

The copyright © of this thesis belongs to its rightful author and/or other copyright owner. Copies can be accessed and downloaded for non-commercial or learning purposes without any charge and permission. The thesis cannot be reproduced or quoted as a whole without the permission from its rightful owner. No alteration or changes in format is allowed without permission from its rightful owner.



**NARRATIVE INQUIRY OF LANGUAGE ASSESSMENT
LITERACY DEVELOPMENT OF PRE-SERVICE ENGLISH
AS A FOREIGN LANGUAGE TEACHER IN CHINA**



GENG JUANJUAN

UUM
Universiti Utara Malaysia

**DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2023**



Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

GENG, JUANJUAN

calon untuk Ijazah
(candidate for the degree of)

PhD

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

“NARRATIVE INQUIRY OF LANGUAGE ASSESSMENT LITERACY DEVELOPMENT OF PRE-SERVICE ENGLISH AS A FOREIGN LANGUAGE TEACHER IN CHINA”

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : **21 September 2022.**

That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on:

21 September 2022.

Pengerusi Viva:
(Chairman for VIVA)

Assoc. Prof. Dr. Mohd Isha Awang

Tandatangan
(Signature)

Pemeriksa Luar:
(External Examiner)

Prof. Dr. Ainol Madziah Zubairi

Tandatangan
(Signature)

Pemeriksa Dalam:
(Internal Examiner)

Dr. Nurliyana Bukhari

Tandatangan
(Signature)

Nama Penyelia/Penyelia-penyelia:
(Name of Supervisor/Supervisors)

Prof. Dr. Nurahimah Mohd Yusoff

Tandatangan
(Signature)

Nama Penyelia/Penyelia-penyelia:
(Name of Supervisor/Supervisors)

Dr. S. Kanageswari a/p Suppiah Shanmugam

Tandatangan
(Signature)

Tarikh:

(Date) **18 September 2022**

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Sebagai unsur yang amat diperlukan dalam pengajaran dan pembelajaran, penilaian bukan sahaja dapat memberikan maklumat mengenai keberkesanan pengajaran kepada guru malah ia dapat membentuk pembelajaran pelajar. Walau bagaimanapun, guru pra perkhidmatan English as a Foreign Language (EFL) di serata dunia dikatakan memiliki tahap language assessment literacy (LAL) yang rendah. Langkah pertama adalah untuk memahami kebimbangan guru dalam LAL. Kajian narrative inquiry ini telah dijalankan bagi meneroka konseptualisasi, tahap kemahiran penilaian sendiri, trajektori perkembangan dan faktor pengantaraan dalam proses perkembangan LAL dikalangan guru pra perkhidmatan EFL.

Enam peserta daripada program pra perkhidmatan EFL di China telah dipilih secara persampelan tujuan bagi kajian ini. Semua temu bual naratif, bahan penilaian dan jurnal telah di kumpul dan dianalisa menggunakan analisis naratif dan analisis tematik. Dapatan kajian menunjukkan bahawa peserta kajian telah mengkonsepsikan LAL sebagai konsep lapan dimensi berskala yang diterapkan dalam konteks sosiobudaya. Bagi kemahiran LAL, peserta kajian cenderung menilai diri sendiri antara kurang layak, sedikit layak dan cukup layak. Selain itu, faktor pengalaman, kontekstual dan pengantaraan peribadi dikenal pasti sebagai penyumbang dalam perkembangan LAL. Faktor pengantaraan menyumbang dalam membina asas pengetahuan guru pra perkhidmatan EFL, membentuk konsepsi penilaian mereka, menapis amalan penilaian dan memudahkan assessor identity construction. Dengan peningkatan LAL, peserta kajian menjadi lebih peka untuk merenungkan pengaruh faktor pengantara. Berkenaan trajektori perkembangan LAL, semua peserta kajian telah melalui tiga tahap penguasaan daripada assessment knowledge base kepada internalized assessment understanding hingga ke assessor identity construction. Kajian ini menyumbang kepada konseptualisasi LAL dengan melibatkan pihak berkepentingan yang terabai, guru pra perkhidmatan EFL dan juga turut memperkukuhkan penggunaan narrative inquiry dalam bidang LAL. Dapatan kajian ini boleh digunakan sebagai panduan penambahbaikan kekurangan LAL dalam kalangan guru pra perkhidmatan EFL dengan menawarkan implikasi kepada pendidik guru dan program persediaan guru.

Kata kunci: Language assessment literacy, Guru pra perkhidmatan EFL, Trajektori perkembangan, Faktor pengantaraan, Konseptualisasi

Abstract

As an indispensable element in the teaching and learning, assessment not only informs teachers about the effectiveness of the instruction, but also frames student learning. However, pre-service English as a Foreign Language (EFL) teachers across the world have been suggested to possess an inadequate level of language assessment literacy (LAL). The first step is to understand their concerns in LAL. To address the paucity, this narrative inquiry is conducted to explore pre-service EFL teachers' LAL conceptualizations, self-evaluated proficiency level, evolvement trajectories, and the mediating factors in the evolvement process.

Six participants from a pre-service EFL teacher education programme in China were purposefully sampled. The narrative interviews, assessment artifacts, and journals were collected and analyzed by narrative analysis and thematic analysis. The findings showed that the participants conceptualized LAL as a scaled eight-dimensional concept embedded in the social-cultural contexts. They tend to self-evaluate their LAL proficiency ranging from insufficiently qualified, through marginally qualified, to satisfactorily qualified. Besides, the experiential, contextual, and personal mediating factors were identified to participate in LAL evolvement. From bottom-up, the mediating factors participated in constructing pre-service EFL teachers' knowledge base, shaping their assessment conceptions, filtering the assessment practices, and facilitating the assessor identity construction. With the enhanced LAL, they were more attentive to reflect on the influence of these mediating factors. As to the LAL evolvement trajectory, all the participants went through three mastery levels from the first assessment knowledge base through internalized assessment understanding to assessor identity construction. This study contributes to LAL conceptualization by involving in the neglected stakeholders, pre-service EFL teachers and enriches the application of narrative inquiry in LAL field. The findings can be used as a guidance for the improvement of deficient LAL among pre-service EFL teachers by offering implications for the teacher educators and teacher preparation programmes.

Keywords: Language assessment literacy, Pre-service EFL teachers, Evolvement trajectory, Mediating factors, Conceptualization

Acknowledgements

First of all, I would like to express my appreciation and many thanks to my dear supervisors, Prof. Dr. Nurahimah Mohd. Yusoff and Dr. S. Kanageswari Suppiah Shanmugam. Without their encouraging support and enlightening comments, this dissertation would be totally impossible. I will never forget their kindness which motivates me to keep on along the research road.

I would like to thank my colleagues in Xianyang Normal University in China who assisted me to sample the proper participants, verify the feasibility of the interview protocol, and share with me the conference information concerning assessment. They are Prof. Huang yongliang, Prof. Xiao chunyan, Prof. Shi yuanhui, Dr. Che ping, Dr. Lu tuanhua, Dr. Gong yun, Dr. Xue xuanzi, Miss. Zhang shasha, and Ms. Chen lin.

I also appreciate the efforts of all participants who volunteered their time to be a member of this study.

Lastly, my gratitude goes to my family members and all other friends who gave me confidence and inspired me throughout my Ph.D. study. I wish to thank my parents and my children, Hao jiajun and Han lijia.

Table of Contents

Permission to Use.....	i
Abstrak	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	x
List of Figures	xi
List of Appendices.....	xiii
List of Abbreviations.....	xiv
CHAPTER ONE INTRODUCTION	1
1.1 Background of the Research	1
1.2 Research Context	4
1.2.1 Introduction of Pre-service Education in China	5
1.2.2 English Assessment in Primary and Middle Schools in China.....	7
1.3 Problem Statement	8
1.4 Research Objectives	13
1.5 Research Questions.....	14
1.6 Background of the Researcher.....	14
1.7 Theoretical Framework.....	16
1.8 Significance of the Study	18
1.9 Definition of Terms.....	20
1.10 Layout of the Thesis.....	23
1.11 Summary	25
CHAPTER TWO LITERATURE REVIEW.....	26
2.1 Introduction	26
2.2 The Assessment Paradigm Shift.....	26
2.3 Rationales of Teachers' AL	32

2.3.1 Mandated Demand of National Policy	32
2.3.2 Professional Requirement of Teachers	35
2.3.3 The Academic Well-being of Learners	37
2.3.4 The Guarantee of Successful Curriculum Reform.....	40
2. 4 Conceptual Understanding of AL and LAL	41
2.4.1 Definitions of AL and LAL	41
2.4.2 Frameworks of AL and LAL	46
2.4.2.1 Competence Lists	47
2.4.2.2 Three-component Models.....	48
2.4.2.3 Scaled Models	51
2.4.2.4 Negotiation Models	56
2.5 Empirical Studies on AL and LAL among In-service Teachers.....	59
2.5.1 AL and LAL Proficiency Level	60
2.5.1.1 AL Proficiency Level	60
2.5.1.2 LAL Proficiency Level.....	65
2.5.2 Factors Mediating AL and LAL Evolvment	70
2.5.2.1 Contextual Factors.....	71
2.5.2.1.1 In the Field of AL	71
2.5.2.1.2 In the Field of LAL	76
2.5.2.2 Experiential Factors.....	78
2.5.2.2.1 In the field of AL	79
2.5.2.2.2 In the Field of LAL	83
2.6 Empirical Studies on AL and LAL among Pre-service Teachers.....	87
2.6.1 AL and LAL Proficiency level.....	87
2.6.2 Factors Mediating AL and LAL Evolvment	90
2.6.3 The Development Trajectory of AL and LAL	97
2.7 Summary	100
CHAPTER THREE RESEARCH METHODOLOGY	101
3.1 Introduction.....	102

3.2 Research Design.....	102
3.3 Research Settings	105
3.3.1 Background Information of X Normal University	106
3.3.2 Background Information of Pre-service EFL Teacher Education Programme	107
3.4 Participants' Selection.....	109
3.5 Data Collection.....	114
3.5.1 Semi-structured Interview	115
3.5.1.1 Development of the Interview Protocol	116
3.5.1.2 Refinement of the Interview Protocol	118
3.5.1.3 Implementation of the Interview Protocol.....	121
3.5.2 Documents	122
3.5.3 Artifacts	122
3.5.4 Journals.....	123
3.6 Research Procedures	124
3.7 Data Analysis Methods	128
3.8 Researcher's Role.....	132
3.9 Ethical Considerations	135
3.10 Trustworthiness of the Study.....	137
3.11 Summary	140
CHAPTER FOUR RESEARCH FINDINGS	142
4.1 Introduction	142
4.2 Findings for Conceptualization of LAL	142
4.2.1 Conceptualization of LAL from Participant 1	145
4.2.2 Conceptualization of LAL from Participant 2	147
4.2.3 Conceptualization of LAL from Participant 3	148
4.2.4 Conceptualization of LAL from Participant 4	150
4.2.5 Conceptualization of LAL from Participant 5	152
4.2.6 Conceptualization of LAL from Participant 6	154

4.2.7 Overall conceptualization of LAL	156
4.3 Findings for Self-evaluated LAL Proficiency	160
4.3.1 Self-evaluated LAL Proficiency	160
4.3.2 Self-diagnosed LAL Improvement	163
4.4 Findings for Interaction between Mediating Factors and LAL Evolvment	165
4.4.1 The Identified Mediating Factors in LAL Evolvment	166
4.4.1.1 Experiential Factors	167
4.4.1.2 Contextual Factors	179
4.4.1.3 Personal Factors	200
4.4.2 Interaction between the Mediating Factors and LAL Evolvment	201
4.5 Findings for LAL Evolvment Trajectory	207
4.6 Summary	213
CHAPTER FIVE DISCUSSION AND IMPLICATIONS.....	216
5.1 Introduction	216
5.2 Discussion of Findings for Conceptualization of LAL	216
5.2.1 Eight Dimensions	216
5.2.2 Scaled Importance	220
5.2.3 Context-dependence	222
5.2.4 Idiosyncrasy	222
5.3 Discussion of Findings for Self-evaluated LAL Proficiency	223
5.3.1 Self-evaluated LAL Proficiency	224
5.3.2 Self-diagnosed LAL Improvement	226
5.4 Discussion of Findings for Interaction between Mediating Factors and LAL Evolvment	228
5.4.1 The Mediating Factors Identified in the Study	229
5.4.1.1 The Experiential Factors	229
5.4.1.2 The Contextual Factors	231
5.4.1.3 The Personal Factors	237
5.4.2 The Interaction between the Mediating Factors and LAL Evolvment.....	238

5.5 Discussion of Findings for LAL Evolvment Trajectory	241
5.6 Implications of the Findings	244
5.6.1 Implications for Participants	245
5.6.2 Implications for Pre-service EFL Teachers	251
5.6.3 Implications for Pre-service EFL Teacher Educators	252
5.6.4 Implications for Pre-service EFL Teacher Preparation Programme	253
5.7 Contribution to the Body of LAL Research	256
5.8 Limitations of the Research	257
5.9 Recommendations for the Future Research	259
5.10 Summary	261
RERERENCES	262
APPENDICES	302
Appendix A: Informed Consent for the Institute	302
Appendix B: Informed Consent for Participants	303
Appendix C: Interview Protocol	304
Appendix D: Interview Transcripts (Excerpt Sample)	310
Appendix E: Participant's Journal (Excerpt Sample)	312
Appendix F: Assessment Artifacts (Sample of Lesson Plan)	314

List of Tables

Table 2.1 Classification of Frameworks of AL and LAL	47
Table 2.2 The Contextual Factors Mediating AL Identified in Previous Studies	71
Table 2.3 The Contextual Factors Mediating LAL in Previous Studies	76
Table 3.1 Outline of Courses Offered in the Pre-service EFL Education.....	107
Table 3.2 Profiles of the Participants	113
Table 3.3 Data Inventory	114
Table 3.4 Structure of the Interview Protocol.....	116
Table 3.5 Refinement of the Interview Protocol.....	118
Table 3.6 Time Duration of Each Interview Round.....	121
Table 4.1 Eight Dimensions of LAL Emerging from the Categories	144
Table 4.2 The Overall Distribution among the Eight Dimensions of LAL	157
Table 4.3 The LAL Proficiency Self-evaluated by the Participants	160
Table 4.4 The Self-diagnosed LAL Improvement among the Participants	164
Table 4.5 The Dimensions of Mediating Factors.....	167
Table 5.1 The Overlaps of Eight Dimensions in the Study with Previous Frameworks	217
Table 5.2 Dimensions on the Two Extremes of Importance Continuum.....	221

