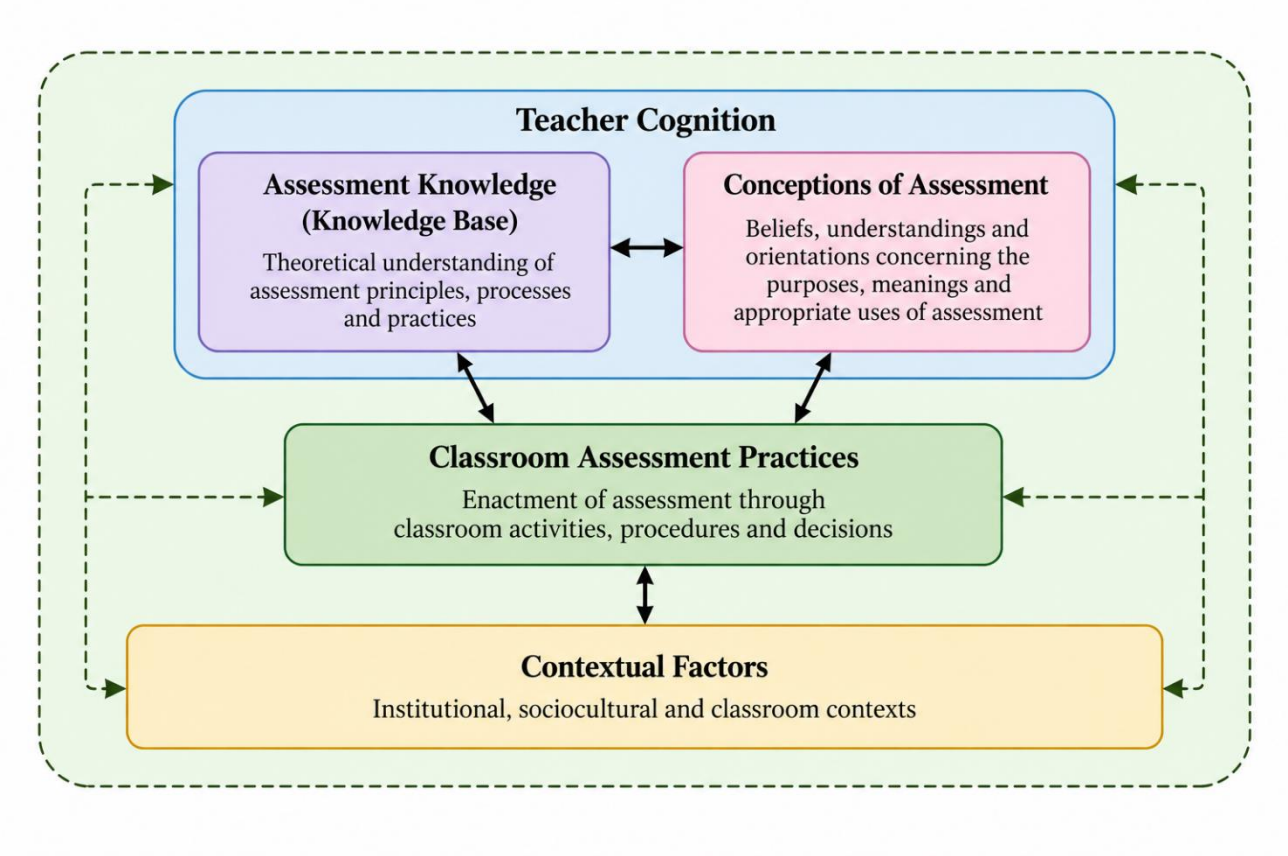
The present study adopted Xu and Brown's (2016) framework as its theoretical foundation but adapted its representation to foreground the relationships that were central to the research questions. Specifically, the adapted framework focuses on assessment knowledge, conceptions of assessment, classroom assessment practices, and contextual factors. Assessment knowledge was conceptualized as teachers' theoretical understanding of assessment principles, processes, and practices. Conceptions of assessment were treated as a related but conceptually distinguishable dimension of teacher cognition concerning the purposes, meanings, and appropriate uses of assessment. Classroom assessment practices were conceptualized broadly as the enactment of assessment in teachers' everyday classroom work, while contextual factors were understood as conditions within and beyond the classroom that could influence this enactment.

The less-hierarchical arrangement of these components in Figure 2.1 is an intentional adaptation for the purposes of the present study. Whereas Xu and Brown represent their six components within a pyramidal framework depicting different layers and stages of TAL development, the present study was specifically concerned with examining assessment knowledge and conceptions as parallel and related dimensions of teacher cognition and their respective relationships with classroom assessment practices. Accordingly, the adapted framework does not represent assessment knowledge as necessarily preceding or determining conceptions, or

conceptions as necessarily determining classroom practice. Rather, it foregrounds the possibility that both dimensions of teacher cognition are relevant to classroom assessment, while contextual factors shape their enactment. This representation should not be interpreted as a claim that Xu and Brown conceptualized knowledge and conceptions as parallel components in precisely this manner; rather, it represents a study-specific adaptation of their framework designed to reflect the conceptual focus of the present research.

Figure 2.1

A Conceptual Framework for Examining ESL Teachers' Assessment Literacy, Conceptions, and Classroom Assessment Practices (Adapted from Xu & Brown, 2016)



The framework therefore provided a conceptual lens through which the relationships among teachers' assessment knowledge, conceptions, classroom assessment practices, and contextual

factors could be examined. Importantly, it was used as a starting theoretical framework rather than as a predetermined explanatory model. The study did not assume that the relationships among these components would necessarily be linear or uniform across teachers or contexts. Instead, the framework provided the initial conceptual structure for investigating how these dimensions of teacher cognition were manifested in classroom assessment and how contextual conditions shaped their enactment.

While Xu and Brown's (2016) framework was particularly useful in moving beyond a static conception of assessment literacy, they themselves acknowledged the need for further empirical research to examine how TAL develops across contexts. They described their framework as an initial attempt to delineate different layers and stages of assessment literacy development and explicitly called for "more empirical studies across contexts" to support the framework. They also called for research examining teacher learning, teacher identity construction, contextual influences, and the complex process through which teachers become assessment literate. This call for empirical investigation was particularly relevant to the present study, which sought to examine TAL through teachers' actual classroom assessment practices rather than relying solely on theoretical conceptualization or self-reported assessment knowledge.

Moreover, although Xu and Brown recognize the dynamic and evolving nature of TALiP, their framework does not fully specify how teachers' real classroom experiences and interpretive processes contribute to the development and refinement of assessment knowledge and conceptions in day-to-day classroom assessment. Their model acknowledges teacher learning, reflection, identity reconstruction, and the reciprocal influence of practice and knowledge; however, the precise mechanisms through which teachers' lived assessment experiences are interpreted and incorporated into the development of their assessment literacy remain an area for empirical investigation. This issue also resonates with Scarino's (2013) emphasis on teachers' interpretive processes and the role

of experience in developing the interpretive frameworks through which assessment is understood and enacted. Xu and Brown themselves cite Scarino in highlighting the importance of teachers' preconceptions, interpretive processes, and identity as assessors.

Accordingly, one goal of the present study was to further examine and potentially expand Xu and Brown's (2016) framework by investigating teachers' actual assessment experiences, assessment procedures, interpretations, and the relationship between their theoretical and practical knowledge in classroom practice. The conceptual framework thus provided the theoretical starting point for the study, while the empirical investigation was intended to provide a more grounded understanding of how assessment literacy is enacted and developed in the ESL classroom. The developmental and practice-based relationships identified through the findings were subsequently used to refine and extend the initial conceptualization rather than being assumed in advance.

CHAPTER 3: METHOD

3.1 Overview of the study

Previous studies have employed various data collection methods to examine teachers' beliefs, practices, and assessment literacy, with many scholars emphasizing the limitations of relying on a single instrument. Phipps and Borg (2009) and Borg (2006) argued that questionnaires often reflect teachers' idealized beliefs rather than actual practices, whereas interviews and classroom observations provide more authentic insights. Researchers such as Campbell (2013), Sanchez (2014), and Hill (2017) similarly stressed the importance of examining teachers' real classroom practices, noting that teachers may possess practical assessment knowledge that they are unable to articulate. Moreover, self-reported data may be influenced by factors such as self-efficacy, leading to over- or underestimation of assessment competence (Campbell, 2013). As a result, several studies have advocated for a multi-method approach and extended data collection to reduce bias and capture a fuller picture of teacher assessment practices (Graham et al., 2014; Rea-Dickins, 2001). Taken together, these studies suggest that a multi-method approach incorporating interviews, observations, and carefully designed questionnaires is essential for obtaining a more accurate and nuanced understanding of teachers' assessment literacy, beliefs and practices.

This study adopted a qualitative, grounded-theory-informed design to explore ESL teachers' classroom assessment practices, assessment literacy, assessment conceptions, and the contextual factors influencing their assessment practices and the development of their AL and conceptions. The study first examined the assessment procedures teachers followed to conduct classroom assessment and ensure quality assessment. It then explored how their assessment conceptions and AL were related to these practices and how their knowledge, experiences, and contextual conditions contributed to the development and enactment of their assessment literacy and conceptions.

Given the multidimensional nature of the constructs examined in this study, multiple data collection methods were used, including a study-specific TAL questionnaire, semi-structured interviews, classroom observations, stimulated recalls, and document collection. Due to the multidimensional nature of TAL, different sources of evidence were used to examine its relationship with teachers' assessment practices. The questionnaire was used to examine teachers' self-reported assessment knowledge across different areas and its alignment with their assessment practices. Interviews provided a more in-depth understanding of how teachers' assessment knowledge and experiences contributed to their assessment decisions and practices. In particular, the interviews explored the role of teachers' past education, theoretical knowledge, and L2 learning experiences as sources of foundational assessment knowledge, as well as the role of assessment experience in shaping, revising, and developing their assessment literacy through classroom practice. Thus, the questionnaire and qualitative data provided complementary perspectives for examining not only the extent to which teachers' reported assessment knowledge was reflected in their practices, but also how different sources of knowledge and experience contributed to the enactment and development of their assessment literacy.

Teachers' assessment conceptions and classroom assessment practices were explored primarily through interviews, classroom observations, and stimulated recalls. The observations provided opportunities to examine teachers' assessment practices as enacted in the classroom, while stimulated recalls provided further insight into teachers' thinking about particular classroom assessment events. In addition, different documents, including course outlines, task guidelines, rubrics, and other assessment documents used by teachers, were examined to provide further evidence of their assessment procedures, criteria, and practices and to examine how their stated assessment knowledge, conceptions, and practices were reflected in their assessment materials.

The qualitative data were analyzed iteratively using thematic analysis informed by grounded theory principles, including constant comparison, while descriptive statistics were used to summarize the TAL questionnaire responses. Evidence from the different data sources was integrated where relevant to provide complementary perspectives and to corroborate, complement, or contextualize teachers' reported assessment knowledge, conceptions, and practices. The data collection tool(s) related to each research question are shown in Table 3.1 below.

Table 3.1

Overview of *the Study Data Collection Tools*

Research Questions	**Tools**	**Explanation**
1-What are ESL teachers' practices and criteria for assessing oral communication in the classroom?	Interview 1 & 3 +Classroom observation + review of course materials	The participants were asked about the procedures they follow for doing CA, what criteria they use for evaluating assessment outcomes, and what methods & goals they follow in CAs. Their reported and observed assessment practices and assessment documents were used for data analysis.
2-What are teachers' L2 assessment conceptions and their relationship with assessing L2 oral communication in the ESL classroom?	Interview 2 + course material/documents used in their CA practices+ observation and SRs	They were asked about their approach to L2 assessment, their beliefs regarding the nature and purpose of CA, and their application in classroom practices and assessment tools.

3- What is the relationship between teachers' L2 assessment literacy and their practices in assessing oral communication in the ESL classroom?	TAL Questionnaire + demographic information + their reported teaching/learning/assessment experience, past education and theoretical knowledge in interview 1	They were asked about their past education, theoretical knowledge, learning experience and teaching/assessment experience and their relationships with their assessment practices were examined. They also completed a TAL questionnaire which targeted different forms of assessment knowledge which seem to be important for effective CA according to the conceptual framework and different models proposed in the field.
4-What contextual factors influence the implementation of teachers' assessment literacy and conceptions in the assessment of oral communication in the ESL classroom?	Interview 2 & 3 + classroom observation+ document analysis	They were asked about influential constructive, and detrimental contextual factors on classroom practices. Also, they were asked about the challenges they encounter in the classroom and how they cope with these challenges.

3.2 Researcher Positionality and Reflexivity

I considered my positionality throughout the study because of my prior professional and academic engagement with ESL teaching, language assessment, and assessment literacy. My interest in the study developed from my engagement with assessment in general and assessment literacy in particular. I had taught ESL, including oral communication, in different contexts for many years and had also conducted research on language assessment and assessment literacy. This background provided me with familiarity with the field and the context of the study, while also

bringing prior understandings and assumptions that could influence how I interpreted participants' experiences and practices.

I occupied a mixed insider–outsider position in relation to the participants. Some participants were my professional colleagues, while others were not previously known to me and were recruited through the study. My shared background in ESL teaching and assessment helped establish rapport and made the interviews more conversational, as participants knew that I was familiar with the area. At the same time, this familiarity required me to remain aware of the possibility of taking meanings, practices, or assumptions for granted and to distinguish participants' understandings from my own. I was also attentive to the possibility that prior professional relationships could influence what participants felt comfortable sharing.

My role was particularly relevant during the semi-structured interviews. I took notes and asked follow-up questions when participants raised interesting, unexpected, or unclear issues. While this allowed participants to elaborate on their experiences and helped clarify their views, I recognized that my questions and interactions could also influence the direction of the conversations. To minimize the possibility of imposing my own views, I used open-ended questions and follow-up prompts primarily to seek clarification or elaboration rather than to suggest particular interpretations. During classroom observations, I adopted a non-participant observer role and avoided becoming involved in classroom activities, while recognizing that my presence could nevertheless influence the observed setting. Similarly, during stimulated recalls, I used follow-up questions to clarify participants' thinking about particular classroom events without suggesting what they might have been thinking or why they had made particular decisions.

I engaged in reflexivity throughout the research process by considering how my background, assumptions, and interactions with participants might shape the research process and interpretation of the data. I remained open to views and practices that differed from my initial expectations and returned to the relevant data when interpretations required further consideration. In analyzing the data, I sought to distinguish participants' accounts from my own prior understandings of assessment. I returned to the interview, stimulated-recall, observation, and documentary data to examine whether interpretations were supported by the evidence, particularly when they differed from my initial expectations. I also used the different data sources in complementary ways to develop a richer understanding of participants' assessment literacy, conceptions, and practices. For example, classroom observations provided concrete examples of practices discussed in interviews, while assessment documents were examined in relation to participants' accounts to explore how their stated understandings were reflected in their assessment tasks, rubrics, instructions, feedback, and other assessment materials. These comparisons provided opportunities to examine convergence and divergence across data sources and to reconsider interpretations when necessary.

3.3 Sampling and participants

This study employed purposeful sampling to collect data from the teachers who were most likely to provide valuable insights into the research problem. According to Dörnyei (2007), qualitative research does not prioritize the representativeness of the respondent sample or the distribution of experiences across the broader population. Rather, its primary focus lies in selecting individuals who can offer rich, meaningful insights into the phenomenon being investigated, thereby maximizing the potential for learning. Dörnyei, whose discussion of purposeful sampling is situated within the context of second language acquisition and classroom-based research, advocated for the use of purposeful sampling as the most appropriate method to achieve this objective. In this study, sampling was done at multiple levels: Institution and program, course and teachers.

In this study, the first level of sampling involved the universities or colleges which offer ESL courses for academic purposes. The second level of sampling involved identifying and selecting ESL courses that include practicing and assessing oral communication along with other skills or separately since the purpose of the study was to examine teachers' AL, beliefs and practices in assessing oral communication. The description of ESL courses is provided in section 3.4.

The final level of sampling pertained to the teachers. The participating teachers for this study were recruited purposively based on the following specific criteria (Dörnyei, 2007). Firstly, teachers who taught oral communication in their courses for academic purposes were recruited, regardless of whether they practiced other skills in the course or not. Secondly, teachers, who had taken an assessment course in their past studies and had experience in teaching and assessing ESL courses and oral communication, were recruited so that I could explore the relationship between their theoretical knowledge and experience with their assessment practices. Teacher's age, educational background, qualifications, and other teaching experiences were not considered during participant recruitment because these factors were deemed part of the teacher's personal context.

To recruit participants for this study, I contacted colleagues who were teaching ESL for communication purposes in academic contexts. I also contacted the English Departments at various universities and colleges via email to request permission for data collection. Upon receiving approval from the department heads, I requested them to share my contact information along with a brief description of the study goals. Subsequently, I introduced myself to the interested teachers, providing an overview of the study and details about the data collection procedures. I explained that I am looking for teachers who are teaching courses that cover L2 oral communication. Interested participants were asked to complete a preliminary questionnaire, which gathered information about their teaching experience, the courses they were teaching at the time of data collection, the courses they had taught previously, and a brief description of the course in which they teach oral

communication (see Appendix A). Fifteen teachers expressed interest in participating, but only ten reported (a) teaching oral communication for academic purposes in their courses, (b) having taken at least one assessment course and (c) having some experience teaching and assessing ESL courses and oral communication. As a result, only these ten teachers were invited to participate in the study. The other five teachers were not invited because they, either were not teaching oral communication or did not take any assessment courses.

Demographic information about the participants is displayed in Table 3.2. It shows that most of the participants had more than 10 years of teaching experience and held an MA or a PhD degree. All of them had taken an assessment course during their undergraduate or graduate studies and most of them had taken a training course such as CELTA or had TESL Ontario accreditation.

Table 3.2

Participants' Demographic Data

Participant pseudonym	Years teaching ESL	Past education	Training/accreditation	Program in which assessment courses were taken	Academic Context
Aly	20	PhD Candidate MA: TESL Med Edu BA: TESL	TESOL Certificate	MA	University
Arman	25	MA-TESL BA-English	CELTA, TESOL Certificate	BA& MA	University
Dave	20	PhD: in TESL MA: in TESL BA: in English	TESOL Certificate	BA & MA	College
Donna	5	PhD: Social Science BA: TESOL MA: Social science?	NA, IELTS Examiner Training	BA	University
Hudson	10	PhD: TESL MA: ESL	NA	MA & PhD	University

		BA: engineer			
Mike	12	PhD: TESL Candidate Med BA: TESL	CELTA-TESOL Certificate	BA, MA, PhD	University
Ray	15	PhD: TESL MA: TESL BA: TESL	TESOL Certificate Webinar, IELTS Examiner Training	BA, MA, PhD	University & College
Reymond	10	PhD Candidate MA: TESL BA: TESL	TESOL Certificate	BA & MA	University
Ryan	15	PhD: TESL candidate MA: TESL BA: TESL	CELTA	BA, MA & PhD	College
Sally	10	MA	CELTA	BA & MA	College

This research was approved by York University Ethics Review Board. Participation in the study was voluntary and participants were asked to sign a consent form (see Appendix B). They were assured of the confidentiality of the information they shared. For confidentiality, pseudonyms were used instead of real names throughout the research.

In the recruitment email as well as teachers' consent form, the purpose and duration of the research as well as compensation for participation were outlined. Participants were informed that the study would pose no risk to them, and that they could withdraw from the research at any stage. Each participant who completed the interviews was given a $ 150 gift card in appreciation of their time. They were informed that if they withdraw halfway through the interviews for any reason, they will be offered this incentive (see Appendix B).

3.4 Study context and programs

This research involved 10 ESL instructors across two adult ESL teaching settings in Ontario Canada: an ESL program at college and a credit-bearing ESL program at university. In both settings, the primary objective of the courses was to equip students with the language skills necessary to excel in their core academic courses and to prepare them for success in Canadian social

contexts. The courses in both programs were content and skill-based with less focus on form. In the ESL courses, different skills, including reading, writing, speaking and listening were taught. Different types of academic oral communication skills such as debating, presentation, group discussions, and summary retelling were addressed in the courses.

In the college setting, ESL programs consist in intensive language training courses specifically designed to equip students with the academic English skills required for success in their education. These programs typically follow a rigorous schedule, for example, four hours of instruction per day, five days a week, over the course of eight weeks, and focus exclusively on reading, writing, listening, and speaking tasks commonly encountered in higher education settings. Unlike the more intensive and tightly structured ESL programs in college, the credit courses in the university setting follow a standard university semester consisting of three-hour classes per week over a term of 12 weeks and are taken as part of a full degree program. They are typically organized around thematic or content-based topics, such as Canadian society and culture and focus on developing academic and communication skills for their ongoing degree program.

3.5 Courses

To meet the goals of the study, this research included two ESL courses in which speaking and listening were taught for communication purposes. The participating teachers used different types of oral communication activities, including watching videos and answering /follow up questions, classroom debates, seminars, group discussions, presentations, oral summaries, and oral narratives. In most of the tasks, listening was integrated with speaking and was assessed in relation to speaking and communication skills.

In both courses, different skills (i.e., listening, speaking, reading and writing) were taught during the semester with more focus on writing and speaking. Accordingly, the courses included different activities such as critical reading, academic writing, watching videos, debates, presentations, and

classroom discussions. Although the courses ostensibly integrate listening, reading, speaking and writing skills, in assessments, listening and reading were often used as input for speaking and writing tasks. A brief description of the courses is presented below. To ensure the confidentiality of the teachers and institutions, I refer to the courses as ESL I and ESL II.

ESL I is a university course that met three hours per week over twelve weeks. The course is designed to develop the academic English proficiency of students from diverse linguistic backgrounds, preparing them for success in their undergraduate studies. Emphasizing critical thinking, effective communication, and cultural awareness, the course combines academic reading and writing with oral communication skills. By exploring social conventions and communication norms within the Canadian context, students strengthen their ability to articulate ideas clearly and confidently in academic, professional, and social settings. Oral communication serves not only as a means of language practice but also as a vital tool for developing argumentation skills, collaborative learning, and cultural competence in Canadian academic life. Accordingly, teachers practice and assess different oral communication activities including oral presentation, group discussion, watching and discussing videos, research-based presentations, and summary retelling.

Depending on the type of task, teachers use a range of criteria in their assessment, including the ability to read or listen to understand content information, use visual aids, organize and present ideas accurately and clearly, use appropriate verbal and non-verbal presentation skills, demonstrate content knowledge, articulate ideas clearly and comprehensibly, use speech acts appropriately, think critically, and use debating skills. Listening is used to teach the content and as input for students to discuss, present and complete other speaking tasks.

ESL II is a college-level course conducted over eight weeks, with classes held for 3-4 hours each day, five days a week. Of the 18 hours of teaching per week, four are allocated to teaching oral communication, while the rest of the hours are devoted to teaching and practicing reading and

writing. This course is designed to enhance students' academic and social communication skills. Students engage in a wide range of interactive activities, including group discussions, debates, presentations, and storytelling, to build fluency, coherence, and confidence in spoken English. They participate in social conversations and academic seminars, and practice strategies for expressing opinions, managing conversations, and supporting ideas clearly and effectively. Listening skills are developed through lecture-based tasks, note-taking, and comprehension of both explicit and implied meaning. Process speaking, narrative tasks, and the use of visuals in presentations further strengthen students' ability to organize and communicate complex information. By combining academic language development with real-world communicative practice, this course prepares learners for success in both university and everyday Canadian contexts. Accordingly, depending on the type of the task, teachers assess different criteria and/ skills including students' ability to present effectively and engage others, use functional language such as expressing opinion, agreement/disagreement, asking questions and responding to others appropriately, debate with each other, narrate a story, use socially appropriate language, summarize information, and understand content and main ideas, take notes, and incorporate and interpret them in their speaking.

3.6 Data collection procedure and tools

As noted above, past research suggests that employing multiple data collection methods is essential for developing an accurate and nuanced understanding of teachers' assessment literacy, beliefs, and practices. As such, this research uses different methods of data collection including interviews, observations, stimulated recalls, document analysis, and questionnaire to explore teachers' L2 oral communication assessment practices, beliefs, AL, and related contextual factors, aiming to understand their relationship and develop a comprehensive model for CAs and TAL.

In this study, in order to understand the participants' assessment knowledge base, each participant was asked to complete a TAL questionnaire before the initial interview. Each participant participated in three interviews addressing different questions of the study. In addition to the

interviews, the classes of some participants were observed to gain a deeper understanding of their assessment practices as well as the implementation of their AL and beliefs in the classroom. Also, course material including course outlines, tasks, rubrics, tests, and feedback provided on student performance were collected.

3.6.1 TAL Questionnaire

Researchers have criticized existing TAL questionnaires on the grounds that they assess prescribed competencies, failing to address teachers' assessment needs (Xu & Brown, 2016), and lack consistency (Fulcher, 2012). Two common tools for assessing teachers' assessment literacy (AL) are the Teacher Assessment Literacy Questionnaire (TALQ) (Plake & Impara, 1993) and the Assessment Literacy Inventory (Mertler, 2004). Xu and Brown (2016) cautioned that while these tools identify AL strengths and weaknesses, their reliability is limited. Fulcher (2012) also criticized their lack of consistency and attention to teachers' needs, advocating for a mix of closed- and open-ended questions. However, his AL questionnaire included only one open-ended item, which asked the participants to report their degree of AL. As such, to better grasp teachers' assessment knowledge in different areas, a TAL questionnaire was specifically designed for this study to cover the different forms of knowledge proposed by researchers in the field (see Appendix C). As noted by Vogt and Tsagari (2014) a good questionnaire should be comprehensive, covering all aspects of language assessment.

For operationalizing the knowledge base and developing the questionnaire, the different forms of knowledge proposed by Xu and Brown (2016) as well as the forms of knowledge proposed in other studies or models were used (see Chapter 2). The following knowledge base components proposed in the literature were added to Xu and Brown's list to develop the TAL questionnaire for the purpose of this study:

- Knowledge of different epistemological views (Cheng et al., 2009; Scarino, 2013).

- Knowledge of students' different assessment conceptions/misconceptions (Brown, 2008; Hill & McNamara, 2012).
- Knowledge of influential contextual constraints and the dynamic nature of CA (Xu & Liu, 2009; Zheng, 2013; Willis et al., 2013).
- Knowledge of informal and formal assessment (Popham, 2009; Hill, 2017).
- Awareness of the distinction between practical and theoretical knowledge (Scarino, 2013; Tsagari &Vogt, 2017).

On the whole, twelve types of assessment knowledge were identified as the knowledge base. Four items were developed to assess teachers' self-perceived knowledge in each area; each item took the form of a statement indicating teachers' awareness in the related area. For example, to evaluate self-perceived knowledge of different purposes of assessment, the following items were designed: How knowledgeable are you about: *a)* using assessment for different purposes, *b)* using assessment to improve teaching and learning, *c)* using assessment for assessing what students have learned, *d)* using assessment to diagnose students' strengths and needs. Each item was measured on a 4-point Likert scale ranging from 0= Not at all; 1= Slightly; 2= Moderately; to 3= A lot. This type of scale better reflects respondents' degree of self-perceived assessment knowledge by avoiding neutral responses.

The initially version of the questionnaire with 48 items was reviewed by an expert researcher in the field of second language assessment, who had experience in teaching assessment courses as well as designing different assessment tools, to ensure that all the items relate to the same construct and are easy to understand by teachers. Based on this assessment, some items were revised to make them clearer in terms of wording, some items were removed or added, and some items were merged. Next, to evaluate the internal consistency of the questionnaire as well as the relevance and contribution of each item to the constructs under investigation (i.e., self-reported assessment knowledge), the questionnaire was administered to 10 ESL instructors, similar to the target

participants of the study, with an average of five years of' experience teaching oral communication courses. Cronbach's alpha analysis was computed using SPSS. The results showed that Cronbach's alpha was around 0.8, indicating that the questionnaire has high internal consistency.

The final version of the questionnaire contained 40 items encompassing the different types of knowledge mentioned above including knowledge of the purpose of assessment, using different assessment tools, feedback provision, assessment interpretation and communication, student involvement in assessment, assessment ethics, epistemological views of assessment, students' conceptions, contextual constraints and dynamics, formal vs informal assessment, application of theoretical assessment knowledge, and assessment complexity and balancing tensions (See Appendix C). An online version of the TAL questionnaire was created using Google Forms Survey and its link was emailed to the participants to answer before conducting the interviews.

3.6.2 Interviews

Each teacher participated in three interviews, addressing different questions of the study. For the overall design of the interview questions, I drew primarily on Barkaoui and Valeo' (2019) research on teachers' beliefs and practices in assessing L2 writing, modifying the questions to fit the focus of the current study on oral communication. The interview questions were reviewed by the supervisory committee members to ensure clarity and relevance to the research goals. Before conducting the interviews, a pilot test was done with an ESL instructor, who had ESL teaching experience for more than 10 years, to identify any ambiguous questions. Following the pilot interview, some questions were revised to make them clearer and some were removed due to their overlap with other questions.

Each interview took 45-60 minutes. In order to elicit more in-depth information, the interviews were semi-structured so that new questions could be asked based on issues and ideas emerging during the interviews. The first interview addressed ESL teachers' past education, training, teaching and assessment experiences, and their assessment literacy (see Appendix D). It

also addressed teachers' L2 oral communication assessment practices, both formal and informal, how they use assessment results in their teaching, and what theories guide their practice.

Interview two included three sections (see Appendix E). The first section addressed the effectiveness of the teachers' CA procedure, and the criteria they use to ensure the effectiveness and fairness of their CAs in assessing L2 oral communication. The questions in the second section investigated different contextual factors and their impact on teachers' classroom assessment practices, and possible tensions/challenges teachers experience in the implementation of their AL or beliefs in assessing oral communication in the classroom. In section three, the teachers were asked about their approach to L2 assessment, their beliefs regarding the nature and purposes of the assessment of oral communication, and their application in the classroom. In order to keep the interview short and gather more focused, and consolidated data about teachers' fundamental beliefs regarding the nature and purposes of assessment, some questions asked teachers about their epistemological views about L2 teaching and assessment on a scale with closed-ended questions designed based on the study's conceptual framework and past studies. This scale addressed ESL teachers' general views of the nature of knowledge and L2 teaching and assessment. It was administered at the end of interview two and took approximately 5 minutes to complete (see Appendix E).

The third interview consisted of three sections (Appendix F). The first section asked about the assessment tasks used by the teachers. The participants were asked to select and talk about two tasks they typically used in their classes for assessing oral communication, including their goals and rationale for using these tasks. The second section addressed teachers' use of assessment tools including rubrics, assessment guidelines, evaluation sheets, tests for assessing oral communication, assessment criteria, and the development and effectiveness of these assessment tools. The third section focused on the feedback teachers provided when assessing L2 oral communication. In this

section, participants were asked about their approach to providing feedback on their students' oral communication and the different types of feedback they provide on different features (Appendix F).

3.6.3 Classroom Observations

Researchers have emphasized that classroom observation is crucial for understanding teachers' AL, beliefs and practices (Borg, 2006; Campbell, 2013; Hill & McNamara, 2012; Hill, 2017; Rea-Dickins, 2001; Sanchez, 2014; Xu & Liu, 2009). Observing teachers' actual practices provides a more authentic and accurate assessment of their AL and the effects of teacher training than relying solely on self-reported data, which may reflect idealized or unexpressed beliefs (Borg, 2006, 2018; Campbell, 2013; Graham et al., 2014, Sanchez, 2014). Rea-Dickins (2001) highlighted the value of extensive, long-term observations, often combined with recorded sessions and interviews, to capture a comprehensive picture of classroom practices. Observation allows researchers to analyze teachers' behaviors in authentic teaching contexts, avoiding biases inherent in self-reported data and providing direct evidence of both practices and underlying beliefs (Li, 2017).

As such, to gain a deeper understanding of teachers' classroom assessment practices, including their formal and informal assessments, and how they relate to their AL and beliefs, some participants' classrooms were observed. The selection of participants for observation was based on their willingness to be observed. Four participants agreed to be observed in person and video recorded while one additional participant shared video recordings of his class. Each teacher was consulted regarding the appropriate session and date of observation. Each observation was conducted during a single three-hour class session and it was scheduled in a session where the teacher was conducting activities related to oral communication. The timing of the observations varied across teachers; however, in most cases, they were planned for weeks 7 to 12, after the teachers had established a routine. This timeframe was chosen because of its relevance to the study's focus on oral communication assessment. The initial few weeks of the course were largely

devoted to introductory activities or assessing other skills, with minimal oral communication assessment activities occurring during this time. The timeline of observations is displayed in Table 3.3.

Table 3.3

Timeline of Classroom Observations

Participant	When	Duration of observation	Context
Hudson	March 2, 2020 (week 8)	3 hours	University
Arman	March 12, 2020 (week 9)	3 hours	University
Donna	March 5, 2020 (week 8)	3 hours	University
Mike	March 2, 2020 (week 9)	3 hours	University
Ryan	Recorded class, April 9, 2020 (week 12)	2 hours	College

Before observing the classes, students were asked to sign an informed consent form (see Appendix G). The purpose of the observation was explained to the students, assuring them that the observation will focus on their teacher, not on them or their performance in the classroom. They were asked for their permission to video record the class and in cases where some students did not give permission, the camera was positioned in a way that they were not captured in the video recording or alternatively only audio recording was used. Students were assured that if their voice was recorded during classroom interaction, their voice would be deleted from the recorded audio files using an audio file editing software. All the data and documents were saved on a secure computer accessed by no one but the researcher.

Semi-structured observations were utilized in this study. I used a checklist in line with the study goals in the classroom observations (see Appendix H). However, in the analysis of the recorded classes, I also considered any emergent event, behavior or assessment opportunity that was not in the predetermined categories. Given that behaviors are hard to predict, semi-structured observation could provide more flexibility in documenting a variety of oral communication assessment practices teachers use. Before proceeding to observe the classes, all the documents related to the course provided by the teacher were reviewed and a preliminary analysis was done on their stated assessment practices, tools, and beliefs to inform the observation. During the observation, I adopted the role of a non-participant observer sitting at the back of the class and taking notes without interacting with the students or the teacher. In addition, the whole session was video recorded. While observing the class, in concert with this study's goals and using the guidelines in the observation checklist, I took notes of the oral communication activities that took place, assessment opportunities, interactions between the teacher and the students, and instances of teachers' feedback to the students. I also marked some segments in the video to discuss later with the participant during the stimulated recalls (See Appendix H).

3.6.4 Stimulated Recalls

After video recording each observation, I chose 4-5 segments of the video to use in follow up stimulated recalls. Each segment was approximately 5 minutes long, and each SR session took approximately 30 minutes. During the SR, the formal and informal oral communication assessment activities of the teachers and the underlying reasons for conducting them were discussed. The participants were also given the option to select and talk about any part of the video recording. Guided questions were used for the SR as outlined in Appendix I to ensure a systematic approach. As an example, teachers were asked about their thoughts regarding the selected assessment

practice(s), if it was an effective way to assess their students' oral communication skills, and how they used the assessment results.

Using video recordings of classroom practices facilitated the SRs, allowing the participant teachers to discuss their assessment practices and the rationale for their decisions without relying solely on memory. SRs provided a concrete context for uncovering teachers' beliefs by grounding them in observed classroom events rather than abstract ideas (Borg, 2018). They also offered qualitative insight into teachers' interpretations, reflections, and decision-making when assessing oral communication, helping me understand the relationship between their beliefs and actual classroom practices. In other words, SRs provided a great opportunity to uncover the underlying rationale for actual assessment practices in the classroom as well as teachers' decision-making processes and their interpretations when employing formal and informal assessments.

3.6.5 Document collection

The participants were asked to share copies of their course outlines, assessment tools (e.g., different rubrics, tests, evaluation sheets, checklists, guidelines) or tasks/activities used for assessing oral communication, and sample feedback they provided to their students. Document analysis was used as an additional data source to better understand the teachers' assessment practices, assessment criteria, and assessment beliefs and AL as well as their application in practice.

The collection of documents was conducted as an ongoing process throughout the data collection phase. At the beginning of the study, participants were asked to provide copies of their course outlines and weekly schedules prior to the interviews. However, documents related to task instructions, and assessment tools were collected after the second interview. This process was guided by the assessment practices, tasks, and tools that the participants discussed during the first two interviews. Regarding document formats, participants either emailed me digital copies or brought printed copies to the interviews.

3.7 Data analysis

Data analysis was conducted iteratively and was informed by grounded theory principles, using thematic analysis as the primary analytical approach. The analysis focused first on understanding ESL teachers' classroom assessment practices and the procedures they followed to ensure quality assessment. It also examined the processes underlying their assessment decisions and the relationships among their assessment literacy, assessment conceptions, classroom practices, and contextual factors. In addition, the analysis explored how teachers' assessment literacy and conceptions developed through their classroom experiences. The research questions and conceptual framework provided initial areas of focus, while the analysis remained open to new codes, themes, categories, and relationships that emerged from the data. Constant comparison was used throughout the analysis to compare emerging codes and categories within and across participants and, where relevant, across data sources. This iterative process allowed emerging interpretations to be continually refined as the analysis progressed.

For the TAL questionnaire, descriptive statistics were used to summarize teachers' self-reported knowledge across the different sections of the questionnaire. General patterns in teachers' reported knowledge were identified, including areas in which teachers reported greater knowledge and areas in which they perceived a need for further knowledge or training. Responses were also compared across participants to identify differences in reported knowledge. For RQ3, questionnaire responses and qualitative data were analyzed as complementary sources of evidence. The questionnaire data were used to examine patterns in teachers' self-reported assessment knowledge across different areas and its alignment with their reported assessment practices. The interview, stimulated-recall, observation and document analysis data were analyzed to examine how teachers' knowledge and experiences informed their classroom assessment practices and how teachers described the development of their assessment literacy through practice. The findings from these

sources were considered together in interpreting the relationship between teachers' assessment literacy and their assessment practices.

The interviews and stimulated recalls comprised approximately 50 hours of audio recordings. All interviews and stimulated recalls were transcribed and summarized. During the initial reading and summarizing of the transcripts, I noted ideas that could potentially serve as codes or themes in relation to the study's research questions. These preliminary codes provided an initial basis for further analysis. The interview and stimulated recall data were then analyzed thematically. The research questions and the conceptual framework served as initial guides for the analysis, while inductive analysis allowed new codes, themes, and categories to emerge from participants' accounts. Codes were continually compared and refined as the analysis progressed. Codes that were relevant to the research questions and conceptually related to one another were examined and developed into themes and subthemes. These themes and subthemes were then organized into broader categories based on their conceptual relationships and relevance to the research questions. Relationships among the emerging categories were examined and revisited throughout the analysis.

I used NVivo version 12 to organize, code, retrieve, and compare the qualitative data and to document the development of codes, themes, and categories. All interviews and documents were imported into NVivo and organized by participant. In the nodes section, separate folders were created for each research question or objective, with parent and child nodes developed and refined as the analysis progressed. Approximately 560 nodes and sub-nodes were created during the analysis. These nodes were used to organize the preliminary codes and emerging patterns in NVivo. NVivo facilitated the organization of a large volume of data and allowed me to examine and revisit coded segments within individual participants as well as across participants. It also enabled me to compare coded segments as themes and categories were refined. The software was used as an organizational and analytical support; the identification, interpretation, and refinement of themes and categories remained an interpretive process undertaken by the researcher.

Following the initial coding and development of themes and subthemes, the relationships among emerging themes and categories were further examined. Codes were continually compared within and across participants to identify patterns, variations, and conceptual relationships. Themes and subthemes were refined through repeated comparison and were organized into broader categories based on their conceptual relationships and relevance to the research questions. The emerging categories were then examined in relation to one another, the research questions, and the conceptual framework. Through this iterative process, relationships among categories were identified, contributing to a broader conceptual understanding of teachers' assessment practices, assessment decision-making, and the development and enactment of their assessment literacy and conceptions. This process also contributed to the development of the classroom assessment procedure model identified in relation to RQ1 and the broader model of the development and enactment of teachers' assessment literacy. The final categories and themes associated with each research question are reported in Chapters 4–7.

As for the documents, both thematic analysis of participants' discussions of the documents and closer analysis of the assessment documents themselves were conducted. During the third interview, participants discussed the documents they had shared, including their assessment goals, criteria, perceived effectiveness, and rationales for using particular materials. These interview responses were coded in NVivo following the procedures described above. The documents themselves were then examined for the criteria and descriptors used by teachers and their weighting, guidelines and instructions, objectives, expectations and prompts, grading systems, and feedback or comments provided to students. The documents were examined in relation to participants' accounts to explore how and where their stated assessment procedures, criteria, assessment conceptions, theoretical knowledge, and experiences were reflected in their assessment materials and practices. Because participants frequently referred to these documents during the interviews, the documents also provided concrete evidence that helped contextualize and elaborate their accounts.

Classroom observations were analyzed to identify teachers' formal and informal assessment practices, the materials they used, their interactions with students, the feedback they provided, and other assessment-related events that could inform subsequent stimulated recalls. Because the classes were video recorded, I was able to revisit classroom events multiple times and examine particular assessment practices in greater detail. Observation data were considered alongside relevant interview and stimulated-recall accounts to provide concrete examples of how teachers' stated assessment procedures and understandings were enacted in classroom contexts. These comparisons were used in a complementary manner rather than as a systematic point-by-point verification of all data sources.

Evidence from the different data sources was integrated where relevant to develop a more comprehensive understanding of teachers' assessment literacy, conceptions, and classroom assessment practices. The sources were used to corroborate, complement, or contextualize interpretations of teachers' reported knowledge, conceptions, and practices. They were not treated as independent measures of the same phenomenon or compared systematically on a point-by-point basis. Rather, they were considered together where relevant to develop a richer and more contextualized interpretation of teachers' assessment decision-making and the relationships among their assessment literacy, conceptions, practices, and contextual conditions.

To enhance the trustworthiness of the analysis, another researcher in the field reviewed the emerging themes, the alignment of selected data segments with themes and sub-themes, and the relationships between themes and categories. Where disagreements arose regarding the alignment of a data segment with a particular theme or sub-theme, the segment was reconsidered and either uncoded or recoded. The emerging categories and their associated themes were also discussed with the second researcher to refine their organization and alignment with the study's research goals. These discussions provided an additional opportunity to reconsider interpretations and identify alternative ways of organizing the emerging findings.