List of Figures

Figure 2.1 Pastore and Andrade's (2019) three-dimensional framework of AL	49
Figure 2.2 Fulcher's (2012) expanded model of LAL.....	50
Figure 2.3 Shin's (2013) AL model for teachers in three different stages.....	52
Figure 2.4 Taylor's (2013) eight-dimensional diagram of LAL.....	53
Figure 2.5 Baker and Riches' (2017) seven-dimensional diagram of LAL.....	54
Figure 2.6 Kremmel and Harding's (2020) nine-dimensional diagram of LAL	55
Figure 2.7 Bøhn and Tsagari's (2021) 10-dimensional framework of LAL.....	56
Figure 2.8 Y. Xu and Brown's (2016) AL framework	57
Figure 3.1 Participants selection procedures	112
Figure 3.2 Translation process of interview protocol	120
Figure 3.3 Procedures of the pilot study.....	125
Figure 3.4 Procedures of the main study	127
Figure 3.5 Procedures of data analysis	129
Figure 4.1 Conceptualization of LAL from P1 (Amy).....	146
Figure 4.2 Conceptualization of LAL from P2 (Betty)	148
Figure 4.3 Conceptualization of LAL from P3 (Carol)	149
Figure 4.4 Conceptualization of LAL from P4 (Daisy).....	151
Figure 4.5 Conceptualization of LAL from P5 (Edwin).....	153
Figure 4.6 Conceptualization of LAL from P6 (Flora).....	156
Figure 4.7 The combined overall conceptualization of LAL	158
Figure 4.8 Self-evaluated LAL proficiency continuum.....	163
Figure 4.9 Assessment experience transfer conditions	168
Figure 4.10 The inter-related network of the mediating factors.....	201
Figure 4.11 Interaction between mediating factors and LAL evolvment	202
Figure 4.12 LAL evolvment trajectory of P1 (Amy).....	208
Figure 4.13 LAL evolvment trajectory of P2 (Betty)	209
Figure 4.14 LAL evolvment trajectory of P3 (Carol)	210

Figure 4.15 LAL evolvement trajectory of P4 (Daisy).....	210
Figure 4.16 LAL evolvement trajectory of P5 (Edwin).....	211
Figure 4.17 LAL evolvement trajectory of P6 (Flora)	212
Figure 4.18 The overall LAL evolvement trajectory among the participants	213
Figure 5.1 Comparison with other models in dimension classifications	217
Figure 5.2 Comparison with other models in scaled importance	220
Figure 5.3 Comparison of LAL conceptualizations among the participants	223



List of Appendices

Appendix A: Informed Consent for the Institute	302
Appendix B: Informed Consent for Participants.....	303
Appendix C: Interview Protocol	304
Appendix D: Interview Transcripts (Excerpt Sample).....	310
Appendix E: Participant's Journal (Excerpt Sample).....	312
Appendix F: Assessment Artifacts (Sample of Lesson Plan).....	314



List of Abbreviations

AaL	assessment as learning
AfL	assessment for learning
AL	assessment literacy
AoL	assessment of learning
EFL	English as a foreign language
ELTS	English language testing service
LAL	language assessment literacy
PCK	pedagogical content knowledge
RQ	research question
TEM4	Test for English Majors-band 4
TESOL	Teach English for Speakers of Other Languages



CHAPTER ONE

INTRODUCTION

1.1 Background of the Research

Assessment can be broadly defined as a collection of evidence about student learning by using a range of techniques for diverse purposes, including making educational judgments, providing feedback for students about their progress, evaluating instruction effectiveness, and informing policy decision (Hamidi, 2010; Norris, 2006; Popham, 2009; Rogers, 1993; Rogler, 2014; Sanders & Vogel, 1993; Scarino, 2010). As indicated in the definition, assessment plays a critical role in the teaching and learning process (Napanoy & Peckley, 2020; Jawhar & Subahi, 2020; Kyttälä et al., 2021; Prasetyo, 2018; Yetkin & Özer, 2020). As an inseparable element in the process, assessment not only informs teachers about the effectiveness of the instruction (Gonzales & Aliponga, 2012; Nimehchisal & Hussin, 2018) but also frames student learning because it is an indication of the priority of institution in decision-making (Bozkurt, 2020).

Historically, assessment has been primarily used for accountability by ranking students according to their performance on traditional tests, most of which are designed by external professionals (Dysthe, 2008; Stiggins, 2007a, 2014). Since the 20th century, a new vision of assessment has emerged to embrace a wider range of purposes of assessment, advocating assessment for learning (AfL) or assessment as learning (AaL)

instead of assessment of learning (AoL) (Dysthe, 2008; McKellar, 2002; Shepard, 2000). The focus of assessment has been shifted from the product of learning to the process of learning (Gipps, 1999). But the point is not to say the traditional AoL has no role to play, it is predicated to undoubtedly co-exist with AfL even the alternative assessment has been adopted (Gipps, 1994; Kahl et al., 2013; Stiggins, 2014). The learning-oriented assessment has placed greater demands on the quality of assessment inside and outside the classroom developed by teachers (Hamidi, 2010; Mertler, 2009). The knowledge and skills teachers need to engage in assessment-related practices are first upgraded by Stiggins (1991) into the level of literacy, as assessment literacy (AL). Later, it has been extended to the field of language assessment, AL for language teachers is particularly described as language assessment literacy (LAL), broadly referring to the literacy linked with language assessment (Brindley, 2001).

It is widely acknowledged that language teachers need adequate LAL to ensure the quality of various assessment activities (Davies, 2008; Giraldo, 2018b; Jin, 2010; Şişman & Buyukkarci, 2019; Y. Xu, 2019a). Language assessment literate teachers are better at using assessment results of student learning to improve the effectiveness of instruction by addressing students' learning needs (Nimehchisalema & Bhatti, 2019; Pastore & Andrade, 2019). By contrast, foreign language teachers with deficient LAL may be more likely to feel challengeable to implement the assessment responsibilities advocated in the curriculum reform (Sultana, 2019). In other words, the inadequate LAL among English as a foreign language (EFL) teacher largely results in the

discrepancy between the guidance from curriculum standards and their own assessment practice. That means the assessment practices mandated in curriculum standard may remain “on paper only” due to their insufficient LAL (Gu, 2014, p. 301). Consequently, such deficiency may account for the declining standards of foreign language teaching (Sultana, 2019). If teachers, the main agent of classroom assessment, are ignorant of assessment, all the expectations derived from promises of classroom assessment may be impossible to be achieved (Islam & Stapa, 2019).

Hence, LAL acts as a prerequisite for becoming a language teacher (Şişman & Buyukkarci, 2019) and constitutes a key dimension in teacher professional development (Davidson & Coombe, 2019; Farhady, 2019; Kavaklı & Arslan, 2019; Popham, 2009; Shepard et al., 2005). Regrettably, however, pre-service English as foreign language (EFL) teachers across the world appear to possess an inadequate level of LAL (Giraldo & Murcia, 2019; Hatipoğlu, 2015; Kavaklı & Arslan, 2019).

In order to improve LAL among pre-service EFL teachers, the first step is to understand the development trajectory of LAL and how the mediating factors interact with LAL during the evolvement. The existing body of research suggests that a variety of mediating factors from external context and internal experience have been identified to come into play to impact on LAL evolvement among pre-service EFL teachers across the contexts (Y. Xu & He, 2019). The contextual dimension encompasses the external social cultural and political factors, for instance, the broader assessment

culture, availability to professional learning resources, government policies related to assessment, and local institutional and community context (Giraldo & Murcia, 2019; Lam, 2015). Besides, personal experience appears to be fundamental in LAL development (Bolívar, 2020; Graham, 2005). The prior learning experience, the experience of being assessed as language learners in schools and current university, the experience of being involved in assessment practices during practicum scenarios, and personal characteristics (i.e., openness and confidence) have been indicated to have a possible framing effect on LAL evolvement among pre-service EFL teachers (Bolívar, 2020; Graham, 2005; Hatipoğlu, 2015; O'Loughlin, 2006). However, the interaction between the mediating factors and LAL evolvement is still under-explored (Bolívar, 2020; Crusan et al., 2016; Giraldo, 2020; Giraldo & Murcia, 2019; Y. Xu, 2019a).

Besides, the voice from pre-service EFL teachers in China is still lacking (Gan & Jiang, 2020). Hence, to address the research gap, this study primarily aims to explore the LAL evolvement trajectory and the interaction between the mediating factors and LAL evolvement among pre-service EFL teachers for primary and middle schools in the context of China.

1.2 Research Context

The present study is implemented in the context of China. The background information related to pre-service teacher education and the status of English assessment in China

will help to understand the educational background of the pre-service EFL teachers.

1.2.1 Introduction of Pre-service Education in China

In May 2001, it was the first time that the concept of “teacher education” was used to replace the concept of “normal education” in China’s education policy, the Decision of the State Council on the Reform and Development of Basic Education, which put forward further improvement of existing teacher education system with normal universities or colleges as the main body and universities of other types participating in preparing pre-service teachers together (Zeng, 2016, p. 472). China’s pre-service teacher education has been an open one with normal universities acting as the main body and other comprehensive universities as the participants (Ding et al., 2013; Zeng, 2016). Later on, a series of national policies related to teacher education were issued concerning pre-service stage, entry certification, and in-service teacher standards to provide guidance on teacher education and to ensure the quality of teachers.

Regarding the initial teacher education, the Ministry of Education (MoE) in 2011 promulgated Teacher Education Curriculum Standards (Trial), which articulated the basic requirements of the nation for teacher education institutions to set up teacher education curriculum, and it also served as an important basis for developing teaching materials and curriculum resources, implementing instruction and evaluation, and identifying teacher qualifications (Zeng, 2016).

As to the entry certification, as early as 2011, the pilot reform by MoE in teacher qualification examination was carried out to change from examinations previously organized by province, municipalities, and autonomous regions to the unified national-level examination. Two years later in 2013, the national qualification examination was formally implemented across China (Zeng, 2016). Only those who have passed the examination can be qualified to teach. For the minimum educational requirements of teachers, those who apply for the qualification of primary school teachers must graduate from junior college or above; those who apply for middle school teachers must graduate from university or college and possess a Bachelor degree or above. The pre-service teacher education programmes in China cultivate pre-service teachers for primary and middle school without differentiation, meaning that the graduates from pre-service teacher education programmes can apply for teaching in primary or middle schools.

In terms of in-service teacher standards, on February 10th, 2012, MoE promulgated The Standards and Guidance of Professional Development for Primary and Secondary Teachers (Trial), which provided an important basis for teacher education, entry, and evaluation (Zeng, 2016). Although they are the standards for in-service teachers, these standards still have a reference value for pre-service teacher education programmes to respond to criteria articulated in the standards by regulating the programmes accordingly.

1.2.2 English Assessment in Primary and Middle Schools in China

The English assessment in the field of China has a long history of regular high-stakes public examinations serving the purpose of selection (Chen & Brown, 2013; Yu & Jin, 2014). English is a compulsory subject in all external high-stakes examinations, such as the middle-school entrance examination at the end of elementary education, the high-school entrance examination at the end of junior middle school education, and the national college entrance examination at the end of the senior middle school education (Jin & Yan, 2017). They are high-stakes because of their significant impact on the students' access to a highly ranked school or university (Qi, 2005).

Although the newly-issued Curriculum Standard of English in Middle School in 2018 advocates EFL teachers to use a variety of assessment approaches for student learning and a balanced assessment system between AoL and AfL or AaL, high-stakes examinations have persistently dominated English assessment in China, especially the entrance examination of senior middle school and university (Liu & Xu, 2017). In such an exam-oriented educational scene, EFL teachers in China tend to prepare their students for those examinations and to ensure students' success by circumscribing teaching and assessment practices (Gu, 2014; X. Yan et al., 2018). To maximize the scores in examinations, Chinese students have a huge academic pressure to master the content examined in the high-stakes examinations (Liu & Xu, 2017).

1.3 Problem Statement

Since it was formulated to integrate AL with the field of language assessment by Brindley in 2001, LAL has become the focal point in the research agenda (Babaii & Asadnia, 2019; Firoozi et al., 2019; Kavaklı & Arslan, 2019; Shahzamani & Tahririan, 2021). Issues in LAL have been widely addressed but not completely resolved yet (Giraldo, 2018b). LAL is still in its infantile stage concerning the number of the theoretical and empirical studies in the field (Fulcher, 2012; Giraldo, 2018b; Hakim, 2015; Jin, 2018). Hence, the theme LAL is under-explored and more research is warranted (Kim et al., 2020; Nimehchisalema & Bhatti, 2019; Yan & Fan, 2020).

Firstly, various attempts have been made to conceptualize LAL from the perspectives of professional researchers and in-service teachers (Butler et al., 2021). For instance, Shahzamani and Tahririan (2021) define LAL as language teachers' understanding of diverse purposes of assessment and knowledge of applying them accordingly. LAL is primarily conceptualized as a multi-dimensional concept shaped by the context where it is embedded, as illustrated in the existing types of LAL conceptualization frameworks: three -component models (Fulcher, 2012; Giraldo, 2018b; Inbar-Lourie, 2008), scaled models (Bøhn & Tsagari, 2021; Kremmel & Harding, 2020), and negotiation models (Y. Xu & Brown, 2016).

While further research is still encouraged to take the perspectives from other stakeholders, especially the pre-service EFL teachers into consideration (Bøhn &

Tsagari, 2021; Nimehchisalema & Bhatti, 2019; Taylor, 2013). It would be intriguing to find out whether the pre-service EFL teachers' LAL conceptualization is in alignment with the existing models. As their voice should be heard and addressed for a more comprehensive and deeper view of the conceptualization of LAL. Moreover, the elaboration of LAL conceptualization may lay a research foundation for subsequent studies, such as describing the LAL profile or selecting priorities in LAL improvement programmes for the target participants. To address the paucity in research, one of the research purposes of the present study is to focus on pre-service EFL teachers to explore their conceptualizations of LAL.

Secondly, there have been inconsistent findings as to both the LAL proficiency level self-evaluated by pre-service EFL teachers and AL by pre-service teachers in general education. The first type is that they seem to evaluate themselves as insufficiently prepared to fulfill assessment tasks in the future teaching career, as evidenced in LAL in Turkey (Kavaklı & Arslan, 2019) and AL in Canada (Volante & Fazio, 2007). On the contrary, the second voice is unrealistic over-estimation that the teacher candidates are reported to over-estimate their abilities to assessment-related tasks with little awareness of the complexity in assessment and demonstration of slightly lower LAL in Turkey (Sahinkarakas, 2012) and AL in USA (Kruse et al., 2020). The last opinion, realistic optimism, which wins support only in the field of AL, asserts although they demonstrate inflated confidence towards assessment, the pre-service teachers realize the practical challenges in using assessment methods for instructional purposes in AL

in USA (DeLuca et al., 2013).

In other words, a few studies have been conducted to explore the proficiency level self-evaluated by pre-service teachers in Turkey, Canada, and USA, though the findings are inconsistent (e.g., DeLuca et al., 2013; Kavaklı & Arslan, 2019; Sahinkarakas, 2012). Regrettably, however, little is known on this unsettled issue from the perspectives of pre-service EFL teachers in the context of China. Such exploration may empower the perspectives of pre-service EFL teachers in China and offer a more in-depth and global understanding of pre-service teachers' self-evaluated LAL proficiency profiles for a holistic picture of LAL evaluation from different stakeholders.