The final analysis was organized according to the sequence of the research questions. Chapter 4 presents findings related to teachers' classroom assessment practices for oral communication and the procedures they followed to ensure quality assessment. Chapter 5 focuses on teachers' L2 assessment conceptions and their relationship with oral communication assessment practices. Chapter 6 examines the relationship between teachers' assessment literacy and their assessment practices, including the role of different sources of knowledge and assessment experience in shaping and developing teachers' assessment literacy. Chapter 7 addresses the contextual factors influencing the implementation and development of teachers' assessment literacy and conceptions in the assessment of oral communication in the ESL classroom.

CHAPTER 4: TEACHERS' CLASSROOM ASSESSMENT PRACTICES

This chapter presents findings related to the first research question. The first research question explores the oral communication assessment practices ESL teachers undertake in the classroom, the procedures they follow when conducting assessment, and the criteria they use to evaluate and ensure the effectiveness of their assessments. In the following sections, assessment procedures are described first, then the criteria for evaluating the effectiveness of classroom assessment are presented.

4.1 Procedure of classroom assessment

The results of data analysis revealed that the participant teachers took different steps to conduct effective oral communication assessment. Specifically, it seems that the teachers went through the following general stages when conducting classroom assessment of oral communication: a) task selection or design, b) rubric selection or design, c) preparation of students for the assessment, d) evaluation of students 'performance, e) responding to the assessment, and using assessment results. In the following sub-section, first a list of stages with their sub-categories is presented, and then each stage is described with related themes and examples from the data.

Table 4.1

Procedure of Classroom Assessment

- **Task selection/design stage**
  - Using tasks aligned with course objectives
  - Using tasks aligned with students 'needs
    - Seeking students 'ideas
    - Using meaningful tasks
  - Sequencing tasks in terms of difficulty and complexity
  - Selecting tasks aligned with target language use situation

- **Assessment rubric selection stage**
    - Considering task objectives when designing/selecting rubrics
        - Tweaking past rubrics to align with task objectives
    - Including important criteria in the rubric
        - Using comprehensive rubrics
        - Giving different weight to different criteria
    - Using clear and/ objective rubrics
    - Selecting rubrics aligned with the construct of interest

- **Preparation stage**
    - Preparing students for task performance
        - Providing background information
        - Brainstorming
        - Providing lists of words/expressions
        - Modelling
    - Sharing assessment rubrics before assessment
    - Teaching rubrics/criteria/descriptors
    - Explaining course/learning expectations
        - Providing clear instruction for completing assessment tasks
        - Explaining the value of tasks/assessments to students
    - Ensuring sufficient time for task performance/preparation

- **Evaluation stage: During the assessment**
    - Diagnosing student needs/problems
        - Understanding the sources of students' challenges and teaching or providing feedback
        - Using scaffolded assessment
    - Using multiple methods of assessment
    - Using and supervising peer assessment

- **Responding to assessment stage**
    - Feedback provision
        - Providing comments to supplement rubrics
        - Providing feedback aligned with task objectives and/ rubric criteria
        - Providing feedback aligned with target language use situation
        - Providing metacognitive feedback about language learning
        - Providing different feedback to different students
        - Conferencing
        - Considering student emotions when providing feedback
    - Using assessment results
        - Sharing rubrics and discussing assessment results
        - Modifying assessment rubrics
        - Modifying assessment tasks
        - Giving students more practice
        - Referring students to support centres
    - Doing trial and error, and reflecting on assessments

Each of these procedures are explained below with specific examples from the data.

4.1.1 Task selection/design stage

Selecting tasks appears to be an important step in the process of CAs. Teachers seemed to attach importance to task selection as a fundamental step in good assessment in the classroom. As outlined in Table 4.1, at this stage teachers considered course objectives and content, tried to sequence tasks in terms of difficulty and connect them so that they build on each other. They chose interesting topics and ensured that tasks are aligned with student interest, context and target language use.

Using tasks aligned with course objectives

Most of the participants (n= 6) in this study considered course objectives an important criterion for the selection of appropriate tasks and a requirement for effective assessment. Aly reported that since she has taken courses in different universities herself and has taught different ESL courses, she is well familiar with the course objectives, therefore she knows how to select tasks that reflect course objectives. Likewise, Arman said that he keeps the course objectives "*in the back of his mind*" when using tasks and assessments and when helping his students to prepare for it.

Dave explained that he looks at the course outline and follows it in his use of tasks, assessments and feedback provision, noting that *"I consider first of all the criteria as set out in the course outline [ ] students should be able to develop some aspect of communication like how they can speak orally to a group of audience, use multimodal aspects such as preparing a PowerPoint slide"* (Dave).

Similarly, Mike reported that presentations in his class represent the goals of the course, which is more content based. He added that at the same time he should pay attention to language elements, so he tries to create a balance between the two in his assessment tasks. Donna also

pointed out that the course she was teaching at the time of data collection was more content based and the focus was on communication and content such as using body language and understanding content, so she was using a presentation task which addressed these criteria. She noted that *"I'm not like looking at ratings for IELTS [ ] the bands are like fluency and coherence, lexical resources, grammar and pronunciation. I'm not looking at any of those during a presentation, but I'm looking at eye contact, are they loud enough? Were they engaging the content"* (Donna).

Using tasks aligned with student needs

According to most participants (n=7) in this study, selecting tasks aligned with student needs is an important step in the process of effective classroom assessment. This alignment involves considering task difficulty relative to students ʾproficiency levels, incorporating student interests, and drawing on various strategies such as seeking students ʾopinions and using meaningful tasks.

Several participants emphasized the importance of adjusting task difficulty based on students' proficiency levels. For example, Dave explained that he assigns simpler questions for lower levels and more challenging ones for higher levels, such as starting with a short, easy two-minute listening task for students new to note-taking. He added that identifying students' zone of proximal development is important, stating, "*if a task is beyond the learning ability of students, it might be counterproductive*." Similarly, Arman, Ryan, and Aly noted that when they find a task to be too complex, they take responsibility by modifying the task or increasing their support to students. Arman stated that he designs different tasks with the same outcomes when his students are not yet ready for student-led discussions. Ray referred to Krashen's i+1 hypothesis and explained that she modifies tasks based on the class's overall proficiency without sacrificing standards. Sally also noted that she considers task objectives and student levels while seeking students' input through surveys before designing different sets of tasks.

A strong consensus also emerged among participants (n=6) on the importance of aligning task selection with students ’interests to increase engagement, motivation, and performance. Several teachers described tailoring tasks based on student feedback and observation. Arman shared that he asks students about their interests and adapts his materials accordingly. For instance, when his students expressed a preference for “*real English like straight talk type of thing*” rather than traditional lectures, he responded by selecting more authentic materials, such as a YouTube video on common Canadian English expressions. Dave and Ray both emphasized that student interests directly affect performance and engagement. Dave explained that he focuses on positive, culturally aware topics that help students learn about each other's backgrounds, while Ray noted that when students express enjoyment in their listening logs, it reinforces the topic's relevance. Similarly, Donna uses her observations of student engagement to guide her task choices, noting that interactive documentaries about challenges new immigrants face are low-stakes, simple, and interesting to students. Mike added that when students are not interested, tasks are less likely to generate any “positive washback.”

Seeking students ideas

One strategy commonly used by participants (n=6) to select tasks aligned with student needs is seeking students' opinions. Several teachers reported using various methods to elicit student input. For example, Ryan conducts stimulated recalls with his students to ask whether the speaker was too fast or if there were too many new words in listening material. He explained that seeking students' ideas gives him a better understanding of task appropriateness and provides students with more agency. Similarly, Sally and Arman both use surveys or feedback forms to gather student opinions. Sally designs tasks based on students' interests and learner types, offering choices so students can opt in or out of certain tasks. As she explained, “*I say: look, we have these two different tasks… Those who like it can do it. And those who are not, they can leave it out.*” Arman gives his students a column to indicate what needs to be changed, continued, or added in his assessment tasks.

Likewise, Ray provides students with a form to share their opinions on useful activities and their needs, and if she notices a pattern shared by many students, she considers it in her teaching and assessment.

Using meaningful tasks

Another strategy used by participants (n=5) to select tasks aligned with student needs is using meaningful tasks. Teachers reported connecting tasks to students' present everyday needs, future workplace needs, and authentic communication. For instance, Dave uses tasks addressing students' everyday needs and interests, such as discussing cultural beliefs, traditions, and celebrations, to help engage them. Mike pointed out that he considers students' future workplace and academic needs when selecting assessment tasks, believing that meaningful tasks increase engagement and interaction. As Mike stated, "*I start from not only present but future needs of students, which means that I always think of their language use in workplace …and academic courses.*" Additionally, Ray and Ryan emphasized the value of authentic and meaningful tasks. Ray provides students with background information to help them connect better and participate more, while Ryan underlined the importance of including an information gap in assessment tasks to create opportunities for students to listen to each other and engage in meaningful speaking practice.

Sequencing tasks in terms of difficulty and complexity

Most participants (n=7) emphasized the importance of sequencing tasks in terms of difficulty or complexity to avoid overwhelming students and to facilitate gradual learning. Many also highlighted the value of selecting tasks that build on each other in a logical progression, allowing students to develop skills step by step.

Participants commonly described adjusting task difficulty based on students 'proficiency levels and using a progressive structure, starting with lower-order and more controlled tasks before moving to higher-order and more cognitively demanding tasks such as discussions or debates. The

rationale behind this approach was to prevent student frustration and support gradual skill development. As Ray explained, grounding her approach in theory: *It s progressive from recognition, and builds up to more cognitively demanding tasks."* She illustrated this by describing how she scaffolds oral summary tasks, first having students practice note-taking and summary writing before expecting them to produce the summary verbally. In this respect Aly described the importance of building foundational skills before advancing to critical thinking, noting that repetition serves as a necessary step in the language learning continuum, explaining that there is a moment where students need to repeat what they hear as *a step forward in forming opinions. And from there we can start discussing about critical thinking."*

Some participants also noted practical limitations. Ryan pointed out that curriculum constraints can restrict the ability to adjust the difficulty of speaking tasks, though adaptations remain possible for listening activities, explaining that *"For listening, I change the type of activity. I try to go for listening that are easier, but in terms of speaking 'No' because you can't."*

Using tasks aligned with target language use situation

Most participants (n = 9) reported that in task selection it is important to consider the target language use situation. Participants emphasized the value of familiarizing students with tasks they encounter in other courses, as well as preparing them for real-life and workplace contexts.

A common rationale was that tasks should reflect how language is used authentically outside the classroom. Several participants therefore favored role-play activities, which simulate real-world interactions such as shopping or opening a bank account. Dave, for example, asks students to do role-play activities where one person is a bank clerk and the other is a client, practicing *what sort of questions are you going to ask?"* Dave criticized the impromptu speaking tasks mandated by his school, noting that they did not represent real-life situations. Ryan also endorsed role-play and presentation tasks on the grounds that students need to speak English in different real-world

situations, emphasizing that *we focus on different functions like persuading, proposing solutions, expressing agreement, and arguing–skills they'll need in real-world situations."*

Others highlighted building skills for future academic and professional contexts. Aly explicitly connected classroom tasks to long-term skill development, telling her students that *this is a skill you need to have in any job like discussing and summarizing something and posing questions."* In addition, Ray described operationalizing authenticity through assessment, noting that she adds sub-scales in her rubric to reflect real-life language use, such as indirect speech acts, and practices note-taking because students will need this skill in meetings.

4.1.2 Assessment rubric selection stage

According to the participant teachers, after selecting tasks for assessing oral communication, they selected assessment rubrics, an important step in the process of effective classroom assessment. In designing or selecting rubrics the teachers considered task objectives. Also, in line with the task and course objectives, they selected/designed clear and objective rubrics, included all relevant and important criteria and gave different weight to different criteria depending on the importance of the criteria for successful task completion. Also, they tried to select or design rubrics that address the construct being assessed.

Considering task objectives when designing/selecting rubrics

The majority of participants (n = 7) reported that they consider rubric alignment with task objectives an important factor for effective classroom assessment. They described starting assessment design with careful consideration of course and task objectives, viewing this as a fundamental step in quality assessment. Participants reported that they design rubrics in line with task objectives, use relevant criteria important for achieving both task and course objectives, and then teach these criteria to students, share them, assign grades, provide feedback, and conduct

follow-up activities accordingly. Attention to task objectives was described as a lighthouse directing teachers to stay on track, directing assessment practices and ensuring effectiveness.

Several participants emphasized the importance of congruence between assessment and course objectives. Ray explained that she keeps the course outline in mind when designing rubrics– for example, if the course objective is producing language, she puts more weight on productive skills in her tests and rubrics: *"In designing a rubric, I consider the target outcome, so matching the target outcome and the rubric and the task itself."* Ryan stressed that if the goal of the course is integrated speaking, then both the task and the rubric should represent this goal to ensure construct validity. He illustrated this by explaining that his rubric for assessing speaking includes cultural appropriateness as a criterion, which reflects both course objectives and real-world speaking demands. Ryan's attention to this factor is evident in the assessment rubric he used for assessing students' overall progress (see *Figure* 4.1). In this rubric, cultural appropriateness is included as one of the assessment criteria: *"One of the linguistic abilities is to be able to persuade and argue and to show disagreement or agreement in an appropriate manner, which is in our rubric. Also, being familiar with Canadian culture and how meetings are conducted."*

Figure 4.1

Ryan's Progress Rubric (Ryan- DOC≠ 3)

Progress Report

- **Initial assessment (Baseline).** **Date:** .....
- **Mid-point progress** **Date:** .....
- **End-point assessment** **Date:** .....

Level of Performance: ● Indicates starting and end points ←indicates progress

Performance Criteria	1	+	++	2	+	++	3	+	++	4
Speaking Accuracy Ability to speak using correct grammar and word choice.										
Speaking Fluency Ability to maintain and direct a coherent flow of conversation.										
Pronunciation Ability to speak clearly and comprehensibly.										
Writing Accuracy Ability to write using correct grammar, discourse style and word choice.										
Writing Fluency Ability to express ideas coherently and logically.										
Listening Comprehension Ability to understand main ideas and Details.										
Reading Comprehension Ability to understand main ideas and relevant details.										
Socio-cultural Appropriateness Ability to use socially acceptable nonverbal and verbal communication according to the situation.										

Observations/Recommendations:

Other participants highlighted how rubrics operationalize task objectives through specific criteria. Sally provided a concrete example, noting that for a narrative task, the rubric should include a criterion related to the use of past tense because that is essential for task completion: *"If you wanted them to use past tenses in the oral narrative task, then this has to be part of it."* She also noted that if there is a criterion in a rubric that is irrelevant to the task and she does not use it in her assessment, she omits it. Mike described how rubrics serve as a guide for both instructor and students, using a presentation rubric that reflects task objectives. For example, he referred to an evaluation sheet (see *Figure* 4.2) he was using for evaluating students' presentations and explained

that in the presentations, he assesses content, how much research students did, organization of slides, and accuracy, which are all task objectives: *"The presentation rubric gives a scheme on what are the most important elements of student speaking competency—a structure for the presentation and a guide that these are the most important elements."*

Figure 4. 2

Mike's Presentation Evaluation Sheet (Mike-Doc 2)

CONTENT	%	PRESENTATION	%
Introduction to presentation.	1	PowerPoint has a readable design.	1
Clear purpose of the research study.	2	PowerPoint has an appropriate amount of text on the slides.	1
Description of the method the researchers used.	2	Presenter "talks" the material and seem to know it well.	2
Description of the results that the researchers found.	2	Presenter speaks in way that is easy to understand in pronunciation and intonation.	1
Conclusion to the topic.	2	Presenter is linguistically accurate in his/her speech.	1

Tweaking past rubrics to align with task objectives

Most participants (n = 7) reported that they modify and tweak rubrics in accord with students' needs and task objectives. Arman highlighted the importance of consistency between rubric criteria and students' needs, explaining that *"I constantly modify rubrics, adjusting them based on students' levels and performance expectations."* Dave reported that when he borrows a rubric from other sources, he adapts it to fit task objectives and student needs, noting that he rarely uses them as-is and that his rubrics are often "*about 60% from existing sources and 40% customized."* He gave the example of adding a teamwork and critical understanding criteria to a school-mandated rubric that lacked them, even though collaboration was essential for the task. *"For example, in a seminar presentation, I included criteria for both teamwork and critical assessment of their readings."* (Dave).

Ray, Sally, Donna, Aly, and Hudson also indicated that they modify rubrics mandated by their institutions or adopted from other sources when the rubrics do not fully reflect course content or task objectives. Ray explained that she might add details or sub-criteria to make a rubric more specific and aligned with her expectations for organization, content, or language. Aly noted that she had tweaked a rubric for student presentations by modifying several criteria and removing one that was irrelevant to the task objectives: *"Somebody gave the rubric for presentation to me… I removed one criterion."*

Including important criteria in rubrics

In line with the goal of teaching and assessing task objectives, a number of participants (n = 6) emphasized the importance of including important criteria in rubrics to guide and engage students and enhance their task performance.

Participants noted that prioritizing criteria based on task goals helps focus both instruction and assessment. For instance, Ryan reported that in a final presentation task, communicating information is a fundamental element, so he places the main weight of his assessment on students' ability to get the meaning across, with appropriateness, accuracy, and pronunciation at the bottom of the hierarchy. When asked about the effectiveness of a checklist he was using to assess his students' progress in speaking (see *Figure* 4.1 above), Ryan stated that since *"it touches upon the most important and essential aspects of assessing speaking skill."*, it is a good rubric:

Others highlighted how including important criteria in rubrics can engage students and make task expectations clear. Dave noted that this strategy is particularly effective for passive students, as it encourages them to pay attention to those criteria and to incorporate them into their presentations, explaining that *building participation into the rubric encourages passive students to engage, especially in seminar presentations."* Mike also described reviewing criteria from different rubrics and selecting the most important ones aligned with his course objectives, noting that these criteria

serve as *a guide for me and the students that these are the most important elements in their presentation."*

Using comprehensive rubrics

Using comprehensive rubrics is an important step in the process of effective classroom assessment according to half the participants (n = 5). They reported that they do their best to use rubrics and criteria which fully represent the target skill.

These participants emphasized that comprehensive rubrics allow teachers to evaluate different aspects of performance required for successful task completion. Ryan shared a rubric he was using for assessing student presentations, which included criteria such as task effectiveness (information transfer), accuracy, fluency, pronunciation, answering questions from the audience, and socio-cultural appropriateness (see Appendix J: Final Assessment Form- Ryan- Doc 1): *"I prefer the rubric that we have, which focuses not only on the linguistic ability, but also effectiveness and cultural awareness and so on."* Sally similarly noted that a good rubric for assessing speaking should include all categories and aspects of the speaking task, such as fluency and accuracy. She added that when school-mandated rubrics are short and not comprehensive, she supplements them with additional comments.

Others highlighted the value of detailed, analytic rubrics for differentiating student performance and providing focused feedback. Ray reported that when she adopts a rubric or uses one assigned by her school, she tries to expand it by adding sub-scales so that it addresses all aspects of the task or skill being assessed. She noted that simple or holistic rubrics do not differentiate between more able and less able students or capture nuanced differences. Regarding the effectiveness of the rubric she was using for oral summary (see *Figure* 4.3), Ray explained that she finds the rubric useful because it has sub-skills and descriptors that elaborate on each criterion

instead of listing only general criteria, noting that *when it's broken down into specific items or skills, it makes more sense. It also makes your comments and feedback more to the point."*

Figure 4.3

Ray's *Oral Summary Rubric (Ray-Doc≠ 1)*

Speaking Assessment – Oral Summary of Presentation

NAME: _____ MARK: _____/60(=_____/3%)

Criteria	Marks	Comments
Organization The summary was well organized and prepared with the audience in mind. There was a clear introduction, body and conclusion.	/10	
Content The summary included all the essential points of the presentation. It did not include any unnecessary details. There were transition words to help coherence.	/10	
Fluency The summary was delivered smoothly with little hesitation. Transition words or other connecting words aided fluent delivery.	/10	
Pronunciation Key words were pronounced correctly. Pronunciation was clear.	/10	
Vocabulary Key words were used appropriately. A wide range of vocabulary was used.	/10	
Grammar Appropriate grammar was used. Grammar used shows understanding of grammar points covered in class e.g. reported speech, past tenses etc.	/10	
TOTAL	/60	

Giving different weight to different criteria

Closely related to including important criteria in rubrics and using comprehensive rubrics, some participants (n = 5) referred to assigning different weights to different criteria as an important step toward effective classroom assessment.

The participants explained that weighting decisions are guided by task objectives and student needs. Aly described adjusting criteria in her rubrics and allotting different marks based on what deserves more focus at a given proficiency level. Referring to a rubric she used for assessing a presentation task (see *Figure* 4.8), she noted that pronunciation may not be a focus at certain levels, so content may carry more weight, explaining that *how much points are allotted to content versus pronunciation… I would say the rubric is rather effective."* Ray pointed out that she assigns

different weights to different criteria. In the rubric for a group discussion task (see *Figure* 4.5), she explained that criteria specific to the task, such as use of functional language and participation, carry more weight than general criteria like fluency and grammar, noting that *functional language skills like interrupting, turn-taking, and expressing opinions are key to discussions."*

Others emphasized aligning weights with course priorities. For example, Sally noted that in courses where the focus is on communication, rubrics should prioritize communication skills over grammar unless grammar affects meaning, explaining that *I don't think grammar has to be that much important in a rubric. If it is in the rubric, then it has to be a small portion unless whatever they're saying changes the meaning."* She added that if a criterion is irrelevant and she does not use it in her assessment, she omits it and redistributes its weight to other criteria.

Using clear and/ objective rubrics

Most participants in this study (n = 8) mentioned the importance of using clear and objective rubrics in their classroom practices for ensuring effective assessment. Several participants argued that objective rubrics benefit both teachers and students. For example, Dave explained that using objective rubrics enables teachers to notice students' needs and makes students aware of areas that need improvement, noting that rubrics should be simple and clear, stating that *objective rubrics help both teachers and students by focusing on specific performance aspects. They clarify students' weaknesses and allow teachers to provide more effective guidance."* Similarly, Ray added that using an objective and detailed rubric allows her to better differentiate students in terms of quality of performance, which clarifies differences in students' marks. She contrasted analytic rubrics with holistic ones, explaining that with analytic rubrics she can highlight criteria and dimensions that match students' performance so they can see where they stand, noting that *the good thing about the rubric (see Figure 4.5) is that it shows the variation of the descriptor. So I kind of highlight whatever matches their performance and you can see where you're standing."*

Many others emphasized that clear rubrics act as a roadmap for students, making instructional goals and expectations transparent. For instance, Arman noted that using rubrics helps students understand their strengths and areas for improvement, explaining that *it acts as a roadmap, helping them clearly understand their strengths and areas for improvement for future tasks."* Sally, however, criticized holistic rubrics, arguing that they do not include all the criteria for assessing speaking effectively and that teachers may need to modify rubrics based on their class.

Additionally, some participants noted that using objective assessment tools enables teachers to explain and justify marks to students. Donna, for example, explained that clear parameters in a rubric help students understand the basis of their evaluation and highlight strengths and areas for improvement, stating that *having some clear parameters in a rubric is important, so students can have an idea of the parameters by which they're being marked is important and fair. I think good assessment would highlight the strengths, weaknesses, places they can work on."*

Selecting rubrics aligned with the construct of interest

Using assessments that reflect the construct of interest appears to be an important step toward ensuring the effectiveness of classroom assessments according to some participants (n = 6) in this study.

These participants emphasized the importance of construct validity, defining good assessments as those that measure what they intend to measure. For example, Hudson noted that assessment should target the construct of interest and that he gears his feedback accordingly, explaining that *"I think it measures what it's supposed to measure. They prepare for it throughout the semester."* Ryan stressed that assessment methods should conform to teaching methods and the related construct being assessed–for instance, if speaking is based on communicative language teaching in class, assessment should reflect this. He explained that he relies on trial and error to

evaluate assessments, modifying them if they do not tap into the intended construct, adding that *"assessments must be valid and align with the content taught."*

Others highlighted the importance of ensuring alignment between rubrics and the construct being assessed. Ray, for instance, emphasized that a good assessment measures what it is supposed to measure, noting that a listening test should not be marked down for grammar unless grammar is part of the assessment criteria: *"It's a listening test, and I marked down for grammar, then there's something wrong with my assessment, unless it's part of the criteria."* She added that she modifies past rubrics and adapts them in relation to the construct of interest, testing them over a term and refining them as needed. Additionally, Sally noted that when she adopts a rubric for speaking assessment, she removes any criterion irrelevant to the speaking construct, explaining that *"when you are assessing somebody's speaking, there is a part in the rubric that is not even applicable. It should not be there, so you remove that part."*

4.1.3 Preparation stage

In this stage of CA, teachers prepared students for the assessment task. The participants reported that before conducting the assessment, they prepared students for task performance and explained the learning expectations, shared assessment rubrics, and taught different assessment criteria to their students to prepare them for the assessment. They also gave sufficient time for the assessment so that students could demonstrate their best performance, which allowed teachers to identify and assess students 'needs better.

Preparing students for task performance

As an important step for effective assessment in the classroom, most of the teachers ($n = 8$) reported that they prepare students for assessments. To prepare students for assessment, they provided background information before conducting oral communication tasks, did brainstorming in class, provided lists of key words for speaking/listening, and did modelling in class.

Providing background information

Some participants (n = 7) pointed to the importance of providing background information to students before conducting an assessment to prepare them for the task. Arman, Sally, Donna, Mike, and David, for example, highlighted that this preliminary knowledge helps contextualize the task and enhance performance.

Several participants described specific strategies for activating students' background knowledge. For example, Mike explained that he shows relevant videos and provides a handout with questions to activate students' schemata before a debate, allowing them to share ideas with peers. Sally described preparing students for narrating stories by starting a story herself and prompting students to complete it before sharing their own, explaining that *"I started with a story… I stopped here, and I was like, 'Okay, you guys, go ahead. What do you think happened next.'"* Arman viewed providing background information as a form of scaffolding, noting that for struggling students during classroom discussions he would give them clues or background information to encourage them to express their ideas, explaining that *"the main idea was to give him something to speak about. And that's why I kind of gave him a little bit of background what the report was about to prompt him to speak."*

Others emphasized the role of background information in contextualization and engagement. During one observed session, Donna gave her students background information before playing a video, explaining in a follow-up stimulated recall that this helps students contextualize the data, noting that *"I think it's just something I would do in any class to just kind of contextualize the data… In this case, it corresponds directly with the kind of information they're reading in the chapter as well."* Ray highlighted that providing background information helps engage students and elicit more answers, stating that *"in terms of delivery, I focus on activating background knowledge*

and making content authentic and meaningful. When students can make connections, they engage more and provide better answers."

Brainstorming

Some participants (n= 5) mentioned that they use brainstorming in their classes to prepare students for assessment. Dave, for example, reported that he prepares his students for successful task performance by allowing them to collaborate with peers to brainstorm the meaning of the words they hear and to find the key ideas so that they can present a better summary.

Arman also reported that he does some brainstorming before asking his students to watch, read, and discuss a topic. He said that it gives a chance to all students, particularly those who are silent and invest less in speaking, explaining that *not everyone is so willing to engage, they just shy away from speaking in front of the whole class… brainstorming gives those students a chance to express their ideas."* On a similar note, Mike mentioned that before debating tasks, he plays videos about the topic and asks students to brainstorm and discuss in pairs. He believed that brainstorming gives everyone a chance to practice speaking, noting that *students listen to others and see how the topic is discussed from different perspectives. It also gives them a chance to practice talking about real events or genuine purposes."*

Providing lists of words/expressions

In order to prepare students for assessment, some participants (n = 7) reported that they give students a list of words and expressions. For example, Aly reported that she has zero assumptions about students' ability and proficiency. As such, she tries to explain everything to prepare students for task performance. Aly said that she provides students with a list of words from Bloom's Taxonomy that they can use to formulate and ask questions during presentations and discussions (see Appendix K), explaining that she had to go through the academic verbs used for questions at

every level of Bloom's model. She further explained that if she plays a video before a discussion, she writes challenging words on the board and discusses them with the students, noting that *I'll provide a list of these words in advance, explain their meanings, and highlight them before the video."*

In similar ways, Sally, Mike, Arman, and Ryan pointed out that they teach a list of key words to their students to prepare them for debating or discussion tasks. For example, Sally described how she gives her students a list of new words before asking them to listen to a lecture and discuss it afterwards, such as providing the word *obesity"* because students might not know its meaning. Ray also stressed the importance of discussing useful words before watching a video and talking about it. She noted that this discussion engages students, elicits more answers, and prepares them for successful task performance, explaining that *we'll introduce a documentary on the topic, provide essential vocabulary, and have students watch it in class, answering related questions."*

Modelling

Some participants (n = 6) reported that they use different sample activities as models to prepare their students for formal assessments. Several participants described using practice activities that build toward summative assessments. For example, Hudson emphasized the importance of preparing students for the final assessment task, explaining that during the semester his class practices debating, which *"builds up and scaffolds toward the final graded debate, which they do formally toward the end of the semester."* Dave reported that he prepares his students by conducting mock assessments before the formal assessment to help them perform better, noting that *"we spent like half an hour to do a mock assessment which replicates the real assessment. So, I do my best to prepare them for the assessment to come."* Ray also noted that she uses ad hoc activities during the semester to prepare students for summative tasks, explaining that before a final graded task

involving giving instructions, *"I provide a sample task and have them practice in groups to prepare."*

Others highlighted modelling as a way to familiarize students with task expectations. Sally explained that in a course where the final task is oral storytelling, she models the task and practices narrative activities in class so that students are prepared to perform at the end, noting that *"the end product is storytelling… We practice and do some models during class."* Aly described using the performances of more proficient students as models for the rest of the class, sometimes asking less confident students to present later in the semester so that they can learn from stronger examples, explaining that *"the first group sets the standard and I want a good group to set the standard because otherwise it goes downhill from there."*

Sharing assessment rubrics before assessment

Most participants (n=8) reported sharing assessment rubrics with students in advance to prepare them for assessment and clarify expectations. For example, Dave, Donna, and Ray all emphasized that sharing the rubric before the task helps students understand what is expected and how they will be assessed. In this respect, Donna noted: *"I share the rubric before performance. That's something pedagogically I find helpful because students know, along with the assignment instructions that are quite explicit, then they know exactly how they're being marked"* (Donna). Sally emphasized that sharing the rubric not only informs students of expectations but also builds their confidence and trust in the teacher, explaining that *"before the task, I share the rubric to show students the purpose behind assessment. When they understand this, they trust you more, knowing everything you do has a reason."*

Ally, Hudson, Ryan, and Mike also reported sharing rubrics before assessment and teaching the steps required to complete assignments. However, Ally noted that while she shares rubrics and

instructions, she is sometimes unsure of their effectiveness in improving performance, as some students may lack motivation.

Teaching rubrics/criteria/descriptors

Closely related to sharing assessment rubrics, some participants (n = 6) reported that they explicitly teach the criteria and descriptors in rubrics to prepare students for assessment. Participants emphasized that teaching rubric criteria and descriptors–beyond simply sharing the rubric–helps students understand expectations, align their learning with assessment goals, and develop practical strategies to succeed.

Several participants highlighted the importance of explicit instruction to ensure students understand assessment expectations. For example, Aly explained that she assumes nothing about students' prior knowledge and thoroughly explains both assignment instructions and the criteria in her rubrics, noting that *"I think the best approach I found in this class is have zero assumptions, and not assuming they understand anything, so I will be explaining it."* Sally emphasized that rubrics are often written for teachers rather than students, so explicit instruction is essential, explaining that *"Rubrics can be difficult for students to understand… I make sure to teach the criteria and descriptors so they know how to achieve a higher mark."*

Others described teaching specific rubric criteria to prepare students for both communication and assessment. For instance, Ryan explained that he teaches the concept of "cultural appropriateness" from his rubric (see *Figure* 4.1 above) by showing students how to express themselves in culturally appropriate ways, noting that they learn to avoid direct confrontational phrases like *"let me finish"* and instead use more polite expressions such as *"could you please hold on for a sec."* Ray also reported that she explicitly teaches assessment criteria and descriptors, though she noted that full comprehension often only occurs during task performance, explaining

that *"I really explain it in class, what this means, what this descriptor is. But based on my experience, students don't understand it and its descriptors fully until they actually work on a task."*

Explaining course/learning expectations

Some participants (n=6) underlined the importance of explaining learning expectations as an important step toward effective assessment in the classroom. Generally, they thought that without ' understanding the course and learning expectations, students would not be able to perform well in the assessment. As such, they ensure that students get clear instructions regarding the learning expectations.

Aly highlighted the importance of providing detailed instructions on course objectives, assignments, and performance expectations. She reported posting clear instructions on the class platform, reviewing them with students, and clarifying the steps required for assessment, explaining that she posted the assignment in simple language, reviewed it with them, and explained the due date, the rubric, and the necessary steps. She also gave specific guidance on task performance, including *do" and don't"* instructions for presentations, such as not reading from notes, noting that *I'm very specific about what I want… I have a long list of don't do and do instructions."*

Sally likened unclear expectations to "*diving into unknown waters*," emphasizing that clarifying learning outcomes at the start of the semester promotes responsibility, engagement, and focused learning, explaining that "*When students know the end goal, they learn more responsibly, listen more carefully, and get more involved, as they see there's a clear outcome to reach*" (Sally).

Hudson and Ray believed that providing clear instructional goals to the students is of utmost importance because students perform better when they understand what is expected. Hudson said that in order to make his students familiar with the learning objectives, he gives clear instruction about what they are expected to do in each task.

Providing clear instruction for completing assessment tasks

In addition to explaining course and learning expectations as described above, some participants (n = 7) mentioned providing clear instructions for completing assessment tasks as an important step for effective classroom assessment. They noted that providing clear, structured, and sometimes scaffolded instructions, both written and verbal, helps students understand task expectations, reduces confusion, and supports effective assessment.

Several participants highlighted the need to address gaps in students' understanding. For example, Donna noted that ESL students' unfamiliarity with certain task types can create confusion, explaining that *"there's often a gap between students' understanding and the instructions, especially since they're not used to different types of assignments. Even with group presentations, some are confused, so I have to clarify the instructions again."* Ryan echoed the value of repeating instructions and providing written guidance for complex tasks, noting that *"after giving instructions, I often see students confused, so I have to repeat them multiple times. It's important to write the instructions down instead of just saying them."*

Others emphasized scaffolding and providing structured support. Mike described offering step-by-step guidance, stating that *"in addition to the course outline, I provide an additional and detailed description of each assignment. They are scaffolded on how to move."* In this respect, Ray noted that she provides templates or guiding questions to help students organize their speaking tasks, explaining that for a mini lecture task on a famous person, *"I provide a few questions or a structure to guide them on what information to gather."* Sally ensured that each task had a one-page instruction sheet on the class platform and repeated key points as needed, noting that *"I think explanation about the task and assignment adds to good assessment… it's easy for students to get lost."* Hudson pointed out that students may struggle with technical terms, such as those used in debating tasks, and explained them alongside task instructions. In the document he shared for a

debating task (see Appendix L), Hudson clearly illustrated the goal of a debate, what students are expected to learn, and how the task should be completed. In the evaluation sheet he was using for this purpose (see *Figure* 4.4 above), he listed three rounds for each debate: claim and proposition stated and repeated, response to the opposite proposition with a clear claim, and closing arguments.

Explaining the value of tasks/assessments to students

Another theme closely related to explaining course and learning expectations is explaining the value of tasks or assessments to students, which was mentioned by a number of participants (n = 6). These participants emphasized that explaining the rationale behind tasks motivates students and increases their engagement. For example, Dave reported that he tries to justify each activity he asks students to do, noting that *"I usually try at various points in my classes to highlight and to justify each and every activity that I ask my students to do."* He illustrated this by explaining that he clarifies the purpose of note taking before a discussion, showing students how it prepares them for university or higher education. Similarly, Sally noted that throughout her teaching career she explains the rationale behind every assessment or task, adding that if a task lacks a solid rationale, she will skip or remove it. She explained that explaining the value of tasks increases students' trust in the teacher and conveys that tasks are not haphazard, stating that *"you talk about the reason behind it. And if you let them know, what is the purpose behind everything, for sure it's going to go more smoothly than we expect."*

Others highlighted using explanations to address student resistance and encourage participation. Ryan noted that before administering a task, he always explains its value and the reason for doing it. For instance, when students dismissed role plays as unimportant, he spent time discussing research on their value to persuade students to take them seriously. He also addressed students' perceptions that presentations are not authentic, explaining that *"presentations tap a wider*

range of skills than, for example, a job interview or role play," including nonverbal communication, body language, and interaction.

Some participants also referred to students' obsession with grades as a barrier to engagement. Aly noted that when tasks are not graded, students often do not pay attention, so she highlights the value of tasks for learning purposes rather than grades, explaining that *"if a task isn't graded, students often don't pay attention. I remind them that at university, they chose to be here, so they should put their best effort into every task, graded or not."*

Ensuring sufficient time for task performance/preparation

Some participants (n = 5) referred to providing enough time for preparation and task completion as a step toward achieving good assessment in the classroom. They mentioned giving opportunities for resubmission, repeating a task, doing more modelling, and listening to video content more than once as ways to give students more time to prepare for successful task completion.

Several participants described strategies for accommodating students who need additional preparation time. For example, Ally explained that she identifies students who are new to the program or unfamiliar with presentations and ensures they present later in the semester so they can learn from their peers' performances first. In this regard, Dave reported that for students' short talks, he provides ample time for them to study, do research, and prepare, explaining that *"I would say 'you have one week to prepare for this. Go and prepare, do some research, educate yourself, read some sources, and come and talk for five minutes about that topic.'"* He believed this allows students to understand the content, review different sources, and improve their grammar and vocabulary.

Others highlighted the importance of flexibility and multiple opportunities for successful task completion. Mike noted that he gives students leeway, such as providing more time or

additional chances to prepare and complete tasks successfully. Ray emphasized that giving enough time is important for successful task completion, explaining that when students watch a video before a discussion, she plays it more than once to help them understand the content better. She also noted that if most students do not perform well on a task, she asks them to work on it and resubmit or present again, stating that *"when they don't meet the criteria sometimes, which means they didn't spend enough time on their task; they just want to get a mark. If it's for the whole class, or most of the class, I usually don't mark it. And I would say submit or present again."*

4.1.4 Evaluation stage: During the assessment

The participant teachers attached importance to evaluation as another important step in the process of ensuring effective CA. According to participants 'reports, in the evaluation stage, which takes place during the assessment, they aim to diagnose students 'problems in order to provide them with appropriate assistance, try different methods of assessment, and use peer assessment.

Diagnosing student needs/problems

Most participants (n=9) reported that classroom assessments serve primarily to diagnose students' needs and address problems immediately through targeted teaching, additional practice, or feedback. This diagnostic work occurs during the assessment itself, as teachers observe performance, note errors, and adapt instruction on the spot.

For Dave, diagnosis is "*the backbone*" of his classroom activities. While students perform tasks such as group seminars, he observes and takes notes, then uses the remaining time to address issues: "*If we have like 10 minutes left, I stop the class and say [ ] these are the problems, and I just try to focus on that*." His notes also inform follow-up lessons tailored to that cohort's specific needs. Hudson linked this process to Vygotsky's Zone of Proximal Development (ZPD), emphasizing the accurate identification of students' "*entry level performance*" and effective mediation to close learning gaps.

Several participants described adapting tasks mid-assessment when they proved too difficult. Aly, for example, shifted from critical thinking to opinion formation after observing her students' struggle: "*I noticed critical thinking is way too high, 'so how about we discuss how you form an opinion'.*" Arman also modified content in real time if comprehension was low, such as replacing a video or adding more explanation.

Others focused on diagnosing and addressing linguistic or pragmatic issues observed during assessment. Donna, for instance, intervened during presentations to model culturally appropriate speech acts when students' responses were abrupt or potentially offensive, elaborating that "*someone asked the presenter, 'Can you speak more about this?' and she simply said, 'No, we talked about it.' I reminded them that we can't just say no — instead, respond politely, for example, 'That's a good question.*'" Ryan similarly walked around during speaking activities, noting errors in both language and cultural appropriateness, then offering targeted feedback and additional materials for practice.

Some participants adjusted assessment criteria or instructions based on early performances. In this respect, Mike explained that "*I learn from the first batch [of presentations]. If that is not working, I devise an instruction that helps them understand their weaknesses. I also may create new criteria*." Ray used short diagnostic talks at the start of the semester to gauge speaking skills and critical thinking, later providing whole-class demonstrations if common gaps were identified.

Understanding the sources of students' challenges and teaching or providing feedback

In line with diagnostic assessment, some participants (n = 5) reported that they aim to identify the sources of students' challenges and, accordingly, adapt their teaching, provide appropriate feedback, or address these challenges in their rubrics.

Several participants described relying on their experience to anticipate and address student difficulties. For example, Arman explained that after twenty years of teaching, he can understand the sources of the challenges that students face and take necessary steps to improve his teaching and students' learning. During one observed class, he-taught vocabulary before a video discussion because he predicted students might struggle with the words, noting that *"I introduced the words beforehand because I felt my students might not know them. This helped with comprehension and gave them confidence in understanding what they were listening to."* On a similar note, Hudson reported that because of his familiarity with Asian students, he can predict the areas of grammar they struggle with, such as relative clauses, and teaches accordingly to fill the gap.

Others emphasized the importance of distinguishing between errors and mistakes to provide appropriate feedback. Ryan explained that based on his experience, he can tell errors from mistakes, providing direct feedback when a student lacks knowledge and indirect feedback or clues when it is merely a slip, stating that *"I can usually tell if a mistake is due to a lack of knowledge or just a slip. I give indirect feedback on errors students likely know and direct feedback on those they don't."*

Some participants also identified cultural differences as a source of student difficulties. Aly mentioned that Asian students have a different interpretation of logical sequence from Canadian logical sequence of events, which can be the source of many errors with regard to tense usage. She noted that logical sequence is one of the criteria in the rubric she uses for assessing students' presentations (see *Figure* 4.8), explaining that *"what's logical sequence to you and what's logical sequence to an Asian are very different concepts. You would naturally start with the past and progress towards the present time and then maybe in the future. Well, I have people starting in the present and then jump to the past."*

Using scaffolded assessment

Most participants (n=9) described scaffolded assessment as a central part of their classroom practice, involving real-time diagnosis of student needs during assessment tasks and adapting support accordingly. Common strategies included moving from implicit to explicit feedback, prompting self-correction, giving graduated clues, modifying tasks to match ability levels, and building on prior knowledge.

Several teachers emphasized using indirect prompts first, encouraging learners to notice and correct errors before providing direct answers. Ryan and Sally, for example, both noted that they collect examples of inappropriate language use during assessment tasks, display them on the board, and prompt students to evaluate or correct them collaboratively. This gradual approach, grounded in ZPD principles, fosters active engagement. Ryan explained that he might put a student's sentence on the board and ask *Do you think this is appropriate?"* to help them discover the correct form.

Scaffolding was also described as building on what students already know. Arman, for instance, begins from learners 'existing knowledge and provides carefully timed clues to move them forward. During observation of his class, he noticed a student struggling to talk about a report that was covered the previous week. He paused the task, provided background context, and gave a specific clue to help the student "*open up—something to bite on*." (Arman, SR)

Many teachers adapted their scaffolding to the level of the student or the difficulty of the task. Ray and Aly both stressed maintaining assessment standards while adjusting support. For lower-level learners, they might introduce more peer work, offer guiding questions, or create mini-lessons targeting recurring issues. For example, Aly illustrated this approach with a presentation activity in which students had to ask questions. Noticing that students did not know what questions to ask, she provided a list of possible questions and followed up with a mini-lesson using Bloom's Taxonomy, explaining that *I found myself they didn t know what questions to ask. So I thought*

okay, let me have a mini lesson on what would be an appropriate [way of asking questions] and how you ask questions in academic environment and I introduced Bloom s taxonomy."

Another shared principle was promoting self-discovery rather than providing answers outright. Dave and Mike both saw the teacher's role as a coach, guiding students to correct themselves. Dave described using clues to help students answer challenging video-based questions, believing that "*the feeling of discovery*" aids retention. Mike similarly guided learners through the correction process, especially in speaking tasks, leveraging his small class size for more personalized support.

Observations confirmed these strategies in action. In Hudson's class, students worked in small groups during an assessed debate. Hudson circulated, observing quietly, and intervened only when groups appeared inactive or off-track. His prompts varied from clarifying task procedures to offering content-based cues, ensuring students stayed aligned with assessment goals. During the SR, he explained that providing some clues helps students to get on the right track and meet the goals of the task, noting that" *when I walk over, it s usually in response to a cue… maybe they seem lost, maybe they re inactive. I address that cue—could be protocol, could be content—to make sure they achieve the goals they re assessed for.*" (Hudson, SR)

Using multiple methods of assessment

Many participants (*n=8*) in this study reported that they use multiple assessment methods, instead of relying on one method. They reported that in addition to formal assessments, they use observation, and informal assessments results to understand students 'achievement, progress, and needs and to be able to address the gaps in their students 'learning and assign accurate grades.

For many, formal assessments were viewed as necessary but limited. Dave, for example, described how he assessed academic seminars and panel discussions. Although he used rubrics, his

grading was not solely based on performance in that single task but also on his ongoing observations of students, explaining that *"it was not 100% based upon their performance in that test, but my assessments were a lot influenced by my general knowledge of their previous performances in class."* Ray reported that while formal summative assessments provide benchmarks, informal assessment during classroom activities is critical for diagnosing needs and preparing students for later evaluations, explaining that *"I observe students to understand their learning styles and needs, but I gain clearer insights after their first or second summative tests."*

Sally emphasized that both assessment types are needed, as formal assessments alone may fail to reflect students' actual progress, noting that *"being a poor test taker doesn't mean you're not learning English."* She uses informal assessments to check understanding, and if students have not learned, *"I reteach the material using a different method.*

Other participants echoed this point, noting that informal assessment allows them to capture aspects of student learning missed in formal assessment. Arman, for instance, explained that he keeps notes on participation during group work, which complements rubric-based grading: "*If there's a group work, I'm hovering over groups, and those who are active in class get a little sign next to their name*." (Arman). Ryan also highlighted the value of informal, formative tasks across the semester, which he often prioritized over high-stakes exams, explaining that "*my focus is on formative assessments throughout the course, not the final exam. Even if the final test counts for 50% of the course, I prioritize the smaller assignments given during the term*." (Ryan)

Several participants further argued that informal assessment can be more valid than formal assessments. Ray explained that tests often generate anxiety, preventing students from showing their true abilities in speaking and listening, noting that *informal assessment is more valid… observing them in non-assessment situations gives the teacher a better opportunity to give feedback in real time."* Mike stressed the importance of informal assessment, especially in content-based courses

where comprehension and participation are crucial, stating that *if [the] course purpose like mine is content-based, yes, informal is more important than formal assessment."*

Overall, participants agreed that relying solely on formal assessments risks overlooking progress and needs. Formal assessments provide structure and accountability, but informal assessments reduce anxiety, capture authentic performance, and allow teachers to adapt instruction. By combining both, teachers create a balanced and comprehensive assessment approach that supports learning while producing fair and accurate evaluations.

Using and supervising peer assessment

Several teachers (n = 7) in this study emphasized the value of peer assessment in oral communication tasks, reporting that they integrate student feedback alongside their own evaluations. They noted that integrating student feedback actively engages students in assessing performance, reinforces learning, clarifies expectations, and guides constructive feedback.

Many participants described using peer assessment to engage students in evaluating specific task elements. For example, Hudson explained that during a debating task, he asks students to give feedback to their peers on elements such as claim, evidence, and warrant as outlined in the evaluation sheet he uses for the debate task (see *Figure* 4.4), noting that *"they do debates in groups or dyads as practice. I listen, the whole class listens, and I provide feedback while also asking the class to give peer feedback on the claim, evidence, and other elements of the debate."*

Figure 4.4

Hudson's *Debate Evaluation Sheet (Hudson-DOC≠2)*

Debate Evolution (15%)

Student name _____**Date** ___ ___ **Time** _____, **Section** _____

Proposition _____

Round 1: Claim/proposition stated and repeated throughout the debate

Evidence

Cites sources:

Warrant

Round 2: responds to opposite proposition with a clear claim

Should respond to _____

Does he/she respond to _____

Evidence

Cites sources:

Warrant

Round 3: closing arguments

Responds to any outstanding arguments

Summarizes evidence and restates proposition

Language and delivery

Maintains eye contact, stands up (not sitting at the computer), does not read from notes, loudness of voice, coherence, clarity, pronunciation, body language, grammar, vocabulary, intelligibility.

Likewise, Ryan described guiding students to provide feedback to their peers during presentations with specific prompts such as *"did he overuse any words?" or "was there distracting body language?"*

Others highlighted using peer assessment to address common errors and help students internalize rubric expectations. Mike reported using peer assessment to address recurring errors in presentations, discussing with the whole class what makes a sentence correct or incorrect and how citation should be handled. He also explained that he prepares peer feedback questions aligned with rubric elements, noting that *"students evaluate other teams' presentations and write notes, which I consider for insights but not for grading purposes."* Ray similarly pointed out that she asks students to listen to their peers' presentations, take notes, and give feedback using the rubric criteria,

explaining that *"peer feedback helps them understand the assessment process better, including expectations and performance standards."*

Several participants also stressed the importance of monitoring peer feedback to prevent misconceptions. Aly stated that during presentations, she allows students to provide feedback to each other and adds her comments at the end, explaining that *"this helps address any incorrect feedback or misconceptions shared among students."*

4.1.5 Responding to assessment stage

Another category that emerged from the data pertains to teachers 'responses to student performance on the assessment. They reported that after each assessment, they provide feedback, discuss the results of assessment in the classroom, tweak the criteria and tasks, provide additional opportunities for practice, hold conferences with students and, when needed, introduce struggling students to support resources.

4.1.5.1 Feedback provision

Providing comments to supplement rubrics

In response to students' performance in oral communication tasks, many teachers (n = 9) emphasized that rubric descriptors alone are insufficient for effective classroom assessment of speaking tasks. They reported supplementing rubric-based scores with written or verbal comments to provide students with specific, actionable feedback.

Several participants described taking notes during student performances and adding personalized comments to highlight strengths and areas for improvement. For example, Ray integrates comments directly into a rubric column dedicated to teacher observations. In the document she shared for assessing group discussions (see *Figure* 4.5), she explained that depending

on what she checks on her observation sheet, she comments on specific aspects such as *"you did a good job on agreeing, disagreeing. However, you need to work on your turn taking."*

Figure 4.5

Ray's *Group Discussion Rubric (Ray-Doc [1]2)*

Speaking Assessment Rubric: Group Discussion

Assessment: Social Interaction: Conversation Management Campus: **XX** Student: **X**

Criteria	5	4	3	2	1-0	Comments
Use of Functional Language (*interrupting, turn-taking, encouraging, giving opinions, agreeing/ disagreeing)	Often uses functional language taught in class. Uses a variety of functional language.	Often uses functional language taught in class. Uses only a couple of structures.	Sometimes uses functional language taught in class.	Rarely uses functional language taught in class or uses some functional language but it is incorrect	Uses no functional language taught in class.	5/ Excellent job in leading your group. It could be useful if you give the organization/steps in the discussion first in introduction
Fluency	Very fluent with no pauses. Speaks at a native-like rate of speed. Easily understood.	Speaks either moderately fast or slow. Some hesitations but do not cause impatience.	Speaks too fast, too slow or with pauses that cause occasional breaks in communication.	Speaks too fast or too slow with many false starts or hesitations. Often affects understanding.	Frequent pauses and hesitations make it very difficult to understand. Cannot articulate ideas.	5 Very natural and communicative
Grammar& Vocabulary Accuracy	Grammatically accurate with good range of structures.	Generally accurate with minimal errors and effort at complex structures	Some grammar errors in basic language. Does not use a range of structures.	Grammar errors that cause misunderstandings in main points.	Grammar errors cause frequent misunderstanding	4.5 some minor errors try to use academic vocabulary
Pronunciation	Clear, mostly error-free pronunciation with very good use of stress, intonation and connected speech.	Few pronunciation errors which do not impede meaning and good control of stress, intonation and connected speech.	Some pronunciation errors that cause misunderstanding. Stress, intonation and/or connected speech have problems.	Speech has many errors and makes speaker difficult to understand.	Often incomprehensible.	5 Excellent
Group Interaction & Participation in Group Discussion	Does not need encouragement to participate and responds very well to other contributions.	Speaks often in group discussion. Usually responds appropriately to other contributions.	Needs encouragement to engage and /or does not respond much and /or often response is not appropriate	Does not actively participate unless directly asked a question. Does not respond to others OR dominates discussion negatively.	Very minimally participates and only when asked OR dominates discussion and responds very inappropriately to others.	5 Excellent job
Overall: Very good discussion/ better to have notes next time for all presentations Introduce the organization or outline of work at the beginning.						Total: /25

*These are sub-topics added by Ray to complement the general criteria. Also, she designed an adjunct rubric (Doc#5) to make it more specific

adapted from EASL 67040 Adv LS (Fejer, A. 2014)

Others described using comments for whole-class feedback. Dave, for instance, writes comments on the board after presentations so that all students can benefit, believing that visual feedback supports second language learners' comprehension. Similarly, during one observed class, Hudson took notes of students' problems during small group debates and later explained them to the whole class, noting that he aligns his notes with the assessment criteria and objectives of the debating task).

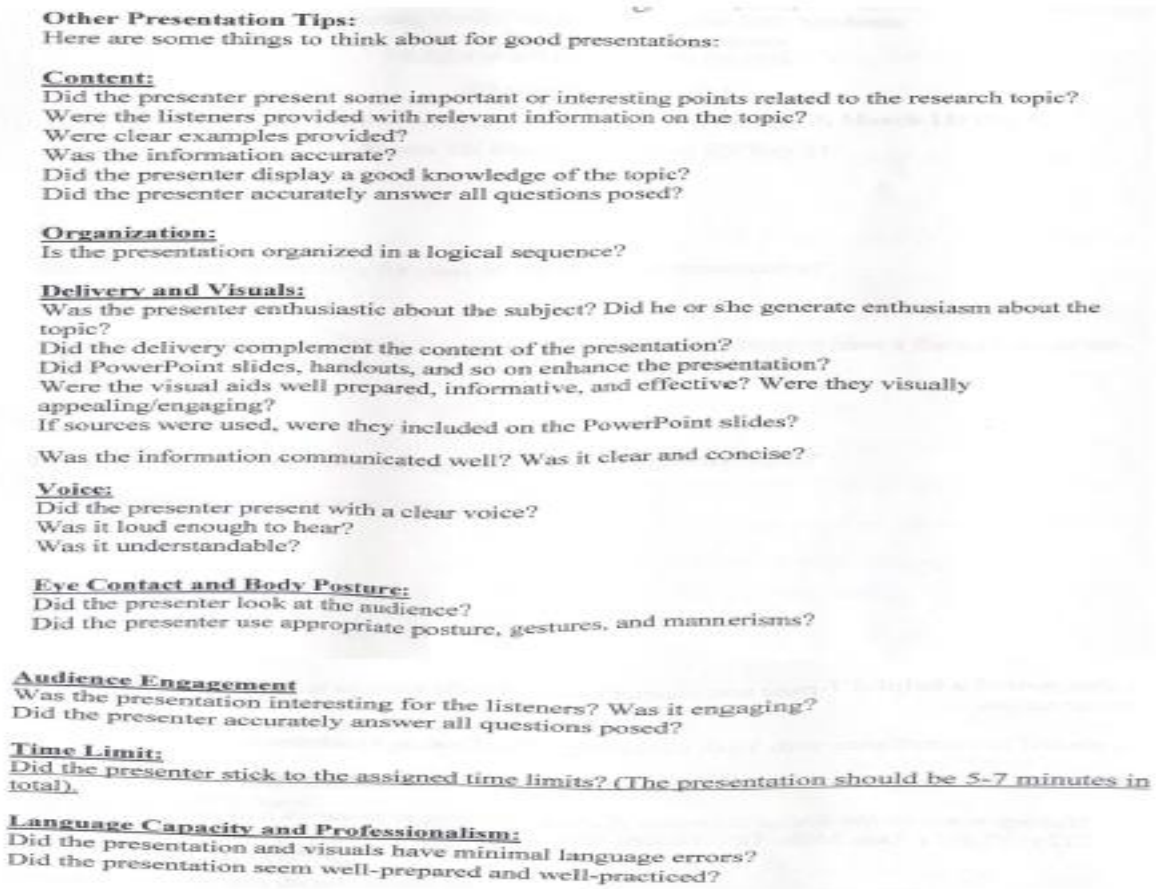
Proving feedback aligned with task objectives and/rubric criteria

In response to student performance, some participants (n = 6) reported that they provide feedback in tune with task objectives and the criteria on the rubric they use for assessing speaking

tasks. Several participants described providing feedback that directly reflects the criteria in their rubrics. For example, Donna explained that during a group presentation, she provided feedback on students' posture and slides, which are among the criteria in the rubric she was using for assessing presentation tasks (Figure 4.6). She noted that after one presentation, when a student did not answer a question appropriately, she provided feedback on answering questions properly–another criterion listed in the rubric.

Figure 4.6

Donna's Group Oral Presentation (Donna-DOC[1] 1)



Other Presentation Tips:
Here are some things to think about for good presentations:

Content:
Did the presenter present some important or interesting points related to the research topic?
Were the listeners provided with relevant information on the topic?
Were clear examples provided?
Was the information accurate?
Did the presenter display a good knowledge of the topic?
Did the presenter accurately answer all questions posed?

Organization:
Is the presentation organized in a logical sequence?

Delivery and Visuals:
Was the presenter enthusiastic about the subject? Did he or she generate enthusiasm about the topic?
Did the delivery complement the content of the presentation?
Did PowerPoint slides, handouts, and so on enhance the presentation?
Were the visual aids well prepared, informative, and effective? Were they visually appealing/engaging?
If sources were used, were they included on the PowerPoint slides?

Was the information communicated well? Was it clear and concise?

Voice:
Did the presenter present with a clear voice?
Was it loud enough to hear?
Was it understandable?

Eye Contact and Body Posture:
Did the presenter look at the audience?
Did the presenter use appropriate posture, gestures, and mannerisms?

Audience Engagement
Was the presentation interesting for the listeners? Was it engaging?
Did the presenter accurately answer all questions posed?

Time Limit:
Did the presenter stick to the assigned time limits? (The presentation should be 5-7 minutes in total).

Language Capacity and Professionalism:
Did the presentation and visuals have minimal language errors?
Did the presentation seem well-prepared and well-practiced?

Hudson ties his feedback to the construct being assessed, particularly in debates. Using his debate rubric in *Figure* 4.4 above, he focuses on clarity of claims, use of evidence, and the connection between the two, explaining that *"my feedback is I write down what I hear… it could be*

pronunciation, it could be grammar, but mostly it's the argument model." Ray also described taking notes and commenting on rubric-based aspects such as appropriate questioning strategies—replacing direct questions with indirect forms—because this is a targeted objective in the task and a listed criterion in her rubric (see *Figure* 4.5 above).

Providing feedback aligned with target language use situation

Several teachers (n=7) reported that they shape their feedback with the target language use (TLU) situation in mind, ensuring that classroom assessment prepares students for real-world communication needs. Ryan emphasized workplace readiness over purely academic skills, tailoring his feedback to the contexts his students are likely to encounter. For example, he advises his students to be concise in meetings, explaining that depending on the job they are preparing for, he gives feedback emphasizing conciseness because *it's crucial to communicate a lot of information in a limited time effectively."*

Donna and Mike also linked their feedback to authentic contexts-academic, professional, and everyday communication so that students can transfer classroom learning to real-life scenarios. Donna provides feedback on nonverbal cues such as posture and gestures in interviews, noting how these can influence employers' perceptions, while Mike highlighted authenticity as central to his teaching philosophy: his feedback focuses on skills students will use beyond the classroom.

Similarly, Ally, Ray, and Sally reported considering the communication demands students may face in future courses or professional settings, and targeting their feedback accordingly. For instance, Sally prepares her students for academic listening by focusing feedback on their ability to extract main ideas from lectures and presentations rather than on isolated vocabulary or pronunciation, explaining that *all these students are going to an [academic context] afterwards… they should be able to get the main idea of a seminar [or] conference."* Overall, teachers viewed

aligning feedback with target language use situations as essential for ensuring learners can successfully apply their language skills beyond the classroom.

Providing metacognitive feedback about language learning

Several teachers (n=8) reported focusing on metacognitive feedback, raising students' awareness of effective communication strategies and broader principles for improving oral communication rather than solely pinpointing specific errors. This approach often involved encouraging students to reflect on their learning habits, understand cultural and contextual nuances of communication, and adopt strategies for long-term skill development. Donna, for example, helps students manage anxiety by encouraging them to practice speaking with native speakers and reflects on appropriate register and formality, explaining that *"rather than saying 'you came from a poor family,' we got into this discussion of register and formality… reinforcing that it's not just in speaking, it's in writing."*

Other teachers described offering concrete strategies and resources that students could use independently. Arman suggests shadowing exercises, recording and repeating sentences, or using curated online materials to improve pronunciation and fluency. Ray engages students in collaborative discussions to identify successful presentation skills, turning these into a shared set of criteria, and also addresses polite language use by highlighting cultural differences in questioning styles.

Ally, Sally, Mike, and Ryan also integrate metacognitive elements into their feedback to help students understand how communication norms vary across cultures and contexts. Ally, for instance, discusses differences in coherence between North American and Asian discourse styles to help students adapt to local expectations. Sally uses analogies to emphasize persistence: *"speaking and listening is like working on your muscles. It takes a very long time… it needs consistency."* Collectively, these practices aim to equip students with the awareness, strategies, and mindset

necessary for ongoing improvement in their communication skills.

Providing different feedback to different students

Many teachers (n=9) emphasized tailoring feedback to each student's specific needs, learning style, and proficiency level to make assessment more effective. They explained that oral communication challenges vary widely among learners, so the type, focus, and level of detail in feedback must be individualized. Ryan illustrated this approach by linking his comments to students' future occupations and linguistic weaknesses, noting that for some he focuses on pronunciation, for others vocabulary. He also adapts delivery style to student personalities-offering explicit criticism to those open to it, and more subtle suggestions to those less receptive-and sometimes records spoken feedback so *"students throughout the semester can keep track of their progress over time."*

Other teachers, such as Ally, Arman, and Dave adjust feedback based on observed performance issues during tasks. Ally described redirecting a student who read directly from slides to instead use notes, while advising another who paused frequently to prepare more thoroughly. Hudson adjusts the depth of feedback according to student motivation and proficiency, providing detailed guidance for engaged, higher-level learners, and briefer notes for those less responsive, noting that *"the level of detail that I provide in the feedback is more than when I'm dealing with a student who I suspect does not even look at my feedback."*

To track and respond to individual needs over time, Mike and Ray maintain logs or portfolios of each student's performance, which they review in one-on-one meetings to pinpoint strengths, weaknesses, and progress. Sally also supplements the school's general rubric descriptors with several paragraphs of individualized commentary, explaining scores for each criterion in detail and highlighting specific areas for improvement in oral communication. Overall, teachers viewed

differentiated feedback as essential for meaningful assessment, enabling them to address the unique challenges and goals of each learner.

Conferencing

Several teachers (n=7) reported holding individual conferences with students after assessments to discuss weaknesses, progress, and strategies for improvement. These meetings allow for personalized, private feedback and targeted guidance. For example, Aly described meeting one-on-one with a student whose presentation skills lagged behind his writing abilities: *"I approached him and encouraged him, saying, 'Let's try again. Watch your classmates first, learn from them, and then give it another shot.'"* Dave avoids public criticism by taking notes during presentations and later meeting individually with students to review rubric scores and discuss performance: *"At the end of class, I held one-on-one tutorials to share and discuss my feedback."*

Other teachers integrate conferencing into their ongoing assessment practices. Donna and Arman use office hours to address students' needs and suggest ways to close gaps in oral communication performance. Mike and Ray keep portfolios of student work and progress, reviewing them mid or end of the semester during individual meetings. For example, Ray asks students to bring back their folders, and rubrics so she can highlight both strengths and areas for improvement, providing constructive feedback to help them focus their efforts in the remainder of the course.

Considering student emotions

Several teachers (n=7) emphasized the importance of addressing students' emotional needs during classroom assessment (CA) and feedback, aiming to build self-confidence, encourage participation, and foster a supportive learning environment. They often use indirect or positive feedback, provide more encouragement in early tasks, and adjust their approach to avoid

embarrassment. Aly noted that she gives struggling students additional opportunities to improve, explaining that *"what we're trying to do with this course is build their confidence and be able to be independent in the university."* Ray likewise offers extra encouragement early on so students feel more confident before tackling more challenging tasks.

Some teachers adapt feedback to student personalities and confidence levels. Ryan avoids singling out individuals when correcting errors, instead writing general points on the board and ignoring minor slips for students with low self-confidence. Arman considers whether students prefer private or public feedback, tailoring his approach accordingly. Hudson and Dave both referred to Krashen's affective filter hypothesis, stressing that lowering anxiety and creating a safe, mistake-tolerant atmosphere is essential for effective learning. As Dave put it *"human beings are not just learning machines, so it's important to create a relaxing, comfortable environment."*

Across accounts, teachers agreed that emotionally aware feedback-delivered in a supportive, respectful manner can reduce anxiety, protect self-confidence, and motivate learners to engage more actively in oral communication tasks.

4.1.5.2 Using assessment results

In addition to providing feedback after conducting assessment, the teachers use their assessment results to make changes to assessment tasks and rubrics, and help students improve learning by sharing rubrics and discussing assessment results with students, providing further practice and introducing them to support centres.

Modifying assessment rubrics

Most participants (n = 9) reported that they might make changes to rubrics as a result of their classroom assessment. For example, Arman said that he keeps modifying his rubrics in different classes in tune with students' needs and his expectations for task performance, explaining that he

constantly tweaks rubrics to align with students' needs, adjusting them for lower-level groups while maintaining clear expectations.

Mike described his approach to assessment as emergentist in the sense that he gets feedback from students' performance and the results of his assessment and makes changes to his rubrics accordingly, stating that *the material is tweaked and at the same time the assessment itself can be tweaked."* He further explained that every time he uses a rubric, he observes students' performance and their common errors and makes changes accordingly. As an example, he mentioned that the rubric he was using for assessing his students ʾpresentations at the time of data collection did not address students' academic word choice and pronunciation, so he was planning to add these criteria to the rubric (see Appendix M), noting that *these are some elements I noticed in these classes which was not addressed in this rubric."*

On a similar note, Dave, Donna, Ray, and Sally reported that for ensuring good assessment and using effective assessment tools, they learn in practice and use the results of assessment to modify rubrics as well as task instructions. For example, Ray explained that whenever she designs a rubric from scratch, she tries it with her students and makes changes using the results of the assessment, stating that *if it doesn't work and doesn't assess this component, or the objective of this skill, I modify it."*

Modifying assessment tasks

Some participants (n=5) reported that they might make changes to assessment tasks or use a new task as a result of assessment. For example, Sally said that if, as a result of assessment, she notices that a task does not lead to learning, she reviews everything and finds another task which is more helpful in terms of learning, stating that *when I see that there is something wrong with the*

task, then I review everything and I find another task, which is more helpful and leads to more learning."

Donna reported that after assessing students' performance in a task, if she sees that students struggle with it, she will add more detailed instruction or discard the task and use another one, explaining that she constantly revises assignments–sometimes refining them, other times completely reworking or replacing them, often focusing on improving instructions and adding more detail to ensure clarity.

Arman and Ryan similarly noted that they evaluate the appropriateness of a task through trial and error and feedback from students' performance to find out whether it taps into the relevant skill or if it needs to be modified. In one case, Arman mentioned that he replaced the "*student-led discussion*" task with an easier one because it was overwhelming to his students.

Sharing rubrics and discussing assessment results

Most of the teachers (n = 9) in this study reported that they share rubrics after doing assessment and discuss assessment results in the procedure of CA. Donna reported that after assessment, she shares the rubric and her feedback/comments with her students and discusses the challenges in their performance in class, explaining how they should have performed. She explained that sharing the rubric aims to highlight students' weaknesses, strengths, and what they need to work on.

Arman, Ally, Sally, Ryan, and Dave indicated that sharing the rubric and results helps students see which areas they did or did not do well and understand why they received a specific mark. Arman described rubrics as roadmaps that guide students' attention to areas needing improvement and help them focus on the next round of tasks. As he noted, *"Using rubric is like roadmap … it gives them a clear sense of what areas did well, and what areas they need to work"*.

He further explained that he goes through the criteria with students after task performance to point out skills they have not yet achieved.

In a similar way, Aly emphasized providing the rubric along with concrete tips for improvement, usually highlighting three main points students need to work on: *"I always have tips for improvement. Three bullet points, work on this, this, that"* (Aly).

Ray and Mike expressed a closely related practice of sharing results and discussing common errors in class. When performance is not satisfactory, Ray shows students a good sample to clarify expectations and discusses students' shortcomings. She also engages peers in identifying strengths and weaknesses of a presentation using the shared rubric.

Giving students more practice

Another follow-up to assessment reported by some participants (n = 5) involves providing additional practice and clearer instruction for completing the task. For example, Ray reported that in cases where many students struggle with a task based on her assessment, she would provide more practice and walk them through the task to fill the gaps. As an example, she explained that when many students struggle with note taking, she would do a demo and take notes herself on the board, noting that *if I noticed that there's something general about the whole class and their performance, I do a demo for them."* In another example, Ray noted that if students' performance in presentation tasks is not satisfactory based on task objectives, she would ask them to practice more and redo the task.

On a similar note, Ryan pointed out that if he feels an activity did not work well and students failed to perform, he will design additional activities for homework so that they practice more and close the gap in their learning, explaining that *if students need more listening or conversation practice, I provide additional materials for independent study. If an activity doesn't go well, I design a follow-up homework task to reinforce learning."*

Likewise, Aly, Mike, and Dave mentioned that as a result of assessing students' presentations, they might ask them to practice further and present again if their presentation is not good enough. For example, Aly reported that she has seen presentations so bad that she asked the student to redo it.

Referring students to support centres

Several teachers (n=6) reported referring struggling students to institutional support centres for additional, individualized assistance in skills such as listening, speaking, and grammar. These referrals are typically made when a student's performance is significantly below class level or when they are at risk of failing. Aly explained that she meets with such students to discuss their situation and then directs them to relevant workshops or tutoring services, noting that *"these are students who are extremely likely not to pass. I tell them there's an ESL open center, sign up for that."* Similarly, Sally noted that very weak students are introduced to tutoring programs to help them catch up with the class pace.

Teachers described using support centres as a complement to their own feedback, especially for students needing sustained practice beyond class time. Ray said she first reviews assessment results with the student, but if the gap remains large, she recommends tutoring around midterm: *"If there is really a concern, I direct them to tutoring."* Donna also emphasized the value of these centres for providing ongoing, personalized feedback especially for lower-level students, which in some cases, according to her, led to significant improvement.

4.1.5.3 Trial and error, and reflecting on assessments

Most of the teachers (*n=9*) in this study pointed to relying on trial and error during assessment and reflecting on their assessments. Through trial and error, they identify the drawbacks of their assessments and take the necessary measures to improve them. Trial and error, and teachers '

reflection on their assessments seem to be an ongoing process in CA. As a result of trial and error, teachers regularly make changes to their plans, curriculum, tasks and assessment tools to enhance the quality of their teaching as well as assessment. To achieve this, they rely on their experience, observation of students' performance, students ʾinput, and fulfillment of instructional goals.

Several teachers described this process in similar ways. Dave viewed trial and error as a form of informal action research in his classroom assessments, where he reflects on whether assessments meet desired goals and modifies them if they do not. As he explained, *"whenever there's a problem, I try to make note of it, think about it and see what went wrong"* (Dave). He added that through trial and error he learns how to better design and adjust tasks to fit students' needs over time. Ryan expressed a closely related idea, noting that he observes students' performance and modifies tasks when they do not tap into the intended skills: *"I think it's trial and error most of the time [ ] sometimes, I modify the task because maybe the task doesn't tap into the intended skill"* (Ryan).

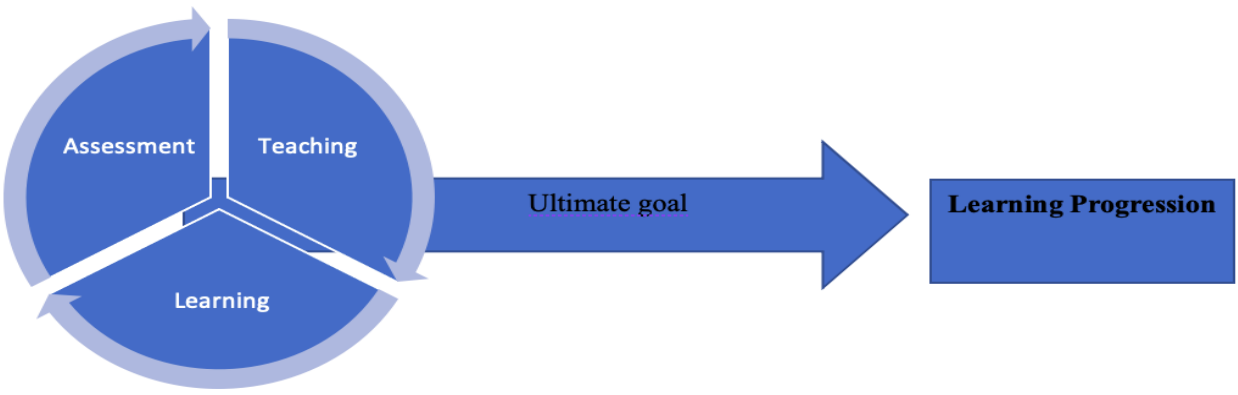
Donna, Ray, and Sally similarly reported piloting and revising rubrics based on students' performance. They test whether the rubric works in assessing the intended components and modify it if necessary. For example, Donna explained, *"I test to see if that works … and make changes according to what the students produce"* (Donna). Ray also noted that she designs her own rubric, tests it over a term, and improves it if it does not assess the skill appropriately.

Arman, Mike, and Hudson likewise stressed that trial and error is central to improving assessments and teaching. Arman emphasized reevaluating failed attempts and taking responsibility for adjustments. Hudson described this reflective process as a continuous improvement cycle: *"Learning by doing basically, this positive loop … you continue doing what worked … and try to modify"* (Hudson).

Overall, the results of data analysis reveal a cycle of assessment, teaching and learning underlying the participating teachers 'assessment practices in the classroom. This cycle, illustrated in *Figure* 4.7, is consistent with formative assessment processes in that teachers consider learning progression in all stages and design or select assessments primarily to understand and address students' needs. In the preparation stage, teachers prepare students for task performance and learning progress. During the evaluation stage, they use multiple assessment methods and reflect on students 'performance and their own practices to enhance the quality of their assessment and provide better assistance to students. After conducting the assessment, they provide feedback, and use assessment results to further support students and help them fill the gaps in their understanding to advance their learning as well as improve the assessments themselves. In sum, as depicted in *Figure* 4.7, in classroom assessment, there is a cycle of assessment, teaching and learning with the ultimate goal of promoting learning progression.

Figure 4.7

Cycle of Assessment, Teaching and Learning in CA



In conclusion, the results show that ESL teachers follow specific procedures in their classroom assessment to ensure effective assessment, especially for L2 oral communication. Teachers first select or design tasks aligned with course objectives and student needs. They then establish appropriate criteria and prepare students for both task performance and assessment. During assessment, they employ multiple methods and diagnose student needs. Afterwards, they provide feedback, use results to improve both their teaching and future assessments, and reflect on their practices to identify and address shortcomings.

4.2 Evaluation of assessment outcome: Effectiveness criteria

To evaluate the outcome of their assessments and their effectiveness, the participants reported relying on their own observation of student performance, student input and grades as sources of evidence, as shown in Table 4. 2.

Table 4.2

Criteria for Evaluating the Effectiveness of Assessments

Teacher observation of student performance - Engagement - Peformance on tasks - Transfer of skills - Progress
Student input and feedback
Grade - Distribution of grades - Valid/accurate grades

4.2.1 Teacher observation of student performance

One main source of evidence for the effective of classroom assessment was teachers' observation of student performance. They relied on different criteria including students' engagement, peformance on tasks, transfer of skills, and progress as source of evidence that their assessments are effective.

Student Engagement

The majority of teachers in this study (n=7) reported that one of the goals of assessment is to engage students, so they consider this factor as one of the main criteria for evaluating the effectiveness of their classroom assessment practices. Aly reported that she understands the effectiveness of a classroom assessment through student engagement. As such, she tries to create every opportunity for her students to participate and share ideas, explaining that to evaluate if something is working, she looks to see if students are interested and engaged, noting that *"my motto is to keep them engaged."* In response to a question about the effectiveness of a rubric she was using for assessing student presentations (*Figure* 4.8), Aly reported that because it encourages participation, it is effective, explaining that "*instead of sitting in class and listening to someone, I am trying to move that to a more student centred approach where students take ownership of learning and generate questions and you drive it and I'm sort of guiding you along"* (Aly). Aly's interest in student participation in presentations can be seen in the rubric she was using for this task. The rubric in Figure 4.8 included a student "*participation*" criterion, which consisted of two descriptors i.e., "*watched and responded to a minimum of three presentations*" and "*responses were detailed and clearly structured*". Both descriptors attest to the importance of student participation and engagement during presentations.

Figure 4.8

Ally's *Presentation Rubric (Ally- Doc 2)*

Category	Scoring Criteria	Total Points	Score
Organization (15 points)	Information is presented in a logical sequence.	5	
	Presentation appropriately cites requisite number of references.	5	
	Formatting and slide design are clear and add to the overall impact of the presentation.	5	
Content (35 points)	Introduction is attention-getting and presents the thesis clearly.	5	
	Concepts and ideas are well-defined in language appropriate for the audience.	5	
	Presentation contains accurate information. Information is properly cited from their sources.	10	
	Material included is relevant to the overall message/purpose.	5	
	There is an obvious conclusion summarizing the presentation.	5	
	Discussion questions are engaging and understandable to the audience.	5	
Presentation (40 points)	Speaker's voice was clear and audible.	10	
	Speaker was well-paced and easy to follow.	10	
	Speaker uses good pronunciation and intonation.	10	
	Speaker is engaging and encourages audience to listen.	10	
Participation (10 points)	Watched and responded to a minimum of 3 presentations.	5	
	Responses were detailed and clearly structured.	5	
Score	Total	100	

Several other participants similarly emphasized student engagement as a key indicator of effective assessment. Arman stressed selecting intriguing topics for speaking activities to motivate peer interaction, explaining that *"for good assessment, the level of engagement on the student side is important, so the topic for speaking should be intriguing"*. Mike described observing students actively engaging with course-related videos and group discussions, while David noted that academic seminars encourage students to participate beyond class time, noting that *"this encourages students to work outside of class hours, do research, and take part in discussions"*. Donna added that low-stakes, interactive tasks, such as discussing documentaries, can effectively maintain student interest and involvement.

Student performance on tasks

According to many participants (n = 8), successful task completion was one of the main criteria they used to evaluate the outcome of their CA and its effectiveness. Overall, they viewed students' ability to fulfil task requirements and meet the criteria as evidence that their assessment and teaching were appropriate and effective.

Ryan explained that student performance in the task is a better source for understanding the appropriateness of assessments than relying on his own assumptions. He considers his assessment effective, *"if students can fulfil the requirements of the task"*. Hudson contended that student progress over time and mastering a task based on the instructions signal effective teaching and assessment. As he noted, if a student initially struggles but later succeeds, he infers that his instruction and assessment played a role. He also added that in tasks such as debates, meeting the criteria for good debating (see *Figure* 4.4) indicates effective assessment.

Aly expressed a closely related view but also highlighted unsuccessful task completion as a sign that revision is needed. She described stopping a group discussion task when she realized students could not perform it well and that it did not trigger the level of discussion intended.

While most participants focused on students' ability to complete tasks successfully, Sally and Arman considered overall class performance or average scores as indicators of successful task performance. In contrast, Donna focused particularly on strong students, noting that confusion among them would signal problems with the rubric or assessment design.

Skill transfer to new tasks

Another theme related to student performance on tasks is students' ability to transfer learned skills to new tasks. Several participants (n = 6) used this criterion to evaluate the effectiveness of their classroom practices. Overall, they viewed students' application of previously taught skills in new classroom, assessment, and real-life contexts as evidence of effective teaching and assessment.

For example, after teaching the passive voice in speaking, Sally observed students successfully using passive sentences in a new task, which she interpreted as effectiveness of her teaching and assessment: *"they had beautiful passive sentences which showed effectiveness of my teaching"*. Dave and Aly emphasized that when students apply learned strategies or skills in presentations and other tasks, it indicates assessment has been effective.

Ray and Ryan further noted that transfer can be revealed through formal assessments designed to differ from practice tasks and require students to incorporate previously learned skills (e.g., Ray's "Oral Summary of Presentation" task).

Beyond classroom tasks, Hudson and Ray highlighted transfer to real-life contexts as an important indicator of CA effectiveness. As Hudson explained, students reported that using Toulmin's model for debating helped them in contexts such as job interviews and resume writing: *"Students have reported back how that model has helped them in their careers"* (Hudson).

Student progress

A number of participants (n = 5) highlighted student progress over time during the semester as an important criterion to evaluate the effectiveness of their teaching and assessment practices. Arman stressed that progress matters more than just grades, emphasizing learning as the key outcome rather than simply passing or failing. He explained how some students who struggled with speaking at the beginning of the semester improved significantly by the end, participating more confidently in discussions and performing better in assessment tasks, stating that *the whole idea for me is just learning. You don't want to assess your students on just passing or failing, but there are students who may not find the perfect answer but they've learned that."*

Mike also uses student portfolios to track progress over time, alongside regular notes on individual student performance and verbal feedback from students. Sally and Ray compare students' performance at various points in the semester to assess their growth and the effectiveness of their

assessments. Sally remarked that she believes her assessment is effective because she has seen progress, noting that "*in just two months, students progressed from struggling with a one-minute audio to handling 20–30 minutes*."

Aly, who has taught the course for many years, relies on her experience by comparing current students to past cohorts. She uses this comparison to identify what defines a successful student and to evaluate her teaching and assessment effectiveness, explaining that having taught the course for years, she can compare students to a large group and identify what defines a successful student versus one who is still developing.

In addition to student progress, some teachers (n=4) also use students' grades in formal, graded assessments during and at the end of the semester as an indicator of effectiveness. Ray noted that if most students achieve good grades, this suggests that assessments have been effective in achieving target outcomes: "*From students' marks, I understand the fulfilment of the target outcomes. If results are good, it's a good assessment*." However, she also noted that if students perform well in class but poorly in summative tests, the assessment design needs revising. Sally, Arman, and Hudson echoed this view. Sally pointed out that a low class average indicates an ineffective assessment, explaining that "*if the average mark is below a borderline, it means the assessment didn't work.*"

4.2.2 Student input and feedback

Another source of evidence for the effectiveness of classroom assessment was student input. Most participants (n=8) considered student input, both verbal and non-verbal, an important criterion for evaluating the effectiveness of their classroom assessment. They reported that they actively seek students' ideas about tasks and observe their classroom behavior to make informed adjustments. Dave stressed the value of student feedback for refining rubrics, reducing ambiguity when students report unclear sections: "*Sometimes they come to me, and say that they do not understand part of the*

rubric. If there are ambiguities, I try to revise the rubric or explain it properly." Similarly, Mike takes student feedback, discusses it with the whole class, and adjusts instruction accordingly, while Ryan incorporates student suggestions when selecting different assessments and tasks.