Thirdly, a wide range of mediating factors, primarily focusing on the effectiveness of the pre-service teacher preparation programmes from various aspects, have been suggested in the literature to interact with LAL evolvement among pre-service EFL teachers (Y. Xu & He, 2019). The underpinning assumption is that pre-service EFL teachers' LAL would improve if all the crucial mediating factors are identified, and their facilitative or inhibitive effects are further clarified. Although the existing identified mediating factors are fruitful, it is still unclear whether the list is exhaustive and how the mediating factors interact with LAL evolvement (Giraldo, 2020; Mellati & Khademi, 2018; Y. Xu, 2019a; Y. Xu & Brown, 2016; Y. Xu & He, 2019). A better understanding of such an interaction may shed light on how mediating factors impact

on LAL evolvment and how LAL evolvment reflects the mediating factors in the process to provide useful suggestions for effective assessment trainings to pre-service EFL teachers. Thus, one of the objectives of this study is to identify the possible mediating factors and their interaction with LAL evolvment.

Fourthly, a large amount of literature focuses on examining the LAL proficiency level and how to improve LAL among pre- and in-service teachers. A wide range of existing studies indicate that both pre- and in-service EFL teachers seem to be inadequately literate in language assessment (e.g., Davidson & Coombe, 2019; Djoub, 2017; Gonzales & Aliponga, 2012; Graham, 2005; Mellati & Khademi, 2018; Saputra et al., 2020; Vogt & Tsagari, 2014; Vogt et al., 2020; Watmani et al., 2020; X. Yan & Fan, 2020). Other efforts aim to find out how to improve the less satisfactory situation by modifying pre-service education programmes (Bolívar, 2020; Coombe et al., 2020; Giraldo & Murcia, 2019; Smith et al., 2014; H. Xu, 2017) or providing more learning opportunities for in-service teachers on language assessment (Baker & Riches, 2017; Djoub, 2017; Jawhar & Subahi, 2020; Saputra et al., 2020; Tsagari & Vogt, 2017; X. Yan et al., 2018). These studies are fruitful in helping to understand the proficiency level of LAL among pre- and in-service teachers and the potential effects of interventions for improving LAL.

Nonetheless, what is unknown is how LAL develops within individuals (Gan & Jiang, 2020; Harding & Kremmel, 2016; X. Yan & Fan, 2020), especially there is scarce

attention on the evolvement trajectory of LAL among pre-service EFL teachers, where the evolvement of LAL starts (Bijsterbosch et al., 2019; Bolívar, 2020; Giraldo, 2018a, Sevimel-Sahin & Subasi, 2019; Ukrayinska, 2018). Additionally, the framework proposed by Y. Xu and Brown (2016) needs more empirical studies to illustrate how each dimension changes during the evolvement (Gotch & McLean, 2019; Y. Xu, 2019a). Tajeddin et al. (2018) apply only one sub-component of the model into their study, and they encourage further study to examine all the components. In response to the call, my study adopts Y. Xu and Brown's (2016) model as the framework so as to explore the evolvement trajectory of LAL among pre-service EFL teachers. Unless more knowledge is gained about LAL development trajectory, more effective interventions or improvement will be possible to facilitate LAL development among pre-service EFL teachers. Additionally, exploration of pre-service EFL teachers' LAL development profiles enables the comparison of LAL development profiles among different stakeholders become possible.

Lastly, in terms of the research methods, a large amount of research is conducted in quantitative ways (Gan & Jiang, 2020; Kremmel & Harding, 2019). Narrative inquiry, focusing on story-telling, is still a less commonly used method and has been rather few applications in LAL research (Cumming, 2004; Gan & Jiang, 2020; Harding & Brunfaut, 2020). The envisioned prospect of LAL lies in exploratory descriptive interpretation rather than prescriptive (Giraldo, 2020; Inbar-Lourie, 2017). Moreover, the qualitative analysis may well address LAL with sufficient context-sensitive

information and differentiate relationships among personal factors, experiential factors, and local contextual factors (Crusan et al., 2016; Nimehchisalema & Bhatti, 2019). Narrative inquiry offers “great promise for tracing the development of language assessment literacy over time, and for exploring convergences and divergences between different narrators” (Harding & Brunfaut, 2020, p. 62). The basis for the approach is the ubiquity of narrative as, according to Barthes (1977), “narrative is present in every age, in every place, in every society” (p.79). Thus, methodologically, narrative inquiry fits in well with the research objectives of the present study. Academically, the adoption of narrative inquiry is also in response to the call of more application in LAL studies.

Therefore, to bridge the research gap, the current study adopts narrative inquiry as the research method to explore LAL evolvment among pre-service EFL teachers in China. A more detailed description of research objectives and questions is mentioned in the next section.

1.4 Research Objectives

The central objective of the current study is to explore the LAL development of pre-service EFL teachers for primary and middle schools in a selected Normal University in western region of China. The following are the specific research objectives in the current study:

- (1) To understand the conceptualization of LAL among the participants.

- (2) To explore the LAL proficiency level self-evaluated by the participants.
- (3) To explore the interaction between the identified mediating factors and LAL evolvement among the participants.
- (4) To describe the evolvement of LAL among the participants.

1.5 Research Questions

Guided by research objectives, the present qualitative study aims to answer the following research questions by narrative inquiry of pre-service EFL teachers for primary and middle schools in a selected Normal University in western region of China.

- (1) What is the conceptualization of LAL by the participants?
- (2) What is the LAL proficiency level self-evaluated by the participants?
- (3) How do the identified mediating factors interact with LAL evolvement among the participants?
- (4) How does LAL evolve among the participants?

1.6 Background of the Researcher

Unlike the positivism paradigm in quantitative research, qualitative research is embedded in interpretative and constructivist paradigms (Tomaszewski et al., 2020). Hence, in qualitative research, the researcher acts as the research tool in data collection and data analysis procedures (Creswell, 2014). The researcher's subjectivity, which means what the researcher sees and understands is a reflection of the researcher's identity and experiences, is not a bias to be dismissed or removed from the qualitative

study (Braun & Clarke, 2013). But rather subjectivity is a contributor to a more passionate and personal study (Glesne, 2011). Therefore, the following paragraphs will introduce the researcher's background experience and subjectivity that inform the study.

Having worked in pre-service EFL teacher education institute for 14 years since graduation in 2007, I am interested in improving the LAL of pre-service EFL teachers to better deal with the challenges in conducting assessment-related activities in becoming teachers in future. At the beginning of my teaching career, I was unfamiliar with LAL. My initial focus of LAL started from my own challenges in assessing the students to improve their learning. After reading the literature, I realized it was a widespread inadequacy of LAL among the practitioners rather than my own difficulty. From then on, I became aware of LAL and started to think about how to improve LAL at the initial teacher education for the pre-service teachers after reflecting on the exposure to assessment in my prior pre- and in-service EFL teacher education.

The past educational experience enables me to get familiar with the pre-service EFL teacher education in China. I have been trained in pre-service EFL teacher education programmes from bachelor to postgraduate education since 2000 to 2007 in China. In the bachelor's education, I was admitted to a 4-year programme in English education in a provincial Normal University in the central region of China in 2000. After graduation, I have continuously studied as a postgraduate student in a highly ranked

Normal University located in the western part of China for three years. That means, I have been trained in two different Normal Universities in China continuously for seven years. This educational background provides me with personal experience of pre-service EFL teacher education in different types of Normal Universities in China.

Besides, I got access to the course related to research methodology in various settings which offered me a solid foundation of research methods. During the Master and Doctoral studies, I learned Educational Research Methodology and Research Methodology respectively in China and Malaysia. Thus, the previous learning experience as a pre-service EFL teacher and teaching experience as an in-service EFL teacher educator offers me rich information in implementing the current study.

1.7 Theoretical Framework

The theoretical framework of the present study is built on the six-component hierarchy model of AL proposed by Y. Xu and Brown (2016). The details of the model are described in Chapter 2 (See 2.4.2); here the rationales of adopting it as the framework of this study are discussed.

Firstly, from the socio-cultural lens, this model reconceptualizes AL as a dynamic social practice. It moves beyond the traditional view of AL as a repertoire of knowledge, skills, and cognition that reside within the individual (Gee, 2003) to the social-cultural view of AL as responsive cultural practices (Y. Xu & Brown, 2016). It

confirms that AL is not just a static or fixed set of capabilities, but a dynamic and ongoing negotiation with the social, political, and cultural contexts where it is embedded (Willis et al., 2013; Y. Xu, 2019a; Y. Xu & Brown, 2016). This may best explain the interaction between AL and context in the present study.

Secondly, this hierarchy is featured by encompassing all phases of teacher education (Y. Xu & Brown, 2016). It integrates the principles of pre- and in-service teacher education and figures out the elements and influences in fostering teachers' AL. In the model, teachers' AL is reconceptualized by bridging the distance between the research in educational assessment and teacher education. It connects the research on educational assessment and teacher education, each with different goals but sharing the same concern for improved teacher AL and high-quality assessment (Y. Xu & Brown, 2016). Thus, it could be applied to the analysis of pre-service teachers' AL during the initial teacher education phase.

Lastly, the model covers overall trajectory of professional development in AL. It sets three tiers of AL improvement which is appropriate for tracing the evolvement trajectory. This model distinguishes three levels of AL improvement. The first level is a mastery of fundamental assessment principles, and then the second level is internalized assessment understandings and skills. Built on the first two levels, teachers are expected to foster a self-directed awareness of the assessment procedures and identity (re)construction as an assessor at the last level. For pre-service teachers,

they may also experience the identity shift from being students as assessee into teachers as assessor, for it offers an overall trajectory of professional development in AL (Y. Xu & Brown, 2016). Thus, it is helpful to reveal the evolvement trajectory of AL according to the identified tiers of the model.

Though it deals with AL, the model also supports LAL among EFL teachers (Y. Xu, 2019a). Y. Xu (2019a), one of the co-authors, elaborates what each dimension in the AL model means in relation to LAL for EFL teachers and provides arguments in favor of the operationalized model for EFL teachers by evidence from a case study in China. To sum up, the model put forwards by Y. Xu and Brown (2016) is appropriate to be adopted as the framework for exploring pre-service EFL teachers' LAL.

1.8 Significance of the Study

This study has great significance in terms of contribution to LAL research by addressing the research gap and providing implications for LAL improvement. Firstly, with regard to the theoretical contribution, it contributes to the theoretical body of LAL as it clarifies the conceptualization of LAL from the neglected stakeholders, pre-service EFL teachers. The existing literature indicates that the conceptualization of LAL primarily derives from either the professional researchers (Giraldo, 2018a; Malone, 2013) or in-service EFL teachers (Fulcher, 2012). Both have excluded the crucial stakeholders, pre-service EFL teachers. On top of that, most of the studies are carried out in western countries; only a few studies focus on the context of China (Gan

& Jiang, 2020; Lin & Wu, 2014). This study, therefore, extends the scope by focusing on pre-service EFL teachers' LAL in China. It helps to clarify the complex components of LAL from pre-service EFL teachers' perspectives so as to provide a holistic picture of the conceptualization of LAL.

Secondly, concerning the methodological contribution, the study enriches the application of narrative inquiry, a less commonly used research method in LAL exploration. Narrative inquiry empowers the participants whose experiences and inner world are the subject of the research (Barkhuizen, 2016). The current study provides an explorative attempt to inquiry the participants narratively in LAL and confirms that narrative inquiry, an important tool to help understand how the changes are occurring, is feasible to describe the evolvement trajectory of LAL.

Lastly, as to the practical contribution, the study sheds light on the improvement of deficient LAL among pre-service EFL teachers. For pre-service EFL teacher education programmes, the findings help to design effective interventions, such as assessment courses, trainings, or practices, by taking the interrelated mediating factors into account. For EFL teacher educators, it gives them a profound knowledge of their students' LAL in terms of conceptualization, self-evaluated proficiency level, and development trajectory. Such understandings may better serve teacher educators to help pre-service EFL teachers develop LAL. For pre-service EFL teachers, the profiles of LAL development delineated in the current study are likely to prompt them to

critically reflect on their LAL evolvement. Such a reflection may make them know where they are, where they need to be, and how best to proceed along the LAL evolvement trajectory.

1.9 Definition of Terms

The definitions of the key concepts mentioned in the thesis are listed below.

Alternative assessment – an umbrella term which shelters any alternatives to standardized testing (Gipps & Stobart, 2003), such as performance assessment, outcome-based assessment, and student-centered assessment (Shepard, 2000).

Assessment – it is the process of collecting information in order to monitor learners' progress and make instructional decisions, including not only a test, but also observations, interviews, behavior monitoring, etc. (Kizlik, 2012).

Assessment content & criterion—knowledge of how to assess the learning goals and specific content being learned (academic achievement or affective performance), and knowledge of rationale for grading or rubrics.

Assessment interpretation & communication—knowledge of ways of interpreting evidence generated from assessment, and ways of communicating assessment results to stakeholders such as students, parents, managers/administrators, and the general public.

Assessment literacy – it is far more than the accumulation of knowledge and skills related to assessment, but rather it is a complex socio-cultural professional disposition tied to teacher identity and other aspects of assessment practices (Coombs et al., 2018).

Assessment methods & implementation—knowledge of a wide range of assessment strategies and competence of applying and carrying out them for the target learners.

Assessment principle & ethics—knowledge of why to assess (i.e., formative, summative) and understanding legal and ethical responsibilities concerning the design, use, storage, and dissemination of assessment.

Assessment washback—knowledge of potential influence of assessment, whether beneficial or damaging, on teaching and learning.

Classroom-based assessment – any reflection by teachers (and/or learners) on the qualities of the learners' work and the use of that information by teachers (and/or learners) for diverse purposes (K. Hill & McNamara, 2012).

Disciplinary & cross-disciplinary competence-- knowledge of language, English curriculum, and other related disciplines, as well as a command of English.

In-service EFL teacher – the full-time teacher teaching English as a foreign language in primary or middle schools in China.

Language assessment literacy – a dynamic context-dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledges with others (including learners) in the entire process of assessment guided by their principles and oriented by historical, social, political, and philosophical frameworks so as to achieve the learning goals (Fulcher, 2012; Willis et al., 2013).

Language assessment literacy evolution – the process in which the practitioner's language assessment literacy changes to a more advanced level.

Mediating factors – factors (contextual and experiential) mediating or shaping LAL evolution (Crusan et al., 2016).

Narrative inquiry – research that views stories, whether gathered through field notes, interviews, oral tales, blogs, letters, or autobiographies, as fundamental to human experience (Tracy, 2013).

Pedagogical content knowledge – knowledge of how to teach the curriculum-based content to learners and the competence to sustain learning.

Personal conception & attribute – knowledge of how one's own preconceptions, understandings and opinions may inform one's conceptualizations, interpretations, judgments and decisions in assessment, and the individual's personality.

Pre-service EFL teacher – an undergraduate student who is majored in English education of teaching English as a foreign language, intending to work with students in primary or middle schools in China.

Primary and middle schools – the elementary schools, junior middle schools, and senior middle schools in China.

Traditional assessment – the standardized and classroom achievement tests that are predominated by close-ended items, e.g., multiple choice, true/false, fill in the blanks (Bol et al., 1998).