Three participants emphasized the importance of non-verbal cues in judging CA effectiveness. Ryan explained that he watches for signs of confusion, such as gazing, to detect when material is "*absolutely baffling*." Aly monitors whether students appear "*interested*" and "*engaged*," while Donna observes whether they respond positively to activities, noting, "*I don't have the sense that they hate these activities*."

Six participants reported eliciting verbal feedback from students through structured methods like questionnaires, checklists, or needs analyses. Sally distributes her own end-of-course questionnaire to encourage honest responses for curriculum planning: "*It's for me to be better… try to tell me if there was any part in my class that you didn't like.*" Arman's evaluations ask students which assignments should continue or change, and Ray integrates widely shared feedback into her teaching and assessment. Mike uses post-assessment discussions to check whether students are truly learning, while Ryan conducts weekly written needs analyses, asking students to specify what should be kept, removed, or added.

Aly also values student feedback to understand her teaching and assessment effectiveness and identify misunderstandings and areas needing further explanation. However, she noted that cultural factors sometimes make it difficult for students to question their teacher or verbalize concerns: "*There's so many students [who] come from this culture where they will never question the teacher. It's hard to even interpret their look*."

4.2.3 Grades

Distribution of grades

Distribution of grades and accurate grading was another source of evidence for classroom assessment effectiveness. Some teachers in this study (n = 3) defined good assessment as assessment which shows the real ability and achievement of students and differentiates students in terms of ability and achievement, especially at the end of the semester. Aly believed that accurate grades should be assigned to students, differentiating them in terms of ability and learning. She criticized other teachers who award higher grades to all students, noting that *we need to have A's and B's and C's, like normal distribution of the grades."* In this regard, Donna pointed out that normal distribution of scores shows that her assessment decisions are valid and accurate, explaining that a typical range would include a handful of A's, a larger range of B's, some C's, possibly some D's, and very few F's.

Ray said that she evaluates the effectiveness of a rubric based on its potential for differentiating students in terms of ability and learning. If it does not, she explained that she tries to add sub-criteria to make it more detailed so that the nuances between students' ability can be shown better, stating that *if it doesn't differentiate a student's ability, I make it more specific and include some sub-topics."* She mentioned the IELTS rubric for speaking as a good example of an effective assessment tool since its criteria and scoring bands are detailed, demonstrating nuances between students' proficiency better.

Valid/accurate grades

Accurate or valid grading was mentioned by some participants (n= 6) as another criterion for evaluating the effectiveness of their classroom assessments. These teachers reported developing unbiased assessment, targeting the construct of interest in their assessments, using the performance

of top students as a reference, and relying on the results of both formal and informal assessments to assign accurate grades.

Aly and Sally emphasized that effective assessment requires grades that accurately reflect students' true achievements rather than inflated marks. Aly criticized grade inflation, noting it can create a false sense of ability: *"Three quarters of the class ending up with an A is unrealistic. It gives them a false understanding that they are really good"*. Sally expressed a similar view, highlighting honesty in grading, explaining that *"I never bump their marks. I only do it when the student is on the verge of failing and deserves a point or two just to pass"*.

Several participants stressed the importance of aligning assessments with the intended construct to ensure accuracy. Mike noted that assessments should measure exactly what they are intended to, while Ray emphasized differentiating students by ability: *"A good assessment assesses what it is supposed to measure. [It] should differentiate or have the capacity to differentiate between less able and more able students"* (Ray).

Consistency between observed classroom performance and formal assessment grades was another key indicator. Donna and Reymond noted that when strong students perform as expected on formal tasks, it validates assessment accuracy. Ray additionally highlighted reflecting on discrepancies when high-performing students score lower than expected.

Finally, several participants, emphasized combining results from informal and formal assessments to improve grading accuracy and understanding of students' learning. Ray explained: *"…throughout the term you see the students in action while they practice; therefore, you have a better understanding of students and their ability*" (Ray).

All in all, the above results show that in classroom assessment, teachers rely on three main sources of evidence concerning the effectiveness of their assessments: their own observations and reflections, student input, and grades.

CHAPTER 5: TEACHERS ʾCONCEPTIONS OF CLASSROOM ASSESSMENTS

This chapter reports results related to the second research question regarding ESL teachers' assessment beliefs and their relationship with oral communication assessment practices in the classroom. During the interviews teachers were asked about their beliefs regarding effective L2 teaching and assessment, goals and methods of assessment, and characteristics of classroom assessment. Table 5.1 lists the themes and categories that emerged as a result of data analysis. Each of the beliefs in Table 5.1 is discussed in the following sections with examples as well as its relationship with practices i.e., its implementation in CA with concrete examples from reports and documents showing their implementation in classroom practices.

Table 5.1

Teachers' Conceptions of Classroom Assessment

1-Teachers ʾgeneral beliefs about teaching & assessment of L2 oral communication
- Communicative language use is central (n=5)
- Communicative competence should be prioritized (n=5)
- Engaging students is crucial (n=8)
- Teacher as facilitator scaffolding assessment (n=6)
- Assessment and teaching are inseparable (n=7)
- Target language use is foundational (n=8)
2- Teachers ʾbeliefs about assessment goals
- Assessment aims to diagnose student needs and scaffold learning (n=8)
- Assessment primary goal is improving student learning (n=8)

3- Teachers' beliefs about the characteristics of good classroom assessment

- Consistency between assessment and construct is foundational (n=6)
- Fair assessment and grading are vital (n=8)
  - A) Fair assessment is accurate and differentiates between students (*n=7*)
  - B) Fair assessment accounts for student effort (*n=7*)
  - C) Fair assessment is unbiased and objective (*n=5*)
    - Teacher should use objective tools (*n=5*)
    - Teacher should give equal opportunity to all students (*n=3*)

    D) Fair assessment follows the same guidelines and grade breakdown for all students (*n=4*)
  - E) Fair Assessment is equitable (*n=6*)
  - F) Fair assessment is transparent *n=8*)
- Good assessment has positive washback (n=6)
  - Use assessment tools and criteria for teaching purposes (n=6)
  - Use formal assessment results for formative goals (n=5)

4- Teachers 'beliefs about methods of assessment of oral communication

- Listening is integral to speaking assessment (n=8)
  - Implicit assessment of listening skills (n=6)
- Use of multiple assessment methods and informal assessments (n=9)
  - Informal assessment as a space for enacting teacher beliefs (n=6)

5.1 Teachers' general beliefs about teaching and assessing oral communication

5.1.1 Communicative language use is central

One of the commonly held beliefs by the participants (n=5) about L2 teaching and assessment was that language should be used for communicative purposes. Teachers who adhere to this view believed that L2 classroom's focus should be on the message and meaning; they put less weight on language accuracy.

Arman stated that his teaching philosophy is based on the communicative approach, prioritizing communication and fluency over accuracy, explaining that *"at this level, I don't care about accuracy, fluency is number one priority."* He noted that he does not put much weight on grammar in his assessments, emphasizing that students' participation and talk–even a few words–is of utmost importance in his formal and informal assessment and grading. During one of his classes that I observed, Arman asked his students to jot down a few words or whatever caught their ear and

talk to a partner. In the stimulated recall, he reiterated that students' participation and talk, regardless of accuracy, is central to his assessment and grading. The rubric Arman used to assess student oral presentations in *Figure* 5.1 demonstrates how he prioritized communicative language use over accuracy; the criterion about proper language use emphasizes meaning and appropriate vocabulary, only considering global grammar errors that might interfere with meaning.

Figure 5.1

Arman's Oral Presentation Rubric (Arman- Doc2)

Oral presentation rubric Name _____

Criteria	Notes	Points	
Preparation – presenter demonstrates evidence of preparation (handouts, .ppt, poster, etc.); there's no reading; note cards used effectively			/10
Visual aids – visuals and props (if any) used effectively and appropriately			/5
Organization – the talk was well-organized, coherent, unified, with no parts missing			/20
Verbal articulation – presentation was smooth and without pauses; ideas were clearly presented			/20
Proper language use – presenter was easy to understand, used appropriate and interesting vocabulary, made few, if any, errors that interfered with understanding			/30
Audience engagement – presenter kept the audience interested and engaged, used eye contact and gestures			/10
Time management – presentation was timed correctly			/5
Total score:			/100
Comments:			

In a similar way, Hudson described himself as "*a child of the communicative approach*," prioritizing meaning over form. He explained that in assessing oral presentations, he only considers

global errors or vocabulary issues that impede communication, noting that he marks performance "*poor*" only when it is so inaccurate or unintelligible that this gets in the way of communication. A look at the criteria in the rubric he used for assessing student presentations (see *Figure* 5.2) shows that most criteria relate to communication–knowledge of topic, clear presentation, use of examples and evidence, non-verbal behavior, enunciation, answering questions, and interacting with others– suggesting that communication is his main concern in assessment.

Other participants expressed similar beliefs aligned with the communicative approach. Mike noted that he prepares students to communicate properly, providing feedback only *"when it impairs communication."* This belief is evident in the rubric he used for a presentation task (see Appendix M), where most criteria relate to the use of source materials, visuals, and non-verbal behavior for successful communication, without any criteria related to grammar. Ray stressed the importance of communicative language use in her practice. She referred to a rubric she used for assessing group discussions and/student interactions with each other (see *Figure* 4.5). She explained that criteria related to the *Use of Functional Language* (interrupting, turn-taking, encouraging, giving opinions, agreeing/ disagreeing), *Group Interaction* and *Participation* are the most important criteria in this rubric as they are more specific to the discussion task.

Donna also expressed a strong belief in the communicative approach, noting that she focuses on helping students *"survive university to be able to communicate."* In the rubric she used for assessing student presentations (*Figure* 4.6), she pays more attention to delivery of content and prosodic features such as body language for successful communication.

5.1.2 Communicative competence should be prioritized

Attention to communicative competence and the different forms of knowledge a language user has, was another theme that emerged from participants' responses regarding L2 teaching and

assessment. Some participants (n = 5) believed that addressing multiple forms of competence in general and pragmatic competence in particular is essential in L2 teaching as well as assessment. Accordingly, in their practices, they reported they make sure to assess different forms of knowledge, include them in their assessment tools, and provide feedback on them.

Ray believed that communicative competence has different components, which are all important, hence rubrics should be comprehensive enough to capture students' different competencies, explaining that *the current optimal methodology is Communicative Language Teaching, an integrated approach that develops various language competencies beyond grammar, including discourse and communication skills."* She noted that if a rubric is not detailed enough for a speaking task, she would add sub-criteria to obtain a full picture of a student's speaking ability–for example, adding criteria such as using speech acts and speaking strategies like formulaic expressions or indirect speech acts. In terms of providing feedback, she explained that she uses a rubric for group discussion tasks (*Figure* 4.5) to provide feedback on different aspects of student performance.

Dave reported that he has been influenced by different models of communicative competence, including Hymes' model and Canale & Swain's more elaborate concept containing sociolinguistic, strategic, discourse, and linguistic competence, noting that *all those different beliefs have influenced me in designing speaking and listening activities."* He added that he also believes in Halliday's functions of language–ideational, textual, and interpersonal.

Other participants expressed similar views on addressing multiple competencies in assessment. Ryan noted that language ability includes different types of competencies beyond linguistic knowledge. He criticized courses and assessments that focus only on content and communicative skills while disregarding linguistic knowledge, explaining that *even in communicative activities, I ensure language instruction is included, never relying on purely communicative tasks."* He argued that presentations are a good assessment tool as they tap into

different aspects of competence, and the checklist he uses for assessing student presentations (see Appendix N) addresses verbal skills, appropriate use of language, non-verbal behavior, and interaction with the audience. Hudson pointed out that in presentation tasks, he emphasizes linguistic, paralinguistic, and debate protocol aspects. In tune with this view, the criteria in his rubric, as shown in *Figure* 5.2, use of language, connectors, gesture, volume of voice, knowledge of content, and arguments–target these three aspects.

Figure 5.2

Hudson's Oral Presentation Rubric (Hudson Doc3)

Presentation scoring rubric for Oral presentation

Student name _____

Title of presentation _____

Date _____

	Excellent 15	Good 11	Adequate 7.5	Poor 5<
1. Speaker demonstrated **knowledge** of the topic.				
2. The presentation had a **clear beginning**, middle, and end				
3. There were **connectors** to link parts together				
4. There are **list** words to indicate the sequence of information				
5. **Ideas** and substance of the presentation were easy to comprehend				
6. Speaker used clear **examples, evidence** that were connected to the claims				
7. Speaker's **use of language** was accurate and appropriate				
8. Speaker used suitable **gestures** and body movements				
9. Appropriate **eye contact** was maintained.				
10. Speaker **articulated** clearly.				
11. Speaker used suitable **volume and voice** variation.				
12. Speaker stayed **on topic.**				
13. Speaker responded suitably to **questions**				
14. Speaker **interacted** with others				
15. The talk finished within allotted **time**				
16. Suitable **closure** was brought to the speaker's portion of the presentation.				
17. Grammar and vocabulary was appropriate				

-Main Argument: Did the student state the thesis? Y N
-Did the student provide evidence for the thesis? Y N
-Rebuttal: Did the student state the counter-claim? Y N
-Did the student provide evidence for the counter-claim (Rebuttal)? Y N
-Response to Rebuttal: Did the student respond to rebuttal? Y N
-Did the student provide evidence for response to rebuttal)? Y N

5.1.3 Engaging students is crucial

Most of the participants (n = 8) pointed to the importance of student engagement as one of their fundamental beliefs for effective teaching and assessment. They reported that they design or select tasks and assessments in a way that involves and engages all students.

Aly emphasized that student engagement is her motto in teaching and assessment, explaining that she does everything to engage all students by directly putting them on the spot and asking questions to those who are silent, stating that *my motto is keep them engaged."* Following this goal, in the rubric she was using for evaluating student presentations (see *Figure* 4.8), one criterion relates to student response to presentations to encourage participation. She awards 10 points to those who give a detailed response to a minimum of three presentations, noting that *it encourages participation and supports a student-centered approach include [ing] a criterion for audience engagement."*

Arman believed that what matters most in students 'performance is that they participate and engage regardless of whether they provide the right answer or not. He pointed to the "*audience engagement*" criterion in a rubric he was using for assessing student presentations (*Figure* 5.1) to highlight the importance of engagement in his assessments, explaining that *there is no right or wrong answer as long as you have something to say… that's how I like to evaluate my students."*

Other participants emphasized similar strategies for fostering engagement. Dave noted that he tries to engage all students, including those who tend to be silent, by calling names and selecting leaders for group discussions to ensure accountability, explaining that *I appoint a leader in each group, making them responsible for reporting back to the class, ensuring that the onus is on all of them."* Ray described selecting authentic and meaningful tasks to activate background knowledge, noting that *when students can make connections, then I can elicit more answers, and they get more engaged."* She also requires students to listen and participate during group presentations, using a rubric (*Figure* 4.5) to give marks and comments on asking questions and participating

actively. Mike and Sally similarly emphasized selecting interesting topics and including participation criteria in rubrics to ensure all students are engaged.

5.1.4 Teacher as facilitator scaffolding assessment

According to participants (n = 6), the teacher's role in teaching and assessment is primarily that of a facilitator who scaffolds learning, a notion similar to that of the ZPD. Sally believed that teachers are facilitators who are expected to diagnose students' needs and tailor their teaching and assessment for optimal outcomes, explaining that she uses SCT and ZPD theories in her classes and therefore sees herself as a facilitator who helps students. In accord with this belief, she scaffolds students by providing framed speaking tasks, such as starting a narrative task herself and then asking students to continue the story, and then provides metalinguistic information regarding the target structure if students struggle. She noted that she identifies students' needs, tailors her teaching accordingly, and scaffolds them to fill gaps in their current understanding.

Similarly, Hudson expressed his belief in scaffolded assessment, explaining that effective assessment involves *correct determination of where to start and proper mediation throughout."* He described that as he teaches and assesses simultaneously, he identifies the appropriate ZPD of students and closes the gap by giving appropriate feedback, characterizing his approach as *the small steppingstones that lead to learning, which eventually will culminate in that final."*

Other participants similarly emphasized scaffolded approaches to teaching and assessment. Arman described helping students build on what they already know to communicate better, explaining that *learning to speak is just like trying to help students with what they already have and bring them to surface."* During one observed class, he gave clues or sample answers to encourage students to express their ideas. Mike viewed teachers as *game keepers or coaches"* who mediate student performance. He provides lists of useful expressions to scaffold discussions (see Appendix O), explaining that these functional language inputs help *provoke their memories."* He also encourages self-correction through elicitation, using backchanneling so students can recognize and

correct their own mistakes. Ray reported that she attempts to be indirect in providing feedback, scaffolding students to self-correct, and becomes more explicit only when necessary– *especially if it's about the specific requirements that brings down the quality of their performance and mark."* Dave emphasized student self-regulation and the teacher's role as scaffolder, explaining that he keeps himself away as much as possible to make all group members responsible and engaged, but pops in to give support when students struggle.

5.1.5 Assessment and teaching are inseparable

Many participants (n = 7) in this study referred to the interrelationship between teaching, assessment, and learning. In accord with this belief, they adopt a step-by-step approach and use multiple assessments integrated with teaching to meet learning goals.

Ray believed in a cycle of assessment, teaching, and learning, stating that this belief guides all her practices in the classroom, explaining that *teaching, learning and assessment should stay as a chain… these three should be aligned."* She emphasized that assessment should be used for formative purposes, noting that a loose connection between them would not lead to desired outcomes. In the case of creating rubrics, she relies on trial and error, testing a rubric for a term and refining it to ensure it accurately assesses intended skills, adding that *if one is weak, the entire process is compromised."*

Dave maintained that assessment is an extension of teaching, regarding assessment as another opportunity to help students achieve their goals, explaining that *assessment is not something external to teaching; it is an extension of teaching… When I assess, I just want to ensure that students achieve them with my assistance."*

Other participants similarly emphasized the close relationship between assessment and teaching. Mike noted that assessment and teaching work hand in hand, explaining that his assessment criteria are always adjusted based on instruction, and instruction is adjusted based on previous assessment results. He added that when the first group of students presents, he assesses

their performance, identifies drawbacks, and makes changes accordingly, stating that *if this is not working, I will devise an instruction that helps them understand this area of their weaknesses."* Hudson used the analogy of a *positive loop"* to describe the close relationship, explaining that as a reflective practitioner, he thinks about what he did properly and what he did less appropriately, and tries to modify his practice accordingly. Aly described teaching a point several times if she does not see progress—for example, when students struggled to ask good questions, she provided a taxonomy with sample questions, then did more modeling and practice when they still did not master the skill. Sally emphasized reflecting on student performance, diagnosing needs, and *tweak[ing] things, change[ing] stuff, go[ing] over the things that didn t work."* Ryan described teaching strategies such as interrupting politely, and when students failed to implement them, he extended instruction and provided more explicit feedback. Arman highlighted the link by explaining that in an oral summary task, when many students struggled to identify main ideas, he adjusted his approach by replaying the video and guiding them through key points, recognizing that *if 12 out of 17 haven t been able to get the main idea, probably there s an issue with the content."*

5.1.6 Target language use is foundational

Most of the participants (n = 8) emphasized that target language use (TLU) is a foundational factor in their teaching and assessment practices. Aly expressed her belief in the importance of TLU, explaining that she considers what will be useful for students in terms of understanding academic tasks and building skills they might need in future jobs, such as summarizing, discussing, and posing questions. She noted that in presentation tasks, she puts more weight on presentation and organization skills than content because *students can learn about content on their own"* through various sources. The rubric she used for assessing presentations, reproduced in Figure 4.8, included organization, content, presentation, and participation, with presentation and organization skills receiving more emphasis in her feedback.

Mike believed that students' current and future needs should be considered to make teaching and assessment effective, explaining that he always thinks about *their language use in the workplace… and academic language courses."* He selects meaningful tasks based on TLU to engage students, noting that his goal is to prepare students for university academic learning. In a watch and discuss task he shared (see Appendix O), the assessment starts with listening to factual information (e.g., *who is the speaker, what is his job*) and gradually moves to deeper understanding of speakers' views (e.g., *what issues they raised; what are their views on multiculturalism*), followed by sharing personal positions and rebutting–smoothly preparing students for debating in academic contexts.

Other participants similarly emphasized preparing students for real-life tasks. Dave adjusted his teaching to focus on employability skills based on cohort needs, emphasizing notetaking and speaking skills for students pursuing business or accounting, explaining that *they need to become familiar with a couple of standard formats for notetaking."* Sally believed that speaking and listening should be assessed together because *that's what happens in real life… in a natural way, we listen and speak."* In her seminar task rubric (Appendix P), she focused on understanding main ideas rather than micro skills like vocabulary or phonemes, explaining that *in academic setting they should be able to get the main idea of like a seminar, conference."* Donna emphasized preparing students for future courses and real-life situations, stating that her philosophy is *to give them the tools to succeed in university, or in life… building towards the other courses and their future work."* Ryan focused on preparing students for the workplace, designing role plays and jigsaw tasks to simulate real-world scenarios, explaining that *they are going to take part in workplaces where they would have to speak English in different situations."* Additionally, Ryan focused on job interview preparation, using interviews as a model for assessment and feedback, ensuring students practice key skills such as persuasiveness, which they would need in their careers.

5.2 Teachers' beliefs about assessment goals

5.2.1 Assessment aims to diagnose student needs and scaffold learning

According to most participants (n = 8), one of the main goals of classroom assessment is to diagnose students' needs in oral communication and scaffold them to close gaps. Ray emphasized that student needs are the foundation of her teaching, explaining that she adapts her teaching and assessment to each cohort, noting that *"teaching is not one size fits all… each time my approach to the whole group might be different."* She highlighted the importance of conducting placement assessment at the beginning of the semester to understand students' needs and adapt classroom practices accordingly. For heterogeneous classes, she uses different tasks and provides more scaffolding if a task is difficult such as giving more practice, providing a list of challenging words, and brainstorming before task completion.

Aly also stressed the importance of diagnosing students' needs and assisting them to fill gaps in learning. She explained that when a task does not correspond to students' level, she increases assistance, explanation, and practice, noting that *"if you find yourself in a class where students are lower than expected, you can scaffold a little bit more… I design a mini lesson."* A good example of this belief in practice is her use of Bloom's taxonomy for posing academic questions. When she realized students were not familiar with how to ask questions in presentation tasks, she introduced them to Bloom's taxonomy (see Appendix K) and prepared a mini lesson with prompts organized into six categories (remembering, understanding, applying, analyzing, evaluating, and creating) which students could use as a frame for academic questions and discussions.

Other participants similarly emphasized diagnostic assessment and scaffolding. Dave explained that depending on students' level and needs, he provides more support if a task is challenging–for example, giving clues or replaying challenging parts of a video to help students discover answers on

their own. Hudson defined good assessment as accurately identifying students' level and needs and mediating to close the gap, explaining that during mock debates he takes notes of students' challenges and provides feedback either on the spot or after they finish speaking, inviting the whole class to learn from each other's presentations. Arman also expressed his belief in diagnosing student needs and addressing them through targeted support.

5.2.2 Assessment primary goal is improving student learning

Improving teaching and learning, as one of the main goals of classroom assessment, was a belief held almost by all participants (n = 8). Dave, for example, highlighted student learning as the goal of all classroom practices, including assessments. He noted that he conducts assessment to identify students' needs and provide required support to help them advance their learning. In line with this belief, he explained that he uses rubrics for improving students' learning, not just as measurement tools, noting that *"the questions and rubrics will guide students to form a genuine conversation [ ] And I guide them through the questions."* He added that if students still do not start the conversation appropriately, he gives them further instruction.

Ryan expressed his belief in using assessment results formatively to improve his teaching as well as students' learning, explaining that he now views assessment as formative i.e., *"to analyze to see what it is that I need to teach a little bit more or elaborate on."* As an example, he noted that he uses the descriptors in the rubric for a final presentation assessment (Appendix J) for teaching purposes, clarifying course and task objectives to enhance students' learning, such as teaching them about cultural appropriateness in speaking–for instance, avoiding phrases like *"let me finish"* when interrupted.

Other participants similarly emphasized improving student learning as the primary goal of assessment. Sally stated that grades are not important to her; what matters is that *"the student has*

learned what I've taught them… the grades are numbers to me." She noted that students learn better when she shares the rubric with them because it helps them get familiar with instructional goals and move faster toward achieving them. Arman said that the main goal of assessment for him is learning, not providing perfect answers or grades, explaining that *"learning is the outcome as opposed to just answering the right question."* Donna observed that students' obsession with grades distracts them from the main goal of assessment, which is advancing learning and preparing for future task performance, explaining that she shares rubrics with comments, including summary feedback, and invites students to discuss further during office hours.

5.3 Teachers' beliefs about the characteristics of good classroom assessment

5.3.1 Consistency between assessment and construct is foundational

Most participants believed that consistency between assessment methods and the construct being assessed is foundational to good assessment. They believed that the construct being assessed should be reflected in the assessment tool, and that the criteria in speaking rubrics should include features essential to the definition of these skills.

Sally believed that classroom assessments should have construct validity in the sense that the rubric and criteria should represent the target skill being assessed. She explained that a good rubric should have all categories or aspects of the speaking construct, such as fluency and accuracy, and that if the goal is for students to use past tenses, that should be part of the rubric, noting that *it has to have all the aspects of speaking in it."* Her belief in construct validity is reflected in her tool development, where she ensures that the rubric includes all the important criteria for assessing speaking.

Hudson also expressed his interest in construct validity, explaining that he is *a big proponent of construct validity"* and ensures his assessments accurately measure the intended

construct, relying on his professional judgment and research literature to support his practices. He implemented this belief in a debating task by using an evaluation sheet (*Figure* 4.4) that included criteria relevant to debating–such as evidence, warrant, citing sources, responding to opposing positions, and summarizing evidence. He contended that the presentation task rubric (Figure 5.2) is effective because *it measures what it's supposed to measure… they prepare for it all throughout the semester."*

Other participants similarly emphasized construct validity in their assessment practices. Ray defined good assessment as one that *measures what it is supposed to measure."* She explained that if a rubric is too holistic, it is not valid, so she makes it more specific by adding sub-topics. For example, she explained that the criterion "*use of functional language*" in the group discussion rubric (*Figure* 4.5) is not sufficiently detailed to allow assessing the different language functions that the students should use in a group discussion task. As such, as displayed in *Figure* 5.3 below, she developed an alternative rubric to better address the task objectives and the construct of group discussion. In the alternative rubric, she added the following sub-scales to assess students 'use of functional language: 1) *expresses opinion* 2) *asks questions and responds to others* 3) *interrupts to express opinion/ask clarification questions* 4) *expresses agreement/disagreement* 5) *Nonverbal.*

Figure 5.3

Ray's Modified Group Discussion Rubric (Ray-Doc 3)

Adjunct Group Discussion Rubric

Topic:	5	4	3	2	1-0	Comments
good preparation and collaboration 5						
Explains opinion clearly with support and details 5						
good grammar and relevant vocabulary 5						
pronunciation and emphasis on important words 5						
1) expresses opinion 2						
2) asks questions and responds to others 2						
3) interrupts to express opinion/ask clarification questions 2						
4) expresses agreement/disagreement 2						
5) Nonverbal 2						

Mike implemented this belief by including criteria that reflect the relevant construct components in his Presentation Evaluation Sheet (*Figure* 4.2), explaining that the rubric gives *a schema on what are the most important elements of the students speaking competency."* Dave highlighted that his assessment practices–such as designing rubrics, selecting criteria, and providing feedback–are driven by his personal understanding of constructs based on solid theories and expert opinions, rather than institutional definitions.

5.3.2 Fair Assessment and grading are vital

Fair and impartial assessment was mentioned as one of the main characteristics of good assessment by many participants. However, teachers 'definitions of fair assessment were diverse; they referred to different characteristics of CA and used different criteria to ascertain the fairness of their CA. For example, Arman believed that student participation during the semester is more important than their performance in exams, so to make his assessment valid and fair he gives more weight to information he gleans through observation when assigning grades to his students, in the sense that he assigns bonus marks and is lenient with students who are actively engaged during the

semester. On the other hand, for some teachers, fairness is defined as equality, which they believed they can achieve by following the course grade breakdown and grading guidelines (e.g., Mike, Donna, Sally). These teachers assign specific marks for classroom participation in line with course guidelines and if a student does not do well in the formal graded assessment, they do not grade them leniently or give bonus marks. For example, Donna reported that participation for her is class attendance and when it comes to assigning grades, she does not consider student participation or being active in the classroom in awarding grades. Only in rare cases when a student is on the border line, she might bump the grade up a little.

Some teachers defined fair assessment as assigning an accurate grade which reflects a student's real ability or performance. Hudson strongly believed in valid or accurate assessment as a criterion for fair assessment. Following this belief, Hudson said he is strict in assessment and does his best to assign a valid score to each student in accord with their performance in the formal and informal assessments. Aly also reported that she is not lenient in her grading because she wants to give a valid evaluation of students' real ability. She emphasized she does her best to assign a grade that is reflective of students' real ability and /learning. On a similar note, Ray explained that good and fair assessment should show the real ability or /competence of students and differentiate them in terms of ability.

To ensure fairness, teachers referred to several characteristics of assessment as shown in Table 5.1 above.

A. Fair assessment is accurate and differentiates between students

A number of participants (n = 7) believed that fair assessment differentiates between students who learn and make progress and those who do not. They also emphasized the importance of assigning accurate grades that reflect students' real learning in ensuring fairness.

Ray emphasized that assessment should differentiate between students in terms of degree of ability or skill, noting that *if it doesn't, for example, distinguish between A, B and C, that's not a valid or reliable measure."* She explained that for assessment to be fair, rubrics should differentiate students in terms of their ability; if they do not, she adds criteria that target the different components of the relevant skill to validly assess it. Aly believed that good and fair assessment is not awarding high grades to all students; rather, fairness is giving a whole range of scores so that grades are normally distributed. She explained that *we need to have A's and B's and C's indeed, normal distribution of the grades."* If grades are not normally distributed, she noted that some adjustment should be made to the assessment.

Other participants also emphasized differentiation through accurate grading. Hudson defined fairness as differentiation based on ability, explaining that *you need to give some student more because they deserve it and they're ready for it, and some student less because they are not ready for it, or they don't need it."* Dave stated that distinguishing between students who put in effort and show progress and those who do not is *one of the functions of a test"* and is crucial in shaping his assessments. Mike stressed that good assessment must accurately measure student progress while maintaining fairness and *human accommodation."* Donna highlighted the importance of differentiating students through accurate grading based on effort and performance, criticizing lenient grading practices by noting that some instructors give 70 on everything regardless of quality, while others fail students who deserve it.

B. Fair assessment accounts for student effort

Some teachers believed that for assessment to be fair, teachers should take into account student effort in their assessment and grading. To incorporate this belief, some participants (n = 7) rely on observations and informal assessments in the class in addition to formal assessment results

when assigning course grades. They also award bonus marks to hard-working students, especially if they are on the borderline in terms of grades.

Arman explained that he tracks participation throughout the semester, noting that *I have my little note. I m going around, getting a little sign next to their name."* He added that engaged students may receive leniency if their answers are nearly correct, while less active students are marked more strictly. Ray emphasized effort and improvement, explaining that while she follows her rubric, *effort and improvement matter. If a student struggles but shows significant progress and determination, it influences my grading."* She noted that a student with persistent weaknesses who makes no effort will not receive the same consideration.

Other participants similarly highlighted the value of considering effort in assessment. Ryan noted that he prefers informal assessment because it provides a clearer picture of development than a single test, explaining that if a student fails a summative test but shows consistent effort and development, *I would definitely change the score and help the guy pass… enthusiasm matters."* Hudson explicitly connected effort with what he called the *Halo effect,"* rewarding classroom engagement and persistence, stating that *I m not marking you based only on what you produce. A genuine effort will be rewarded."* Sally and Donna both considered effort when students were on the borderline. Sally explained that she would give extra points only if a good student was on the verge of failing, not for small increments like 78 to 80. Donna admitted that despite strict guidelines, she uses participation marks to *bump up"* borderline grades if a student actively participates in class.

C. Fair assessment is unbiased and objective

Some participants (*n=5*) defined fair assessment as being unbiased and objective. They believed that teachers should use objective assessment tools and give equal opportunity to all

students. Regarding the general importance of unbiased assessment, Dave maintained that assessment should be fair and as unbiased as possible, explaining that he tries to be *100% objective"* and assesses students based on their product regardless of their background or individuality, noting that *I assess them for their work."* Donna stressed the importance of objective assessment, explaining that she tries to be impersonal in her assessment and grading, noting that a student's active participation during the semester does not affect their marks on regular work but only their participation marks.

In line with their beliefs about unbiased assessment, some participants (n = 5) reported that they try to use rubrics to ensure fairness in classroom assessments. Arman pointed to the use of objective tools for fair assessment, noting that the rubric acts like a roadmap, adding to assessment fairness and giving students a sense of their strengths and areas for improvement. Dave also noted that using a rubric contributes to fair assessment, explaining that *I always try to use the rubric as an important tool in order to make sure that my test is fair… regardless of who the student is, I just look at the rubric."*

Ray pointed out that she uses a rubric not only to better differentiate students in terms of speaking ability but also as an objective way to explain why students received specific marks. In a similar vein, Aly said that using rubrics adds to assessment transparency, enabling her to explain why a student received a specific mark.

Closely related to unbiased assessment, a few participants (n = 3) mentioned giving equal opportunity to all students as another characteristic of fair assessment in the classroom. Mike defined fairness as giving equal opportunity to all students, especially regarding assessment timing, extensions, or second chances, explaining that *when I give a student a chance to redo an assignment, presentation, or a test, the same opportunity will be given to all students."* Ray also considered equal opportunity an indicator of fairness; however, unlike Mike, she noted that if a student misses a test, she does not allow retakes, as giving a second chance to some would mean

depriving others. Donna similarly stressed the importance of equal opportunity, explaining that she withholds feedback on presentations until all students have finished so that those who present first are not at an unfair disadvantage, stating that *I hold back my comments until they're all finished because I don't want them to think they're at an unfair advantage going first."*

In contrast, Hudson defined fair assessment as equality of opportunity as well as the extent to which the student is able to take advantage of that opportunity, explaining that *fairness is not about an equal distribution of resources. It's about providing instruction and resources to students as you see fit as a professional… based on what you know they know, and what you know they can know."*

D. Fair assessment follows the same guidelines and grade breakdown for all students

Closely related to objective assessment, some teachers (n = 4) believed that using the same guidelines for grading all students contributes to fair assessment. For example, Mike pointed out that he distinguishes between students' performance in different sections of the course and assigns different grades without mixing them together or allowing himself to be affected by students' performance in other areas. He explained that he does not consider participation marks when grading formal summative assessments, noting that *if you are active in the classroom, doesn't mean that you are good in other skills too."*

Donna noted that she follows the guidelines and grade breakdown set for the course to ensure fair assessment and grading. Following the grade breakdown, she allocates only 15% of the course grade to participation and does not refer to information from her informal assessment, such as observation, when marking assignments or formal assessments, explaining that *if I see that a student is active during semester, that's not going to affect their regular work, it would only affect their participation."*

Other participants expressed similar views. Ray mentioned that she tries to follow the rubric and grading guidelines, assigning separate grades to student performance in different areas such as participation, quizzes, and final assessments, noting that this contributes to accuracy as well as fairness. She added that after summing up students' scores, she does not bump up their grades, as this might have negative consequences for students when they pass to higher levels, explaining that *if a student s final grade is 70… and the student does not have the ability to go a little bit further than this, we feel that this student will have a hard time next level."*

E. ***Fair Assessment is equitable***

Another characteristic mentioned by some participants (n = 6) as an indicator of fair assessment was equity. Teachers acknowledged that students do not have the same circumstances, and that this should be taken into account in classroom assessments and grading.

Ryan explained that he takes into account students' circumstances when assigning grades, noting that he may assess students who face greater challenges, such as international students working long hours, more leniently. He argued that *if two students produce work of the same quality, the international student may receive a higher score, acknowledging the additional challenges they face."*

Ray emphasized that equity is important for fair assessment, explaining that she tries to accommodate and support students with special needs. She noted that when students have learning disabilities, she shares concerns with coordinators to connect them with available support. For struggling students, she provides more support in the form of additional instruction and practice, and she may reduce the assignment load or make assessments easier while *never lower[ing] the standard."*

Other participants highlighted the importance of equity in their assessment practices too. Mike noted that he accommodates students' needs by giving weaker students a chance to resubmit assignments or present again. Sally reported that she would give extra time for submission and presentation if students request an extension a few days before the due date. Aly referred to the conflict between equity and equality as a challenge in her work, explaining that it is unfair to evaluate all students using the same criteria given their different proficiencies and past learning experiences, questioning *am I assessing them in the same way? I have to like I'm using the same rubrics?"*

Participants thus implemented their beliefs regarding equity in different ways, some by providing additional support to weaker students, others by increasing students' grades given student's circumstances, still others by giving students extra time to submit their work or another chance to resubmit, while others give bonus marks to those students on the verge of failing the course.

F. Fair assessment is transparent

Most of the teachers (n = 8) in this study considered transparency as a key criterion for assessment fairness. For this purpose, they share assessment tools and results with students and explain to them their instructional goals and expected learning objectives.

Dave explained that he provides rubrics at the start of the semester so students *know how they are going to be assessed and what criteria are important for completing a task."* Hudson stressed openness, stating that educators should be transparent about what is assessed and the assessment criteria, and that *you make it transparent to your students, and they prepare for it."*

Several teachers echoed this link between sharing criteria and fair assessment. Ray noted that evaluation sheets tell students *the way you re going to be assessed… these are things that you ve got to follow."* Sally connected transparency to trust, explaining that *if they dive in unknown*

water, they don t know what to expect. When you do that, they trust you more." Donna added that rubrics enhance fairness, noting that clear instructions and expectations ensure fairness and help students follow guidelines effectively. Mike also practices transparency by sharing the course outline and rubrics at the semester's start; for a task involving watching videos and taking notes (Appendix O), he gives students the questions beforehand to make expectations explicit and help them prepare. Ryan and Arman similarly emphasized that sharing assessment criteria contributes to fairness and helps students understand what is expected of them. Overall, participants consistently framed transparency, especially through sharing assessment tools, as a key way to ensure fairness, build trust, and clarify expectations for students.

5.3.3 Good assessment has positive washback

A number of participants (n = 6) stressed the importance of assessment consequences as one of the important characteristics of assessment and reported that assessment should have positive consequences. Mike highlighted the importance of positive consequences of classroom assessment, noting that the feedback he gets from assessment results helps him align assessments with students' needs to improve their oral communication skills. During the stimulated recall, he explained that if topics are not interesting or relevant, students will not retain the information, and therefore *the washback effect would be very minimum."* Similarly, Dave believed that assessment should have a positive washback effect on students' current and future performance, explaining that rubrics should be designed in a way that they have *washback effects on students learning and future performance."*

Other participants emphasized the role of positive washback too. Ray pointed to the importance of positive washback on students' learning and progress, noting that assessment is not done only for grading but should help students advance their learning, explaining that *every assessment and feedback, big or small, is meant to help them improve."* Hudson highlighted that for

him, the positive effect of classroom assessments on learning is evidence supporting their effectiveness, stating that effectiveness comes down to an argument based on evidence that the assessment *assesses what it's supposed to assess and it has led to learning."* Ally expressed her belief in the positive washback effect of assessment on student learning, noting that feedback provision should lead to filling gaps in student learning. Ray also mentioned that she provides more detailed feedback in the rubric and adds comments to increase the positive washback effect of her assessments.

The participant teachers reported using two strategies to implement their belief about the importance of positive washback: a) using assessment tools and criteria for teaching purposes, and b) using summative assessments for formative goals. In line with formative assessment goals, some teachers (n = 6) in this study reported that assessment tools should be used for teaching purposes in addition to being used as measurement tools. These teachers think that since the criteria in rubrics represent task objectives and instructional goals, they can be a useful guide for directing students toward these goals, preparing them for better performance and enhancing their learning.

For example, Dave believed that rubrics and their criteria can be used as guidelines for students to complete speaking tasks successfully, adding to the positive washback of classroom assessment. He used rubrics as an outline to teach different skills as well as how to complete speaking tasks, explaining that in role play instructions, he uses questions and criteria as *a kind of guideline… think of possible ways to start the conversation."* Ryan noted that the criteria on the rubrics he used reflect course content and act as a guide for successful communication in his class. He referred to the socio-cultural appropriateness descriptor in the rubric (Appendix Q) he was using for assessing students' progress in speaking, explaining that this criterion serves as a guide for appropriate communication in addition to being used as a measurement tool. The descriptor addressed commonly accepted social behaviors, conventions, and norms relevant to occupation-

specific contexts, including organization of content, non-verbal communication, and use of formal and informal language appropriate to the communication task.

Most teachers in this study expressed their support for the use of formative assessment, believing that classroom assessment should have positive consequences in terms of student learning. In keeping with this belief, some teachers (n = 5) reported that they use results of formal assessments for formative purposes to promote positive washback. Sally noted that after doing a formal graded assessment, she uses the results to understand students' needs and assist them to close gaps in their knowledge, explaining that *if students performance or the class average is below expectations, it signals problems with either the assessment or the teaching. You then need to identify the issue and adjust your teaching or tasks accordingly."* Ryan also reported that he uses the results of formal assessments for formative purposes by discussing exam results and conducting follow-up activities, stating that assessments help him analyze *what it is that I need to teach a little bit more or I need to elaborate on."*

Finally, some teachers (e.g., Aly, Ray, Dave, Hudson) reported that they provide comments in the rubric in addition to reporting marks to enhance students' learning. (See Section 4.1.5 on feedback provision in assessment procedures for examples.)

5.4 Teachers 'beliefs about methods of classroom assessment

5.4.1 Listening is integral to speaking assessment

Most of the teachers (*n=8*) in this study expressed a strong belief in integrated assessment in general and joint speaking and listening assessment in particular. Analysis of their reported and observed classroom practices also showed that most of them integrated different skills as well as speaking and listening. In most cases, listening tasks were accompanied by discussion, debate, role play, peer feedback, and/or questions and answers. There were only few cases where listening was assessed separately from speaking or writing. The participants referred to several reasons for

integrating speaking and listening in their classroom assessment such as authenticity, facilitating communication in class, and the nature of content-based courses and content reinforcement.

Sally believed that speaking and listening should be assessed together because they are used together in authentic contexts as well as in the classroom. She also noted that when these skills are blended, it is easier for students to communicate since listening provides prior knowledge and prepares them for speaking. Sally's belief in integrated assessment is demonstrated in the rubrics she uses for assessing students' communication. For example, in the rubric for a discussion seminar task (Appendix P), listening, reading, and speaking are assessed together. The rubric includes criteria related to reading and listening comprehension, such as "comprehension of main ideas" and "connection and support of ideas," alongside speaking skills like "summary and interpretation," "fluency and coherence," "discussion strategies," "language use," and "clarity." Students are asked to read two texts and listen to two texts on similar topics, then discuss them in groups of 4 to –5 students, identifying main ideas, making connections, and using appropriate discussion strategies.

In this respect, Ryan rejected discrete-point testing in favor of an integrated approach, explaining that *since integrating speaking and listening represents authentic language use, I teach and assess them together."* He uses listening prompts for speaking assessment tasks such as interviews and role plays based on videos, using listening as input for activities with information acquired from the listening part.

Other participants similarly emphasized integrated assessment. Aly supported assessing speaking and listening together, noting that listening gives students the background and context needed for higher-level discussion. In her courses on Canadian culture, she frequently provides a short video introduction to each topic as a *springboard"* for discussion, ensuring students receive multiple types of input before discussion. Ray argued that speaking and listening are a continuation of each other, explaining that *any vocabulary, any pieces of information that is received from the*

listening part is transferred into speaking activities." Mike reported that he combines listening with speaking in his assessment practices as it fosters communication. During one observed class, he asked students to watch videos about multiculturalism and take notes of key ideas, then prepared questions targeting students' understanding before asking them to argue for or against the ideas expressed in the videos (Appendix O), noting that *listening isn't a primary assessment tool in my class but is used for formative assessment, it fosters discussions on Canadian culture."* Donna expressed her belief in assessing skills together, explaining that in her course listening reinforces content rather than being a separate skill, and she uses videos with subtitles as *a reinforcement in a different format"* to help students gain knowledge and prepare for speaking tasks.

Implicit assessment of listening skills

The results of interview and document analysis point to the dominance of the speaking part in integrated assessments. According to some participants (e.g., Ray, Ryan, Hudson, Mike, Dave, Ally), speaking and language production are predominantly emphasized in their assessments, so listening is not assessed in depth. The assessment of listening seems often peripheral to the assessment of speaking with listening included as a supporting skill for primarily speaking tasks. Consequently, it is either not evaluated at all or is assessed implicitly and indirectly through speaking.

Aly pointed out that, due to the nature of content-based courses, listening is typically embedded in speaking tasks, making it hard to recognize students' listening needs. She explained that she uses videos as input material in speaking task in which students are asked to identify, discuss, and write down ideas, which gives her only a partial sense of their listening needs, noting that *"it's hard to know if students are struggling with listening because we don't test it directly."*

Hudson pointed out that in content-based courses, although students are required to listen, identify the structure of information, and use it in their speaking and discussions, listening is not assessed formally. A look at the rubrics he used for assessing student presentations (*Figure* 5.2) and debates (*Figure* 4.4) reveals that listening is assessed as part of speaking and receives little weight if at all. Most criteria in the presentation rubric relate to knowledge of topic, body language, comprehensibility, prosodic features, and language use, with no criterion for listening comprehension. Similarly, his debate evaluation sheet focuses on evidence, claim, rebuttal, warrant, and citation, with no listening-related criteria. Hudson explained that he does not explicitly teach or formally assess listening aside from ensuring students understand him and their classmates, noting that *"the focus is on recognizing structure rather than direct listening instruction."*

Other participants also reported that listening is not directly assessed. Ray explained that in most tasks in her class, listening is embedded in speaking tasks and is not assessed directly. In her Oral Summary of Presentation task (see *Figure* 4.3), the rubric contains no criteria related to listening skills; all criteria pertain to organization, content, fluency, pronunciation, grammar, and vocabulary. In her group discussion task (*Figure* 4.5), she noted that listening is assessed informally, explaining that when students respond to each other with phrases like *"I agree," "I don't understand,"* or *"Could you repeat that?"* their comprehension is assumed. Donna stated that she indirectly assesses students' listening comprehension through their speaking, noting that she selects videos that are slow, clearly spoken, and often subtitled, making them easily understandable so that *"students at this level generally have no comprehension difficulties."* Ryan**,** Mike**,** and Sally also reported that listening is assessed indirectly through speaking rather than directly for comprehension.

5.4.2 Use of multiple assessment methods and informal assessments

Formal assessments, including rubrics, are planned and are either mandated by the school or designed or selected by the teachers. They are mostly graded and count toward the students' final grades. Teachers use the results of formal assessments, for formative purposes, summative purposes or both. However, as noted above, most of the teachers in this study reported that they use the results of formal assessment for formative purposes in addition to grading.

As noted above, in addition to formal assessments, teachers reported using informal observation of students' behavior, performance, participation, effort, needs, progress and personality. Informal assessments are either graded or ungraded and their results are used for formative purposes as well as when assigning grades.

Almost all the teachers (n= 9) in this study endorsed the use of both formal and informal assessments for evaluating students 'performance. They often use informal assessments to bolster formal assessments and to better understand students' learning, progress, needs, and/engagement, understand task difficulty, adapt assessment criteria, provide further instruction, reduce anxiety, fine- tune their feedback and assign accurate and fair grades. Some participants (n= 6) also noted that informal assessments are more useful for implementing their assessment beliefs in the classroom, especially in cases where school policies and school mandated assessments are restrictive and can prevent teachers from enacting their beliefs in the classroom.

Dave stressed the importance of using both formal and informal assessment for assessing his students' performance on speaking tasks. For example, when assessing panel discussions or group debates, he does not merely rely on rubrics or formal assessments for checking and evaluating each group's discussion; he also uses informal observation of students 'performance and provides feedback accordingly. Mike, likewise, expressed a belief in elective assessment for ensuring

effective assessment. He reported that he considers both his informal assessment and formal assessments of course content mandated by the school. He explained that he tries to keep a balance between course content or objectives and his informal assessment practices. Based on observing students 'responses to assessments, he might make changes to the criteria in the rubrics and provide further instruction to meet students' needs.

Ryan reported that he regularly uses informal assessment in addition to formal assessment to improve the quality of his teaching and students' learning, explaining that he asks quick questions and checks students' facial expressions to gauge their understanding, noting that *"I get some sort of impression by their facial expression that some sort of listening or speaking activity was vague."* Donna reported that besides rubrics, she uses observation to fine-tune her teaching to students' needs, criticizing her school's emphasis on formal assessments and noting that *"if I had to construct this course on my own, you wouldn't see as many assignments. You would see more informal feedback assignments rather than formal assessments."*

Sally pointed to the importance of using both formal and informal assessment in her classroom and thought that using both is a great combination to understand students' performance and needs: "*I think the only way to make formal assessment better is to pair it with informal assessment"* (Sally). She mentioned that she uses observation of students 'performance to understand task difficulty, tweak tasks, and teach accordingly. As an example, she described how in listening tasks, she observes student performance to evaluate task difficulty and align them to students' needs: "*If they grasp the main idea after one or two plays, they're improving; needing three or four plays shows the material is still challenging*."

Arman, Hudson and Ray reported that in addition to formal assessment, they consider informal assessment when evaluating students 'performance and assigning grades at the end of the semester. For example, Arman noted that informal assessments during the semester influence his

being lenient or severe in grading students. He explained that since he is in direct contact with the students during the semester, his observations give him a better picture of their performance. In one of his classes that I observed, before the students watched a video for pair discussion, Arman wrote a list of difficult words and explained their meaning. During the SR, Arman explained that as a result of experience and informal observations, he could predict which words they might find challenging. Hudson explained that he makes mental notes of students' participation and considers them when assigning participation marks, stating that *"the results of my informal assessment count toward their participation marks."* Aly also stressed the importance of informal assessment besides formal assessment in her CAs to better understand her students' performance and engagement. As an example, she described how in a presentation task, she gives more weight to students' engagement with the task and time spent for preparation than the grades. Thus, although there is no criterion for evaluating preparation in her rubric for assessing a presentation task (see *Figure* 4.8), Aly reported using observation to evaluate students' effort and preparation through the way they present, the amount of information they have, their behavior and the way they answer questions. "*valuing effort over grammar or points.*"

Informal assessment as a space for enacting teacher beliefs

A number of participants (*n=6*) reported that informal assessment lends itself better to implementing their beliefs in the classroom than formal assessment. For example, Ryan criticized the formal assessments mandated by his school, complaining that sometimes they are not reflective of his beliefs regarding good assessment. Therefore, he trusts his own informal assessments as they involve more frequent and extended contact with students compared to one-time tests, explaining that *"I prefer informal assessment because it draws on multiple observations and experiences, giving a better sense of student development than a single summative test."*

Similarly, Sally stated that she implements her assessment beliefs via informal assessments. For example, in line with her beliefs, she pays less attention to grading and shows more leniency in her informal assessments and marking in-class activities, whereas in formal assessment she has to be more stringent given that the tools and criteria are designed and mandated by the school, leaving less leeway for teachers. She explained that she uses formal assessments to assign grades and meet school requirements, but for learning purposes and providing feedback she uses informal formative assessments.

Other participants similarly emphasized the value of informal assessment for enacting their beliefs. Ray and Mike reported that in cases where curriculum or school assessment policies are restrictive, they rely on informal assessment. Ray explained that through observation of student behavior she can gain a better understanding of students' learning patterns and needs, stating that *"I trust my observation for assessing students… you get to understand student's patterns of learning."* She noted that some rubrics assigned by the school fail to capture all task objectives, so she adds comments and sub-criteria to them based on her observations. She believed that informal assessment is more valid than formal assessment because formal assessments might make students anxious, impinging on their performance and giving an inaccurate picture of their real abilities, explaining that *"observing them in real and non-assessment situations definitely gives the teacher a better opportunity to give them feedback in real time."* Arman reported that he enacts his belief regarding the importance of student engagement via his informal assessments and observations rather than summative formal assessments. During the stimulated recall, he explained that he regularly uses observation, and if he notices a student is not engaged, he gives them clues to elicit responses, adding that if a student makes effort and participates well in informal assessments, he tends to be lenient when grading them.

These examples suggest that teachers are influenced by their assessment beliefs in using both formal and informal assessments in the classroom., They seem to use informal assessment, besides formal assessment, to ensure the appropriateness of assessment tasks and tools, understand students 'performance, progress and needs, provide appropriate assistance, and tune their feedback to students' needs. In the case of contextual constraints, they seem to resort to informal assessments to enact their beliefs in CA.

All in all, based on the results reported in chapter 4 regarding teachers' use of multiple assessment methods and the results reported in this chapter about teachers' beliefs and practices concerning the use of different assessment methods and the importance of formative assessment , classroom assessment practices can be seen as constituting a continuum ranging from formal to informal, with the results of all assessments used for formative and/or summative purposes as illustrated in *Figure* 5.4 below. In the case of formal assessment, as reported in section 5.3.3 about the importance of positive washback, some teachers use its results for formative purposes in addition to grading. Similarly, in the case of informal assessment, the teachers use its results mainly for formative purposes as well as for assigning grades. As such, dichotomizing CA in terms of assessment goals does not seem to be well supported by these findings.

Figure 5.4

Classroom Assessment Methods and Goals



Overall, the results in this chapter suggest that there is a close alignment between teachers' stated beliefs and their classroom practices in assessing oral communication. In their practices, the teachers are informed by their general beliefs about the teaching and assessment of L2 oral communication, as well as their beliefs about assessment goals, the characteristics of good assessment, and assessment methods. They implement their beliefs mainly through the selection of appropriate assessment tools and criteria, employing appropriate assessment methods, using assessment results formatively, and providing feedback. However, sometimes contextual factors may moderate the implementation of their beliefs in their assessment practices. In such cases, teachers may resort to informal assessment to circumvent contextual constraints or create a balance between their beliefs, contextual constraints, and their practices. The effect of different contextual factors on the relationship between teachers' assessment beliefs and practices is discussed in detail in chapter 7.

CHAPTER 6: THE RELATIONSHIP BETWEEN TEACHERS' ASSESSMENT LITERACY AND PRACTICES

This chapter addresses research question 3 which examines the relationship between teachers' assessment literacy and their practices in assessing oral communication in the ESL classroom. To answer this question, data from the TAL questionnaire and the interviews about teachers 'past education, theoretical knowledge, and teaching, learning and assessment experiences were used. First results from the TAL questionnaire are presented, then results concerning teachers' past education, theoretical knowledge, and learning, teaching and assessment experiences are presented and discussed.

6.1. Findings from the TAL questionnaire

The analysis of data from the TAL questionnaire (Appendix C) shows that participants in this study have adequate self-perceived AL in different areas with the mean score ranging from 2.18 to 2.94 out of 3. Teachers 'overall self-reported literacy, based on the items in the questionnaire, shows that their average AL is 2.60 on a scale of zero to three. The results concerning teachers' self-reported assessment knowledge are displayed in *Figure* 6.1.

Figure 6.1

Teachers' Self-reported Assessment Knowledge in Different Areas

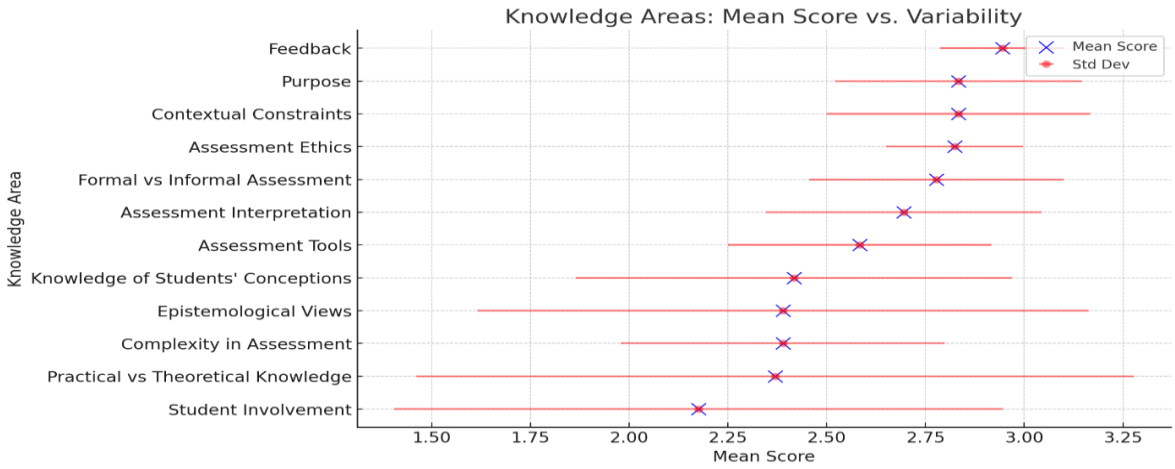


Figure 6.1 shows the mean score for each knowledge area along with its standard deviation, indicating the relative strength of each knowledge area and the variation in responses. A comparison across the different knowledge areas shows that the majority of participants reported higher level of knowledge about feedback (*M*= 2.94), contextual constraints (*M*= 2.83), purposes of assessment (*M*= 2.83), assessment ethics (*M*= 2.82) and formal vs informal assessment (*M*=2.78). Low variation of these forms of knowledge (*SD*=.25) indicates strong consensus among teachers in having assessment knowledge in these areas.

On the other hand, they reported having less knowledge regarding student involvement (m= 2.18), practical vs theoretical knowledge (m= 2.37), epistemological views (m= 2.39), complexity in assessment (m=2.39) and knowledge of students' conception (m=2.42). High variation of these forms of knowledge (SD=.71) suggests mixed opinions among participants.

Overall, the results suggest that most participating teachers felt more confident about the practical aspects of assessment but struggled with theoretical aspects. Comparison across participants showed that Donna reported the lowest level of knowledge overall, likely due to a lack of formal training and not taking any assessment courses in her graduate studies.

As for the application of their knowledge, the results in the previous chapters (interviews, SR, observation and documents) revealed that the participants seem to be aware of different forms of assessment knowledge and implement them in their practices. For instance, most of the participants reported higher level of knowledge about different purposes of assessment in the questionnaire. Accordingly, when describing their practices in the interviews, they reported using assessment for different purposes. As displayed in *Figure* 5.4, in their formal assessments, the teachers utilize assessment results not only for grading but also for formative purposes. Similarly, in their informal assessments, teachers reported using assessment results primarily for formative objectives while also ensuring accurate grading. In tune with their knowledge about the purposes of

assessment, most of the participants (n=8) emphasized the importance of formative assessment and mentioned various activities they carry out to promote students learning and enhance their teaching. Likewise, they referred to the importance of diagnostic assessment (n=8) and noted that, in their practices, they try to diagnose students' needs and provide appropriate assistance to help students advance their learning.

Even in the cases where teachers reported having moderate knowledge in the questionnaire, they seemed to still use this knowledge in their practices. For example, compared to other forms of knowledge in the questionnaire, some participants (e.g., Ray, Ally, Mike, Raymond, Sally) reported having less knowledge regarding peer assessment and involving students in assessment. However, in their reported practices during the interviews they seemed to be familiar with these areas and to incorporate peer assessment in their practices. Mike and Sally, for example, reported that they take notes on common errors during student presentations and encourage the class to provide feedback to each other on word choice, structure, and citations. Mike and Ray ask students to use rubric criteria to assess their peers' presentations. Mike asks his students to take notes and evaluate their peers' presentations based on task guidelines and rubric criteria, while Ray asks her students to listen, take notes, discuss the rubric, and provide feedback to their peers accordingly. Similarly, Aly asks her students to comment on each other's presentations, but she reviews the comments to ensure accuracy and prevent misleading feedback.

Even when the teachers' self-reported low levels of awareness about practical vs theoretical knowledge and the application of theoretical knowledge in assessment practices on the questionnaire, they seemed to enact it in their assessment practices. For instance, several participants (Ryan, Ray, Sally, and Mike) mentioned the concept of the ZPD and reported that they are informed by it in their assessments. Hudson referred to *construct validity* noting that he considers it in his actual practices and assessment rubrics by aligning assessment criteria in the

rubric with the target construct. In this respect, he referred to the rubrics he used for a Debate Task (*Figure* 4.4) and a Presentation Task (*Figure* 5.2) which align with this principle. Meanwhile, Dave and others referred to communicative competence and language functions, highlighting their consideration of various forms of competence in their assessments.

In contrast, the areas that some participants (e.g., Donna and Dave) reported having no knowledge or little knowledge about in the questionnaire were not incorporated or considered in their assessment practices as revealed by the interviews, observations, SRs and document analysis. For example, Donna reported the lowest level of knowledge (less than 1 out of 3) regarding theoretical vs practical knowledge, student involvement in assessments), student conceptions, and epistemological views). When describing her assessment practices, she did not refer to any of these forms of knowledge. Similarly, Dave reported little knowledge regarding involving students in assessments (1.66) and in the interviews he did not mention any examples of using peer assessment and no example was observed.

All in all, the results show that there is consistency between teachers' self-reported AL in the questionnaire and their classroom assessment practices. Their practices seem to be informed by their AL. In order to gain a deeper understanding of the relationship between teachers' AL and their classroom practices, information gained via the TAL questionnaire, together with information from the interviews about teachers 'past education, theoretical knowledge, L2 learning experience, and teaching and assessment experiences are used. The following sections describe the relationship between these elements and the participants' assessment practices.

6.2. The relationship between teacher past education, theoretical knowledge and classroom practices

According to most participants (n = 7), their past education and their theoretical knowledge are important factors in shaping their classroom assessments. Dave, for instance, reported that his theoretical knowledge and conceptions of different constructs are a strong force in shaping his assessment practices. For example, in assessing coherence and cohesion, he explained that he relies on definitions presented by experts such as Halliday and Hassan rather than those used by institutions like IELTS, stating that *"when I describe and discuss cohesion and coherence, I wouldn't refer or use the definitions as set out by IELTS, I would refer to Halliday and Hassan."*

Similarly, Ryan stressed the role of his theoretical knowledge as an important factor in shaping his practice. While acknowledging the importance of teaching experience and student input, he asserted that he mostly relies on his theoretical knowledge in his assessment practices. For example, he explained that he applies his knowledge of task-based language teaching by selecting authentic tasks with information gaps so that speakers can interact and perform specific speech acts, noting that he draws on Ellis (2003) regarding task features. He also noted that when he learns about a new approach or theory, he experiments with it in his classes, stating that *"I read a paper for dynamic written corrective feedback two months ago, and then I implemented it in the course."* In applying his knowledge of SCT and ZPD, Dave provides scaffolding by moving from more implicit to more explicit feedback as needed.

Other participants similarly emphasized the role of theoretical knowledge in shaping their practices. Sally referred to the importance of her past education and her belief in the ZPD, noting that *"one thing that stayed with me since my graduate degree, this ZPD, and sociocultural theory of mind changed my way of teaching, and helped me a lot in assessment."* Mike highlighted the importance of studying assessment theories during his PhD, noting that an assessment course gave

him insight into how to evaluate the effectiveness of his classroom assessments and how to *"defend your assessment criteria and procedures."* He also noted that during CELTA and other graduate courses, teachers were considered scaffolders who diagnose students' needs, and he now plays the role of a *"game keeper or coach"* by giving clues to elicit answers and engaging students in self-discovery.

Ray noted that based on her knowledge of Krashen's i+1 hypothesis and the ZPD, she adapts task difficulty to students' level, simplifies rubrics, and chooses easier topics for oral communication tasks. She also pays particular attention to consistency between assessments and the construct being evaluated, criticizing a Group Discussion Rubric (*Figure* 4.5) assigned by her school for missing pragmatic aspects such as appropriate speech act use. In response, she added sub-criteria (i.e., expressing opinion, asking questions, interrupting, expressing agreement/disagreement, and nonverbal behavior) in a modified version of the rubric (*Figure* 5.3). She also noted that her limited attention to grading and her prioritization of learning and progress stem from her PhD studies, explaining that *"seminars and courses have emphasized tracking student progress through formative assessments, which has influenced my approach."*

The results of data analysis showed that some teachers (n=5) are guided by their knowledge of communicative competence and different forms of knowledge required for successful communication. This theoretical knowledge is manifested in their attention to different forms of knowledge in their assessment practices and the tools they use for assessing oral communication. For example, they unanimously agreed that pragmatic competence and sociocultural awareness are important aspects of assessment, so they include them in their rubrics for assessing speaking.

Dave reported that his practices have been influenced by his knowledge of communicative competence and different models proposed for it. He explained that when assessing different skills, he considers different forms of knowledge a good language user should have, implementing his

knowledge even if it is not reflected in the course outline. For example, in an academic seminar presentation task, he added appropriate use of *"how to ask the question politely, how to criticize and respond to a criticism in a polite way"* to his teaching and assessment, as well as using transitions when moving from one subject to another.

Similarly, Ryan expressed a strong belief in communicative competence and sociocultural appropriateness as an important component of it. In concert with this belief, he thought that presentations are a good assessment task as they tap into students' appropriate verbal and non-verbal skills. Following this belief, he teaches and assesses the use of appropriate speech acts. The rubric he shared for assessing students' progress (*Figure* 4.1) indicates that *"sociocultural appropriateness"*–the ability to use socially acceptable non-verbal and verbal communication according to the situation–is one of the key criteria he uses for assessing students' speaking ability.

Hudson expressed an interest in Halliday's functions of language, which he had studied, and noted that he provides feedback on three aspects of speaking: linguistic, paralinguistic, and debate protocol. In the evaluation sheet he used for assessing a debate task (*Figure* 4.4), these three aspects are addressed, though the focus is on debate protocol. He also noted that he is informed by his studies on Toulmin's model of argumentation, and in the Debate Task he highlighted the importance of evidence, claim, warrant, and rebuttal, which are reflected in one of the assessment sheets he uses (*Figure* 4.4). Based on his knowledge of construct validity, he noted that both in the Debate Task (*Figure* 4.4) and the Presentation Task (*Figure* 5.2), the rubrics he uses address the intended construct.

Some participants (*n= 5*) subscribed to the principle in the *communicative approach of* prioritizing meaning over accuracy in their teaching and assessment. As a result, in the rubrics they used for assessing speaking, more weight was placed on getting meaning across rather than accuracy. In terms of accuracy, they mainly considered and assessed global grammatical errors

which might interfere with meaning. For example, Sally reported that she now believes that "*how well you can put your point across*" and "*that you can communicate and say what you want is the important thing, not focusing on very small points*" or "*those small, teeny tiny things*." In this regard, Arman said that he was educated following the communicative approach, in which the focus was on meaning and communication rather than form. He said that in his current assessment practices, he does not put much weight on accuracy; he just tries to create a situation in which students interact with each other. Following his interest in the communicative approach, Mike also reported that he only assesses global errors and provides feedback on errors that interfere with the conveyance of meaning.

In short, the above examples reveal the importance of teachers 'past education and theoretical knowledge in shaping their assessment beliefs as well as practices.

6.3 The relationship between teacher learning experiences and classroom practices

Some participants (*n=5*) referred to their own language learning experiences as another influential factor in shaping their assessment practices. This factor, like their teaching experience, seems to equip teachers with hands-on knowledge in assessment matters in the classroom. This type of knowledge seems to enable teachers to anticipate students 'pitfalls and sources of errors, so they can provide appropriate and effective assistance.

Ryan said that he can anticipate his students' challenges and needs because of his past L2 learning experience, allowing him to design more appropriate tasks in accord with students' needs. He explained that he sometimes anticipates what problems students might encounter and prepares what he calls a *"cheat sheet"* accordingly. He also noted that he teaches students metacognitive and other strategies for improving speaking, such as memorizing words and thinking aloud in English,

explaining that *"I always encourage them to think in English—an approach that helped me as a student."*

Similarly, Aly referred to her learning experience as a source of inspiration for her practices regarding using different activities that build on each other to prepare students for academic tasks. She explained that she conducts assessment step by step and prepares students for target tasks. For example, in a critical thinking and discussion task, because her students were not able to argue or engage in critical thinking, she asked them to express their general opinion and learn how to express it before proceeding to engage in critical thinking. She noted that based on her L2 learning experience, she sees repetition as *"an essential step toward forming opinions and critical thinking,"* so she encourages students to start using academic concepts, even through simple repetition, as a foundation.

Other participants similarly acknowledged the influence of their learning experiences. Dave reported that he is guided by his past learning experiences when identifying the sources of students' learning difficulties and providing feedback, noting that his L2 learning experience helps him understand issues like pronunciation, palatalization, and phoneme use; he stresses to students that *"a foreign accent isn't a problem—only not being understood is."* Sally also noted that she can predict students' pronunciation difficulties and what keywords or expressions students need in order to talk about a specific topic based on her own learning experiences. Donna acknowledged the influence of her experience learning a second language on her classroom practices, explaining that learning another language and living abroad gave her intercultural awareness which has greatly strengthened her work as a second language instructor.

These examples highlight the importance of teachers 'past L2 learning experiences as an important factor contributing to the formation of their knowledge and beliefs, which in turn seem to

influence their classroom practices. It enables them to diagnose and predict their students ’needs and take the necessary actions to provide appropriate assistance.

6.4. The Relationships between teaching and assessment experiences and classroom practices

The assessment practices of most of the participating teachers (*n=9*) in this study appear to be informed by their teaching and assessment experiences. They seem to resort to their experience as a reference and a practical form of knowledge to conduct effective assessment in the classroom. Among all the factors (i.e., L2 learning experience, past education/training, theoretical knowledge, assessment/teaching experience in the classroom) that they reported have shaped their assessment practices, teachers referred to their assessment experiences in the classroom as the main factor forming and reshaping their practices. Also, it seems that the theoretical assessment knowledge they acquired during their past education and training courses is redefined in light of their assessment experience in the classroom context. It is through experience that they come to a better understanding of students ’needs, the amount of assistance students require, the appropriateness of their assessment tools, student diversity, and meso contextual factors such as the curriculum and school policies. Furthermore, through trial-and-error in the classroom they can figure out useful assessment strategies and procedures.

Aly emphasized the superior role of teaching experience over past education, learning experience, and theoretical knowledge for effective classroom assessment on the grounds that each classroom context is unique. She pointed to the dynamic nature of the classroom and explained that teachers learn by doing, with experience giving them a better understanding of the classroom context, learners' needs, and what works best for each class. According to Aly, sometimes there is a gap between her past education, the theories she studied, and her practices, so she relies more on her practical experience over the years as a trustworthy source for effective assessment, noting

that *they cannot prepare you for the workplace because you will find a completely different reality."*

Similarly, Dave acknowledged that his past education, training, and teaching and assessment experiences all have had an impact on his classroom assessment practices, but he attributed most of his practices to his teaching and assessment experiences, as it is in the actual classroom (in contact with students and their needs) that theories get their actual, practical shape. He underlined the importance of assessment experience for making solid assessment decisions, explaining that he can gauge the average level of students and the appropriateness of task difficulty based on his past experience. He recalled that as a novice teacher, he could not recognize the complexity of a task and match it to students' level, stating that *as a new teacher, I didn't have that gut feeling. I made a lot of mistakes, so I gave something which was too challenging or too easy for students. So that particular type of gut feeling comes with experience."* As an example, he said that based on his experience, if his students' level is low for note-taking, he will use shorter videos with simple language to make note-taking easier. He added that it is years of assessment and teaching experience that guide him to diagnose students' problems and assist them accordingly, noting that *having the scars on my own arms, teaching in different contexts, I know where the problem areas are."*

Sally reported that her teaching experience has had the greatest impact on her assessment beliefs and practices. She explained that she has learned a lot from the experience of teaching and assessment in the classroom, noting that past training and education sometimes do not work in the classroom. She pointed out that *while training influenced me, my real growth came from classroom experience. Practicing what I learned showed me which methods work and which don t, helping me refine my assessment practices."*

Ryan contended that it is important that teachers have adequate teaching experience for providing effective feedback. He believed that because of his past teaching experience he can

distinguish errors from mistakes and provide feedback accordingly, providing direct feedback when a student lacks knowledge and indirect feedback or clues when it is merely a slip. He explained that *with experience, I can usually tell if an error is due to lack of knowledge or just a slip. I see this as knowledge, not intuition."* He added that because of his teaching experience he can evaluate students' performance in speaking and provide feedback intuitively without a rubric, relying on his head notes, mental scale, or intuition.

Donna pointed out that early in her teaching career she was not sure, but now, after many years of teaching EAP, she knows what her students in ESL courses need to prepare for their core courses. She noted that her experience with target language use needs for academic courses guides her assessment practices, explaining that *when I started, I only had a vague idea of what students needed for university success. Now, as a university professor, I know exactly what they need and teach the course accordingly."*

In the case of contradictions between theory and practice, some teachers (e.g., Aly, Mike, Ray) seem to rely on their experience to achieve a compromise. Aly pointed to the key role of her assessment experience in the classroom and said that there is a gap between theory and practice due to diverse context requirements, so she uses her assessment experience to balance the contradiction and ensure the effectiveness of her assessment. She used the metaphor of *driving,* that "*the more you drive, the better you drive*" to underline the importance of teachers 'assessment experience in the classroom for effective assessment. In a similar vein, Mike said he *"tries to create a balance between course syllabus and his own formative assessment beliefs"* by using his past and current experience in the classroom context. Ray also referred to the drawbacks of syllabus and assessment tools which are not in line with her beliefs about effective assessment. She explained that she uses her assessment experience to create a balance between course syllabus and her beliefs by adding comments or sub-criteria to the rubrics mandated by her school. Ray explained that she tries to

create a balance between a syllabus that emphasizes grading and her belief concerning students ' progress which is more important in her opinion.

All these examples of contradiction and the way teachers deal with them demonstrate the role of teachers 'assessment and teaching experience as a key determining factor in forming, reshaping and amending their theoretical knowledge, and, accordingly, their TAL. Furthermore, as noted in the section on CA procedure, many teachers reported that they regularly reflect on their practices and rely on trial-and-error to improve their practices and amend their previously held beliefs and AL. This suggests that teachers 'experience plays a pivotal role in modifying their 'AL and optimizing their assessment practices. All in all, by comparing all the three factors and their effect on classroom practices, it seems that teachingexperience in the classroom plays a central role according to the participants. The key role of teaching experience in reshaping TAL can be seen in the reasons reported by the participants above: the dynamic nature of the classroom, the various contextual factors involved in the classroom, the contradictions they encounter in classroom practices, their reflection, and trial-and-error.

Taken together, it seems that teachers 'past and current teaching and assessment experience as a form of practical knowledge plays a critical role in their assessment practices and decisions. They use it as a reliable source for ascertaining effective assessment in their classes. It appears that classroom experience as an asset adds a new aspect to teachers' theoretical assessment knowledge, reshaping and improving it to make it applicable in the classroom. In other words, in the classroom teachers gain a deeper understanding of various contextual factors (i.e., *constructive factors*) such as students 'needs, course and task objectives, and task complexity, that shape their assessment beliefs and literacy, thus developing a revised version of their theoretical AL which is more practical. This has implications for defining TAL and TAL models in which teachers 'own experience seems to go

unnoticed. Teachers' actual teaching and assessment experience can be an important piece of the puzzle for defining TAL and its relationship with beliefs and classroom practices. (See *Figure 8.1*)

CHAPTER 7: CONTEXTUAL FACTORS THAT INFLUENCE TEACHERS' ASSESSMENT LITERACY, BELIEFS AND PRACTICES

This chapter reports findings related to research question 4 regarding the effect of different contextual factors on the implementation of teachers' AL and beliefs. During the interviews, the participants were asked about different macro and micro contextual factors that can affect their assessment practices as well as the implementation of their assessment beliefs and AL in classroom assessment. The analyses identified seven such factors as listed in Table 7.1.

These factors, which can be described as constructive or detrimental, operate at the micro level (student needs, student performance, student motivation, students' conceptions of assessment, assessment timing, and graded vs ungraded assessment) and the meso level (school policies, teacher's freedom in selecting or designing assessments, content-based syllabus, and curricula). As noted in the section on the procedure of CA, teachers' practices seem to be positively influenced by constructive factors such as student needs, task objectives, instructional goals, student input, and student performance. On the other hand, detrimental factors such as school policies, inflexible curricula, students' different assessment views, content-based approaches, students' lack of motivation, the stakes associated with the assessment, and the timing of the assessment seem to impede effective assessments in the classroom. (See Figure 8.1)

Table 7.1

Contextual Factors Influencing the Implementation of Teachers' Beliefs and AL

- School policies (n=8)
- Teacher autonomy in selecting or designing assessments (n=9)
- Content-based syllabus (n=5)
- Student motivation (n=8)
- Students' conceptions of assessment (n=7)
- Assessment timing (n=7)
- Graded vs. ungraded assessment (n=5)

7.1 School policies

Many participants (*n* = 8) pointed to the detrimental role of assessment guidelines and criteria mandated by their institutions, which can constrain teachers from implementing their beliefs and AL in their assessment practices. According to the participants, detrimental policies include restrictive curricula, assessment criteria and grading breakdown, all set by the school.

Sally reported that some of the theories she studied during her MA cannot be applied in the classroom due the program's policies. She explained that due to a set curriculum, curriculum developers do not incorporate theories, and teachers have less freedom to implement their knowledge and beliefs. She noted that she tends to be more lenient in informal assessments than in formal assessments where she follows school guidelines and policies, stating that *when I do informal assessment, I'm more lenient. But when I'm doing formal assessments, I'm totally a different person, I follow guidelines and rubrics."*

Similarly, Ryan criticized some of his school's policies as being detrimental to good assessment. He noted that sometimes the curriculum is restrictive when, based on his observation of

students' needs, he feels some skills should be practiced more often, but he has to follow the curriculum to meet course objectives. He explained that due to time constraints and school policies, he does not provide feedback on micro-level and linguistic errors; instead, he focuses on course content and fluency in assessing speaking. He complained that his school does not allow multiple graded assessments during the semester, stating that *the main issue is that we re not allowed to have graded summative assessments or quizzes during the semester; only one final assessment counts for the whole score. I don t think this is effective, because without ongoing assessments I can t motivate students to study or review the material I teach."* He also noted that retention rate is important to his school in evaluating teachers, which affects his assessments and makes him more lenient in grading.

Other participants similarly expressed concerns about restrictive school policies. Ray criticized the rigid syllabus set by her institution, noting that the school has used the same syllabus for different groups of students over the years despite their different needs. She expressed her belief in authentic tasks such as role play, but noted that the rubric assigned by the school does not include important criteria for assessing such tasks, focusing instead on language and pronunciation rather than whether students can actually handle real situations. Dave and Hudson noted that rubrics are sometimes mandated by the school, so teachers have to follow the criteria. However, they noted that if there is any room for changing the assessments, they will do so. Dave stated that *My personal philosophy of learning combines a social and critical understanding of learning, shaped by study and experience. I try to apply this philosophy in rubrics whenever possible, but sometimes college policies and fixed syllabi limit its implementation."* Donna pointed to the detrimental role of her school curriculum in evaluating students' progress rather than achievement or proficiency, explaining that marks and grade weights are fixed, and she cannot adjust grades due to school

policy. She noted that *more one-on-one interaction would make assessing skill levels clearer, but the course structure doesn t allow it."*

In sum, these examples show that restrictive school assessment policies can prevent teachers from implementing their knowledge and beliefs regarding effective assessment. Despite the restrictions, sometimes if there is wiggle room, teachers try to tweak assessment tools to some extent by adding sub-criteria or comments in the rubric. Also, as explained in previous sections, despite restrictive policies, some teachers use informal assessment to enhance the effectiveness of their classroom assessments. For example, Dave and Arman used observation of students in addition to formal graded assessment results to provide feedback and assign grades at the end of the course.

7.2 Teacher's autonomy in selecting or designing assessments

Most of the participants (n = 9) pointed to the key role of teachers' autonomy in designing the syllabus, assessment tasks, and practices for effective assessment and the implementation of their beliefs. This seems to be one of the key factors in the application of teachers' assessment literacy and beliefs in classroom assessment. According to the participants, this autonomy enables them to design tasks in tune with students' needs, change school-mandated tools, consider their observation of students in their assessments, and include relevant criteria consistent with the relevant constructs to conduct effective assessment.

Mike pointed to the importance of teachers' autonomy in developing or selecting assessments. He explained that when he has the choice, there is a smaller gap between his assessment beliefs and what he practices in the classroom. In the university course he was teaching at the time of the interview, he had more freedom in making assessment decisions, whereas in his previous school, assessment was standardized, so he could not implement his assessment conceptions. He stated that *at this university, instructors have freedom in their classes, which*

reduces conflict between their beliefs and practice. While assignment definitions and weights are set by the institution, how I assess speaking and listening skills is up to me."

Similarly, Hudson reported that in the content-based course he was teaching at the university, although the program gave general guidelines for teaching and assessment, he could still implement his beliefs about effective assessment. Hence, he could develop assessments based on his students' needs, course objectives, and task objectives to promote the quality of his assessment.

Other participants similarly emphasized the importance of freedom in assessment. Dave criticized school-mandated assessments for limiting teachers' flexibility, noting that some tasks, such as an Impromptu Task, lacked authenticity but were non-negotiable, explaining that "*in real life, no one hands you a cue card and asks you to speak for 2-5 minutes with only two minutes to prepare. But since it was mandated by the college and part of the course outline, my hands were tied.*" Ryan argued that formal summative assessments often fail to reflect actual student progress, preferring informal methods, and stressed that teachers should have the freedom to design assessments that align with their teaching. Arman highlighted the importance of teachers' freedom to design tasks in accord with the needs of the class, explaining that he adapts tasks when students are not at the right level, stating that *I adapt the student-led discussion in my program because my students aren t yet able to sustain a conversation at that level. Instead, I design different tasks that align with the same outcomes but through a different path."*

The examples above show how teachers underlined the importance of their autonomy in changing course syllabus, tasks, and assessments to achieve effective assessment in tune with the needs of their students. Some teachers seem to go around restrictions and implement their beliefs and assessment literacy anyway (e.g., Arman, Ryan), while some adhere to the syllabus, tasks, and criteria set by the school despite being against their beliefs (e.g., Dave, Sally). Teachers noted that in cases where they have more freedom, there is a smaller gap between their beliefs and practices

(e.g., Mike, Hudson), as they can develop or select tasks and assessments based on their conception of effective assessment.

7.3 Content-based syllabus

Closely related to school policies, the type of the course also seems to exert influence on teachers' assessment practices. In content-based courses, the main focus is on students' understanding of the content and their ability to discuss or present in relation to the content. According to the participants, in content-based courses, the assessment criteria mostly revolve around content and communication skills rather than linguistic features. They noted that linguistic functions lag behind other aspects, but they expressed a belief in the importance of linguistic features and said that they try different strategies to address them in their assessment practices.

Hudson, for example, said that in his content-based syllabus, the focus is on debating content rather than form, so his assessment and feedback mainly revolve around these skills, not linguistic aspects. He explained that *during presentations or debates, I give on-the-spot feedback for serious errors, but because the focus is on communication rather than form, there's limited opportunity to address linguistic aspects in speaking."*

Similarly, Mike noted that in content-based courses, the focus of teaching and assessment is on academic skills. He pointed out that content-based courses are more about content and using language functions appropriately and successfully. Regarding listening, he explained that listening is assessed implicitly and informally by observation and the notes he takes while students discuss videos and their content, stating that *listening isn't an explicit part of the syllabus, so I assess it only through observation and notes during presentations or discussions. Speaking is easier to assess, as discussions reveal the depth and range of their language abilities."*

Other participants similarly described the influence of content-based syllabi on their assessment practices. Ryan noted that in content-based courses, the focus of teaching and assessment is on academic skills and using language functions appropriately. Donna emphasized that the structure of the course she was teaching is designed in a way that there is little room to assess and discuss language skills; her assessment criteria revolve around content and the use of academic skills such as reading and watching videos and then presenting or discussing that content. Dave pointed out that at colleges and universities, teachers are restricted when it comes to assessing listening, which is mostly assessed through note-taking and integrated with speaking and writing. He noted that he does not evaluate students' notes; instead, students are assessed indirectly through the answers they provide in speaking.