1.10 Layout of the Thesis

The thesis is composed of five chapters and presented as follows. The first chapter begins with the introduction of the research background, especially the assessment context and pre-service teacher education in China. Then the research objectives and questions are presented. It ends with the significance of conducting the study and an explanation of key terms.

Chapter 2 reviews the literature pertaining to AL and LAL. It first discusses the rationales of AL for teachers, including the necessity of LAL for EFL teachers against the background of assessment paradigm shift. Followed this significance, a range of definitions and frameworks of the concept are analyzed to clarify their meanings. It also sorts out the empirical evidence focusing on both pre- and in-service teachers. In each target population, their proficiency level and mediating factors identified by prior studies are presented to set a foundation of analysis and discussion of the present study.

Chapter 3 outlines the research method adopted in the present study. It first justifies the rationale of narrative inquiry, and then followed by the research design. The data collection and analysis procedures are described in a detailed way for a better understanding of the whole process of implementation. Moreover, the ethical considerations in this narrative inquiry are depicted throughout the entire research process and the reporting stage. Meanwhile, the trustworthiness of the study is also analyzed.

Chapter 4 consists of the research findings according to the sequence of research questions put forward in the current study. The findings first present the conceptualization of LAL and then the self-evaluated proficiency level by the participants. The identified network of mediating factors in LAL evolvement are illustrated. Finally, LAL evolvement trajectory profile of each participant is analyzed to draw the possible similarities.

Chapter 5 discusses the research results in relation to the previous studies. It also presents the implications for the parties involved, like pre-service EFL teachers, teacher educators, and pre-service teacher education programmes to further improve LAL. The enlightening contributions to the existing body of LAL literature and limitations of the study are also analyzed. At the end of this chapter, recommendations for further studies are listed to direct the subsequent efforts in LAL field in the future.

1.11 Summary

This section starts with the importance of becoming language assessment literate for both pre- and in-service EFL teachers and the information pertinent to assessment education and English assessment practices in China as the background knowledge of the research context. After the research gaps are identified in the field of LAL research, research purposes and specific research questions in the field of LAL are put forward. It aims to focus on pre-service EFL teachers in China to explore their LAL conceptualization, self-evaluated proficiency level, the evolvement trajectory, and interaction between the identified mediating factors and LAL evolvement.

It also provides the rationales of setting Y. Xu and Brown's (2016) model as the theoretical framework in terms of four advantages: its socio-cultural perspective, the combination of educational assessment and teacher education, tiered mastery level, and appropriateness to be applied in LAL research. Besides, the significance of the study is analyzed. Finally, key terms are defined for a clear and unified understanding.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

A comprehensive review of the literature related to AL and LAL may provide a basis for understanding what has been researched and what still remains the gap waiting to be filled in. Thus, the chapter will first introduce the assessment paradigm shift as background knowledge to understand the significance of AL and LAL. Then the rationale of AL among teachers is listed from diverse aspects. After the discussion of the importance of AL among teachers, the definitions and conceptual frameworks of AL and LAL are presented together, due to the fact that the large body of literature on LAL derives from that of AL in general education (Inbar-Lourie, 2017). At last, the empirical studies related to both pre- and in-service teachers are categorized into AL and LAL proficiency level and mediating factors identified in the literature so far.

2.2 The Assessment Paradigm Shift

The paradigm is defined as “a set of interrelated concepts which provide the framework within which we see and understand a particular problem or activity” (Gipps, 1994, p.1). The paradigm in which we work determines what we intend to, how we construct, and how we solve emerging problems (Gipps, 1994). The practice and philosophy in the field of assessment have undergone a dramatic change, which is described as a paradigm shift (Gipps, 1999). It has been shifted from a positivistic

paradigm prioritizing traditional psychometrics to an interpretative paradigm emphasizing qualitative socio-cultural perspectives, from a testing-oriented examination culture to a learning-oriented assessment culture (Gipps, 1994, 1999; Scarino, 2009; Shepard, 2000).

Closely linked to the positivist lens of knowledge and reality, the traditional paradigm views knowledge as an object which is transmitted from instructor to learner and later measured by norm-referenced standardized tests (e.g., large-scale test) as the primary tool at the end of the instruction to give a reliable and valid measurement of what students have learned (Klenowski, 2009; Serafini, 2000; J. R. Wang et al., 2010), to rank students in accordance with certain criteria (L. Wilson, 1994), or to provide accountability (Broadfoot & Black, 2004). Besides, sometimes it is inside the classroom that the evidence teachers gathered is applied to determine a student's learning grade or unit final test (Stiggnis et al., 2004). Teachers always put the validity and reliability of the assessment at priority to guarantee the accurate judgment of students' academic achievement (L. Earl & Katz, 2006). The students' role in the assessment is rather passive, either the object of assessment or the recipient of assessment results. All these characteristics can be encompassed under the umbrella term AoL, interchangeable with the concept of summative assessment (DeLuca & Klinger, 2010; Hughes, 2014).

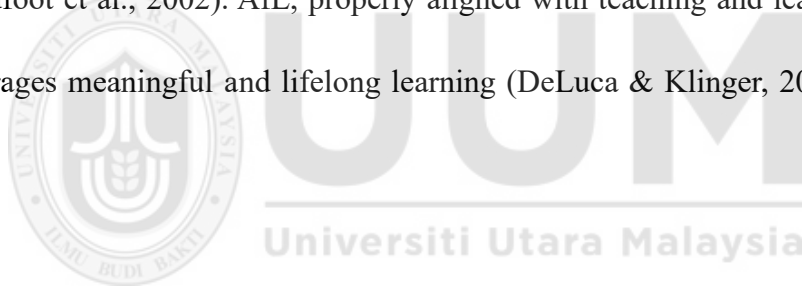
The deficiency of the old model is obvious that the scores achieved from the test can

only provide a facet but not the overall picture to serve students better (Stiggins, 1994). In addition, it is ineffective to assess students' high-order thinking skills, real-world problem-solving abilities, and communicative competence which is crucial to prepare students to be critical thinkers and lifelong learners in the contemporary 21st century (Koh & Luke, 2009). Consequently, it would definitely go through a shift if the old one fails to resolve the prominent issues (Gipps, 1994).

The new emerging paradigm views assessment within the social-cultural framework. As an integral part of teaching and learning, assessment is taken as a dynamic, interactive, and collaborative process built within the network of the social and cultural life of the classroom, with a focus on assessment of learning process, elicitation of elaborated performance, and highlight of collaboration (Gipps, 1999; Han & Kaya, 2014). To this end, alternative assessment methods have achieved great attention from teachers who are expected to apply a wide range of both quantitative and qualitative assessment skills to help to inquire about students' learning process (Han & Kaya, 2014; Serafini, 2000; Stoyanoff, 2012), to make judgments on the nature of a student's construction (L. Wilson, 1994), to provide feedback to empower learning, and to make informed decisions related to teaching (Klenowski, 2009; J. R. Wang et al., 2010).

To be specific, this new paradigm advocates AfL or further AaL. AfL, which is aimed to enhance student learning, is different from AoL designed to provide accountability or ranking (Black et al., 2004). AfL covers a more extensive meaning than the familiar

concept of formative assessment (Hughes, 2014; Stiggins et al., 2004), for AfL underscores teachers' statement of descriptive information instead of evaluation data of their students and emphasizes students' active engagement into a series of assessment chains ranging from clarification of assessment targets to self-assessment to communication with other stakeholders concerning their achievement progress (L. Earl & Katz, 2006; Stiggins et al., 2004). The classroom assessment information is used by both teachers and learners to adjust the teaching and learning process (Chappuis & Stiggins, 2002) and to judge where the learner is in the process of learning, where the learner needs to go, and how to reach the destination in the best way (Broadfoot et al., 2002). AfL, properly aligned with teaching and learning practices, encourages meaningful and lifelong learning (DeLuca & Klinger, 2010; Klenowski, 2009).



L. Earl (2003) steps further to identify a subset of AfL, the concept of AaL, which advocates instructors and learners to utilize assessment as a metacognitive instrument which allows learners to self-monitor their learning activities so that learners are able to select most appropriate strategies to satisfy their learning needs (DeLuca & Klinger, 2010; L. Earl, 2007; Lee & Son, 2015; Rodríguez-Gómez & Ibarra-Sáiz, 2015). Learners play the significant role of critical agents linking assessment and learning together. They are seen as active and concentrated assessors, who figure out the meaning of the assessment information, associate it with the previous knowledge base, and apply it for acquiring new knowledge and learning new skills. This process

depicted above is how regulatory activities work in metacognition (L. Earl & Katz, 2006). AaL also helps students to cultivate lifelong and self-directed learning competence through their active participation in self-assessment and peer-assessment activities (Dann, 2014; K. Earl & Giles, 2011).

What is needed for teachers to be prepared for changes in paradigm shift goes beyond the surface behavior of observable classroom assessment practices, but also should involve a fundamental shift in underlying values and philosophies to reconsider the relationship between assessment and teaching-learning process (L. Earl, 2013; Gielen et al., 2003; Pedder & James, 2012; Stoyanoff, 2012; L. Wilson, 1994) and to review the dynamic power relationship between teachers and students, who have been empowered more responsibilities to impact the process of learning (Gipps, 1999; Inbar-Lourie & Donitsa-Schmidt, 2009; Pedder & James, 2012). Instructors are supposed to change their roles from assigning the tasks for students to complete to “a kind of orchestration of the learning” (James et al., 2007, p. 217). Meanwhile, students are not merely confined to the role as the passive receiver of their teacher’s behaviors, but are entitled as co-constructors of collaborative teaching and learning activities (Pedder & James, 2012). As Moss et al. (2006) explain that discrepancy between paradigms lies in the philosophical level, teachers must embrace a principle that collaborates assessment theories and skills with the teaching-learning process (Klinger et al., 2012; Scarino, 2013).

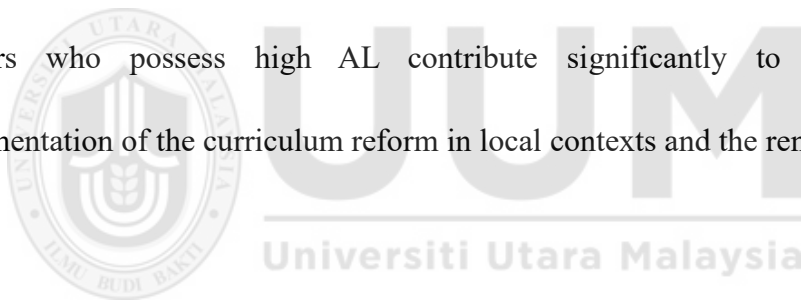
It is widely agreed that the assessment paradigm has been shifted from AoL to AfL or even AaL (K. Hill, 2017; Torrance, 2007), from quality control to quality assurance in the process of learning, and from the product of learning to the process of learning (Wiliam & Thompson, 2008). However, this is not to mean that AoL anchored in traditional standardized tests is useless nowadays (Gipps, 1994), just as East (2008) contests that no paradigm is better or worse than another, nor right or wrong. Each is built on the basis of unique underpinning premises about what we intend to measure and globally both traditional and alternative assessment methods co-exist (Redecker & Johannessen, 2013). It is plausible to assume that two paradigms are not regarded as an opposite but as a continuum from viewing knowledge as a commodity or an object external to learning at one extreme to viewing knowledge as an activity or a process co-constructed by learners and teachers at another extreme (Hargreaves, 2005).

The testing and assessment cultures, each embedded in different epistemological paradigms, call for reformulating and reorienting the assessment competencies required for teachers to engage in assessment in educational contexts (Inbar-Lourie, 2008; 2017). A high literacy of teachers in terms of using traditional (objective) assessment techniques would not ensure they can understand how to deal with alternative (subjective) assessment methods (Quilter, 2000). Thus, there is a consensus among scholars that AoL has to be supplemented with AfL and AaL in a balanced way to aid teachers to develop a more comprehensive and accurate understanding of student learning so as to serve students better in their learning process (Berger, 2012; Chappuis

& Stiggins, 2008; Kahl et al., 2013; Shepard, 2000; Soh & Zhang, 2018; Stiggins, 2004; Tran, 2012).

2.3 Rationales of Teachers' AL

Under the background of learning-oriented assessment culture, teachers' AL is put under the spotlight. It is crucial for teachers to be assessment literate for the following four reasons. AL is the mandated demand of the assessment reform policies worldwide, including China. Teachers' AL is the essential aspect in their professional development and one of the most influential factors in students' academic performance. Lastly, teachers who possess high AL contribute significantly to the successful implementation of the curriculum reform in local contexts and the renewal of school.



2.3.1 Mandated Demand of National Policy

In response to the paradigm shift in assessment, assessment-related policies worldwide have already undergone reform with an inclusion of the latest research findings to challenge a balanced teacher's AL constituted by both traditional and alternative assessment skills in conducting assessment in educational activities, such as in the USA (DeLuca et al., 2016b), Finland (Hildén & Fröjdendahl, 2018), Singapore (Shin, 2015), Israel (Inbar-Lourie & Levi, 2020), New Zealand (Smith et al., 2014), North Africa (Davidson & Coombe, 2019), and China (Gu, 2014), to name just a few. To stay focused, the policy concerning the issue of assessment within the context of China will

be analyzed in detail.

In China, there are two main national policies involving the professional responsibilities of primary and middle school EFL teachers in assessment practices. One is professional standards for primary and middle school teachers issued in 2012 and the other one is Curriculum Standards for the English Course in Middle Schools by MoE of the Peoples' Republic of China in 2018. The first document, the professional standards for teachers in primary and middle schools, differentiates assessment requirements for teachers in three distinctive stages: novice-skillful, skillful-proficient, and proficient-excellent. For the novice and less experienced EFL teachers, they are expected to know how to use multiple assessment techniques to monitor learning. For proficient teachers in the second stage, the assessment skills and competence are notably underscored ranging from engaging students in peer-assessment and self-assessment activities and utilizing processive assessment methods to appropriately design end-of-unit or end-of-term tests for the purpose of checking learning outcomes. For excellent instructors, they are suggested to take the feeling of students through and after the assessment into account. Excellent teachers should also help students to experience the facilitative function of assessment for learning and provide learning improvement suggestions accessible to students. Through the list of the varying requirements for teachers in different career stages, what can be seen clearly is the upgrading complexity of LAL requirements advancing from possession of assessment awareness and knowledge for novice teachers, through the attainment

of assessment competence for proficient teachers, to internalization of assessment ethics for excellent professionals.

With regard to curriculum reform, the policy issued by MoE (2018) suggests six assessment standards in the design and implementation of English course in middle school:

- (1). put core literacy in prominent place in assessing student's academic performance with a focus on student's overall development and growth;
- (2). treat students as the subject in the assessment to serve student's overall development and progress;
- (3). focus on the teaching process within the classroom to implement various assessments through activities carried out in English;
- (4). focus on the rationale and variety of assessment methods to effectively conduct formative assessments;
- (5). handle well the relationship between daily assessment and periodical assessment to select appropriate paper-pencil tests; and
- (6). utilize the washback effect of assessment productively to realize assessment for teaching and learning. (p. 80)

Thus, according to the policy, LAL is an integral part of EFL teacher's daily teaching work. A language assessment literate teacher is supposed to select AoL properly, optimize AfL effectively, and practice AaL productively.

A trend is increasingly evident that classroom teachers have shouldered much greater assessment responsibilities both in amount and sophistication (Lan & Fan, 2019; Leahy et al., 2005; Puspawati, 2019; Wise & Lukin, 1993). The purpose of assessing is not merely restricted to the measurement of student learning but also creating more accessible learning opportunities (Djoub, 2017). This challenges teachers to be competent in various aspects of the assessment process, such as comprehending the

rationale underpinning the learning-oriented assessment (Brown et al., 2011), having confidence in their abilities to be engaged in the whole process of assessment (Kruse et al., 2020), mastering the assessment principles and practices (Brindley, 2001; Lan & Fan, 2019; Lee & Son, 2015), collecting information about students' cognitive, affective, and psychomotor abilities (Puspawati, 2019), applying assessment results to adjust teaching, and giving proper feedback for students to improve learning collaboratively (Leahy et al., 2005). Thus, all these assessment tasks recommended in the policies set a higher expectation of EFL teachers' LAL.

2.3.2 Professional Requirement of Teachers

It has long been agreed that assessing student's academic performance is one of the most crucial parts of school teachers' daily work (Bachor & Baer, 2001; Mertler, 2003; Mertler & Campbell, 2005; Napanoy & Peckley, 2020; Plake, 1993; Shulha, 1999; Suah & Ong, 2012; Zulaiha et al., 2020). A dominant proportion of teacher's time is devoted to the assessment-related activities (Cheng et al., 2004; McMillan & Nash, 2000; Mertler, 1998). As estimated by Stiggins (2007b, 2014), this percentage may amount to a quarter or even a third of their available time in dealing with issues around assessment. Teachers make nearly hundreds of decisions and judgments concerning what to assess, how to assess, and how to teach in the classroom every hour, or at the rate of one every two or three minutes on average (L. Earl, 2007; Stiggins & Conklin, 1988; Tatto, 2006).

AL of teachers is not only a critical guarantee of high-quality assessment, but also an indispensable feature of an effective teacher. Teacher's expertise in assessment is undoubtedly essential to the sophisticated judgments with regard to the validity, reliability, etc. of the assessment practices in concrete contexts (Kane, 2006; Messick, 1989; Y. Xu, 2019b). Moreover, it is also a key element in successful and effective teaching as teacher's effectiveness partially depends on the teacher's interpretation and use of all types of assessment results to respond to diverse students' learning needs and inform instructional decisions (Daniel & King, 1998; Gareis, 2007; Joachim et al., 2020; Pastore & Andrade, 2019; T. H. Wang et al., 2008).

For today's EFL teachers, LAL has also been a key factor in the success of instruction (Büyükkarçı, 2016; Munasinghe, 2020) and an integral part of their ongoing professional development (Coombe, 2010; Harding & Kremmel, 2016; Scarino, 2013). Such a language class where no assessment occurs would not be possible to exist (Gonen & Akbarov, 2017). Assessing EFL learners' performance is one of the most essential roles of EFL instructors in the language classroom (Watmani et al., 2020). The EFL instructors, often formally or informally, monitor the students' performance and continue the lesson according to the responses from learners; they usually play the double role of the teacher and assessor (Hakim, 2015; Jin, 2010). Thus, it is undeniable that language teachers need a solid assessment knowledge to direct teaching in accordance (Berger, 2012) and the deficiency in LAL may be one of the contributing factors linked to the declining English teaching standards (Sultana, 2019).

To sum up, being a central part of teacher's professional improvement, AL seems like a necessary commodity for teachers in all subjects across the contexts (Bracey, 2000; Kelly et al., 2020; Newfields, 2006; Weigle, 2007). It is imperative for teachers to acquire an adequate level of AL to engage in their professional activities (Fan et al., 2011; Stoyanoff, 2012; Zhang & Yan, 2018) and language teachers are no exception. The significance of LAL for EFL teachers cannot be minimized (Giraldo, 2018b, 2020). Those who are still ignorant about the assessment in education and overlook the significance of AL would commit "professional suicide" (Popham, 2011, p. 269), and AL is even perceived by Bachor and Anderson (1994) as both an academic skill and a life skill for instructors to survive.

2.3.3 The Academic Well-being of Learners

Teachers' AL is not only beneficial for them to fulfill their professional responsibilities, but also conducive for the educational well-being of their learners (Lian et al., 2014; Koh & Velayutham, 2009). It is indicated that teacher quality seems to be the strongest impactful factor in student achievement (Darling-Hammond, 2000). As a powerful tool to orient student learning, assessment cannot be undervalued if teachers intend to change the way of learner's learning (McKellar, 2002).

Teachers who are equipped with sufficient knowledge and skills to use assessment for various purposes could succeed in motivating students to learn and improving their

academic performance (Campbell & Collins 2007; Luthfiyyah et al., 2020; Roeber, 2016; Rogler, 2014; Stiggins, 2007b; White, 2009). Coombe et al. (2012) confirm that teachers with a higher level of AL may be more likely to help students to attain better academic performance. Evidence has also shown that assessment literate teachers' classroom assessment practices are greatly different from those implemented by assessment illiterate teachers in that the former is more facilitative for students' learning and motivation (Mellati & Khademi, 2018). Furthermore, students' AL is more possible to be developed under the professional guidance of assessment competent teachers who are skillful at passing the assessment responsibilities to their students by modeling and scaffolding (McMillan & Hearn, 2008).

However, if there is a deficiency in teachers' assessment knowledge and practices, the potentially disastrous effect on learners appears to be unescapable. It would be, more often than not, the students rather than the teachers who experience the dire results of the unsound assessment activities (Quilter, 2000). For instance, an essential component of AL is to use assessment results to feedback (Rogler, 2014). Brown et al. (2012) reveal that feedback, as a crucial element in AfL, provided by teachers is strongly focused on improving students' learning rather than their well-being (e.g., in encouragement and protection of learners' self-esteem). Consequently, student may not learn how to give effective feedback to themselves and their classmates without experiencing the growth-oriented feedback from their teachers. Indeed, students may have chances to escape from the undesirable teaching through their efforts, but they

seem to be trapped by the consequences of unqualified assessment if they are to graduate (Boud, 2006). Research has indicated teachers' weakness in developing high-quality tests (Williams, 2015; Zulaiha et al., 2020). The potential consequences of the poor teacher-devised end-of-term examinations seem impossible to be removed from the test-takers who need to pass the examination to be promoted into higher grade.

In the research of language testing and assessment, there is no doubt that LAL is a must for EFL teachers if they would like to help students to attain improved performance (Şişman & Buyukkarci, 2019). Meanwhile, EFL learners' empowerment as active agents in classroom-related language assessment seems to rely on to what extent the instructors have a mastery of LAL and learners' awareness of AfL may be fostered by being involved in assessment activities initiated by highly assessment literate teachers (Vogt et al., 2020). Additionally, with regard to the improvement in the four skills (i.e., listening, speaking, reading, and writing), much evidence reveals the significant positive relationship between EFL teachers' LAL and students' achievement in reading (Braney, 2011) and writing (Mellati & Khademi, 2018).

In a word, the existing studies have associated AL in general education or LAL in language assessment with noticeable benefits in improving teaching and enhancing students' achievement (Nimehchisalema & Bhatti, 2019; S. Wilson et al., 2001). It may be unlikely for instructors to identify their students' learning needs, keep track of their learning progress, evaluate the impact of teaching, giving grades, and communicate

the assessment results with the absence of AL (Stiggins & Conklin, 1988). Hence, AL is critical in the process of both teaching and learning (Volante & Melahn, 2005), serving as a dual tool for the professionalism of the teachers at first, and then for the student's improvement in learning at the second step (Amirian et al., 2016).

2.3.4 The Guarantee of Successful Curriculum Reform

Teachers' AL has been regarded as one of the key factors in curriculum reform (Gu, 2014). In educational reforms or innovations, assessment acts as a cornerstone (Dierick & Dochy, 2001). Any changes in curriculum and learning objectives are fruitless when assessment practices inside and outside the classroom remain nothing different from the old (Cachia et al., 2010). Teachers' conceptions and practices of assessment play a decisive role in how the curriculum is implemented inside the classroom (Chew & Lee, 2013; Laren & James, 2008). In other words, whether the curriculum reform is a success or a failure is largely determined by the assessment practices enacted by classroom teachers (Gu, 2014).

An additional benefit of possession of AL among teachers is for the school's success (Noonan & Renihan, 2006). Any endeavors towards school improvement must include such an element as teachers are acquainted with the knowledge on how to use the large amount of classroom assessment time effectively and productively (Stiggins, 1993). L. Earl (2007) argued that assessment is viewed as one of the most influential levers for renewing school and that teachers with a good mastery of assessment can be perceived

as the powerful agents to push the assessment lever for school refreshment.

To draw the above discussion about the rationales of teachers' expertise in assessment-related activities into a conclusion, teacher's AL or EFL teachers' LAL is doubtlessly positioned at the very center of the educational assessment and the overall education quality (Y. Xu & Brown, 2017). The critical importance of teachers' AL cannot be under-emphasized (Soh & Zhang, 2018). The growing expectation of educators' AL will continue to be greater given the pervasiveness of tests and daily assessments (Bracey, 2000) and ongoing educational reforms (Kahl et al., 2013).

2. 4 Conceptual Understanding of AL and LAL

The above discussion of AL and LAL has been progressed without formally defining AL and LAL. The priority for teachers to develop AL has been illustrated, yet what AL and LAL mean for teachers as well as what components the terms contain need to be clarified in this section.

2.4. 1 Definitions of AL and LAL

AL, first coined by Stiggins (1991) in general education, was defined from the description of the assessment literates who could discern the high-quality assessment from unsound ones. Nearly a decade later, one of the earliest attempts to integrate AL with language assessment was Brindley (2001), who argued for a curriculum-based

LAL, although he did not specifically address LAL (Fulcher, 2012; Hildén & Fröjdendahl, 2018; Inbar-Lourie, 2017).

The subsequent defining efforts on AL and LAL could be roughly evolved from traditional instrumental understanding into a modern social-cultural perspective (DeLuca et al., 2019). Initial definitions on AL, both in general education (e.g., Alkharusi, 2011; Brookhart, 2011; Koh, 2011; Mertler, 2009; Mertler & Campbell, 2005; O'Sullivan & Johnson, 1993; Paterno, 2001; Plake et al., 1993; Popham, 2009; Stiggins, 1995) and in language assessment (e.g., Brindley, 2001; Malone, 2013; Vogt & Tsagari, 2014) concentrated on identifying the componential elements of assessment knowledge and skills primarily required of teachers. For example, Paterno (2001) conceived AL as “the possession of knowledge about the basic principles of sound assessment practice, including terminology, the development and use of assessment methodologies and techniques, familiarity with standards of quality in assessment” (p. 2). Similarly, Popham (2009) asserted genuinely assessment literate teachers should be “conversant with a wide array of potential assessment options” beyond the knowledge of developing appropriate assessments (p. 7). Influenced by the view, LAL was also approached in the same way. According to Vogt and Tsagari (2014), LAL was mastery of knowledge, skills, and familiarity with principles, procedures, and practices especially in language assessment. LAL was the combination of AL skills in general and language-specific skills related to language assessment (Inbar-Lourie, 2017; Malone, 2013).

The later broader conceptualizations of AL (e.g., Gareis & Grant, 2015; Kim et al., 2020; Lam, 2015; Zolfaghari & Ahmadi, 2016) and LAL (e.g., Davies, 2008; Fulcher, 2012; Giraldo, 2018a; Hassan & Coombe, 2020; Inbar-Lourie, 2008; Shahzamani & Tahririan, 2021) were extended to contain the awareness of the potential function and impact of assessment on individuals and society at large. For example, according to Gareis and Grant (2015), assessment literate teachers needed to use assessment practices to enhance students' learning. Zolfaghari and Ahmadi's (2016) definition took the understanding of the emotions of the learners into account. Kim et al. (2020) advocated to include the social context of the assessment into the definition apart from a set of knowledge, skills, and competencies needed in assessment.

In the same vein, Davies (2008) incorporated the awareness of the proper use of language tests, fairness, and potential impact on a range of stakeholders into the core elements of LAL. For language assessment literate teachers, they should acknowledge the reasons or rationales for assessment (Inbar-Lourie, 2008; Shahzamani & Tahririan, 2021). Unlike the previous definitions, Fulcher (2012) provided a more detailed and comprehensive working definition of LAL by highlighting the wider social-political context where assessment embedded, though the element of language was excluded from the following description

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 understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals. (p. 125)

Over time, these instrumental conceptualizations of AL and LAL were shifted towards a recent application of social-cultural lens which was linked to practitioner's evolving professional identities (e.g., Coombs et al., 2018; Cowie et al., 2014; Looney et al., 2018; Scarino, 2013; Willis et al., 2013; Y. Xu & Brown, 2016). AL was far more than the acquisition of assessment-related knowledge and skills, but rather a social-cultural professional disposition tied to teacher identity. It was reconceptualized by Willis et al. (2013) as "a dynamic context-dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledges with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students" (p. 242). Looney et al. (2018) echoed the perspective with updated teacher assessment identity involving

not only a range of assessment strategies and skills, and even confidence and self-efficacy in undertaking assessment, but also the beliefs and feelings about assessment that will inform how teachers engage in assessment work with students, and focuses not simply on what teachers do, but on who they are. (p. 446)

Therefore, more than an accumulation of knowledge and skills mandated in assessment activities, AL was contemporarily reconceptualized as a developmental process that was mediated by a variety of external and internal factors, including but are not limited to context, opportunities to learn, personal preferences, and system culture (Coombs et al., 2018; DeLuca et al., 2016a).

In this study, which aims to describe the trajectory of pre-service EFL teacher's LAL involvement, the latest social-cultural stance is more suitable for the developmental perspective towards LAL and the acknowledgement of the mediating factors. Thus, based on the social-cultural reconceptualization synthesized from Fulcher (2012) and Willis et al. (2013), LAL is considered to be a dynamic context-dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledges with others (including learners) in the entire process of assessment guided by their principles and oriented by historical, social, political, and philosophical frameworks so as to achieve the learning goals.

To summarise, the conceptualizations of AL and LAL have been evolved from an instrumental perspective to a social-cultural lens. However, maybe largely due to its multifaceted and context-sensitive nature (Esfandiari & Nouri, 2016; Giraldo, 2019; Kaur et al., 2018; A. Kim et al., 2020), no consensus on the conceptualizations of AL and LAL has been reached yet (Deygers & Malone, 2019; Giraldo, 2020; Shahzamani & Tahririan, 2021). Its concrete meanings and specific requirements depend on the varying needs of different stakeholders in diverse educational contexts (Pill & Harding, 2013; Y. Xu & Brown, 2016). But so far, the existing conceptualizations of LAL are derived either from academic researchers or in-service teachers, little is known from the pre-service EFL teachers' perspectives, an important stakeholder in LAL.