7.4 Student motivation/participation

Most of the participants (*n*= *8*) in this study mentioned student motivation as one of the central factors affecting the effectiveness of assessment in the classroom. Specifically, students' motivation affects the amount of effort teachers put into preparing assessments, sharing and discussing assessment criteria and providing feedback. Lack of motivation can impact students' engagement in assessment activities, thus impacting the effectiveness and success of CA. As noted in the section on the procedure of classroom assessment above, many participants mentioned student engagement as one of the main criteria for evaluating the effectiveness of CA.

Dave believed that student motivation is an important factor mediating his assessment practices and the effort he puts into preparing tasks and assessments and providing feedback. When students are not motivated, he said he would provide quick feedback without elaborative comments. Conversely, if his students are motivated, he would spend more time and would provide more feedback and comments. On a similar note, Aly explained that students 'motivation and their response to her assessment results and feedback play a key role in the way she conducts assessment

and provides feedback. She noted that earlier in her teaching career she used to provide elaborative feedback to her students, but nowadays she provides less feedback due to the lack of motivation among students. For example, she used to share rubrics with her students and explain the assessment criteria with the purpose of fostering students' performance, but students did not show interest and her rubric sharing did not make a difference in their performance: "*I tried discussing printed rubrics with students, but I found it time-consuming, and the results were similar to before, so I no longer use that approach.*" (Aly).

Ray also reported that if the class is motivated and asks many questions eagerly, she would increase the workload, adding more tasks and giving more feedback. If a student does not uptake feedback, she will give the same feedback. Conversely, if a student is motivated and puts in effort but does not make progress, she will provide more feedback or prompts to guide them. She explained that giving feedback is progressive *"when students use it to improve, I'm motivated to provide more specific comments but if students ignore feedback and show no improvement, I invest less time in detailed comments."*

Indeed, students 'lack of motivation and engagement in assessment practices can remarkably affect the outcome of CA especially formative assessments in which students 'engagement and motivation can play a crucial role in their learning and progress. Apparently, as reported by some participants (e.g., Aly, Dave, Ray), teachers 'motivation to conduct effective assessment is under the influence of students 'motivation and their response to CA.

7.5 Students 'conceptions of assessment

Some teachers (*n=7*) highlighted the difference between teachers' and students' conceptions regarding the assessments and tasks used in the classroom. Students seem to hold different views in term of the importance of grading, the value of assessment tasks, the value of teaching assessment criteria, and their response to feedback. According to the participants, student's different views of

assessments, tasks and grades can lower students' motivation and engagement with the practices, resulting in ineffective CA.

One main factor relates to students 'focus on grades in contrast to teachers who seem to be more concerned with learning and progress. This gap between students' and teachers' views can affect students 'engagement in assessment, thus affecting the effectiveness of these assessments. According to Ray and Ryan, students 'obsession with grades is manifested in students' evaluations of their teachers at the end of the semester. They explained that students evaluate the performance of their teachers based on the grades they receive.

Ryan, Dave and Aly mentioned that students excessive 'attention to grades directly impacts their participation and engagement in assessment tasks. For example, Aly explained that if a task is not graded, students do not pay attention to it, noting that *if it's not graded, they don't pay any attention to it."* Similarly, Dave noted that if an assessment task is outside the course syllabus and not graded, students do not take it seriously and may not participate actively, explaining that *assessment feels like a business contract—regardless of personal teaching beliefs, students prioritize what counts toward their grade."*

Donna similarly expressed her belief in attention to students' learning rather than grading, but noted that most of her students are more concerned with grades than learning and improving their English, noting that *grades matter more to them than learning."* She explained that students are unfamiliar with the type of tasks done in the Canadian context and do not understand the learning goals for completing a task, especially if it is different from tasks they did in past assignments. Sally also pointed to students' different understanding of the value of rubrics and the instruction she provides on assessment criteria. She noted that students do not seem to pay attention to her instruction regarding assessment criteria, explaining that even when she goes through the

rubric and underlines different parts, *they don't listen, I don't know why they're not interested in going over the rubric."*

Ryan and Sally pointed to students' and teachers' different views concerning assessments and grades and said that they try to downplay the role of grades and draw students' attention to learning and progress. For example, Sally explained that she talks to her students to change their misconceptions about grading, stating that *I tell students not to focus on grades but on learning. Over time, they shift from aiming for 100 to understanding the lesson and grasping the concepts."* In this respect, Aly pointed to students' expectations about getting higher grades because of their performance in other courses. To deal with this, she said she would explain the difference between a non-credit and credit course, where grades should be more valid.

Ryan and Dave said that they explain the value of assessment to their students to justify their practices when they notice a conflict between their assessment conceptions and those of their students. Ryan contended that if students are not convinced of the value of a task, they will not participate actively, so he explains the underpinning reasons or even research findings to highlight its importance, stating that *I actively persuade them about the importance of activities like roleplay, sometimes spending significant time explaining research findings to counter their initial dismissal."* Dave similarly noted that students often ask him to make tasks easier for higher marks, but he tries to explain the value of the assessment to increase students' motivation and engagement, explaining that *I listen and appreciate their feedback, but I usually justify my* approach to motivate them to engage with the assessment."

Aly, Sally, and Donna pointed to students' different understanding and use of feedback. To address this conflict, Aly said she would explain the value of feedback and what students need to do after receiving it. Sally reported that some students just check their mark without paying attention to their performance on different criteria, so she explains the different criteria and the importance of

using rubrics. Donna noted that some students show negative reactions to direct feedback during presentations because they do not know the rationale, so she explains the value of providing feedback at the beginning of presentation tasks, stating that "*I'm not trying to pick on you, but there's something we can all learn from this for the purposes of the presentation.*"

As another solution to the above conflict between teachers' and students' views of assessment, some participants (*n= 3*) mentioned that they might incorporate some of their students' views in their assessments in addition to adhering to their assessment beliefs to keep students motivated and engaged. Aly sometimes compromises and scales down her assessment tasks and criteria to adapt them to students' views if their views are different and they do not participate well. In one case, she explained "*Realizing that students face multiple barriers—cultural and otherwise—and that my efforts don't always lead to improvement, I've scaled back my expectations and approach.*"

Ryan also pointed out that sometimes he scales down his beliefs about effective assessment and incorporates his students' views to make the assessment task less challenging to keep students motivated. "*Even if I believe an assessment criterion is effective, I avoid using it if it risks overwhelming students and demotivating them, causing dropouts. Keeping them engaged in the course matters more.*" However, as noted above, if he feels that an assessment is very useful to students and for their future performance in the workplace, he will insist on the task and criteria and will try to explain the value of the task to convince them.

On a similar note, Hudson noted that the majority of his students are Asian who are obsessed with grades, so he tries to incorporate their perspective by being less severe in grading, explaining that *I try to understand their perspective and incorporate them into my assignment of marks. Basically, into rater severity, so I become less severe."*

All in all, the above results point to the detrimental role of the conflict between students' and teachers' views concerning assessment, which can impinge on students' motivation and engagement with the assessment and, subsequently, the effectiveness of CA. In the case of such a conflict, teachers either try to compromise by incorporating students' views or try to explain the value of the assessment to keep students motivated and engaged, contributing to the effectiveness of CA. This suggests that despite this class-related contextual factor, teachers seem to go to great lengths to enact their established beliefs regarding effective assessment, however they sometimes compromise and consider students' views to keep them engaged with the assessment.

7.6 Assessment timing

The timing of conducting assessments seems to influence the implementation of teachers' beliefs (n = 7) in their assessment practices as well as assessment effectiveness. According to some participants (e.g., Dave, Aly, Ryan, Mike), if an assessment is done at the end of the semester, the washback effect will not be notable as teachers do not meet students to conduct follow-up activities, discuss results, or give further feedback. However, if the assessment is conducted at the beginning or during the semester, they can make it formative, increasing its positive washback by diagnosing students' needs and providing appropriate support to fill gaps in student understanding.

Dave, Aly, and Ryan reported that for formal summative assessments done at the end of the semester, they only give general feedback to the class because they do not have time to discuss results in more detail. Dave explained that when an assessment is done after the final exam, students do not take it seriously and do not participate actively. He referred to a group seminar task and noted that since it is done after the final exam, some students do not present or participate, negatively affecting the quality of the assessment. He explained that in his college, *"my hands are tied when it comes to changing assessments, especially their timing. I tried requesting adjustments, but decisions are made at the administrative level."*

Mike noted that if an assessment is done at the beginning of the semester or midterm, he can make changes and adapt his teaching to students' needs to enhance their learning, stating that *if it is in midterm or in the first month of instruction, then there is a chance to adjust my instruction and maybe gradually make this skill more feasible for students."* Donna also pointed out that in the final presentation there are no opportunities to check student uptake of the feedback she provides, explaining that *there is not a practice presentation… there's not a follow up. What they take away from it is really on their own."*

7.7 Graded vs ungraded assessment

Another factor that seems to impact teachers' implementation of their beliefs in classroom assessment as well as the achievement of effective assessment is the stakes associated with classroom assessments. Some teachers (e.g., Donna, Hudson) objected to graded assessment on the grounds that they could better coach students and provide more constructive feedback to promote student learning instead of assigning grades. On the other hand, some teachers (e.g., Ryan, Aly, Dave) thought that graded assessments make students take assessment seriously and engage with assessment tasks. As noted above, sometimes students are too obsessed with grades, and when an assignment is not graded, they do not engage and thus do not benefit from the teacher's instruction and assessment results.

Donna expressed a belief in ungraded assessment, noting that she wished assessments were ungraded so that she could provide more constructive feedback to her students. However, this belief is not implemented due to the curriculum and assessment guidelines mandated by her school. She explained that in a different context, she sometimes only took notes and gave detailed feedback, which proved valuable by letting students refine their work before final presentations. She added that in courses with graded assignments, she does not do brainstorming and discussions because she thinks students might borrow ideas from each other, which can impact their grades, stating that *if it*

wasn't an evaluation of ideas, it would be great to discuss what comes out in their essays or presentation." Specifically, regarding a graded presentation task, Donna noted that she tries not to give public feedback to those who present first because future presenters might benefit from such feedback to obtain higher marks, which could undermine the fairness of the grades.

Similarly, Hudson advocated for ungraded assessments, explaining that they allow him to provide more qualitative feedback on different language skills throughout the semester, noting that *feedback you get from students is even more positive"* when assessments are not graded.

On the other hand, some participants (e.g., Ryan, Aly, Dave) reported that graded assessments are necessary because students do not take ungraded assessments seriously. Ryan noted that some assessments conducted during the semester are not graded and only the final assessment is graded, so students do not take the former assessments seriously and do not learn the nuances important for L2 speaking development. Due to a restrictive curriculum, he could not assign multiple graded assessments, stating that *we're not allowed to have [graded] summative assessment, we can't have quizzes [ ] because there's no way I can somehow motivate them to study throughout the semester."* On a similar note, Aly pointed out that ungraded assignments are not appealing to students, explaining that *if it's not graded, they don't pay any attention to it."* She referred to a small group discussion activity that was low stakes (5% of the grade) and students did not put in full effort, so it did not work well. Dave also mentioned that when a task is not graded or its mark is not important for passing the course, students are not motivated enough to participate. He gave an example of a group seminar held after the final exam when students already know if they have passed, noting that *many students with passing marks, like 80%, skip it because they don't need the extra 10%."*

All in all, the above examples show that teachers deal with different factors that might affect their assessment practices. Sometimes, they go around the constraining factors and implement their

belief anyway, sometimes they balance their beliefs or AL with the contextual factors, and sometimes act against their beliefs and yield to the constraints.

Collectively, the above factors and those discussed in the section 4.1 about procedure of CA can be categorized under *constructive* and *detrimental* factors. As noted in the section on the procedure of CA, teachers' practices seem to be positively influenced by constructive factors such as student needs, task objectives, instructional goals, student input, and student performance, which make teacher reflect on their practices and beliefs and use trial and error to improve their assessment practices. These factors seem to help teachers revise their preliminary beliefs and theoretical assessment knowledge and consolidate them in practical ways in the classroom. On the other hand, some factors seem to be detrimental such as school policies, inflexible curricula, student different assessment views, content-based approaches, student lack of motivation, the stakes associated with the assessment, and the timing of the assessment, which can impede' effective assessments in the classroom. (See *Figure* 8.1)

As for the implementation of their beliefs in the case of macro contextual detrimental factors, or class related micro detrimental factors, teachers either comply with the constraints and act inconsistently with their AL and beliefs, thus sacrificing the quality of their CA; balance their beliefs with the constraints, or just follow their beliefs in classroom practices. In the case of detrimental factors, they seem to balance the conflict by using informal assessments in addition to formal assessments. The relationship between teachers' belief, different contextual factors and practices is displayed in *Figure* 8.1.

CHAPTER 8: DISCUSSION AND IMPLICATIONS

This chapter presents a summary of the key findings of the study and relates them to existing research on CA, teacher cognition and assessment literacy. First, based on the synthesis of all findings, a model is proposed depicting how teachers' AL, beliefs, and practices interact with each other and with various contextual factors, highlighting the developmental nature of TAL and beliefs. Next, the discussion focuses on assessment practices in CA and their relationships with TAL and beliefs. The chapter also presents key implications for theory, policy and curriculum development, teacher training programs, and classroom practices. It concludes by addressing the study's limitations and directions for future research concerning teachers' beliefs, AL and classroom assessment practices.

8.1 Summary and discussions of the findings

8.1.1 A model of the relationship between teacher AL, beliefs, context and practices

As explained above, the dynamic nature of the classroom, the contradictions experienced by some teachers between theoretical knowledge and classroom practices, the use of trial and error in the classroom and the influence of different contextual factors, all point to the central role of teachers' actual assessment experience which can bridge the gap between theoretical and practical AL, and help shape their assessment literacy to become more applicable in the dynamic classroom context.

In order to better understand the effects of contextual factors on the implementation of TAL and assessment beliefs in CA, a distinction should be made between *constructive* and *detrimental* contextual factors. *Constructive* factors, such as student needs, student engagement, course and task objectives, and student diversity support the revision and improvement of teacher theoretical AL and preliminary beliefs, which in turn, help teachers improve the effectiveness of their CAs. On the

other hand, *detrimental* factors, which include institutional factors (such as school policies, strict curriculum, institutional assessment guidelines, assessment timing, stakes associated with the assessment), and student related constraints (such as student motivation, and student assessment conceptions) can impede the implementation of teachers' AL and beliefs as well as undermine CA practices.

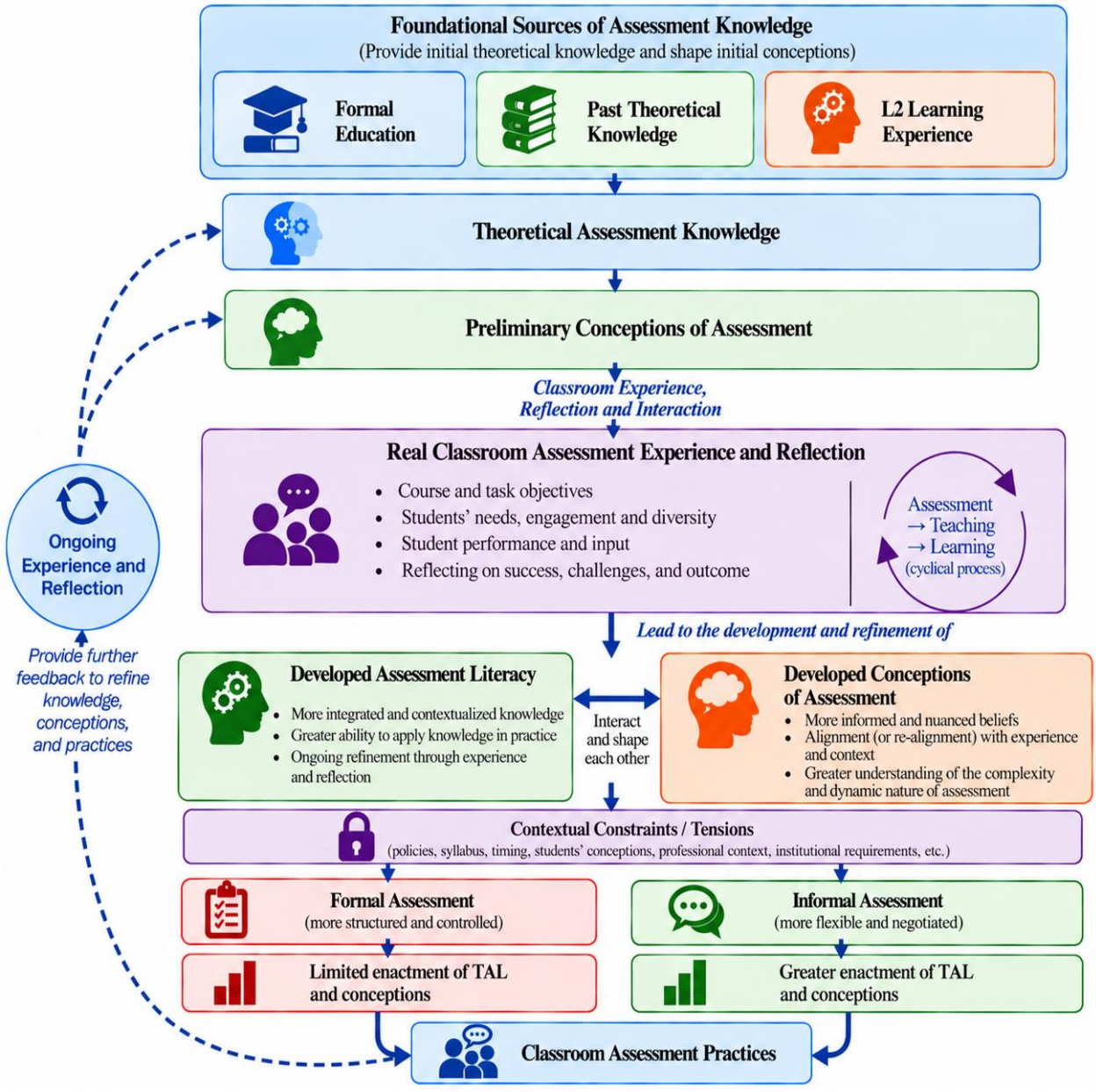
Teaching' experience plays a central role in shaping teachers' assessment practices, leading them to revise their theoretical knowledge as well as the beliefs they had developed as a result of this theoretical knowledge. Teachers 'theoretical knowledge and their past education and learning experiences seem to shape their *preliminary beliefs*; however, these preliminary beliefs are revised in light of their experiences in the classroom and in contact with constructive contextual factors, becoming *internalized beliefs and practical literacy* (see *Figure* 8.1). These established beliefs and literacy are more likely to be implemented in assessment practices. When faced with *detrimental* contextual constraints, teachers either go around the constraints and find ways to still implement their beliefs and AL in their practices or they act inconsistently with their established AL and beliefs. In either case, it is unlikely that there will be any change in the teachers 'established beliefs and literacy that have already become persistent beliefs, if not to say ideologies. In other words, in the face of detrimental constraints, teachers' established beliefs and AL do not seem to waver or change.

As for the implementation of established beliefs, a distinction should be made between formal and informal assessments. As explained above, generally, teachers (e.g., Ryan, Sally, Dave, Arman, Ray) seem to implement their beliefs more through informal assessment than formal assessment practices. Therefore, it can be argued that teachers 'established beliefs are more likely to be realized when using informal assessment methods. In the case of formal assessment, teachers either act against detrimental contextual factors and enact their beliefs in their practices or go

against their beliefs and yield to the detrimental factors. Figure 8.1 shows these relationships between theoretical TAL, practical TAL, preliminary beliefs, established beliefs, contextual factors, and classroom practices.

Figure 8.1

A Developmental Model of the Relationships between Teachers' AL, Conceptions, Contextual Factors and Classroom Practices



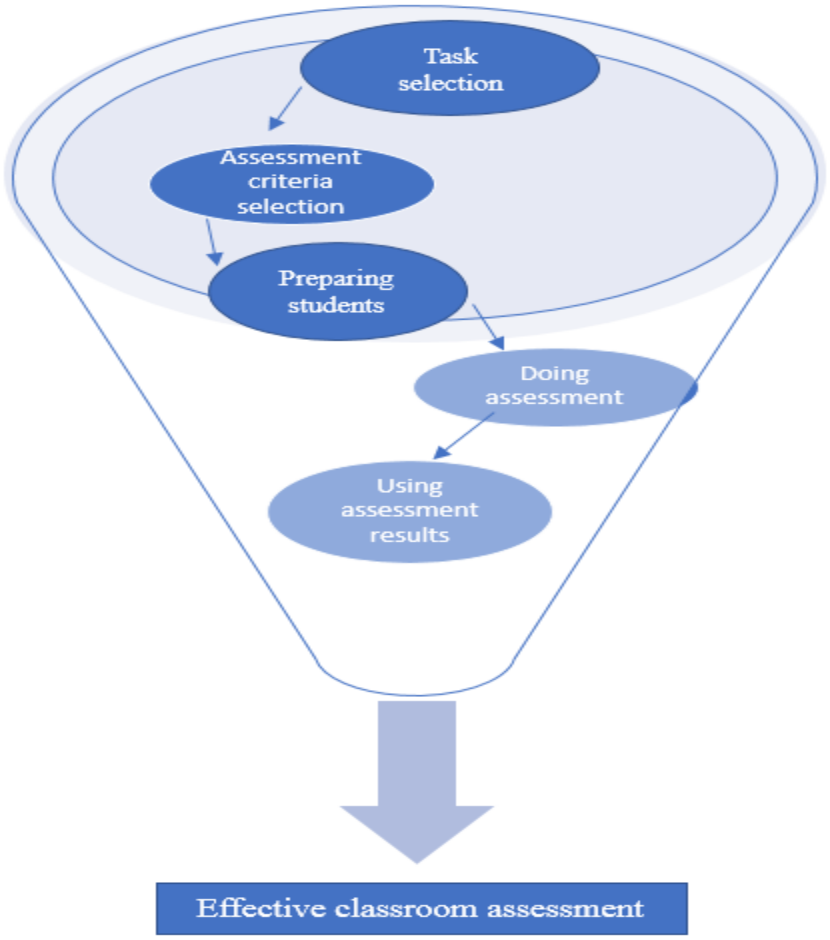
8.1.2 Classroom assessment processes

This study set out to explore the relationship between teachers' AL, their beliefs and assessment practices and the possible effect of contextual factors on these relationships. To understand these relationships, it was necessary to gain a clearer understanding regarding teachers' actual assessment practices and processes. The importance of studying teachers' actual practice in the classroom as a more authentic way to examine teachers' assessment literacy, their beliefs and their relationship with classroom practices has been emphasized by several researchers (e.g., Campbell, 2013; Hill & McNamara, 2012; Xu & Brown, 2016; Xu & Liu, 2009). As such, teachers' actual CA procedures were examined first.

The results in this respect revealed that ESL teachers follow several procedures in their CA to ensure effective assessment in general and assessments of L2 oral communication in particular. After selecting or developing appropriate assessment tasks in line with course objectives and students' needs, teachers use appropriate assessment criteria and prepare students for task performance and assessment. During the assessment, they use multiple assessment methods and diagnose students' needs. Finally, in response to the results of their assessment, they provide appropriate feedback and use the results to improve their assessments as well as their teaching. They also reflect on their assessments to identify the drawbacks and improve them. This process is depicted in *Figure* 8.2.

Figure 8.2

Classroom Assessment Processes



Given the unique characteristics of CA compared to psychometric testing, some researchers have raised the need for the development of models specific to classroom practices (e.g., Brookhart, 2003; Heritage, 2018; Moss, 2003; Moss et al., 2005; Shepard et al, 2018; Wilson, 2018). However, empirical research on CA procedures remains limited. Reviews by Lee (2017) and Wiliam (2011) highlighted the use of formative, task-based assessments, continuous feedback, and teacher adaptation based on student performance. While these studies underscore the value of feedback, reflection, and adaptability in CA, they do not offer a comprehensive, step-by-step model outlining the different stages of assessments teachers use to ensure CA effectiveness.

Overall, the results of this study show that classroom assessment is context-dependent, and that teachers use assessment to support student learning, which they see as evidence of the effectiveness of their CA. Classroom assessment seems to follow a continuous cycle of assessment, teaching, and learning, ultimately aiming to promote student progress as illustrated in Figure 4.7. The findings indicate that teachers integrate learning progression into all stages of classroom assessment, aligning with formative assessment principles. They choose appropriate assessment tools to identify and address student needs. During preparation for assessment, they help students improve their performance and progress in their learning. Throughout the evaluation process, they use diverse methods to identify gaps in student learning and provide appropriate assistance to bridge these gaps. They also reflect on student performance, and their own practices and refine assessments accordingly.

Another finding in this part relates to the importance of teachers' informal assessment in their classroom practices in assessing oral communication. As shown in *Figure* 5.4, teachers use informal assessment throughout the assessment process including tool selection and adaptation, diagnosing student needs and tracking their progress, providing feedback, and using assessment information to improve teaching. This finding addresses the call by several researchers (e.g., Brown, 2008; Hill & McNamara, 2012; Kane & Wools, 2020; Rea-Dickins, 2001; Rea-Dickins & Gardner, 2000) to investigate informal assessment in CA. In addition to using informal assessment for formative and summative purposes, the teachers use it as a workaround strategy to circumvent contextual factors that constraint the implementation of their AL and beliefs in their assessment practices as discussed below.

8.1.3 Fair classroom assessment

The study' results show that, in addition to attending to construct validity and the positive consequences of assessment, teachers consider fairness a key characteristic of good classroom

assessment. The results, however, highlight the complexity of fairness, indicating that teachers' definitions of fair assessment are diverse. Some teachers, like Arman, viewed fairness as giving more weight to student participation, using observational data and offering bonus marks for engagement. Others, such as Mike, Donna, and Sally, defined fairness as equal treatment, adhering strictly to grading criteria without adjusting marks beyond formal assessments. Despite the diversity in the definitions of fairness, the participants seem to see it as related mainly to grading. Teachers reported using objective rubrics, applying the same criteria to all students, and sharing criteria with students to achieve fair grades. They aim to assign accurate grades that reflect each student's real ability or performance. Some reported taking into account student effort and questions of equity to increase the fairness of their grading. While these findings are consistent with previous studies that show the complexity and diversity of fairness definition among teachers (e.g., Tierney, 2012), they also indicate that the participants in this study tended to focus mainly on grading as the area of most importance for fair assessment.

In the literature, fairness has been defined differently by different researchers who have put forward different criteria for ensuring fair assessment such as accurate grading and positive consequences of assessment (Messick, 1994), unbiased assessment and providing equal opportunities for all students (Kunnan, 2004), improving student learning (Black & Wiliam, 1998), justification for assessment use in the context of learning (Kane, 2010) and providing all students with the support and conditions they need to succeed (Tierney, 2012). Tierney (2013), after reviewing the literature on educational assessment, acknowledges the complexity of fairness due to diversity of classroom contexts, and notes that there are multiple and conflicting interpretations of the concept of fairness.

8.1.4 The relationship between teacher assessment beliefs and practices

The results of this study revealed that, generally, there is consistency between teachers' assessment beliefs and their classroom practices in assessing L2 oral communication. They show that teachers are substantially informed by their general beliefs about teaching and assessment of L2 oral communication, the goals of assessment, the characteristics of the assessment, and methods of assessment in their assessment practices of oral communication in the classroom. However, as noted in chapters 5 and 6, some contextual factors mediate the implementations of teachers' beliefs in their classroom practices.

The findings of this study support the results of the past few studies on oral communication assessment (Özdemir-Yılmazer and Özkan, 2017; Ahmad, 2021), which found that teachers prioritize formative assessment, meaning over accuracy, peer assessment, and diagnosing students' needs and providing appropriate feedback both in their stated beliefs and their speaking assessment practices. Similarly, this study lends support to past research findings on writing assessment (e.g., Wang et al., 2020; Guadu & Boersma, 2018; Lam, 2019; Montgomery & Baker, 2007) in which participant teachers reported strong support for formative, learning-oriented assessment; however, they struggled to align these beliefs with their actual assessment practices due to factors like lack of understanding, institutional constraints, or inadequate training.

However, these studies were limited in scope, examining only teachers' beliefs regarding assessment goals, accuracy and feedback. To gain a clearer picture of the relationship between teachers' beliefs and their assessment practices, this study addressed a broader area of teachers' beliefs including their general beliefs about teaching and assessing oral communication, assessment goals and methods and the characteristics of good assessment. This study also made a distinction between teacher formal and informal assessment practices in relation to the implementation of teachers' beliefs. As the results revealed, teachers behave differently when it comes to informal

assessments especially when they encounter contextual constraints and experience tensions. There seems to be more consistency between teachers' assessment beliefs and their informal assessment practices. Especially in the case of tensions, teachers resort to informal assessments to implement their beliefs concerning effective assessment.

It is noteworthy that teachers' attention to informal assessment besides formal assessment in this study can be due to the nature of oral communication. The interactive nature of oral communication gives teachers more room to use observation and informal assessment compared to other skills such as reading and writing assessment that rely on tangible written texts.

Another finding relates to the affective dimension of teachers' beliefs. Some researchers (e.g., Borg, 2011; Scarino, 2013) posit that teachers' cognition is not purely logical or knowledge-based; rather, it is emotional and interpretive, shaped by past experiences, personal values, and interactions with students. While prior research acknowledges the emotional dimension of teachers' cognition, the application of these affective aspects in practice, especially in defining equity and equality, remains underexplored.

The findings of this study indicate that the affective dimension plays an important role in teachers' cognition in relation to assessment fairness and equity. Teachers seem to experience conflict when it comes to equity and equality in classroom assessment. On one hand, they aim for trustworthiness by applying the same criteria to all students when assigning grades. On the other hand, they try to take into account students' circumstances which is reflected in such practices as giving bonus marks to students who are on the verge of failing and providing additional opportunities and support to weaker students. This creates conflicts, as teachers strive for both objectivity in grading and flexibility to support struggling students. It can be argued that equality and using the same standards, guidelines, or grade breakdown for all students seem to be more consistent with teachers' knowledge base than considering students' personal circumstances which

seems to be based on teachers' personal values. Therefore, it is safe to say that understanding these subtleties between the affective and cognitive dimensions of teachers' beliefs can help us better understand the relationship between knowledge and beliefs.

8.1.5 The effect of contextual factors on the relationship between beliefs and practices

As reported by Wang et al. (2020) contextual factors at the micro, meso, and macro levels can impact the implementation of teachers' beliefs. In this study, at the micro level student motivation and their different conception of assessment created challenges for the teachers. At the meso level, school policies, content-based syllabus, the time of assessment and the stakes associated with assessment caused tensions and challenged teachers' implementation of their beliefs.

At the micro level, teachers seem to have an easier job in balancing tensions and coming to a resolution than in the case of meso level factors. For instance, in order to increase students' motivation and change their misconceptions of assessments, some teachers explain the value of assessments, tasks, and feedback and downplay the importance of grades, while others try to incorporate some of their students' views in their assessment practices to encourage more participation and increase the effectiveness of their assessment. However, in the case of meso level factors such as rigid school assessment guidelines, teachers may try to balance the tensions but ultimately tend to withhold their beliefs and follow the mandated guidelines, which they believe would not lead to effective assessment.

By exploring the contextual factors, the tensions teachers experience and how they respond to them, we can gain a deeper understanding of the relationship between beliefs and practices and teacher AL development. As Xu and Brown (2016, p.157) pointed out, AL can be better understood in practice when teachers experience contradictions and balance their classroom practices with contextual factors and limitations. When teachers navigate and resolve these challenges, the nature

of AL becomes clearer. Therefore, by highlighting the central roles of contextual factors, the challenges teachers experience and their reaction to these tensions, this study provides a better picture of the relationship between teachers' practices, AL and beliefs and their development. The study also proposes a model for TAL that highlights its developmental nature by focusing on contextual factors and the tensions teachers experience and the balances they make in response to these tensions (see Figure 8.1).

As illustrated in the proposed model in *Figure* 8.1, a distinction should be made between preliminary and established beliefs. The results suggest that teachers' preliminary beliefs develop in light of their experience in the classroom in contact with contextual factors. Preliminary beliefs undergo transformation based on real-world teaching experiences. Some beliefs may solidify and become more stable, while others are revised or even replaced, depending on teachers' experiences in the classroom. This process better illuminates the developmental and dynamic nature of teachers' beliefs in contrast to some past studies which have dichotomized beliefs into core and peripheral beliefs (e.g., Zheng, 2013b; Richardson, 1996; Phipps & Borg, 2009). For example, Zheng (2013) distinguishes between core and peripheral beliefs, emphasizing their interaction, with core beliefs often acting as the dominant beliefs guiding practice. He argues that multiple beliefs underpin teachers' practices, either complementing or contradicting each other, which defines the distinction between core and peripheral beliefs. However, this distinction overlooks the developmental dimension of beliefs.

All in all, this study's findings challenge the dichotomous view of the core-peripheral belief framework and instead propose a developmental model where teachers' beliefs start as preliminary, and then evolve through lived experiences, engagement with the classroom environment and reflection. This perspective provides a nuanced understanding of how teachers' beliefs shape and adapt to their teaching practices over time.

8.1.6 Teachers' knowledge base, assessment literacy and practices

Following the conceptual framework of the study, the relationship between teachers' knowledge base and their assessment practices was examined through a TAL questionnaire developed specifically for this study. The findings indicated a strong alignment between teachers' self-reported assessment knowledge base and their classroom practices. Generally, participants demonstrated awareness of various forms of assessment knowledge and applied them in both their formal and informal assessment practices. However, when teachers reported little to no knowledge in certain areas, practices related to these areas were largely absent.

The analysis of data from the interviews, observations, SR sessions, and documents indicated that teachers' assessment practices are informed by their teaching and assessment experiences, prior learning, past education, and theoretical knowledge. The analysis of these influencing factors indicated that classroom teaching and assessment experience is regarded as the most influential factor in shaping teachers' assessment practices. This prominent role is attributed to the inherently dynamic nature of classroom settings, the complexity of contextual variables, the contradictions encountered in daily practice, and the reflective processes including trial and error through which teachers continuously adapt and refine their assessment practices and literacy.

Some previous studies acknowledge the importance of past learning experience (e.g., Borg, 2003; Farrell & Guz, 2019; Yilmazer & Ozkan 2017) and past education and training (e.g., Scarino, 2013; Volante & Fazio, 2007; Vogt & Tsagari, 2014) in shaping teachers' cognition as well as their practices. According to the results of this study, although teachers' past education, learning experiences and theoretical knowledge play an important role in TAL and shaping classroom practices, their actual assessment experience in the classroom has the most prominent role, overshadowing the formation and development of TAL and assessment practices. Assessment experiences serve as a critical medium through which theoretical knowledge and initial beliefs are

refined and solidified. Through ongoing interaction with contextual variables and reflective practice, teachers consistently reevaluate and adjust their prior understanding and beliefs about assessment. Although theoretical knowledge and learning experience are the foundations of a teacher's assessment literacy, teachers' experiential knowledge plays a crucial role and should be considered in developing assessment literacy.

8.1.7 Integration of listening and speaking

The majority of teachers in this study expressed strong support for the combined assessment of speaking and listening. In most cases, listening activities, used as prompts, were paired with tasks such as discussions, debates, role plays, peer feedback, and question-and-answer sessions. Teachers cited several reasons for integrating speaking and listening, including task authenticity, the promotion of classroom communication, the demands of content-based courses, and the reinforcement of course content. However, in their assessment practices, listening was either not assessed or assessed indirectly through speaking. Likewise, in their assessment rubrics no criterion was dedicated for assessing listening. Consequently, there seems to be an inconsistency between integrated tasks and assessment rubrics. Integrated tasks, thus, might reduce assessment effectiveness because they do not include a direct assessment of listening, which means that teachers cannot properly diagnose the listening needs of their students and, consequently, cannot provide appropriate feedback to address any gaps in their listening comprehension. This finding highlights the need for more direct assessment of listening in integrated tasks and developing criteria that address listening comprehension besides speaking assessment criteria. This lack of attention to listening can be due to the focus on content in academic courses which emphasize communication, taking listening for granted, or it can be due to a lack of proper training for integrated assessment in educational programs.

8.2 Implications

8.2.1 Implication for theory and models of classroom assessment and teacher assessment literacy

The model proposed for the process of CA in *Figure* 8.2 is based on the lived experiences of ESL teachers in actual classrooms. This has some implications for theory. Most of the models in this field are theoretical; not based on the results of empirical research, therefore this model can pave the way for the development of evidence-based models for CAs. The findings of this study call for an expansion of classroom assessment models to more fully integrate different dimensions and stages of the assessment process. They also support the development of more comprehensive, practice-informed theories that reflect the realities of classroom assessment.

The findings related to the importance of informal assessment in the classroom, highlight the need to revise existing models of TAL, which often overlook the significance of informal assessments. A more comprehensive TAL model would better reflect the realities of teachers' assessment practices. Also, the findings related to the relationship between teachers' assessment experience in the classroom and TAL suggest that experiential knowledge should play a more central role in studies on teachers' AL. In defining the assessment knowledge base and developing models of it, both teachers' experiential and theoretical knowledge should be accounted for. The focus should shift to teachers' knowledge in practice, classroom dynamics, the tensions teachers experience, the needs of students, and teacher reflection in response to different contextual factors.

Most of the existing models of the knowledge base and questionnaires used for evaluating it address prescribed competencies and focus on the psychometric properties of testing, failing to address teachers' assessment needs in the classroom (Xu & Brown, 2016, Fulcher, 2012, Volante & Fazio, 2007; DeLuca et al. 2016). In their model of assessment knowledge, Xu and Brown (2016) acknowledge the inadequacy of the components and caution that the forms of knowledge in their

model show general principles without considering the dynamic context which can cause the knowledge base to change. As such, the different forms of knowledge identified in this study can inform the development of a more comprehensive knowledge base model as well as questionnaires for evaluating it, encompassing both theoretical and practical aspects of teachers' assessment knowledge, including their interpretive and adaptive role.

Another implication of this study concerns the evolving nature of teachers 'beliefs. The results of this study offer an alternative to the binary core- peripheral conceptualization, supporting a developmental model in which teachers 'beliefs begin as preliminary and gradually transform through lived experiences, interaction with the classroom environment, and reflective practices. This conceptualization offers a more nuanced understanding of how teachers 'beliefs evolve and adjust to their teaching practices over time.

8.2.2 Implications for policy

The findings concerning tensions and teachers' approaches to balancing them carry important implications for policy. The study underscores the importance of creating assessment frameworks that recognize and support teachers' professional judgment. When accountability systems are overly prescriptive, they can conflict with what teachers believe is best for their students. The study suggests that a more balanced and flexible policy approach, one that aligns institutional expectations with pedagogical realities, could ease this tension and promote more student-centered assessment practices. These reforms should emphasize teacher autonomy and uphold pedagogical integrity, shifting away from rigid institutional structures to better support the complex goals of education (Bonner, 2012; Brookhart & McMillan, 2020; Kane & Wools, 2020).

The implications in this respect extend to curriculum design and institutional support. Schools and curriculum developers can benefit from these insights by offering more flexible and supportive assessment structures that empower teachers rather than constrain them. Creating

professional learning communities and collaborative spaces where teachers can share strategies and experiences may also enhance their capacity to navigate assessment tensions effectively.

The findings of this study also underline the critical role of informal assessment in classroom practices in assessing oral communication, particularly in how teachers implement their beliefs. For policymakers and school administrators, the study suggests the importance of granting teachers greater autonomy to design their own syllabi and incorporate informal assessments. Such flexibility would empower teachers to rely more on their professional judgment, leading to more effective and context-sensitive evaluations of student learning.

8.2.3 Implications for teacher education and training

The assessment process outlined in this study underscores the fundamental role of selecting or using appropriate tasks as the foundation of effective assessment. Task objectives function like a lighthouse, steering teachers toward effective assessment practices. By keeping task objectives in focus, teachers select appropriate assessment criteria, and provide feedback and use task completion as an important criterion to evaluate their assessment effectiveness. This suggests that more attention should be paid to task-based assessment in teacher education programs. Highlighting the importance of using appropriate tasks in accord with curriculum and students' needs and then selecting or designing appropriate assessment criteria for them can give pre-service and in-service teachers some insights into applying their assessment knowledge and ensuring good assessment in the classroom.

The observed consistency between beliefs and assessment practices in this study highlights the need for teacher education programs to critically address both accurate and misguided beliefs regarding L2 assessment. These programs should prioritize the effective instruction of assessment principles, ensuring that teachers are equipped with accurate and comprehensive knowledge of formative as well as summative assessment, which is essential for cultivating informed, effective

belief systems. By fostering a clear understanding of sound assessment practices, such programs can support the development of well-informed and effective assessment-related beliefs. This, in turn, prepares teachers to implement assessment practices that are pedagogically sound and conducive to successful language learning outcomes.

By uncovering the real-world tensions that ESL teachers experience in classroom assessment, the study highlights the need for teacher education and professional development programs to be more attuned to these practical challenges. Such programs should not only build foundational assessment literacy but also emphasize strategies for navigating contextual constraints while keeping student learning progression at the core of assessment practices.

The findings also highlight the importance of addressing the affective and ethical aspects of assessment in teacher training. Educators need support in navigating tensions such as balancing grade adjustments for struggling students with maintaining grading standards. Training should incorporate case studies of real-world ethical dilemmas, and simulations to prepare teachers for emotionally complex decisions in assessment. Pre-service teachers should be equipped not only with theoretical knowledge but also with strategies for managing emotions and handling ethical challenges. Understanding how to balance equity and equality in assessments is essential. The conflict between ensuring objective grading and meeting individual student needs reveals that assessment is not solely rational but influenced by both cognitive and affective factors.

Furthermore, the results highlight the need to emphasize assessment literacy that balances formal and informal methods. This need is particularly urgent given that pre-service language teacher education programs have consistently demonstrated a recurring gap in assessment literacy development (DeLuca & Klinger, 2010; Vogt & Tsagari, 2014). Research has shown that teacher candidates demonstrate overall weakness in assessment knowledge and skills (Brookhart, 2003, 2011; Mertler, 2004). Scarino (2013), Xu and Brown (2016), Vogt and Tsagari (2014), and Graham

(2005) all highlight a common concern: theoretical knowledge of assessment does not guarantee its implementation in practice. This suggests that AL does not merely consist of theoretical knowledge; it encompasses knowledge of how to implement assessment methods as well. Training should foster reflective, student-centered assessment practices and promote the use of diverse criteria beyond final grades. Teachers should be equipped to integrate ongoing observations, feedback, and student performance indicators to ensure assessments are both fair and trustworthy.

8.2.4 Implications for classroom practices

A persistent challenge for teachers is the gap between theoretical knowledge and its application in classroom assessment. Since the processes revealed in this study are based on experienced teachers' real practices and their interpretations, the findings can inform in-service and pre-service teachers about the procedures that may be important for ensuring effective CA. Being aware of the cyclical nature of the CA process, can heighten teachers' awareness of the dynamic nature of CA, and encourage them to use their reflection throughout the assessment process. By adopting a reflective approach, and through trial-and-error teachers can continuously evaluate the effectiveness of their CA and make adjustments to promote student learning. Also, the interactive nature of beliefs and their development in the classroom context suggests that teachers may benefit from being reflective in their assessment practices, and remaining open to modifying their initial and previously held beliefs as they come into contact with the classroom context.

The assessment practices identified in this study also highlight the importance of considering the alignment among assessment tasks, course objectives, students' needs, and target language use situations. These findings do not suggest that all ESL teachers should follow the same procedures; rather, they may provide useful considerations for teachers when reflecting on the relevance and effectiveness of their own assessment practices. In particular, teachers may consider whether the tasks they select or design are aligned with course objectives and students' needs, provide

meaningful opportunities for language use, and are appropriately sequenced in terms of difficulty and complexity.

The findings also suggest that teachers' assessment tools and practices can develop through classroom experience and reflection. The participating teachers described how they revised their assessment tasks, rubrics, scaffolding, feedback, and other aspects of their practices based on their classroom experiences. For ESL teachers, this highlights the potential role of everyday assessment experiences as opportunities for professional learning and the development of practical assessment knowledge. Reflecting on students' performance and on the effectiveness of assessment practices may help teachers identify areas for revision and make adjustments to their assessment practices over time

Meanwhile, teachers' awareness of contextual factors and the resulting tensions can encourage them to develop adaptive strategies, particularly through informal assessments. These strategies allow teachers to respond more effectively to students' needs, even when faced with external pressures. Encouraging reflective practice and strategic decision-making can thus enhance the quality and relevance of assessment in everyday teaching. It is noteworthy that informal assessment not only supports more accurate evaluations of student performance but also helps teachers cope with institutional constraints and maintain alignment between their assessment practices and professional beliefs. Recognizing the role of informal asseTssment may therefore contribute to more authentic and effective classroom assessment practices.

The research findings underscore a critical gap in the integration of listening and speaking skills. While teachers in the study expressed strong support for combining speaking and listening in classroom tasks, there was a noticeable inconsistency in how these integrated tasks were assessed. This finding suggests a need for greater attention to how both skills are represented in assessment criteria and practices. Teachers may consider whether their assessment tools allow them to evaluate

speaking and listening as distinct but interconnected skills. This could help teachers avoid relying on speaking as a proxy for listening, leading to more precise assessments and better understanding of students' needs.

8.2.5 Limitations and implications for future research

This study, despite being comprehensive in its exploration of CA procedure and the relationships between teachers 'AL, beliefs and classroom practices, is not without its limitations. Firstly, only five participants agreed to be observed during one of their three- hour classes. Nevertheless, most teacher observations lasted three hours, offering considerable depth and detail in each session to offset their limited frequency. This length was deemed sufficient to gain a thorough understanding of the teaching and assessment practices within the observed periods. Additionally, this limitation was addressed through the study's transparent methodology and the comprehensive context it provides, enabling careful transferability to comparable educational settings.

Secondly, the results related to students 'conceptions of assessment was based on teachers ' views through interviews without seeking students 'views. The study could offer more insightful ideas if some students were interviewed regarding their beliefs about good assessment. The findings could shed more light on the discrepancies and tensions teachers experience in the assessment practices. Therefore, future research can explore the differences between teachers 'and students ' conceptions of assessment in the classroom.

Finally, the TAL questionnaire was developed for the purpose of this study. It was based on the conceptual framework and the different forms of knowledge mentioned in existing theoretical frameworks. However, due to the qualitative nature of the current research, it was piloted with a small number of participants, and it was not validated using a solid procedure as is typically done in quantitative studies. Therefore, the findings should be interpreted cautiously. Nevertheless, this can be a valuable step toward the development of practice-based assessment tools for evaluating TAL.

In addition to the forms of knowledge used in the development of the TAL questionnaire, future research can draw on the assessment procedures and criteria identified in this study (see chapter 4) to design more authentic and practice-based instruments for evaluating TAL. Because the forms of knowledge identified in this study reflect teachers' real practices in actual classroom assessment, they offer a rich basis for constructing instruments and models that capture what teachers actually know and do. For example, future tools could incorporate knowledge of students' needs and emotions, aligning tasks and rubrics with relevant constructs and course objectives, preparing students for task performance, employing both formal and informal assessments, using assessment results to improve teaching and learning, providing meaningful feedback, reflecting on and revising assessments, and balancing contextual tensions. Research that develops and tests such tools would help ensure that TAL evaluation is grounded in practice rather than abstract criteria, producing models that are both theoretically sound and practically relevant.

Furthermore, teachers were found to implement their beliefs more effectively through informal assessments than through formal ones. Teachers' ability to integrate informal assessments into their routines allows them to better navigate contextual constraints and classroom tensions, using such assessments as a tool to support student learning and uphold their professional beliefs. This distinction suggests that future research on the relationship between teachers' beliefs and practices should consider and differentiate between formal and informal assessments.

The findings also have implications for research on the relationship between teacher cognition and classroom assessment practices. The findings suggest that both teachers' assessment literacy and their assessment conceptions play a role in shaping their classroom assessment practices. Although these represent distinct dimensions of teacher cognition, examining them together provides a more comprehensive understanding of the factors underlying teachers' assessment decisions and practices. Therefore, future research investigating the relationship between

teacher cognition and classroom assessment practices may benefit from considering both assessment literacy and assessment conceptions rather than examining either dimension in isolation.

Moreover, this study revealed that the affective dimension of teachers ’beliefs plays a role in classroom assessment. Teachers appeared to navigate a tension between applying uniform criteria to ensure fairness and adjusting decisions to support students ’individual circumstances-for example, awarding bonus marks or offering additional support. These patterns suggest an interplay between cognitive (knowledge-based) and affective (value-laden) considerations in assessment decisions. This finding highlights the need for more focused and in-depth research on how the affective and cognitive dimensions of teachers ’beliefs shape classroom assessment practices. Understanding these nuances would clarify the extent to which teachers ’decisions are guided by professional knowledge versus personal values and emotions, thereby advancing our understanding of the complex relationships among teachers ’knowledge, their beliefs, and their assessment practices.

Finally, the findings of this study underscore the complexity and diversity of teachers ’ conceptualizations of fairness in classroom assessment. While some teachers prioritized participation, engagement, and effort, others emphasized equal treatment and strict adherence to grading criteria. Despite these differences, the results indicate a predominant focus on fair grading-using objective tools, consistent criteria, and transparent communication to reflect students ’actual abilities. This highlights that, in practice, fairness in assessment is closely tied to grading accuracy, despite broader and sometimes conflicting theoretical perspectives (Black & Wiliam, 1998; Messick, 1994; Kane, 2010; Kunnan, 2004; Tierney, 2012). Importantly, these findings point to a pressing need for further empirical research in actual ESL classrooms to explore how teachers define and operationalize fairness in assessment, how these interpretations influence classroom assessment practices, and how assessment models can be adapted to support both equitable and context-sensitive evaluation of student learning.

All in all, this study makes a significant contribution to the field of language education by proposing an empirically grounded model of classroom assessment, directly built upon the lived experiences and professional practices of ESL teachers. It also, proposes a model for TAL that illustrates the developmental nature of teacher assessment literacy, showing how teachers' knowledge and beliefs evolve through ongoing reflection and adaptive engagement with classroom contexts. A pivotal finding is the strategic role of informal assessment. Teachers use it not only for evaluation but to navigate constraints and enact their beliefs when formal systems fall short, making it the key mechanism for aligning beliefs and practices amidst contextual challenges. Ultimately, this work highlights the enduring value of research that bridges theory and practice, advocating for more responsive assessment policies, reflective teacher education, and professional development programs that empower educators to conduct fair, effective, and context-sensitive evaluations of student learning.

References

Ahmad, N. S, (2021). *Teachers beliefs about assessment: a multiple narrative case study of internationally educated teachers assessment of speaking skills in Canadian ESL/EAP classrooms* (unpublished master's thesis, University of Toronto)

Alonzo, A. C. (2020). Defining trustworthiness for teachers 'multiple uses of classroom assessment results. In Brookhart, S. M., & McMillan, J. H. (Eds.) *Classroom assessment and educational measurement* (pp. 120-145). New York: Routledge.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for Teacher Competence in Educational Assessment of Students.* Washington, DC: Authors. Retrieved from www.unl.edu/buros/bimm/html/article3.html

Andrade, H. L., & Brookhart, S. M. (2021). Classroom assessment as co-regulated learning: A systematic review. *Frontiers in Education*, 6. p. 751168 https://doi.org/10.3389/feduc.2021.751168

Arkoudis, S., & O'Loughlin, K. (2004). Tensions between validity and outcomes: Teacher assessment of written work of recently arrived immigrant ESL students. *Language Testing*, *21*(3), 284-304.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly: An International Journal*, *2*(1), 1-34.

Bailey, A. L., & Durán, R. (2020). Language in practice: A mediator of valid interpretations of information generated by classroom assessments among linguistically and culturally diverse students. In Brookhart, S. M., & McMillan, J. H. (Eds.) In *Classroom assessment and educational measurement* (pp. 46-62). New York: Routledge.

Baker, A., & Burri, M. (2016). Feedback on second language pronunciation: A case study of EAP teachers' beliefs and practices. *The Australian Journal of Teacher Education, 41*(6), 1-19 http://dx.doi.org/10.14221/ajte.2016v41n6.1

Barkaoui, K., Valeo, A. (2019). *Writing Assessment in the ESL Classroom: Teachers Beliefs and Practices,* Unpublished research project, SHHRC

Basturkmen, H. (2012). Review of research into the correspondence between language teachers' stated beliefs and practices. *System*, 40, 282-295.

Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). TOEFL 2000 listening Framework (TOEFL monograph series, MS-19). *Princeton, NJ: Educational Testing Service*.

Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., ... & Urzua, A. (2004). *Representing language use in the university: Analysis of the TOEFL 2000 spoken and written academic language corpus.* Princeton, NJ: Educational Testing Service.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education. *Principles, Policy & Practice*, 5(1), 7–74.

Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability: The 103rd yearbook of the national association for the study of education, Part II* (pp. 20–50). Chicago: NSSE: Distributed by the University of Chicago Press.

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers ' summative assessments. *Assessment in Education: Principles, Policy & Practice*, *17*(2), 215-232.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8–21.

Bonner, S. M. (2012). Validity in classroom assessment: Purposes, properties and principles. In McMillan, J. H. (Ed.), *SAGE handbook of research on classroom assessment* (pp. 87-106). Los Angeles: Sage.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), 81–109.

Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London: Bloomsbury.

Borg, S. (2011). The impact of in-service education on language teachers 'beliefs. *System*, 39, 370–380.

Borg, S. (2018). Teacher's beliefs and classroom practices. In P. Garrett & G. M. Cots (Eds.), *The Routledge Handbook of Language Awareness* (pp. 75–91). Routledge. https://doi.org/10.4324/9781315676494-5

Boubris, A., & Haddam, F. (2020). Reading assessment: A case study of teachers 'beliefs and classroom evaluative practices. *Arab World English Journal (AWEJ) Volume, 11*.

Brindley, G. (2001). Language assessment and professional development. In C. Elder et al. (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126–136). Cambridge University Press.

Brookhart, S. M. (1993). Assessing student achievement with term papers and written reports. *Educational Measurement: Issues and Practice, 12*(1), 40–47.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment. *Educational Measurement: Issues and Practice,* 22(4), 5–12.

Brookhart, S. M., & McMillan, J. H. (2020). Perspectives on teachers ’assessment decisions. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 11–26). Routledge.

Brown, G. T. L. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*, (3), 301-318.

Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York, NY: Nova Science.

Brown, G. T., Gebril, A., & Michaelides, M. P. (2019). Teachers' conceptions of assessment: A global phenomenon or a global localism. *Frontiers in Education, 4*(16), 1–13. https://www.frontiersin.org/articles/10.3389/feduc.2019.00016

Buck, G. (2001). *Assessing listening*. Cambridge University Press.

Butler, F. A., Eignor, D., Jones, S., McNamara, T., & Suomi, B. K. (2000). *TOEFL 2000 speaking framework* (Vol. 20). Princeton, NJ: Educational Testing Service.

Campbell, C. (2013). Research on teacher competence in classroom assessment. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 71–84). SAGE.

Cheng, L. (2001). An investigation of ESL/EFL teachers’ classroom assessment practices. *Language Testing Update, 29*, 53-83.

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors ’classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21, 360–389.

Cheng, M. M., Chan, K. W., Tang, S. Y. F., & Cheng, A. Y. N. (2009). Pre-service teacher education students' epistemological beliefs and their conceptions of teaching. *Teaching and Teacher Education*, 25(2), 319–327.

Crusan, D., & Matsuda, P. K. (2018). Classroom writing assessment. *The TESOL Encyclopedia of English Language Teaching*, 1-7.

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Donald, E. P. (2005). *A teacher-verification study of speaking and writing prototype tasks for a new TOEFL* (TOEFL-MS-26). Princeton, NJ: Educational Testing Service.

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25, 327–348.

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305–334.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251-272.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies* (2nd ed.). Oxford University Press.

Farrell, T. S. C., & Guz, M. (2019). 'If I wanted to survive I had to use it': The power of teacher beliefs on classroom practices. TESL-EJ, 22(4), 1–17. https://tesl-ej.org/pdf/ej88/a4.pdf

Farrell, T. S., & Ives, J. (2014). Exploring teacher beliefs and classroom practices through reflective practice: A case study. *Language teaching research*, *19*(5), 594-610.

Fives, H., & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers ’beliefs. In K. R. Harris et al. (Eds.), *APA Educational Psychology Handbook*: Vol. 2 (pp. 471–499). American Psychological Association.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132.

Fulcher, G. (2003). *Testing second language speaking*. Longman.

Fulcher, G. (2015). Assessing second language speaking. *Language Teaching*, *48*(2), 198-216.

Ginther, A. (2013). Assessment of speaking. *The Encyclopedia of applied linguistics*, 1-8.

Golombek, P., & Doran, M. (2014). Unifying cognition, emotion, and activity in language teacher professional development. *Teaching and Teacher Education*, 39, 102–111.

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21, 607–621.

Graham, S., Santos, D., & Francis-Brophy, E. (2014). Teacher beliefs about listening in a foreign language. *Teaching and Teacher Education*, 40, 44–60.

Guadu, Z. B., & Boersma, E. J. (2018). EFL instructors 'beliefs and practices of formative assessment in teaching writing. *Journal of Language Teaching and Research*, 9(1), 42–50. https://doi.org/10.17507/jltr.0901.06

Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research papers in education*, *20*(3), 245-270.

Harris, L. R., & Brown, G. T. (2009). The complexity of teachers 'conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 365–381. https://doi.org/10.1080/09695940903319745

Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs*. http://www.ealta.eu.org/resources.htm

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin Press.

Heritage, M. (2012). Gathering evidence of student understanding. In McMillan, J. H. (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 179-195). Los Angeles: Sage.

Heritage, M. (2018). Making assessment work for teachers. *Educational Measurement: Issues and Practice, 37*(1), 39-41.

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational measurement: issues and practice*, *28*(3), 24-31.

Hill, H. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment*, 6(1), 1–17.

Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420.

Hoy, A. W., Davis, H., & Pape, S. J. (2006). Teacher Knowledge and Beliefs. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 715–737). Lawrence Erlbaum Associates Publishers.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, *25*(3), 385-402.

Isaacs, T. (2016). Assessing speaking. In D. Tsagari & J. Banerjee (Eds.), *Handbook of Second Language Assessment* (pp. 131–146). De Gruyter Mouton.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17–64). Praeger.

Kane, M. (2010). Validity and fairness. *Language testing*, *27*(2), 177-182.

Kane, M. T., & Wools, S. (2020). Perspectives on the validity of classroom assessments. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom Assessment and Educational Measurement* (pp. 11–26). Routledge.

Kunnan, A. J. (2004). Test fairness. *European language testing in a global context*, *18*, 27-48.

Kuzborska, I. (2011). Links between teachers' beliefs and practices and research on reading. *Reading in a Foreign Language*, 23 (1), 102-128.

Lado, R. (1961). *Language Testing: The Construction and Use of Foreign Language Tests*. A Teacher's Book.

Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81, 78–89. https://doi.org/10.1016/j.system.2019.01.006

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Springer. https://doi.org/10.1007/978-981-10-3924-9

Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly: An International Journal*, *1*(1), 19-41.

Leung, C. (2005). Classroom teacher assessment of second language development: Construct as practice. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 869-893). Mahwah, NJ: Lawrence Erlbaum.

Leung, C. (2014). Classroom-based assessment issues for language teacher education. In A. J. Kunnan (Ed.), *The Companion to Language Assessment* (pp. 1510–1519). Wiley-Blackwell.

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Singapore: Springer Singapore.

Li, L. (2017). *Social interaction and teacher cognition*. Edinburgh University Press.

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

McMillan, J. H. (2012). Why we need research on classroom assessment. In McMillan, J. H. (Ed.), *SAGE handbook of research on classroom assessment* (pp. 3-16). Los Angeles: Sage.

McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language testing*, *18*(4), 333-349.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational researcher*, *18*(2), 5-11.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational researcher*, *23*(2), 13-23.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(1), 49–64.

Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing*, *16*(2), 82–99. https://doi.org/10.1016/j.jslw.2007.04.002

Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13–25. https://doi.org/10.1111/j.1745-3992.2003.tb00140.x

Moss, P. A., Pullin, D., Gee, J. P., & Haertel, E. (2005). The idea of testing: Psychometric and sociocultural perspectives. *Measurement: Interdisciplinary Research and Perspectives*, 3, 63–83.

National Education Association. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30–32.

Özdemir-Yılmazer, M. & Özkan, Y. (2017). Speaking assessment perceptions and practices of English teachers at tertiary level in the Turkish context. *Language Learning in Higher Education, 7(2), 371-391. https://doi.org/10.1515/cercles-2017-0015*

Parr, J. M. (2012). Classroom assessment in writing. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 527–544). SAGE Publications

Phipps, S., & Borg, S. (2009). Exploring tensions between teachers ʾgrammar teaching beliefs and practices. *System*, 37(3), 380–390.

Plake, B. S., & Impara, J. C. (1993). *Teacher assessment literacy questionnaire*. University of Nebraska-Lincoln. In cooperation with the National Council on Measurement in Education and the WK Kellogg Foundation.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48, 4–11.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18, 429–462.

Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing*, 17, 215–243.

Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning-focused interpretation. *International Journal of Applied Linguistics* 16: 163–88

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. *Handbook of research on teacher education*, *2*(102-119), 273-290.

Sanchez, S. H. (2014). The impact of self-perceived subject matter knowledge on pedagogical decisions in EFL grammar teaching practices. *Language Awareness*, 23(3), 220–233.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309–327.

Sheppard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.

Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1066- 1101). Washington, DC: AERA.

Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 624–646). Westport, CT: Praeger.

Shepard, L. A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice, 28*(3), 32-37.

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Classroom assessment principles to support learning and avoid the harms of testing. *Educational Measurement: Issues and Practice, 37*(1), 52–57. https://doi.org/10.1111/emip.12195

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.

Skott, J. (2014). The promises, problems and prospects of research on teachers' beliefs. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 13–30). Routledge.

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5–14.

Stiggins, R. (2010). Essential formative assessment competencies for teachers and school leaders. In *Handbook of formative assessment* (pp. 233-250). Routledge.

Stiggins, R. J. (2012). *Classroom assessment competence: The foundation of good teaching*. Retrieved from http://images.pearsonassessments.com/images/NES_Publications/2012_04Stiggins.pdf

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36.

Taylor, L., & Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalising the test construct. *Journal of English for Academic Purposes*, 10, 89–101.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp.127-146). New York: MacMillan.

Thompson, I. (1995). Assessment of second/foreign language listening comprehension. In D. Mendelsohn & J. Rubin (Eds.), *A Guide for the Teaching of Second Language Listening* (pp. 31–58). Dominie.

Tierney, R. D. (2012). Fairness in classroom assessment. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 125–144). Sage.

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41–63.

Vandergrift, L. (2011). Second language listening: Presage, process, product, and pedagogy. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 455–471). Routledge.

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402.

Volante, L., & Fazio, X. (2007). Exploring teacher candidates ʾassessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749–770.

Wagner, E. (2010). Test-takers ʾinteraction with an L2 video listening test. *System*, 38, 280–291.

Wang, L., Lee, I., & Park, M. (2020). Chinese university EFL teachers' beliefs and practices of classroom writing assessment. *Studies in Educational Evaluation, 66*(1), Article 100890. https://doi.org/10.1016/j.stueduc.2020.100890

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

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualizing teachers ʾassessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241–256.

Wilson, M. (2018). Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues & Practice*, 37(1), 5–20.

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162.

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 493–513.

Zheng, H. (2013a). Tensions and relations: Exploring the dynamic relationship between teachers ' EFL teaching and learning beliefs and practices. In X. Zhu & K. Zeichner (Eds.), *Preparing Teachers for the 21st Century: New Frontiers of Educational Research* (pp. 395–407). Springer.

Zheng, H. (2013b). The dynamic interactive relationship between Chinese secondary school EFL teachers 'beliefs and practice. *The Language Learning Journal*, *41*(2), 192–204. https://doi.org/10.1080/09571736.2013.790133

Appendix A

Demographic questionnaire

Name:

Preferred pseudonym:

Level of education:

Please list all the degrees you are holding. Please start from the most recent one.

What ESL/EFL teacher education program did you complete? When did you complete it and where?

Teaching experience/ number of the years you have taught ESL:

What course(s) are you teaching currently?

What courses have you taught in the past?

How many years/terms have you taught speaking & listening?

Please briefly describe the program where you currently work.

Please briefly describe the course (the listening & speaking course/the course that includes oral communication) you have been teaching.

What is your current employment status?

Appendix B – Teacher Consent Form

Participant information sheet for the Research Project:

Exploring the relationship between ESL teachers' assessment literacy, conceptions and practices in assessing oral communication in the ESL classroom

Researcher:
Amir Rezaei, Ph.D. Candidate, Department of Languages, Literatures and Linguistics,
York University
Email: arezaei@yorku.ca

You are invited to take part in this research study. The following explains why the research is being done and what it involves.

What is the purpose of this study?

The general goal of this study is to explore what assessment practices ESL teachers do in the classroom for assessing oral communication. It further explores the relationship between their assessment knowledge, their assessment conceptions and practices in assessing L2 oral communication in the ESL classroom.

The research will be presented to York University as for the fulfilment for the degree of Doctor of Philosophy. The results of the study will be submitted for publication to a number of peer-reviewed journals. Also, the findings will be presented at some conferences in the field of second language teaching and assessment.

What will you be asked to do in the research?

After signing the consent form, you will be asked to participate in 4 interviews about your assessment literacy, your conception of L2 assessment and your assessment practices in assessing L2 oral communication in the classroom. Each interview will take about 40-60 minutes. You are also asked to allow the researcher to observe 1-2 of your classes. After each observation, you will be asked to share your thoughts about your classroom practices with the researcher. The estimated time for the follow-up interview (after observation) is 30 minutes. You are free to withdraw from this research at any time.

What are the likely benefits and costs of taking part?

Each interview will take approximately 40-60 minutes. The interaction and discussion with you as participant will be useful for you by helping you become aware of your assessment literacy and your conception of second language assessment. Your awareness of your assessment conceptions can help you understand your classroom practices better and can help you to understand the reasons

behind most of your classroom assessment practices. In appreciation of the time you spend for participating in this research, you will be given a 150-dollar gift card.

Voluntary participation
Your participation in the research is completely voluntary and you may choose to stop participating at any time. Your decision not to continue participating will not influence your relationship or the nature of your relationship with researcher, faculty or with staff of York University either now or in the future.

Withdrawal from the study
You may stop participating in the study at any time, for any reason, if you so decide. Your decision to stop participating, or to refuse to answer particular questions, will not affect your relationship with the researcher, York University, or any other group associated with this project. In the event that you withdraw from the study, all associated data collected will be immediately destroyed wherever possible.

Confidentiality

Confidentiality will be provided to the fullest extent possible by law. Your contributions to this research will be treated and stored confidentially as required by the Research Ethics Code of York University. Any publications from this study will protect participants' anonymity. Throughout the interview a pseudonym will be used instead of your real name and in any report or publication of the results your real name will not be used. All the data recorded will be safely stored in the researcher's personal computer only accessible by the researcher. After analysing the data and reporting the results, all the stored data (rubrics, assessment samples and any other relevant data source) will be destroyed by 10/01/2024.

Questions about research

If you have questions about the research in general or about your role in the study, please feel free to contact Dr. Khaled Barkaoui either by telephone at (416)736-2100, extension 33209 or by e-mail (kbarkaoui@edu.yorku.ca). This research has received ethics review and approval by the Human Participants Review Sub-Committee, York University's Ethics Review Board and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines. If you have any questions about this process, or about your rights as a participant in the study, please contact the Sr. Manager & Policy Advisor for the Office of Research Ethics, 5th Floor, Kaneff Tower, York University (telephone 416-736-5914 or e-mail ore@yorku.ca).

Legal Rights and Signatures:

I……consent to participate in ……conducted by …… I have understood the nature of this project and wish to participate. I am not waiving any of my legal rights by signing this form. My signature below indicates my consent.

Signature_____ Date_____

Participant

Signature **Date**

Principal Investigator

Additional consent (where applicable)

1. **Audio recording**

☐ I consent to the audio-recording of my interview(s).

2. **Video recording or use of photographs**

1. I _____ consent to the use of images of me (including photographs, video and other moving images), my environment and property in the following ways (please check all that apply):

In academic articles	[ ] Yes	[ ] No
In print, digital and slide form	[ ] Yes	[ ] No
In academic presentations	[ ] Yes	[ ] No
In media	[ ] Yes	[ ] No
In thesis materials	[ ] Yes	[ ] No

Signature: Date:

_____ _____

Participant: (name)

3. **Consent to waive anonymity**

I, _____ , consent to the use of my name in the publications arising from this research.

Signature: Date:

_____ _____

Participant: (name)

Appendix C: Assessment Literacy in Practice Questionnaire (ALiPQ)

Thank you for taking part in this survey. Your answers are very valuable for completing this important study on teachers' assessment literacy. The main purpose of the survey is to understand your assessment literacy, conceptions of assessment and assessment practices in the ESL classroom. The questionnaire will take approximately 45 minutes of your time.

Your individual privacy and confidentiality of information you provide will be maintained at all stages of the study.

Assessment literacy

How knowledgeable do you think you are about each aspect of second language assessment below. Please rate your skill and awareness of each concept on a 4-point Likert scale. Use the following scale for responding:

0= Not at all; 1= Slightly; 2= Moderately; 3= A lot;

Knowledge base	Questionnaire Items	Rating
1: Purpose (formative, summative, etc.)	- Using assessment for different purposes - Using assessment to improve teaching and learning - Using assessment for assessing what students have learned. - Using assessment to diagnose students' strengths and needs	0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3
2: Assessment tools (Rubrics, objective testing, holistic vs analytical assessment, observations)	- Designing and using rubrics for assessing L2 learning - Using different tests including multiple choice and open-ended tests to assess L2 students' learning - Using statistics to analyze assessment results - Selecting and designing appropriate test items or tasks for a specific assessment purpose	0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3
3: Feedback (methods/types of feedback)	- Using different types of feedback (e.g., direct/indirect, implicit/explicit, etc.) - Giving useful feedback on the basis of an assessment	0 1 2 3 0 1 2 3
4: Assessment interpretation and communication (ability to use different evidence to assess students' progress and ability	- Interpreting students' scores when assessing their progress or their ability - Using different types of evidence to assess students' learning and progress - Using assessment data/evidence to help students learn - Using assessment data/evidence to modify my plans for instruction	0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3

5: Student involvement in assessment (student engagement, self-regulation, self/peer assessment)	- using peer assessment and self-assessment in the classroom - Using assessment to engage students and help them become self-responsible - Helping students self- monitor and, subsequently, self-regulate	0 1 2 3 0 1 2 3 0 1 2 3
6: Assessment ethics (how to deal with plagiarism, reporting results, equity and social justice)	- Dealing with plagiarism - Reporting assessment results and decisions to the students - Understanding assessment ethics and ethics in my profession as an assessor - Considering the consequences of my assessment decisions	0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3
7: Knowledge of different Epistemological views of L2 teaching and assessment	- Understand the philosophy behind the design of quality language assessments - Understand different assessment theories	0 1 2 3 0 1 2 3 0 1 2 3
8: Knowledge of students' conceptions	- Awareness of the differences between students' and teachers' beliefs about assessment - Awareness of students' preconceptions/misconceptions about assessment - Addressing students' beliefs for effective assessment - Awareness that my beliefs about assessment might conflict with students' beliefs	0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3
9: Contextual constraints and classroom dynamics	- Awareness of the influence of contextual factors on assessment practices in the classroom - Awareness of different contextual constraints and how they might influence a teacher's assessment practices	0 1 2 3 0 1 2 3
10: Formal vs informal assessment	- Using observations to assess my students' progress and weaknesses - Using my intuition for diagnosing my students' weaknesses and strengths - Developing formal exams (e.g., mid-term, final exams) - Using assessment tasks embedded in instruction that are not graded.	0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3
11: Practical vs theoretical knowledge (implementation of knowledge in the classroom, gap between these two, etc.)	- Awareness of the gap between theoretical and practical knowledge of assessment. - Awareness of the challenges in implementing theoretical knowledge in the classroom. - Applying my theoretical knowledge in my assessment practices in the classroom.	0 1 2 3 0 1 2 3 0 1 2 3

12: Knowledge of complexity in assessment practices and ability to balance the tensions	- Dealing with the challenges encountered in L2 assessment in the classroom - Balancing the conflicts in teaching, learning and assessment - Addressing the conflicts between my assessment knowledge, my beliefs and contextual constraints - Balancing the conflict between my assessment beliefs and the beliefs of others	0 1 2 3 0 1 2 3 0 1 2 3 0 1 2 3

Appendix D: Interview 1

Past education:

Can you tell me about your past education? What undergraduate or graduate degree(s) related to L2 teaching do you hold?

Did you take any course related to teaching oral communication?

If yes,

- What aspect of teaching oral communication was emphasized during the course?
- What approach(s) was/were practiced in the class?

Did you take any course related to second language assessment/evaluation during your past studies? Please describe the course(s) and explain what it/they addressed.

Did you study any theories/approaches of L2 assessment during those courses?

- If yes, do you remember some of the theories/approaches you studied?
- Which type of assessment or assessment approach was emphasized in the course?
- Do you follow or practice any of these approaches in your ESL classes?

Did any of the courses include assessing L2 oral communication?

-If yes, what type of assessment was addressed for assessing L2 oral communication?

Have you taken any training course on language assessment? (e.g., TESL Canada, TESL Ontario, etc.)

If yes,

- Did it address different approaches to assessment? (e.g., Formative vs. summative assessment)
- Did it address assessing L2 oral communication in the course?
- What aspects of assessing L2 listening/speaking were addressed?

Did you take any practicum during the program?

If yes,

- Was assessment addressed during your practicum? What aspect of assessment was practiced during your practicum?

Did the program have any impact on your beliefs and knowledge about language teaching in general and language assessment in particular?

Do you think that teacher training programs prepare teachers for effective assessment in the ESL classroom?

- What are some of the drawbacks of such programs?

Do you have any suggestions for the improvement of assessment courses in teacher training programs?

Teaching experience and courses:

About the L2 oral communication course you teach currently, please tell me:

- What are the goals of the course?
- Does the course include a final and mid-term exam?

  If yes, please explain what assessment tools do you use in the exams?

What specific goals do you follow to help your students to improve their oral communication in this course?

What kind of oral communication tasks do students do in this course?

Are there any specific language skills that you want your students to follow? Please explain what they are and why.

How do you help your students improve their oral communication in English?

Do you follow a specific approach for teaching oral communication in this course? Please describe your approach.

Do you combine SP & LS in your teaching or teach them separately?

- Do you assess them in a joint way or separately? Why/why not? How?
- If yes, please explain how you combine SP & LS in your teaching and your assessment?

Assessment practices in the classroom

How often do you assess your students' speaking/listening in the class? Please explain why.

What do you assess when you are assessing your students' oral communication? Please briefly describe any formal method/strategy you use for assessing L2 oral communication.

- Do you assess specific structures, features, strategies students use to listen/speak?
- Do you assess specific sub-skills or strategies that students use?
- Do you assess top-down/bottom-up strategies students use to process listening texts?

Do you assess your students' SP & LS diagnostically? (to identify their listening/speaking strengths and needs)

- How do you use the results of your diagnostic assessment for teaching SP & LS?

What methods or evidence do you use to assess your students' progress in speaking/listening in your class? Please provide specific examples.

Do you use the results of assessment for the teaching of listening/speaking?

- If yes, how do you use the result of your assessment for your teaching of listening/speaking? (lesson planning, teaching methodology, classroom activities, focusing on specific sub-skills, structure, etc.)

Is there any conflict/challenge between your ability to assess listening/speaking and your ability to use the results in teaching them?

Do you share the assessment results with your students, colleagues, coordinator, etc.?

- If yes, how do you share them? Why?
- Do you do any follow up activities?

Do you evaluate your students' oral communication during class? How often do you evaluate them? When do you evaluate them?

- How do you assess their performance or progress during classes? Please explain the strategies/activities you use.
- Does this evaluation influence your grading when you are assigning grades to their assignments, and exams?

Do you ever use informal assessment? (spontaneously or intuitively)

- If yes: Please provide specific examples (explain when, how and why you use them).

What strategies/activities do you use to assess your students' oral communication informally? Please provide concrete examples.

- How does it help you to assess your students' oral communication ability and their needs?
- How do you use the result of your informal assessment?

Do you think that informal assessment is an effective way of assessing L2 oral communication? Please explain why/why not?

Are you guided by any theory or belief for using informal assessment in your class?

Do you allocate any marks for classroom participation?

If yes:

- What does participation mean for you?
- Do you consider your classroom observation and/or informal assessment in assigning grades for classroom participation?

Do you teach SP & LS separately or together? Please explain why/why not.

Do you assess them together or separately? Please explain why you assess them separately or together? Are there times when you assess them separately and times when you assess them together? Please explain when and why.

If you assess them together,

- How do you assess these skills together?
- What sub-skills/features do you assess in each skill?
- How do you use the results of your assessment for teaching?

In the case of integrated assessment, how does integrated assessment help you understand the SP & LS needs of your students?

- How do you ensure the effectiveness/usefulness of your assessment?
- Do you use other evidence besides students' interaction for assessing listening/speaking or diagnosing their listening/speaking problems?

Is there a specific theory that guides you in assessing listening & speaking together?

What challenges have you experienced in assessing listening & speaking together?

Do you think that your assessment practices of speaking/listening have changed over time?

- If yes, please explain how and what factors have caused the change.

Which of the following factors have had a great impact on changing your assessment of oral communication practices? Teaching experience, training courses, education.

Appendix E: Interview 2

Effectiveness and validity of teachers' assessment practices:

How do you assess the effectiveness, and fairness, of your assessment methods and strategies in the classroom?

- Do you use any evidence to evaluate the effectiveness of your assessment in the classroom? Please explain it with an example.

How do you ensure that you have made the right decisions in your assessment practices in the classroom?

- What sources of information/evidence do you use to ensure the validity/credibility of your assessment?
- Please describe each source/evidence briefly and explain how it helps you to come up with a valid assessment.

What beliefs guide you in ensuring the validity of your assessment decisions in the classroom?

Have you ever changed your assessment decisions in different classes? What caused the change?

In general, how effective do you think the listening/speaking assessment procedures and tools you are using are?

- Why do you think they are effective or not effective?

In general, how do you evaluate, and ensure the credibility/ appropriateness of the listening/speaking assessment tools and procedures you use and their results? Please explain and give specific examples.

Are there any other methods or strategies that you believe are effective to assess L2 listening/speaking?

Contextual factors:

Does your institution have a specific policy and guidelines you are required to follow for assessing oral communication? Please explain.

Are there any tools, rubrics or other tools you are required to follow for teaching and assessing oral communication? Please explain.

Does your institution consider students' evaluation of teachers in offering positions to different instructors?

- Does students' evaluation of you/your class affect your assessment and grading? Please explain.

What contextual factors influence your assessment decisions the most in the classroom? (e.g., classroom size, students' diversity and level, school policies and standards, students' evaluation of teachers, etc.)

Tensions:

Have you encountered any challenges in assessing your students in general?

- How did you react to/cope with the challenge(s)?

Have you ever experienced any conflict between your beliefs in assessment and your classroom practices?

- What was it?
- How did you address the challenge?

Have you practiced something that was against your beliefs about assessment?

- What was the reason you did not implement your beliefs in the classroom?
- Is there any policy at your program/institution you find contradictory to your beliefs in assessing speaking/listening? What is it? Please give an example and explain.

Please tell me if you have experienced any other conflicts between your assessment knowledge, beliefs, teaching and your assessment practices?

Have you ever felt challenged when assessing listening/speaking in your class?

If yes,

- What was the challenge?
- What caused the challenge?
- How did you deal with it?

Have you experienced any challenges in applying the results of your assessment of listening/speaking in teaching them?

- How did you react to it? What decision did you make?

Have you been challenged by specific factors when making assessment decisions? Please give specific examples. (e.g., contextual factors, individual differences, your assessment beliefs, etc.)

Have you ever been challenged by the complexity and interaction between different factors for practicing a specific method in assessing oral communication?

If yes,

- What was the challenge?
- How did you compromise the complexity or balance the interplay between different factors?

Have you experienced any challenge in providing feedback or getting feedback from your students in oral communication class? Please explain

- What decision did you make?

Do you consider your students' ideas in assessing oral communication? Please provide some examples from your classroom practices.

Have you experienced any conflict between your assessment approaches and students' interests or conceptions of assessment?

If yes,

- what was it?
- What decision did you make?

Have you experienced any conflict(s) with other teachers at your institution regarding assessment beliefs, approaches and practices?

Do you see any conflicts in your dual role as a teacher and an assessor? Please explain.

Teachers' conceptions

Do you follow a specific approach for teaching ESL? Please describe it.

- Is it guided by any theory?
- Do you think it is an effective approach? Please explain why.

Do you have a specific philosophy in teaching a second language? Please explain.

What do you think the roles of the teacher and the learner are in relation to learning a second language? Please explain the role of each separately.

What does learning to speak and listen in a second language involve?

How would you describe your personal philosophy of teaching L2 oral communication?

What aspects of oral communication do you think teachers should teach in ESL classes?

What is the role of the teacher in relation to assessment?

Do you have a specific philosophy for assessing L2?

- Is it based on a specific theory? Pleas explain

Do you follow a specific approach in assessing L2? If yes, please describe it.

What goal(s) do you follow in assessing oral communication? (e.g., improving learning, making pedagogical decisions, meeting the requirements of your institution, etc.)

In your view, what is the relationship between learning, teaching, and assessment?

Have your beliefs about assessing oral communication changed over the years?

- If yes, please provide concrete examples and explain how they have changed.

What experiences (including ongoing experiences) have influenced your beliefs about assessment and your assessment practices the most?

- What experience has influenced your beliefs and practices in assessing oral communication the most?

Which of the following have had the greatest impact on your beliefs about L2 assessment in general and the assessment of oral communication in particular? Your teaching experience, past education, or learning.

Teachers' beliefs about L2 teaching and assessment

The following questions address your beliefs about teaching and assessment. Please mark your answer using the following 4-point scale:

*Strongly Disagree= **SD**; Disagree= **D**; Agree= **A**; Strongly Agree= **SA***

Please indicate whether you agree or disagree with each of the statements below:

a. Knowledge is stable SD D A SA
b. Knowledge changes and evolves SD D A SA
c. Knowledge is made of discrete and simple facts SD D A SA
d. Knowledge is complex and comprises interrelated concepts SD D A SA
e. Knowledge resides outside the 'self' and is transmitted SD D A SA
f. Knowledge is constructed by the 'self' SD D A SA
g. What makes a knowledge claim sufficient is the authority of its source SD D A SA
h. What makes a knowledge claim sufficient is observation and evaluation of expertise SD D A SA

Please indicate whether you agree or disagree with each of the statements below about the role of the ESL teacher in the classroom:

a. The ESL teacher should be a facilitator SD D A SA
b. The ESL teacher should be an authority SD D A SA

c. The ESL teacher should be a source of knowledge SD D A SA

Please indicate whether you agree or disagree with each of the statements below concerning the relationship between assessment and teaching:

- A. assessment and teaching are interrelated SD D A SA
- B. teaching should be adjusted to the goals of assessment SD D A SA
- C. assessment should be adapted to teaching goals SD D A SA
- D- assessment is the end point of pedagogy SD D A SA
- E. assessment provides evidence of knowledge and competence. SD D A SA

Please indicate the extent to which you agree or disagree with each of the statements below concerning the main purpose of assessment.

a. Assessment should be used to support/promote learning SD D A SA
b. Assessment should be used to evaluate learning SD D A SA
c. Assessment should be used for comparing students in terms of learning SD D A SA
d. Assessment should be used for making fail/pass decisions SD D A SA
e. Assessment should be used for diagnosing students' needs SD D A SA

Please indicate the extent to which you agree or disagree with each of the statements below concerning the nature of assessment?

a. Multiple assessments lead to more valid decisions SD D A SA

b. Language skills should be broken into smaller components SD D A SA
and assessed separately.

c. Different language skills should be assessed together SD D A SA

Appendix F: Interview 3

Assessment tools

Tasks:

Generally, what type of tasks do you use for assessing oral communication?