2.4.2 Frameworks of AL and LAL

Alongside the progression of conceptualizations of the terms, the frameworks of AL and LAL have also evolved from a mere focus on assessment-related knowledge and competence to a broader focus on the context-dependent interactive process. The representative influential models are listed below in a chronological order in Table 2.1.

As illustrated in Table 2.1, there are primarily four types of models, focusing on AL or LAL, namely, competence lists, three-component models, scaled models, and negotiation models. The first type is characterized by listing the most necessary competence a teacher needs to possess to engage in assessment activities. This kind of expectation is usually set by authorities or institutions, for example AFT, NCME, and NEA.

The second type conceptualizes it as a three-tiered concept, which is mediated by a variety of factors, for instance, Fulcher (2012) and Inbar-Lourie (2008). But this type fails to distinguish the depth of LAL required for teachers in different professional development stages. Thus, the scaled models are put forward with varying importance in multiple dimensions, as evidenced in Shin (2015), Taylor (2013), and Kremmel and Harding (2020). The last type highlights the negotiation of AL with the contexts (Y. Xu & Brown, 2016). More detailed descriptions are provided in the following sub-sections.

Table 2.1

Classification of Frameworks of AL and LAL

Type	Characteristics	Focus	Author (Date)
1	Competence lists	AL	American Federation of Teachers, National Council on Measurement in Education, and National Education Association (AFT, NCME & NEA, 1990), Stiggins (1993), and Brookhart (2011)
2	Three-component models	AL LAL	Lam (2019) and Pastore & Andrade (2019) Davies (2008), Inbar-Lourie (2008), Fulcher (2012), and Giraldo (2018b)
3	Scaled models	AL LAL	Shin (2015) Pill & Harding (2013), Taylor (2013), Baker & Riches (2017), Kremmel & Harding (2020), and Bøhn & Tsagari (2021)
4	Negotiation models	AL	Y. Xu & Brown (2016) and E. Huang & Yang (2019)

2.4.2.1 Competence Lists

Early research attempted to list the core knowledge and competence a teacher needed to become literate in educational assessment (e.g., AFT et al., 1990; Brookhart, 2011; Stiggins, 1993). It was acknowledged that one of the earliest and most authoritative documents is Standards for Teacher Competence in Educational Assessment of Students under a joint effort by AFT et al. (1990). They proposed seven items of assessment competence covering the whole process of assessment activities from assessment design to interpretation and communication of the assessment. The list was so influential that it was adopted as a blueprint for subsequent questionnaires to investigate to what extent teachers were assessment competent in the assessment activities (Y. Xu, 2013).

Another list of five aspects developed by Stiggins (1993) was loosely convergent with the 1990's standards though some competencies were reclassified. Along with the learning-orientation in assessment culture, Brookhart (2011) criticized AFT et al.'s (1990) standards for over-reliance on summative assessment and being outdated due to failure to include the formative assessment and accountability context, so he put forward a repertoire of 11 core competencies, in which the content of understanding learning in the subject and articulating clearly the learning intentions to students were newly added to reveal the teaching-contextualized nature of assessment (Y. Xu, 2013).

2.4.2.2 Three-component Models

Different from these models in the first type with a restricted focus on knowledge and skills, several three-dimensional models were developed with different labels and emphases by incorporating a range of contextual factors (e.g., Davies, 2008; Fulcher, 2012; Giraldo, 2018b; Inbar-Lourie, 2008; Lam, 2019; Pastore & Andrade, 2019).

The two models focusing on AL in the second type were developed in response to Scarino's (2013) calling for an inclusion of the teacher's interpretative framework in conceptualizing AL. Lam's (2019) conceptual framework encompassed teacher conception (belief systems) besides the assessment knowledge base and practices. The social-emotional dimension (the teacher's ability to manage the social and emotional effects of assessment) was also highlighted in Pastore and Andrade's (2019) conceptual framework, in which the other two dimensions were conceptual and

praxeological (See Figure 2.1).

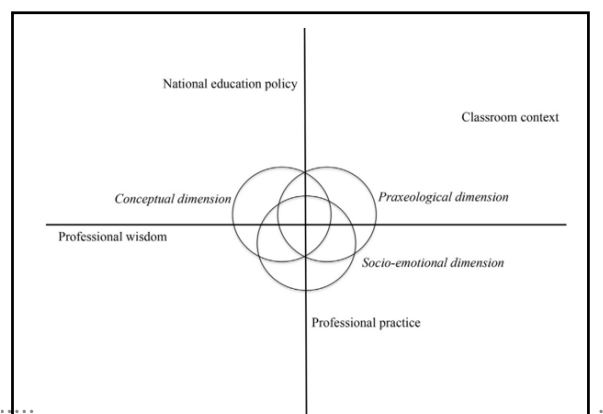


Figure 2. 1 Pastore and Andrade's (2019) three-dimensional framework of AL

In the conceptual framework of LAL, four models were proposed by diverse methods. After examining the changes in language testing textbooks, Davies (2008) noted the emerging importance of principles and supplemented it into “Knowledge + Skills + Principles” model (p. 335), in which knowledge referred to the information about assessment and language; skills contained the assessment strategies or expertise; and principles entailed concepts underlying testing such as validity, reliability, and ethics as well as professionalism. This model has been still prevalent in more recent studies (e.g., Giraldo, 2018b; Lin, 2019).

Likewise, based on the elaboration of Brindley's (2001) model suggesting a modular professional development programmes, Inbar-Lourie (2008) established a framework for courses in language assessment, labeling “Why + What + How” model (p. 390). To be specific, the three components were the purpose or rationale of assessment (the “why”), the theoretical basis of assessment (the “what”), and the performance of the assessment process (the “how”).

Based on the need analysis of in-service EFL teachers, Fulcher's (2012) hierarchy clarified the inter-related relationship among the three elements and addressed the issue of LAL in a much wider context (See Figure 2.2). In his three-tiered Practices + Principles + Contexts framework, the bottom was the practice of language assessment (knowledge, skills, and abilities) with principles (processes, principles, and concepts) positioned in the middle serving as the guidance of the practice, and the top was historical, social, political, and philosophical contexts (origins, reasons, and impacts).

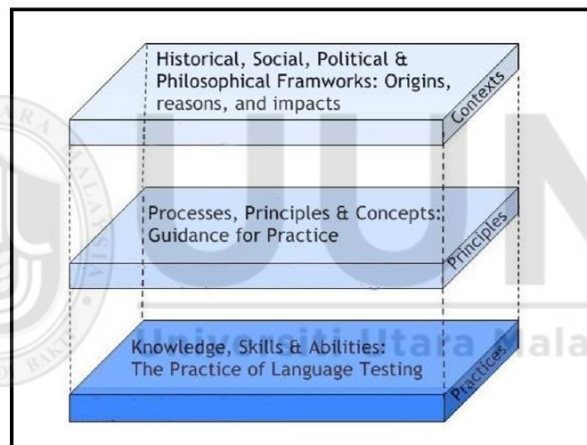


Figure 2.2 Fulcher's (2012) expanded model of LAL

Though helpful in identifying the key components in complex contexts, the three-element models had the deficiency that to what extent teachers should be assessment literate was not mentioned, just as what Taylor (2013) commented that the depth of LAL required for teachers was unclear. Thus, to address the gap, the third type of models characterized by the multi-dimensional continuum emerged.

2.4.2.3 Scaled Models

The third type was featured by multi-dimensional scaled models with varying degree of expectations for teachers, specifically Shin's (2015) seven-dimensional progressive model in AL and in LAL were Pill and Harding's (2013) continuum scale, Taylor's (2013) eight-dimensional model, Baker and Riches' (2017) seven-dimensional model, Kremmel and Harding's (2020) nine-dimensional model, and Bøhn and Tzagari's (2021) ten-dimensional model.

Rather than treating teachers as a whole group, the study contributed by Shin (2015) was characterized by distinguishing AL requirements for teachers in three stages of professional career from pre-service through the novice to experienced. The framework elaborated a progressive requirement of assessment practices from seven dimensions covering (1) developing proper assessment activities, (2) planning assessment in alignment with learning, (3) communicating the assessment purpose and standards, (4) cultivating learners' self-assessment ability, (5) giving feedback, (6) interpreting the outcomes, and (7) recognizing the unethical and improper assessment process (See Figure 2.3). Teacher candidates in the initial teacher preparation were suggested to reach the most basic level of assessment-related knowledge, comprehension, and practice, while beginning teachers (less than three years of service) were required to master a more advanced level of being skillful at application and analysis of the assessment, and experienced teachers were expected to fulfill the most complicated task to conduct effective synthesis and evaluation in assessment.

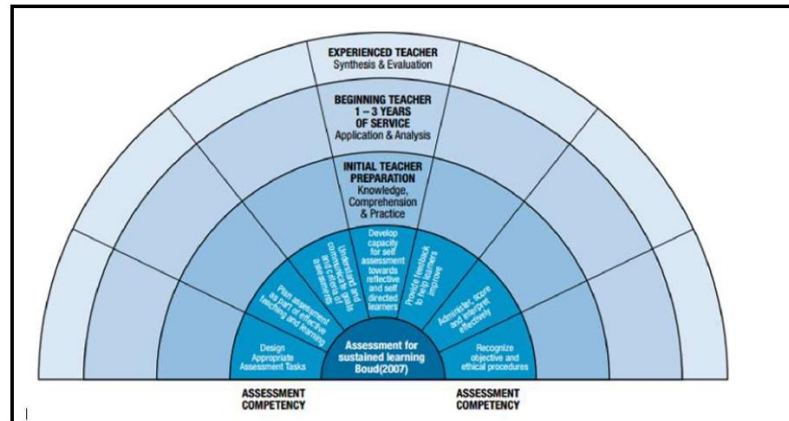


Figure 2. 3 Shin's (2013) AL model for teachers in three different stages

In the efforts to explore LAL framework, Taylor's (2013) eight-dimensional spider diagram might be the most impactful and subsequent studies still attempted to validate the classifications. Inspired by Pill and Harding's (2013) continuum of LAL in rejection of dichotomy of literate or illiterate with a preference to view it from a continuum, from "illiteracy", through "nominal literacy", "functional literacy", "procedural and conceptual literacy", to an expert level of knowledge (p. 183), Taylor (2013) hypothesized a model for teacher's LAL, involving (1) "knowledge of theory", (2) "technical skills", (3) "principles and concepts", (4) "language pedagogy", (5) "socio-cultural values", (6) "local practices", (7) "personal beliefs/attitudes", and (8) "scores and decision making"(p. 410). The teachers were suggested to be assessment literate to varying degrees among the eight aspects with the highest mastery level in language pedagogy and the lowest mastery level in three dimensions: knowledge of theory, principles & concepts, as well as scores and decision-making (See Figure 2.4)

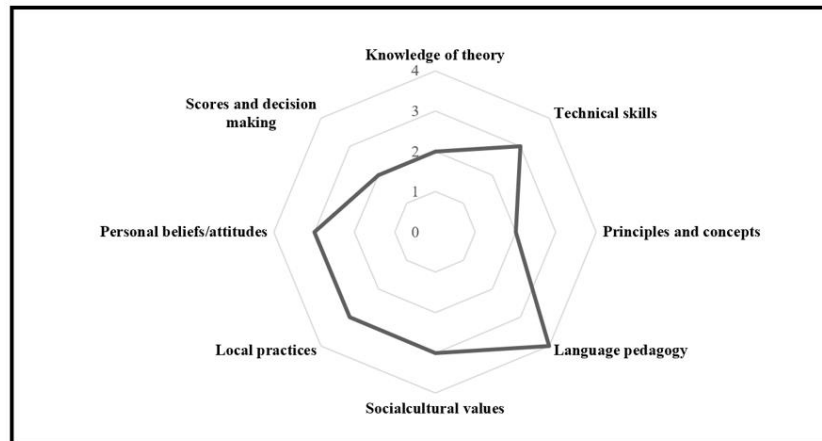


Figure 2.4 Taylor's (2013) eight-dimensional diagram of LAL

This model was recently reclassified and modified by Baker and Riches (2017), Kremmel and Harding (2020) and Bøhn and Tzagari (2021). According to Baker and Riches (2017), the dimension of socio-cultural values suggested by Taylor (2013) might best be considered as background or contextual influence implicitly subsumed within all other components instead of an independent one as depicted in Figure 2.5. They relabeled as (1) “theoretical and conceptual knowledge”, (2) “task performance”, (3) “language pedagogy”, (4) “collaboration”, (5) “awareness of local practices”, (6) “awareness of personal beliefs/attitudes”, and (7) “decision making”, with all the dimensions positioned under the impact of the social-cultural values (p. 99). Another difference from Taylor's (2013) description was that language teachers were supposed to be most capable in three dimensions in total (i.e., language pedagogy, awareness of local practices, and awareness of personal beliefs/attitudes) and to be least capable in theoretical and conceptual knowledge.

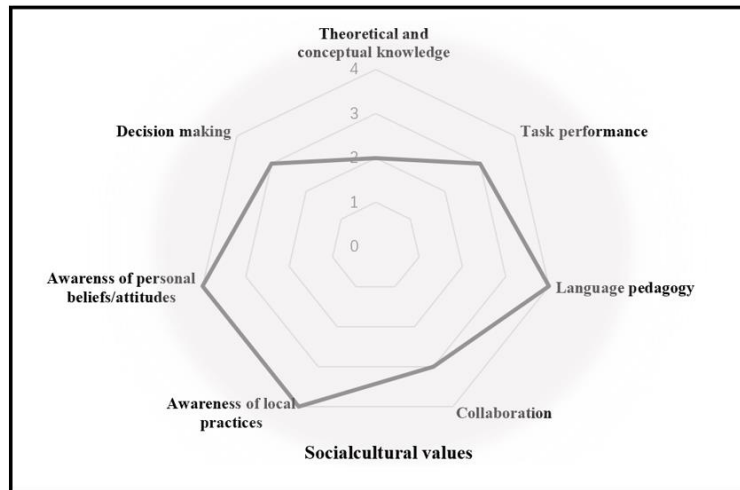


Figure 2.5 Baker and Riches' (2017) seven-dimensional diagram of LAL

Another attempt was made by Kremmel and Harding (2020) who intended to validate Taylor's (2013) LAL profile model, which was speculative in nature (Stabler-Havener, 2018). Kremmel and Harding (2020) revealed that there might exist a nine-component construct of LAL, which was generally in alignment with Taylor's (2013) diagram but adding some expansions: (1) "developing and administering language assessments", (2) "assessment in language pedagogy", (3) "assessment policy and local practices", (4) "personal beliefs and attitudes", (5) "statistical and research methods", (6) "assessment principles and interpretation", (7) "language structure, use and development", (8) "washback and preparation", and (9) "scoring and rating" (p. 111). Consistently, in their proposal language teachers were described to be least capable in the dimension of statistical knowledge (See Figure 2.6).

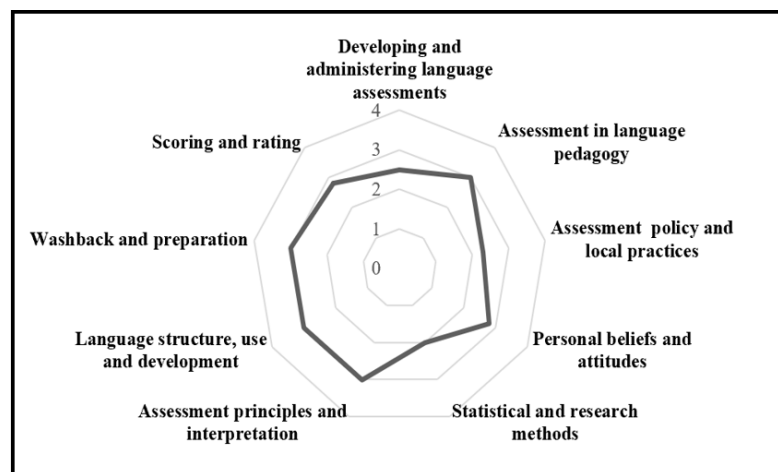


Figure 2.6 Kremmel and Harding's (2020) nine-dimensional diagram of LAL

Likewise, Bøhn and Tsagari (2021) criticized Taylor's (2013) model for not specifying the components or dimensions in any great detail. Thus, they provided an extended version of Taylor's (2013) scaled model with great details to explain the meaning of each revised dimension. As depicted in Figure 2.7, they suggested there were 10 dimensions in teachers' LAL: "knowledge of assessment theory, technical skills, principles, language pedagogy, social-cultural values, local practices, personal beliefs/attitudes, scoring, disciplinary competence, and collaboration competence" (p. 231). Teachers should be most proficient in three dimensions (i.e., principles, language pedagogy, and disciplinary competence) and least proficient in both technical skills and collaboration competence.

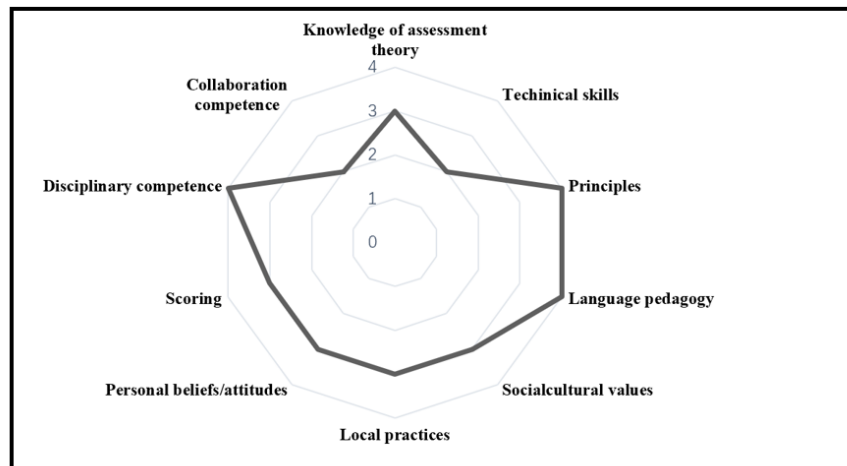


Figure 2.7 Bohn and Tsagari's (2021) 10-dimensional framework of LAL

2.4.2.4 Negotiation Models

Regrettably, however, these multi-dimensional models did not cover the dynamic and complex attribute of the concept (Vogt et al., 2020) and neither integrated AL with teacher's professional development (Y. Xu & Brown, 2016). Because of these limitations of the third type, the fourth type of AL framework characterized by negotiation between AL and contexts was proposed. A six-component negotiation pyramid was reconceptualized as teachers' AL in practice by Y. Xu and Brown (2016) based on the inter-relatedness between educational assessment and teacher education (See Figure 2.8). The six elements from bottom to top respectively were: (1). "The knowledge base", involving seven subcategories including knowledge of the discipline and Pedagogical Content Knowledge (PCK), knowledge of assessment principles and strategies, knowledge of scoring and effective feedback, knowledge of peer- and self-assessment, knowledge of interpreting and communicating assessment, and knowledge of ethical assessment; (2). "Teacher conceptions of assessment", including

cognitive and affective dimensions as well as views on learning and beliefs; (3). “Macro social-cultural & micro institutional contexts”, referring to the broad social and cultural assessment environment and local educational assessment contexts; (4). “Teacher assessment literacy in practice”, showing the constant compromises in making decisions and taking actions in assessment; (5). “Teacher learning”, denoting teachers seeking for a better understanding of assessment; (6). “Assessor identity (re)construction”, indicating the teacher identity constructed or reconstructed as assessors (p. 155).

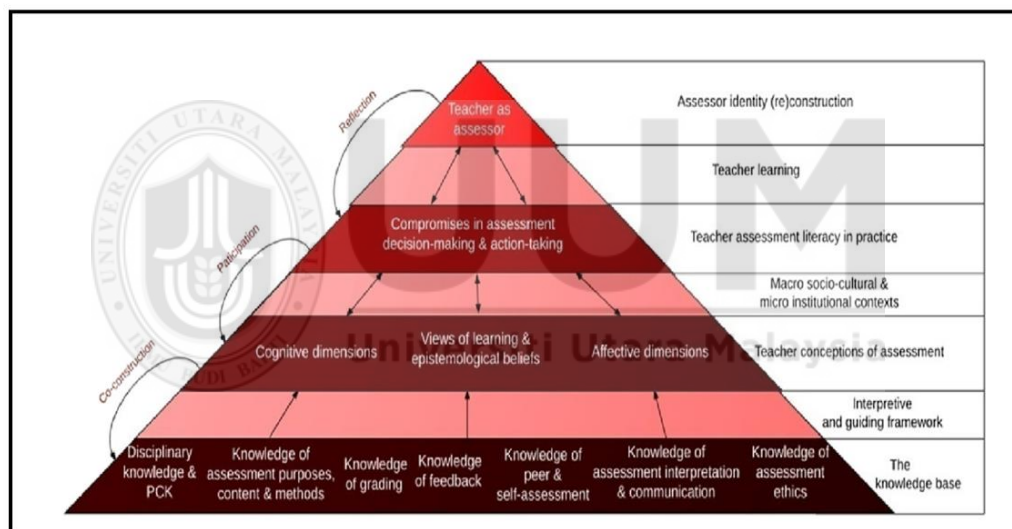


Figure 2.8 Y. Xu and Brown’s (2016) AL framework

They further clarified the functions of these constituents. Serving as a necessary yet insufficient condition, the knowledge base sitting at the bottom was interpreted and mediated by the teacher’s conception of assessment. The macro- and micro- contexts located in the third level set the boundaries where AL was viewed as compromises made among competing tensions. To be specific, the equilibrium of AL among various tensions was temporarily arrived at through the constant negotiation between the

instructor's conception of assessment and the local contexts as well as the assessment knowledge base. The compromises might be implicit and internal in assessment conception or explicitly and external in assessment actions. With teacher learning acting as the driving force, the advancement of AL might be possible to be oriented towards the final goal of assessor identity (re)construction.

The relationship among them was featured by multi-directional flows. Changes in one component were impactful on others. On the one hand, theoretical assessment knowledge and principles grew from the bottom (the knowledge base) to top (teachers as assessors); on the other hand, the assessment practice tested by the local contexts flew from top to bottom to update and to reconstruct the assessment knowledge system.

Rejecting the dichotomy of literacy or illiteracy, this model differentiated three levels of AL improvement. The first level was an elementary mastery of assessment principles. Built on the knowledge base as a foundation, the second level was an internalized body of assessment understanding and skills in alignment with teaching and learning. The first two levels set the prerequisite for the third, in which teachers were expected to develop a self-oriented awareness of the assessment procedures and identity (re)construction as an assessor. This hierarchy of AL development provided a route for teachers to improve their AL gradually and cyclically along the continuum.

In a more recent discussion by Y. Xu (2019a), one of the co-developers of the model,

elaborated what each layer of AL model meant to EFL teachers for understanding LAL. The model could be applied to EFL teachers' LAL with all the six dimensions remained the same. Just in the knowledge base dimension, she specified the knowledge for English teachers in LAL with detailed discipline-based knowledge in the three sub-categories (i.e., Disciplinary knowledge & PCK, Knowledge of assessment purpose/methods, Knowledge of feedback) while the rest of sub-categories in knowledge base remained the same. Furthermore, the model was validated to be feasible to describe the trajectory of an EFL teacher's LAL evolvement in China.

Being the key focus of Y. Xu and Brown's (2016) framework, assessor identity (re)construction also received great concern in another model formulated by E. Huang and Yang (2019), who underscored AL development in the community of practice and tracked the identity changing trajectories of two EFL teachers in learning community with the support from academic researchers. Their model echoed Y. Xu and Brown's (2016) illustration that the teacher's identity (re)construction from traditional instructors to learning-oriented assessors resulted from active recognition and meaning negotiation. Specifically, the identity shift was the outcome of the interplay among teacher professional learning, assessment cognition, assessment practices, and interaction with the researcher team.

2.5 Empirical Studies on AL and LAL among In-service Teachers

This section primarily concentrates on the empirical studies conducted about AL

among in-service teachers in general education and LAL among the in-service EFL teachers of both primary and middle schools across the world from two aspects pertinent to the research focus of the present study: teachers' AL and EFL teachers' LAL proficiency level and mediating factors identified in the AL or LAL evolvement. In general, the available studies in the literature suggest that the research focusing on issues related to AL in generic education outnumbers that of LAL in the language field. What's more, the discussions regarding the issues around AL and LAL in China are relatively under-explored both in quantity and complexity compared with the research conducted abroad.

2.5.1 AL and LAL Proficiency Level

As to whether teachers are sufficiently literate to fulfil the required assessment responsibilities is one of the primary concerns, scholars utilize a variety of data collection methods including questionnaires, classroom observation, assessment quiz, and interviews to discover teacher's strengths and weaknesses in AL or LAL based on the components articulated in the assessment documents or standards. Studies on the investigation of the proficiency level of AL are introduced first, followed by the studies in LAL.

2.5.1.1 AL Proficiency Level

It has been well documented the inadequate AL among teachers across school levels

has experienced little improvement in the past two decades and has still been prevalent nowadays (Kalajahi & Abdullah, 2016; Kim et al., 2020; Levy-Vered & Alhija, 2015; Napanoy & Peckley, 2020; Pfeiffer-Hoens, 2017; Prizovskaya, 2018; Talib et al., 2014). Most of the studies are conducted in quantitative way, either self-report questionnaires or the tests of assessment-related knowledge. A few studies focus on qualitative data through interviews with teachers or the collection of teacher's assessment-related artifacts.

The multiple data collection and analysis lead to the consistent statement that overall teachers have a poor accumulation of knowledge both in testing and assessment. They displayed a deficiency in knowledge concerning the psychometric features of tests and statistics related to assessment (Daniel & King, 1998; Montee et al., 2013), had a less satisfactory mastery of assessment principles and techniques (Alkharusi et al., 2012; Howley et al., 2013; Maclellan, 2004), and experienced difficulty in interpreting technical terms in the score report (Kim et al., 2020). They were also reported to lack the understanding of a sound assessment practice (Ch'ng & Rethinasamy, 2013) and to seldomly consider the authenticity in assessment design, which revealed their shortcomings in developing well-constructed test and rubrics (Armstrong et al., 2004; Birenbaum et al., 2015; Prizovskaya, 2018; Williams, 2015). Moreover, they appeared to be fully ignorant of the adverse influences that their unsound assessment activities might have upon students (Pfeiffer-Hoens, 2017; Vandeyar & Killen, 2007).

Additionally, teachers tended to feel less comfortable with the emerging learning-oriented paradigm than with the traditional measurement-oriented paradigm (Hargreaves, 2005; Koh et al., 2012; Maclellan, 2001; Muñoz et al., 2012; Portelli & O'Sullivan, 2016). Evidence showed teachers' over-reliance on traditional summative tests or quizzes to assess student's recalling of factual knowledge with rare coverage on higher-thinking skills, such as comparison and evaluation (Birenbaum et al., 2015; Koh & Luke, 2009; Laren & James, 2008; Stiggins et al., 1989).

They were challenged to apply or modify AfL methods into classroom contexts (Koh et al., 2012; Kuze & Shumba, 2011; Marshall & Drummond, 2006; Stewart & Houchens, 2014). Integration of assessment with classroom instruction and learning remained inconsistently (Hudson, 2017; Portelli & O'Sullivan, 2016). Although some teachers combined it superficially (assessment formats) without difficulty, they were less successful in carrying out the deep core aspects (cognition and purpose) of the assessment (Choi, 2017). Moreover, self-assessment as the main tool for AaL was least frequently used by teachers within the classroom activities (Bol et al., 1998; Volante, 2010; Young & Jackman, 2014).

With regard to the interpretation and communication of assessment results, they seemed to be less certain about how to explain the information they gathered (Bachor & Anderson, 1994), how to utilize the assessment information (Plake, 1993; Rogers et al., 2007; Wicking, 2017; Yamtim & Wongwanich, 2014), how to provide feedback for

the purpose of student well-being improvement (Brown et al., 2012; Kanjee & Mthembu, 2015), how to share and explain the assessment process with students (Andrews et al., 2018), and how to communicate assessment results with other stakeholders (Napanoy & Peckley, 2020; Plake, 1993).

In the field of China, what seems consistent with the studies abroad on teacher's AL proficiency level is that qualitative studies are fewer and quantitative surveys are more favored, either large-scale involving thousands of participants or small-scale involving dozens of samples. The research results confirm that AL proficiency level among teachers in primary and middle schools in China remains expectedly less desirable.

Zheng (2010) investigated to what extent the teachers of all subjects in primary and middle schools were assessment competent according to the seven standards issued in 1990 by sampling nearly 1,000 participants across 18 schools in a province in China through the test adapted from Plake et al. (1993). The findings revealed that teachers' overall AL was rather low with an average accuracy of less than 50%, even less than 10% on several items. They were less sure about how to use assessment results to make instructional decisions and were least certain about the issues of ethics and fairness involved in the assessment.

With the mere focus on classroom AL among primary teachers, Zhao (2020) utilized a classroom AL questionnaire devised by DeLuca et al. (2016a) to survey AL proficiency

of 1,032 primary teachers in the Mainland of China. Similarly, he argued that there still existed a certain gap between the participants' AL and the ideal level. Besides, their performance did not reach the basic requirements in assessment communication, especially in ethics, which meant that their awareness of assessment ethics was relatively weak.

Utilizing a large-scale survey from over 1,000 participants from Hong Kong and Guangzhou Province located in the southern part of China, Brown et al. (2011) showed that teachers tended to strongly hold the conception that the examination or test was a powerful way to improve students learning. Their research also revealed that teachers felt challengeable to implement AfL advocated in assessment policies and curriculum standards.

Unlike the above quantitative analysis, qualitative data were collected and analyzed by Chai (2020) through an interview with 30 novice primary teachers in Shanghai, a rather developed city in China. The research discovered the contradiction between the requirement of AL among classroom teachers and inadequate AL among them. Such a contradiction was manifested in their confusion in assessment beliefs, insufficiency in the assessment knowledge base, and unbalanced use of assessment methods. Novice teachers in this study seemed to consider the assessment as the main means for educational management rather than a means for promoting students learning. They also viewed assessment feedback for the sole purpose of strengthening desirable

students' behaviors rather than for the purpose of enhancing students' cognition and learning autonomy. In addition, they appeared to have great difficulty in adjusting the assessment plan dynamically in the process of instruction on the basis of the constant interaction with students in the classroom.