- Please briefly explain why you use each task.

Do you have any beliefs about how to select or design tasks for assessing oral communication?

How did you develop the ideas for designing/selecting this task?

What were your goals in using this task?

How did you select this task for assessing oral communication?

a. What factors, principles, and criteria influenced your use of this task?

How did you evaluate the effectiveness of this task for assessing oral communication? (limitations, student feedback, their progress)

Do you consider students' future use of language or target language use when you are using or selecting tasks for assessing oral communication?

Have you encountered any challenges in using this task in the classroom?

- If yes, please explain what it was.
- How did you react to the challenge?

Tests:

What assessment tools or procedures do you use for assessing your students' listening/speaking?

- Please describe different tools you use at the beginning, during the term or at the end of the term.

Why do you select and design these assessment tools for assessing speaking/listening?

Do you use any standardized tests, tests prescribed by your institution or your own tests for assessing speaking/listening?

- Please describe them and explain why you use them.
- Please describe the procedure for the development or selection of each of these tests.

Speaking particularly of this test, did you design this test yourself or adapted/adopted it from other sources?

- Why did your select/adopt this test? Please explain.

- If you developed it yourself, please explain how you developed it.

How does this help you to come up with fair marks for your students' listening/speaking?

Have you ever changed the test? Please explain how and why.

Do you think that it is an effective tool to assess your students' listening/speaking abilities and their learning needs? Please explain how and why.

How did you develop the ideas for using/developing this test?

Rubrics:

Do you use any rubrics for assessing your students' oral communication? Please explain why, when and how you use them?

Did you design this rubric yourself or adapted/adopted it from another source?

- If you developed it yourself, please explain how you developed it and why.

Please describe the criteria on the rubric briefly and explain why you use this rubric for assessing oral communication?

How does the rubric help you evaluate your students' oral communication in a fair and valid way?

Do you discuss their performance after assigning scores using this rubric?

- Please explain why/why not.

How did you develop the ideas and practices for developing and using the rubric?

Do you think that it is an effective tool to assess your students' oral communication ability and their learning needs? Please explain how and why.

Is there any belief and/or theory which guide(s) you in relation to how you assign final grades?

Feedback:

In this part of the interview, I would like to know more about your approach to providing feedback on your students' oral communication.

Do you have a general approach to providing feedback on your students' oral communication?

- If yes, please explain what it is and why do you think that it is a good approach.

How has your approach to responding to students' oral communication been formed? (e.g., past education, training courses, experience, community of practice).

- Is there any theory/belief underpinning your approach to providing feedback?

What features/criteria are important for you while assessing and providing feedback in oral communication tasks?

Do you use different types of feedback for different students in different classes? (e.g., explicit, implicit, prompts, recasts, ZPD etc.,)

Is there any specific type of feedback you prefer to use when assessing oral communication?

- What type of feedback is it?
- Why do you think this type of feedback is effective?

What challenge(s) have you experienced in providing feedback to your students in oral communication class?

Rate your needs for further training in the following assessment approaches and practices on the 3-point scale below.
0= Not at all, 1= Somewhat, 2= A lot

- Training on Portfolio assessment 0 1 2
- training on performance assessment 0 1 2
- training on peer and self-assessment
- training on implementation of assessment theories 0 1 2
  in the classroom
- training on test and task design 0 1 2
- training on using statistics to compute the validity 0 1 2
  and reliability of tests
- training on observation techniques/ informal assessment 0 1 2
- Any other approach? Please add other approaches that you need more training.
  …….

Suggest 1-2 way(s) that teacher education programs or university instruction can help you improve your assessment literacy. …….

Appendix G

Informed Consent Form (Student participants)

Study name: Exploring the relationship between ESL teachers' assessment literacy, conceptions and practices in assessing oral communication in the ESL classroom

Researcher name: Amir Rezaei, Ph.D. Candidate, Department of Languages, Literatures and Linguistics, York University

Contact: South Ross building, S402A, Office Phone # 416-736-5134 Ext 20270, Email:arezaei@yorku.ca

Purpose of the research:

The general goal of this study is to explore what assessment practices ESL teachers do in the classroom for assessing oral communication. It further explores the relationship between their assessment knowledge, their assessment conceptions and practices in assessing L2 oral communication in the ESL classroom.

The research will be presented to York University as for the fulfilment for the degree of Doctor of Philosophy. The results of the study will be submitted for publication to a number of peer-reviewed journals. Also, the findings will be presented at some conferences in the field of second language teaching and assessment.

What will you be asked to do in the research?

I am doing a study on teachers' assessment practices in the classroom; therefore, your teacher will be the focus of research. As a part of this research, I will observe, and video-/audio record your class. Your teacher has agreed to be observed while she/he teaches. While the focus of observation is your teacher, I need your permission too to observe and record the class.

The data will be used for research purposes only and your name will not appear in any part of the research. Meanwhile the recorded video-/audio file will not be shared with anyone.

If you feel uncomfortable with video recording, I can place the camera in a place that you are not video-taped.

Risks and discomforts:

The focus of the study is on your teacher and it will not impose any risks on you. If you feel uncomfortable with video recording, I will place the camera in a place that you are not video-taped.

Also, if you are not comfortable with audio recording and your voice is recorded during classroom interaction, I will delete your voice from the recorded audio files using an audio file editing

software. All the data and documents will be saved on a secure computer accessed by no one but the researcher.

Benefits of the research and benefits to you:

Your contribution will be useful for improving the quality of ESL teaching which directly impacts ESL students' learning. The findings will also make teachers more aware of their beliefs, assessment literacy and classroom practices; therefore, they can help teachers improve their assessment practices in the classroom.

Voluntary participation:

Your participation in the study is completely voluntary and you may choose to stop participating at any time. Your decision not to volunteer, to stop participating, or to refuse to answer particular questions will not influence the nature of the ongoing relationship you may have with the researchers or study staff, or the nature of your relationship with York University either now, or in the future.

Confidentiality:

Confidentiality will be provided to the fullest extent possible by law. Your contributions to this research will be treated and stored confidentially as required by the Research Ethics Code of York University. Any publications from this study will protect participants' anonymity. All hard copies of the data (assessment samples, interview transcripts and consent forms) will be shredded by a shredding machine by 10/01/2024. All stored electronic data (video/audio recordings and any other relevant data source) will be destroyed and deleted from the USB key and any other software used to facilitate data analysis by the date set above.

Questions about the research:

If you have questions about the research in general or about your role in the study, please feel free to contact Dr. Khaled Barkaoui either by telephone at (416)736-2100, extension 33209 or by e-mail (kbarkaoui@edu.yorku.ca). This research has received ethics review and approval by the Human Participants Review Sub-Committee, York University's Ethics Review Board and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines. If you have any questions about this process, or about your rights as a participant in the study, please contact the Sr. Manager & Policy Advisor for the Office of Research Ethics, 5th Floor, Kaneff Tower, York University (telephone 416-736-5914 or e-mail ore@yorku.ca).

I ……. , consent to participate in the study: *Exploring the relationship between ESL teachers' assessment literacy, conceptions and practices in assessing oral communication in the ESL classroom* conducted by Amir Rezaei. I have understood the nature of this project and wish to participate. I am not waiving any of my legal rights by signing this form. My signature below indicates my consent.

Signature _____ Date _____

(Participant) *Your email address: ……..*

Signature _____ Date _____

(Researcher)

Additional consent (where applicable)

Audio recording

☐ I consent to the audio-recording of my interview(s).

Video recording or use of photographs

I _____ consent to the use of images of me (including photographs, video and other moving images), my environment and property in the following ways (please check all that apply):

In academic articles	[ ] Yes	[ ] No
In print, digital and slide form	[ ] Yes	[ ] No
In academic presentations	[ ] Yes	[ ] No
In media	[ ] Yes	[ ] No
In thesis materials	[ ] Yes	[ ] No

Signature: Date:

_____ _____

Participant: (name)

Appendix H

Observation guidelines

Name of participant:

Time and date of observation:

Activity/task:

What assignment/lesson/activity is the teacher practicing?

What material/tools is the teacher using?

What does the teacher do before the activity?

- How does he elicit speaking performance?

What is the teacher doing while students are doing the task?

- Does he/she monitor students 'performance?
- Does he/she assign any marks for the activity?
- How do students respond?

What discussions are taking place?

- What caused the interaction between them?
- How did the teacher behave?

Assessment:

What does the teacher do after the activity?

Does he do any formal/informal evaluation?

- Formal:
- Informal:

Does he provide any assistance or feedback while they are performing or after performance?

Instances of assessment opportunities:

- Formal
- Informal

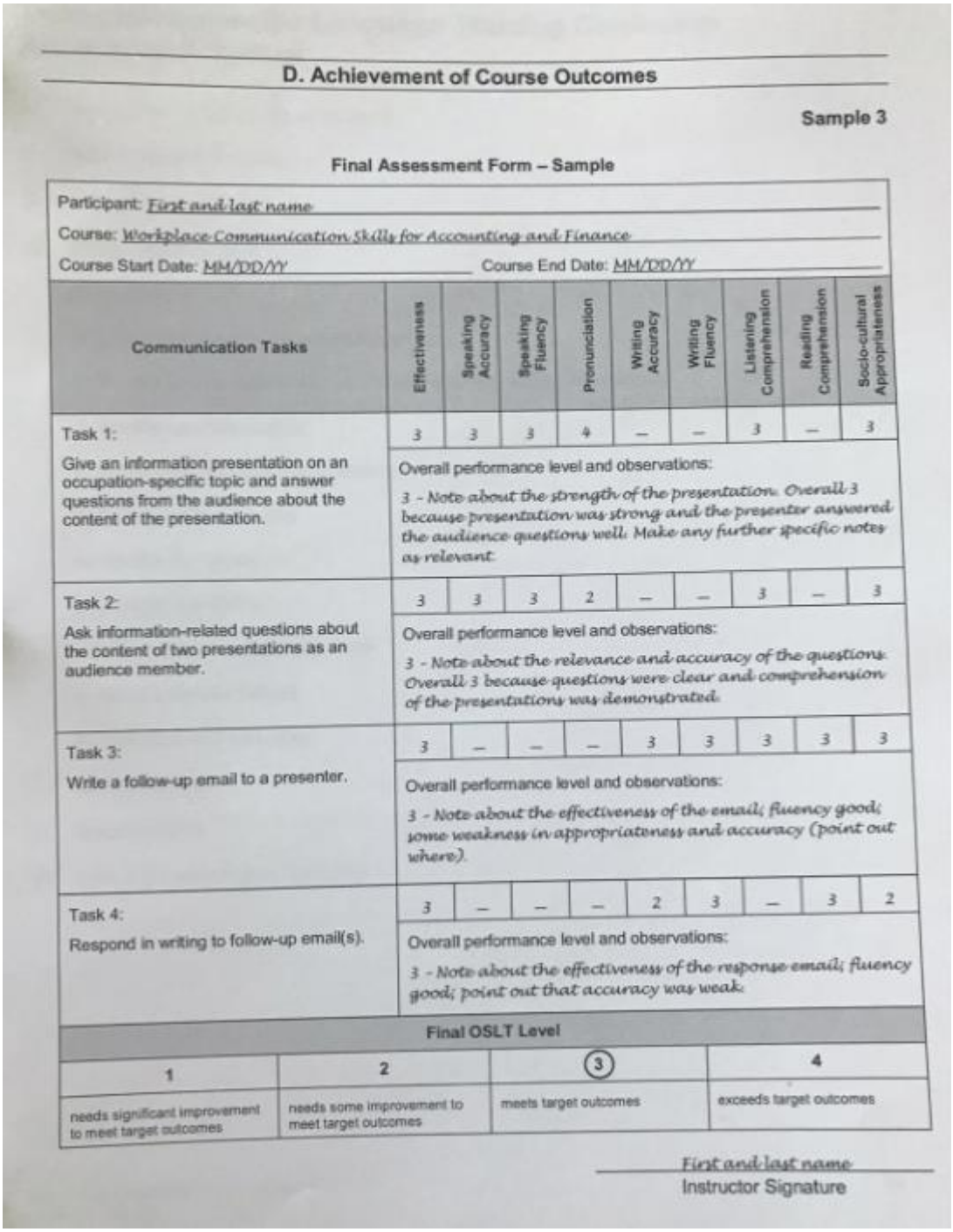
Instances of possible connections between the teacher's stated beliefs during interviews and his/her practices? (To be furthered inquired during SR)

Appendix I

Stimulated recall questions

- Can you tell me what were the general goals of the lesson?
- Can you tell me about this part of the lesson?
- What were your goals in doing this task/activity/assignment?
- What does this task/activity/assignment tell you about your students' oral communication ability?
- Did the segment work out the way you had planned? Please explain why/why not?
- Do you think it is an effective way to assess your students' oral communication? Why/why not?
- How do you use the results of your assessment in this assessment/task in your teaching of oral communication?
- Do you have any other thoughts that you would like to share with me about the selected segment?

Appendix J: Final assessment form (Ryan-Doc≠ 1)



D. Achievement of Course Outcomes

Sample 3

Final Assessment Form – Sample

Participant: *First and last name*

Course: *Workplace Communication Skills for Accounting and Finance*

Course Start Date: MM/DD/YY Course End Date: MM/DD/YY

Communication Tasks	Effectiveness	Speaking Accuracy	Speaking Fluency	Pronunciation	Writing Accuracy	Writing Fluency	Listening Comprehension	Reading Comprehension	Socio-cultural Appropriateness
Task 1: Give an information presentation on an occupation-specific topic and answer questions from the audience about the content of the presentation.	3	3	3	4	–	–	3	–	3
Overall performance level and observations: *3 - Note about the strength of the presentation. Overall 3 because presentation was strong and the presenter answered the audience questions well. Make any further specific notes as relevant.*									
Task 2: Ask information-related questions about the content of two presentations as an audience member.	3	3	3	2	–	–	3	–	3
Overall performance level and observations: *3 - Note about the relevance and accuracy of the questions. Overall 3 because questions were clear and comprehension of the presentations was demonstrated.*									
Task 3: Write a follow-up email to a presenter.	3	–	–	–	3	3	3	3	3
Overall performance level and observations: *3 - Note about the effectiveness of the email; fluency good; some weakness in appropriateness and accuracy (point out where).*									
Task 4: Respond in writing to follow-up email(s).	3	–	–	–	2	3	–	3	2
Overall performance level and observations: *3 - Note about the effectiveness of the response email; fluency good; point out that accuracy was weak.*									

Final OSLT Level

1	2	(3)	4
needs significant improvement to meet target outcomes	needs some improvement to meet target outcomes	meets target outcomes	exceeds target outcomes

First and last name
Instructor Signature

Appendix K: Questions for Bloom's Taxonomy (Aly- Doc≠ 1)

QUESTIONS FOR THE REVISED BLOOM'S TAXONOMY

(from Quick Flip Questions for the Revised Bloom's Taxonomy)
EDUPRESS EP 729 – www.edupressinc.com

LEVEL 1 - REMEMBERING		LEVEL 2 - UNDERSTANDING		LEVEL 3 - APPLYING	
Exhibit memory of previously learned material by recalling facts, terms, basic concepts, and answers.		Demonstrate understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating main ideas.		Solve problems to new situations by applying acquired knowledge, facts, techniques and rules in a different way.	
Key Words	**Questions**	**Key Words**	**Questions**	**Key Words**	**Questions**
choose define find how label list match name omit recall relate select show spell tell what when where which who why	What is ...? Where is ...? How did ___ happen? Why did ...? When did ...? How would you show ...? Who were the main ...? Which one ...? How is ...? When did ___ happen? How would you explain ...? How would you describe ..? Can you recall ...? Can you select ...? Can you list the three ...? Who was ...?	classify compare contrast demonstrate explain extend illustrate infer interpret outline relate rephrase show summarize translate	How would you classify ...? How would you compare ...? How would you contrast ...? State in your own words ...? Rephrase the meaning ...? What facts or ideas show ...? What is the main idea of ...? Which statements support ...? Explain what is happening ...? What is meant ...? What can you say about ...? Which is the best answer ...? How would you summarize ...?	apply build choose construct develop experiment with identify interview make use of model organize plan select solve utilize	How would you use ...? What examples can you find to ...? How would you solve ___ using what you've learned ...? How would you organize ___ to show ...? How would you show your understanding of ...? What approach would you use to ...? How would you apply what you learned to develop ...? What other way would you plan to ...? What would result if ...? Can you make use of the facts to ...? What elements would you choose to change ...? What facts would you select to show ...? What questions would you ask in an interview with ...?

LEVEL 4 - ANALYZING		LEVEL 5 - EVALUATING		LEVEL 6 - CREATING	
Examine and break information into parts by identifying motives or causes. Make inferences and find evidence to support generalizations.		Present and defend opinions by making judgments about information, validity of ideas, or quality of work based on a set of criteria.		Compile information together in a different way by combining elements in a new pattern or proposing alternative solutions.	
Key Words	**Questions**	**Key Words**	**Questions**	**Key Words**	**Questions**
analyze assume categorize classify compare conclusion contrast discover dissect distinguish divide examine function inference inspect list motive relationships simplify survey take part in test for theme	What are the parts of ...? How is ___ related to ...? Why do you think ...? What is the theme ...? What motive is there ...? Can you list the parts ...? What inference can you make ...? What conclusions can you draw? How would you classify...? How would you categorize...? Can you identify ...? What evidence can you find ...? What is the relationship ...? Can you distinguish between ...? What is the function of ...? What ideas justify ...?	agree appraise assess award choose compare conclude criteria criticize decide deduct defend determine disprove dispute estimate evaluate explain importance influence interpret judge justify measure opinion perceive prioritize prove rate recommend select support value	Do you agree with the actions...? with the outcome...? What is your opinion of ...? How would you prove/disprove ? Assess the value /importance of? Would it be better if ...? Why did they (the character) choose ...? What would you recommend...? How would you rate the ...? What would you cite to defend the actions ...? How could you determine...? What choices ...? How would you prioritize ...? What judgment can you make ..? Based on what you know, how would you explain ...? What information would you use to support the view...? How would you justify ...? What data was used to make the conclusion...? What was it better that ...? How would you compare the ideas ...? people ...?	adapt build change choose combine compile compose construct create design develop discuss elaborate estimate formulate happen imagine improve invent make up maximize minimize modify original originate plan predict propose solution solve suppose test theory	What changes would you make to solve ...? How would you improve ...? What would happen if ...? Can you elaborate on the reason ...? Can you propose an alternative...? Can you invent ...? How would you adapt ___ to create a different ...? How could you change (modify) the plot (plan) ...? What could be done to minimize/max ..? What way would you design ...? What could be combined to improve (change) ...? Suppose you could__what would you do ...? How would you test ...? Can you formulate a theory for ...? Can you predict the outcome if ...? How would you estimate the results for ...? What facts can you compile ...? Construct a model that would change ...? Think of an original way for the ...?

Appendix L: Debate Protocol (Hudson-Doc≠1)

The purpose of the debate is to research, study, collect information, discuss, and argue *for* or *against* a proposition. There are two teams in each group (Government and Opposition) as practiced many times throughout the semester. Each team consists of two members and each group of two teams, unless otherwise noted. In each group, one team (i.e., the Government) will argue *for* the proposition and one team (i.e., the Opposition) will argue *against* the proposition. If there is a three member group, you still divide into Government and Opposition. Please log into e-class to select your debate propositions (and your groups) on a first-come-first-served basis. You will decide among yourselves who will be the Government and who will be the Opposition. Here are the debate topics and the schedules.

1. Educational institutions (i.e., schools and universities) should promote (e.g., give more weight to and assess) the traditional definition of intelligence more than multiple intelligences (see for example, pp. 202-203, 244). July 14 at 11:30 am.
2. Anthropologists who do ethnographic work do more harm than good (see for example, pp. 16, 24-25, 187). July 14, after debate #1.
3. In today's society, family still plays a larger role in children's socialization than online media (e.g., social media, Youtube, Netflix). See the slides. Note the definition of socialization on the slides. July 14, after debate #2
4. Addiction is a choice (not a disease) (see p. 231, for example). July 19, at 11:30 am.
5. Single-gender classrooms improve the educational achievements of both boys and girls (see p. 247, for example). July 19, after debate #1.
6. Legal strikes help improve the working conditions for the entire society (see p. 105 on Marx and conflict and p. 160, for example). July 19, after debate #2.

Once you have made a choice, I will put the group members in touch with one another.

Please note that you may end up arguing *for* or *against* a proposition regardless of your own personal convictions or beliefs (as a lawyer does). It is important that you coordinate your roles with your team member and group members. This is an academic exercise to see both sides of an issue regardless of one's beliefs. It is your command of the argument structure, delivery, and use of language that is being assessed. Please see attached the evaluation sheet.

References. This debate is research-based. For your references, each group finds and uses four scholarly articles and two newspaper articles. You must indicate your source of data/evidence as you speak. When you quote references, you need to indicate the source. If you use direct quotes,

you may use the phrase "quote," to start the quote, and 'end quote" to end the quote. For example, you may say, "According to Smith (2019), *quote … end quote*."

In doing your library search, it is possible that you find articles on the topic, but not necessarily on the same exact proposition. For example, you may be able to find articles on the traditional definition of intelligence, and articles on multiple intelligences, but none on the comparison of the two. In this case, the student should relate the two groups of articles in the argument that he or she makes. In other words, the data is derived from the two sets of articles, but the warrant (links and connections) is supplied by the students themselves. Then, you will be making "warrant establishing" arguments. After coordinating with your group members, one group representative should upload your references (formatted APA style) to e-class.

As each team speaks, the other team should take careful notes so that they will respond to the other team. The rest of class also listens, takes notes and participates in the question and answer. That will be part of your participation mark. All students are expected to participate in the Q/A following each debate. Therefore, you should all be familiar with the topics ahead of time. All students should study the respective slides and pages of the textbook.

The debate protocol

Each debate takes 40 minutes according to the following schedule.
Government: Argues *for* the proposition for 10 minutes (5 minutes each member).
Opposition: Argues *against* the proposition for 10 minutes (5 minutes each member).
Government: Responds to Opposition 2 for 5minutes (2-1/2 minutes each member)
Opposition: Responds to The Government 1 for 5 minutes (2-1/2 minutes each member)
Government: Makes closing arguments for 5 minutes (2-1/2 minutes each member)
Opposition: Makes closing arguments for 5 minutes (2-1/2 minutes each member)

The government should define the key terms in the first round. Both the government and the Opposition should adhere to the same definitions throughout the debate. Each debater will speak for 10 minutes, regardless of group size. I will keep the time and indicate it to you, using colour coded sheets. I may intervene at any time for challenges to each group's argument and delivery. In this role, I will be an adjudicator, not an instructor. Therefore, the usual support that I have provided will be absent (no comments or visual cues from me). There may be questions and challenges from the rest of class (time permitting). Therefore, the class should listen and take careful notes to be ready to ask questions and pose challenges. That will be part of your participation mark.

This is a collaborative project, so you can use your partners' help to prepare for and deliver this debate. Please see the course outline for academic integrity expectations. No outside help for the debate is allowed.

I wish you the best with this assignment.

Debate Evolution (15%)

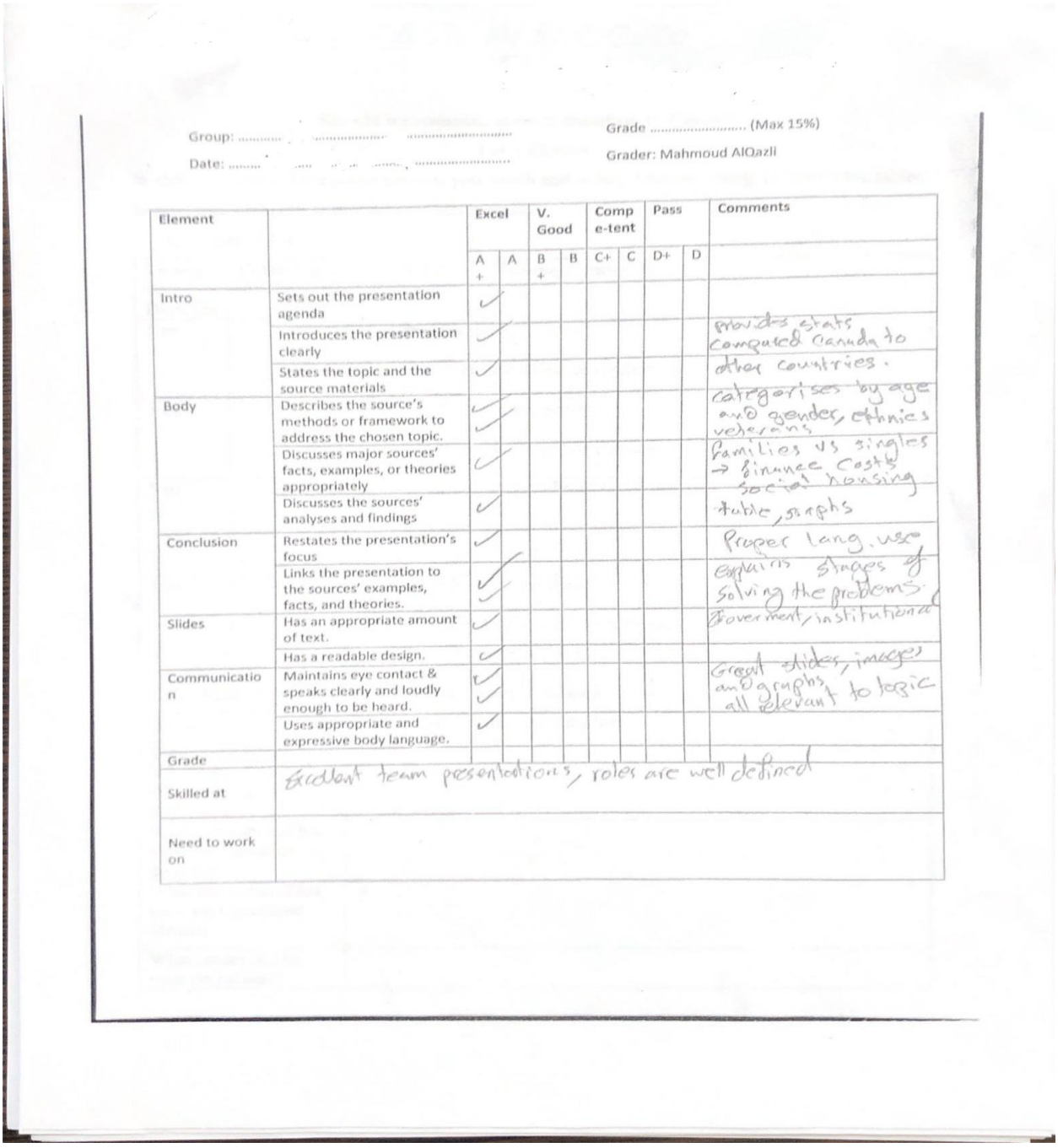
Student name _____Date ___ ___,Time _____, Section ____
Proposition _____
Round 1: Claim/proposition stated and repeated throughout the debate
Evidence
Cites sources:
Warrant

Round 2: responds to opposite proposition with a clear claim
Should respond to _____
Does he/she respond to _____
Evidence
Cites sources:
Warrant

Round 3: closing arguments
Responds to any outstanding arguments
Summarizes evidence and restates proposition

Language and delivery
Maintains eye contact, stands up (not sitting at the computer), does not read from notes, loudness of voice, coherence, clarity, pronunciation, body language, grammar, vocabulary, intelligibility.

Appendix M: Presentation Rubric (Mike-Doc≠1)



Group: ..... Grade ..... (Max 15%)

Date: ..... Grader: Mahmoud AlQazli

Element		Excel		V. Good		Compe-tent		Pass		Comments
		A+	A	B+	B	C+	C	D+	D	
Intro	Sets out the presentation agenda	✓								
	Introduces the presentation clearly	✓								provides stats conquered Canada to other countries.
	States the topic and the source materials	✓								
Body	Describes the source's methods or framework to address the chosen topic.	✓ ✓								categorises by age and gender, ethnics veterans
	Discusses major sources' facts, examples, or theories appropriately	✓								families vs singles → finance costs social housing
	Discusses the sources' analyses and findings	✓								table, graphs
Conclusion	Restates the presentation's focus	✓								Proper lang. use
	Links the presentation to the sources' examples, facts, and theories.	✓ ✓								explains stages of solving the problems.
Slides	Has an appropriate amount of text.	✓								Government, institutions
	Has a readable design.	✓								
Communication	Maintains eye contact & speaks clearly and loudly enough to be heard.	✓ ✓								Great slides, images and graphs, all relevant to topic
	Uses appropriate and expressive body language.	✓								
Grade										
Skilled at	Excellent team presentations, roles are well defined									
Need to work on										

Appendix N: Communicative Task Assessment Criteria (Ryan-Doc ≠ 4)

Ryan- DOC≠ 2B- Communication Task Assessment Criteria

The following elements for effective presentation and questions, all discussed and practiced earlier in the course, will be used as the performance criteria for the final assessment:

Presentations • Introduction: The presenter greets the audience, introduces self and states the purpose of the presentation • Body: The presenter uses signposts to indicate transitions. There is clear, logical flow of information being presenter. • Conclusion: The presenter summarized the key points of the presentation. • Visuals: The PowerPoint slides or other visuals present information in a clear and concise way with no grammatical errors. • Body language: The presenter uses eye contact effectively, appears well prepared, uses notes effectively and shows good time management. • Pronunciation: The presenter uses appropriate volume, pace and enunciation. • Audience participation: The presenter invites and encourages questions and comments. He or she handles questions or comments effectively, confidently and non-defensively.
Questions • The question demonstrates the listener has understood the presentation (as long as the instructor considers the presentation comprehensible) • The questioner uses accurate word order and question formation. • The question is asked appropriately for the context.

Appendix O: Debating-based-on-videos task (Mike-DOC≠ 3)

Should we embrace multiculturalism in Canada?

Let's discuss

Watch the videos. Take some notes as you watch and listen. After watching, complete the tables below, then compare with your shoulder partner.

1. **Video # 1-4:**

Timing + Duration	Video Topic	Your notes **Timing 20 minutes**
5 m	1. Census Canadian diversity	What is the topic about? List three key words or concepts….
5 m	2. Mixed Raced Canadians identity	What is the topic about? List three key words or concepts….
5 m	3. Being Mixed Race	What is the topic about? List three key words or concepts….
5 m	4. 'Where are you from?'	What is the topic about? List three key words or concepts….

2. **Video #5 =13 minutes: by Hon. Ahmed Hussen.**

Questions	Your answers **Timing 25 minutes**
Who is the speaker? Where was he born? What jobs did he undertook?	
What are some of his view on Canadian diversity	
What are some of his view on Canadians identity	
What issues did he raise on racism?	

3. **Video #6: 14 minutes: media interview on Maxime Bernier views of multiculturalism**

Questions	Your answers **Timing 25 minutes**
Who are the speaker? What are they discussing?	
Are they supporter of Canadian diversity? Discuss briefly	
Why are they critiquing **Maxime Bernier's tweet on diversity?**	
List 5 major Key words they discussed	

So, Should we embrace multiculturalism in Canada? Let's debate (timing: 55 minutes):

Multiculturalism is a somewhat Hazy (=cloudy=unclear) term, and as such can be used to mean a variety of things, as we have seen in the vids.

This debate, however, will concern itself with the effectiveness of the Canadian multiculturalism model.

Arguments one (FOR) favors the Canadian multicultural integration model. Argument two (AGAINST) rejects the Canadian multicultural integration model.

1. **How it works:** We are going to be divided in two teams. Team A are supporters (for) supporter multicultural advantages. (FOR). Team B are against the multicultural Canada model (AGAINST)
2. First, each one of you will fill out in bullet points (boxes 1 & 2) with arguments based on our understanding of the previous discussion (**Timing 5 minutes)**

Box (1) FOR

Box (2) AGAINST

3. Now those who found more arguments for box (1) will team up in team A to decide together on two major points to rebut (counter argue) those who favored Box (2) items.

Major Two POINTS FOR (Team A) decision	Major Two POINTS AGAINST (Team B decision)

Each team will take turns to defend their opinions as in the table below:

Timing 15 minutes	Team A (for)	Team B (against)
1 m each	Start: Introduce your position to Multiculturalism	Respond: Introduce your position Against Multiculturalism
2 ms.	Argue: State two arguments for your opinion	Argue: State two arguments for your opinion
2 m	Rebut team B's arguments	Rebut team A's arguments
2 m	Team B's Conclusion	Team A's Conclusion
Points		

State opinion	Sequencing	Responding to the other side
In our opinion... We (don't) think that… The way we see it... If you want our honest opinion.... According to the other side/ our opponents… As far as I'm concerned... Our position is the following…	Firstly…, secondly…, our third point is that… • The first good reason to… is that…; next ; what's more ; moreover…. • To begin, we think that…; in addition, you have to know that… ; last but not least…. • The first point I would like to raise is this… • Here's the main point I want to raise… • I'd like to deal with two points here. The first is…	• I see your point, but I think… • Yes, I understand, but my opinion is that… • That's all very interesting, but the problem is that… • I'm afraid I can't quite agree with your point. • I think I've got your point, now let me respond to it. • We can see what you're saying. Here's my reply… • Excuse me, but that's not quite correct. • Sorry, I just have to disagree with your point. • Let me just respond to that, please. • I'd like to take issue with what you just said. • We said that… but the other side has not replied to our point. • I'd like to focus on two points that the other side has failed to address. • There are two issues our opponents have failed to dispute, namely… • We pointed out that… • Our opponents have claimed that…
Disagreeing	**Agreeing**	**Interruptions**
I don't think so. (strong) No way. I'm afraid I disagree. (strong) I totally disagree. I beg to differ. (strong) I'd say the exact opposite. Not necessarily. That's not always true. That's not always the case. No, I'm not so sure about that.	You're absolutely right. Absolutely. That's exactly how I feel. Exactly. I'm afraid I agree with James. I have to side with Dad on this one. No doubt about it. (agree with negative statement) Me neither. (weak) I suppose so./I guess so. You have a point there. I was just going to say that	Can I add something here? Is it okay if I jump in for a second? If I might add something... Can I throw my two cents in? Sorry to interrupt, but... (after accidentally interrupting someone) Sorry, go ahead. OR Sorry, you were saying... (after being interrupted) You didn't let me finish. Settling an argument Let's just move on, shall we?

Appendix P: Seminar Rubric (Sally-Doc ≠ 1)

Sally- DOC≠ 5A- Seminar Rubric

Excellent 10 points	Good 8 points	Satisfactory 6 points	Needs work 4 points	Unsatisfactory 0–2 points
Comprehension of Main Ideas				
Demonstrated 100% comprehension of main ideas from all the reading and listening texts. Answered 90-100% of the discussion questions accurately and effectively.	Demonstrated over 75% comprehension of the main ideas in both the listening and reading texts. Answered more than 75% of the discussion questions accurately and effectively.	Demonstrated basic comprehension of the main ideas from both the listening and reading texts. Answered at least 60% of the discussion questions accurately and effectively.	Did not demonstrate comprehension of both the reading and listening texts. Answered <60% of the discussion questions accurately and effectively.	Did not demonstrate comprehension of main ideas from the reading and listening texts. Was unable to respond to discussion questions (less than 50%).
Connection and Support of Ideas				
Made complex connections between the reading and listening texts with a high degree of effectiveness. Fully supported ideas and opinions with explanations and details/evidence from both the reading and listening texts.	Made logical connections between the texts with considerable effectiveness. Expressed opinions and ideas supported by adequate details/evidence from both listening and reading texts.	Made simple connections between texts with some effectiveness and offered simple but accurate responses. Provided some supporting explanations and evidence from at least one reading and one listening text though more support or clarification was needed at times.	Attempted to make connections between the texts with limited effectiveness and needed some coaching to offer responses. Did not use supporting evidence from the listening and reading texts and/or the support was inaccurate or irrelevant.	Was unable to make connections between the texts and needed lots of one-on-one coaching and prompting to respond. Did not use evidence from the listening and reading texts to support ideas and opinions.
Summary and Interpretation				
Skillfully summarized, interpreted, reasoned, and explained the topic and themes in a way that required no further clarification.	Competently summarized, interpreted, reasoned, and explained the topic and themes with	Adequately summarized, interpreted, reasoned, and explained the topic and themes with the	Demonstrated limited ability to summarize, interpret, reason, and explain the topic and themes. A	Was unable to summarize, interpret, reason, or explain the topics or themes even with prompting.

	only occasional clarification needed.	need for some clarification.	lot of prompting may have been required.	
Excellent **5 points**	**Good** **4 points**	**Satisfactory** **3 points**	**Needs work** **2 points**	**Unsatisfactory** **0–1 point**
Fluency and Coherence				
Always (100%) stayed on topic. Ideas were well organized and easy to follow. Spoke fluently with minimal hesitation, which did not impact communication.	Mostly (75%+) stayed on topic. Ideas were organized and generally easy to follow. Spoke fluently with occasional hesitation with little to no impact on communication.	Stayed on topic at least 60% of the time. Ideas were somewhat organized but may have taken some effort to follow. Hesitation was noticeable, but it did not prevent the message from coming across.	Frequently deviated from the topic (less than 60%). Ideas lacked organization and were hard to follow. Frequent hesitation had noticeable impact on communication.	Did not stay on topic (less than 50%). Poor organization of ideas and the inability to maintain an uninterrupted flow of speech prevented the student from participating in the discussion.
Discussion Strategies				
Used a wide variety of discussion strategies to respond appropriately, engage others, agree and disagree, ask for and give clarification, and take **turns. Consistently showed interest and participate equally.**	Used a variety of discussion strategies to respond appropriately, engage others, agree and disagree, ask for and give clarification, and take **turns. Generally showed interest and participated equally.**	Used some discussion strategies to respond appropriately, engage others, agree and disagree, ask for and give clarification, and take **turns. Showed some interest and made an effort to participate equally.**	Used a limited range of discussion strategies to respond, engage others, agree and disagree, ask for and give clarification, and take turns. May not **have shown interest or participate equally.**	Used no or very few discussion strategies. Showed no interest and made little to no effort to participate.
Language Use				
Used a range of vocabulary sufficient to discuss all topics at length. Inaccuracies in word choice/form and grammatical errors were minimal and did not affect communication.	Used a range of vocabulary sufficient to discuss all topics. Inaccuracies in word choice/form and grammatical errors occurred but only occasionally impacted communication.	Vocabulary range was basic but adequate to discuss all topics. Inaccuracies in word choice/form and grammatical errors were noticeable and may have had some impact on communication.	Vocabulary range was limited and prevented the student from discussing all topics. Inaccuracies in word choice/form and grammatical errors were frequent and had a significant impact on communication.	Vocabulary range was inadequate to discuss the topics. Errors in vocabulary and grammar impeded communication.
Clarity				
Student spoke very clearly and used appropriate volume and pace. Pronunciation errors had no impact on intelligibility.	Student generally spoke clearly and used appropriate volume and pace. Pronunciation errors had little impact on intelligibility.	Student spoke with occasional lack of clarity and some minor problems with volume and pace. Pronunciation errors had occasional impact on intelligibility.	Lack of clarity was frequent with noticeable problems with volume and pace. Pronunciation errors had significant impact on intelligibility.	Speech was mostly unintelligible

Appendix Q: Target Performance Descriptors (Ryan-Doc≠ 2)

Ryan-Doc≠ 5- Target Performance Descriptors

A. Target Performance Descriptors

This chart describes the level against which performance will be assessed.

Target Performance Descriptors (Level 3)

Rating Criteria	Target Performance Descriptors (Level 3)
Effectiveness	The participant is able to complete the communication task according to the task requirements and the message is successfully conveyed (speaking or writing) or comprehended (listening or reading). There may be aspects of the communication that might still require attention.
Speaking Accuracy	The speaker uses a variety of grammatical forms and a range of appropriate vocabulary. Some grammatical and vocabulary errors may be evident in the discourse, but most people can understand the message.
Speaking Fluency	The speaker uses appropriate connectors and linking words (such as "then," "next," "on the other hand," etc.) as well as communication strategies (such as asking questions when something is not clear, etc.) to maintain and sometimes direct a comprehensible and logical flow of speech.
Pronunciation	The speaker's message is easily understood, although there may be many individual sounds, and some rhythm and intonation patterns, that could be causing difficulty and occasionally impeding communication.
Writing Accuracy	The writer uses a variety of grammatical forms and a range of relevant vocabulary to produce meaningful text. Although errors may be evident (grammar, spelling, capitalization, punctuation, etc.) the overall meaning is clear.
Writing Fluency	The writer organizes the content and uses transition words (such as "however," "but," "and," etc.) to produce comprehensible and logical text, and follows standard occupation-specific conventions and formats (email, work forms, etc.) to effectively convey information and ideas within a reasonable amount of time as dictated by the communication task.
Listening Comprehension	The listener follows the main points and relevant details in face-to-face and telephone interactions by using some listening comprehension strategies, such as active listening, vocalizations, clarification questions, requests for repetition, etc., and responds as appropriate to the communication task.
Reading Comprehension	The reader follows the main points and relevant details within a reasonable amount of time as dictated by the communication task.
Socio-cultural Appropriateness	The participant uses commonly accepted social behaviours, conventions and norms relevant to an occupation-specific context in Ontario. This includes effective organization of content for the target audience (such as introductions and conclusions in speaking and writing), context-appropriate use of non-verbal communication (body language, tone and vocalizations) and use of formal or informal language appropriate to the communication task.

Level	1	2	3	4
Descriptor	needs significant improvement to meet target outcome	needs some improvement to meet target outcome	meets target outcome	exceeds target outcome