2.5.1.2 LAL Proficiency Level

Similar to the deficient AL proficiency among teachers in other subjects, in the specific field of language assessment, it has been universally acknowledged that EFL/ESL teachers ranging from the beginning (Mellati & Khademi, 2018) to much more experienced (Babaii & Asadnia, 2019) lack the satisfactory knowledge and skills to be effectively engaged in assessment activities (Büyükkarçı, 2016; Mede & Atay, 2017; Vogt & Tsagari, 2014; Watmani et al., 2020).

There might exist a gap between assessment activities advocated in assessment policy and assessment practices implemented by EFL teachers in the classroom (Newfields, 2007; Shahzamani & Tahririan, 2021). EFL teachers tended to have inadequate knowledge to develop sound language tests, to make effective evaluations of their own tests, and to utilize the results of tests for diverse purposes (Davidson & Coombe, 2019; Farhady & Tavassoli, 2018; Mede & Atay, 2017). Even the experienced EFL instructors, their understandings of language assessment were restricted to the rudimentary theories (Babaii & Asadnia, 2019). Also, the assessment principles, such as ethics, legitimacy, and appropriateness were usually ignored by language teachers

in the process of assessing students (Aria et al., 2021; Luthfiyyah et al., 2020). With regard to the assessment purpose and methods, they were reported to have an unclear vision of the purpose of the assessment practices initiated by them in the class and to apply a limited range of assessment techniques in assessing students learning (Shahzamani & Tahririan, 2021). They also seemed to have fuzzy ideas of implementing innovative reforms to assess students' language proficiency, although learning-oriented assessment had been highlighted in various documents and policies, which encouraged instructors to actively use alternative assessment techniques (Firoozi et al., 2019; Shahzamani & Tahririan, 2021; Vogt & Tsagari, 2014).

One of such assessment illiterate behaviors was teaching to test. Assessment illiterate teachers tailored language teaching to the examination content covered in the external tests and the tests developed by EFL teachers were only confined to a very limited range of language skills measured in the external tests whereas assessment literate EFL teachers preferred non-washback assessment activities in their classrooms (Díaz et al., 2012; Kiomrs et al., 2011; Ragchaa, 2019; Shahzamani & Tahririan, 2021; Tsagari, 2016). The significant discrepancy between the expected LAL proficiency level and real LAL proficiency level of teachers was also perceived by teachers themselves who were involved in daily routine classroom assessment activities. Generally, they felt not confident in developing and designing assessment activities (Berry et al., 2019; Tsagari & Vogt, 2017).

Numerous studies pointed out EFL/ESL teachers primarily used assessment for the summative purpose not formative purposes (Büyükkarcı, 2014; Giraldo, 2019; Gonzales & Aliponga, 2012; Hidri, 2016; Luthfiyyah et al., 2020; Portelli & O’Sullivan, 2016) and they preferred traditional assessment methods to alternative assessment methods (Kiliçkaya, 2016; Kirkgoz et al., 2017; Shahzamani & Tahririan, 2021; Tsagari, 2016; Tsagari & Vogt, 2017). Their traditional assessment activities were primarily focused on learning outcomes rather than the learning process (Djoub, 2017; Sultana, 2019). Most of the assessment feedback they provided was in the form of grades usually issued at the end of the assessment activities for measuring students learning rather than functioned as a means of monitoring teaching and learning (Berry et al., 2019; López & Bernal, 2009). EFL teachers seemed to pay less attention to AfL, such as sharing learning goals and involving students in the whole process of assessment (Büyükkarcı, 2014; Seden & Svaricek, 2018). It seemed as if the AoL prototype was “deeply ingrained in their assessment DNA” (Levi & Inbar-Lourie, 2020, p. 177).

In China, the research on LAL proficiency level among primary and middle school in-service EFL teachers has shown similar results with studies in other countries. EFL teachers were reported to have low LAL to deal with the strong washback of the external high-stake tests and lack of skills to transfer the knowledge to assessment practices in the Mainland of China (Lan & Fan, 2019; Lin & Su, 2015) while teachers in Taipei (Chan, 2008) and Hong Kong (Lam, 2019) performed much better in

assessment.

Collecting and analyzing the questionnaires of a sample of over 300 EFL teachers in middle schools in China, Lan and Fan (2019) found that participants possessed deficient LAL to carry out the classroom assessment activities and knew little about how to apply the assessment terminologies (e.g., validity and reliability) into the classroom-based language assessment activities. Additionally, teachers' teaching and assessment process were strongly influenced by the high-stake test mandated at the end of the last year in middle school. More specifically, teaching to the test played a dominant role in their classroom teaching and assessment activities, which were exam-oriented both in the assessment formats and contents to some extent.

The study conducted by Lin and Su (2015) utilized a test, not the self-report questionnaire, to investigate the status quo of a small sample of 39 middle school EFL teachers. They confirmed the low proficiency level of LAL, which was manifested in participants' poor understanding of the key statistic terms and rare consideration of authenticity and fairness in assessment development. The importance of students' active agent role in assessment did not receive due attention from EFL teachers.

Unlike previous studies with a focus on quantitative data, X. Yan and Fan (2020) relied on interviews, the qualitative data only, to describe LAL profiles across several stakeholder groups. With regard to the EFL teachers, the results showed that the

participants heavily concentrated on issues closely related to traditional language tests, for instance, how to develop language test items, how to use the test results, and how these items influenced the teaching and learning process. They seemed to attach little importance to the alternative assessment techniques.

In Taipei, over 500 EFL teachers in elementary schools were investigated through a questionnaire to reveal their beliefs and practices towards multiple assessments (Chan, 2008). The findings were encouraging that nearly all of them held a rather positive attitude to the multiple assessments. The overwhelming participants indicated that they had a good knowledge of alternative assessment methods, which were used more frequently than traditional tests in their actual assessment practices. They seemed to embrace the belief that assessment was a key way to evaluate the effectiveness of instruction and that students were also main assessors in the assessment activities.

In Hong Kong, Lam (2019) examined to what extent the 66 English teachers in secondary school were assessment literate by adopting the framework of Y. Xu and Brown (2016) to collect multiple data from a questionnaire, interviews, and classroom observation. Although there was still much to be improved to achieve a comprehensive mastery of AL, the participants were found to possess a much better understanding of AfL and AoL than AaL. In their AaL practices, they just imitated the steps rather than completely internalizing the nature. Contrary to the long weakness of assessment ethics and fairness among teachers in Mainland of China, teachers in Hong Kong were

found to have considerable knowledge about these assessment principles. However, they were weak in reflecting on the use of assessment results to improve teaching and learning.

In a word, based on the larger amount of quantitative research, it is notably evident that the teachers including EFL teachers, across the world, do have an undesirable proficiency level of AL or LAL. However, what has been presented here is not to blame teachers for their insufficient AL or LAL, their conceptions towards assessment are “ecologically rational” within their own local surroundings (Brown, 2011, p. 70). Furthermore, their varying proficiency levels are plausible to be interpreted as the different stages during the AL or LAL developmental process and how to facilitate teachers through the continuum of AL or LAL needs more attention and further exploration (Choi, 2017). Thus, the following part is aimed to sort out factors mediating AL and LAL evolvement in order to understand how to effectively improve AL among teachers and LAL among EFL teachers.

2.5.2 Factors Mediating AL and LAL Evolvement

Crusan et al. (2016) classified factors mediating AL evolvement into two general types: contextual and experiential factors. Contextual factors were described as the broader political, historical, cultural, and educational factors which jointly shaped the assessment culture in a certain context while experiential factors referred to the teachers’ educational background, on-the-job-training experience, teaching experience,

and their inner world (e.g., awareness or conception towards assessment).

2.5.2.1 Contextual Factors

The contextual factors identified in the literature cover a wide range from macro social culture and policies through meso school-based environment to micro classroom context. The details in this theme are illustrated in AL from general education first, and then followed by LAL in the language assessment field.

2.5.2.1.1 In the Field of AL

In AL-related literature, three levels of contextual factors have been identified: macro, meso, and micro (See Table 2.2). The sources are listed in a chronological order.

Table 2.2

The Contextual Factors Mediating AL Identified in Previous Studies

Three levels	Identified factors	Sources
Macro-level	Educational policies	Azis (2012), Abrams et al. (2016), and Koh & DePass (2019)
	Historical culture	Black & Wiliam (2005), Brown (2011), and Gebril & Brown (2014)
	Social and cultural norms	Seden & Svaricek (2018)
Meso-level	School type	Quilter (2000) and Taber et al. (2011)
	School effectiveness	Prizovskaya (2018)
	School culture	Allal (2013), Djoub (2017), Inbar-Lourie & Levi (2020), and Luthfiyyah et al. (2020)
	School policy	Chew & Lee (2013), Tsagari & Vogt (2017), Lam (2019), and Zulaiha et al. (2020)
	Grade level	Cheng & Sun (2015) and Zhang & Burry-Stock (2003)
Micro-level	Workload	Alkharusi et al. (2012)
	Class size	Cheng et al. (2008) and Cheng & Sun (2015)

From the macro perspective, educational policies and the historical cultural background of the country framed teachers' conceptions and practices of assessment (Black, 1993; Koh & DePass, 2019). The distinctive culture in each site might be a great contributor to the discrepant assessment conceptions among teachers from different countries (Black & Wiliam, 2005). A series of evidence came from the assessment conception survey among teachers in various nations. In the western results (e.g., Australia, New Zealand), teachers tended to endorse assessment for improvement strongly (Brown, 2011; Brown et al., 2011) whereas in Iran assessment for accountability was endorsed more strongly than improvement (Pishghadam & Shayesteh, 2012). In low-stakes assessment contexts (e.g., New Zealand), low inter-correlation between assessment for accountability and assessment for improvement was reported (Brown, 2011; Brown et al., 2011).

By contrast in a high-stakes assessment environment (e.g., China), a high correlation was proven to exist between assessment for accountability and improvement purposes (Brown et al., 2009). Besides, conceptions towards assessment in high-stakes societies shared great similarities (Gebril & Brown, 2014).

Besides, the social and cultural norms also shaped teachers' assessment activities (Seden & Svaricek, 2018). Against such backdrop, teachers' assessment practices were manipulated, including selecting assessment methods, adopting assessment tools,

interpreting the results, and deciding the usage of assessment outcomes (Black & Wiliam, 2005). The national assessment policies and curriculum standards were powerful levers on teachers' use of the assessment data (Abrams et al., 2016). This may be due to the fact that assessment-related policies oriented teachers to a different focus in instruction and assessment (Azis, 2012).

At the meso level, schools, the main workplace of teachers, exerts a significant mediating effect on teachers' AL. It is quite evident that a great difference exists in assessment activities across schools (Taber et al., 2011). The types of school, school culture, school policy, and grade level are evidenced to be the major factors.

In terms of types of school, the assessment-related activities initiated by teachers varied largely between primary and middle schools, however, as to which group was more proficient in AL than the other, there were inconsistent findings (Abrams et al., 2016; Suah & Ong, 2012). The study from Quilter (2000) showed that teachers in middle schools overall had a better mastery of knowledge about educational assessment than teachers in elementary schools. On the contrary, Prizovskaya (2018) revealed that it was elementary school teachers who performed better than teachers from middle schools. Additionally, the effectiveness of school was also another mediating factor. Teachers working in high achieving schools were more likely to outperform those from low achieving schools (Prizovskaya, 2018).

The culture, embedded in the school, affects all school-based activities including assessment. The way that a school chooses to approach assessment, whether a test-driven one for assessing students' performance or a learning-oriented assessment for improving students' academic achievement, constitutes assessment culture in a school (Inbar-Lourie & Levi, 2020). The assessment culture with different orientations and foci across schools might influence teachers' assessment conceptions and practices (Abrams et al., 2016; Allal, 2013). In general, change-oriented schools were more willing to redirect their assessment culture in assessment understandings and activities than centralized schools which seemed less innovative in pedagogy and more likely to be driven by tests (Inbar-Lourie & Levi, 2020). In cooperation-directed settings, novice teachers could gain a substantial amount of assistance from senior experienced teachers during the application of assessment knowledge into practice (Luthfiyyah et al., 2020). It seemed not conducive to improve teachers' AL in schools where the institution played a dominant role in teachers' entire assessment process (Djoub, 2017).

Apart from school culture, the school policy also shaped teachers' classroom-based assessment practices (Tsagari & Vogt, 2017; Zulaiha et al., 2020). The support from colleagues and school leaders, including access to training and resources in assessment, tended to positively impact instructors' assessment activities (Chew & Lee, 2013). Otherwise, the institutional constraints might hinder teachers' initial attempts to try out non-traditional assessment methods (Lam, 2019).

In the school context, grade level had long been identified as a crucial variable in teachers' understanding of classroom-based assessment especially in making decisions about scoring and choosing appropriate assessment tools for specific purposes (Cheng & Sun, 2015; Stiggins & Bridgeford, 1985). Evidence showed that with the increasing of grade level, teachers in higher grade tended to be more dependent on objective tests and teacher-made tests rather than the tests adopted from published textbooks or other printed materials in the classroom assessment (Stiggins & Bridgeford, 1985; Zhang & Burry-Stock, 2003).

Moving from meso to micro level at classroom, assessment practices might differ considerably across classrooms even in the same school (Taber et al., 2011). Teachers' teaching workload and class size were possible contributing factors to such a variation. Specifically speaking, the study suggested that weekly teaching load had a negative correlation with the use of classroom tests by teachers for diagnosing learners' weaknesses (Alkharusi et al., 2012). The class size affected various aspects of teachers' decisions in grading activities. Teachers who worked in a larger size classroom were more possible to utilize paper-and-pencil tests as a more convenient way to assess learners while those who taught the smaller class were more willing to create assessments themselves (Cheng & Sun, 2015). The relatively larger classes to some extent directed teachers to favor less labor-intensive assessment methods, which were easily implemented and scored (Cheng et al., 2008).

2.5.2.1.2 In the Field of LAL

In the same vein, research approaching contextual factors in LAL roughly follows the same categories (See Table 2.3). However, due to the relatively limited number of studies, the categories identified is not as rich as AL.

Table 2.3

The Contextual Factors Mediating LAL in Previous Studies

Three levels	Factors identified	Source
Macro-level	National policy	Jia et al. (2006), Portelli & O’Sullivan (2016), and Sultana (2019)
Meso-level	School type	Farhady & Tavassoli (2018) and Tavassoli & Farhady (2018)
	School culture	Jia et al. (2006) and Seden & Svaricek (2018)
Micro-level	Class size & workload	Han & Kaya (2014)
	Other stakeholders	Jia et al. (2006) and Rogers et al. (2007)

At the national policy level, the high-stake test mandated by the country usually overwhelming directed classroom teachers to be test-driven in classroom teaching and assessing, especially on what and how teachers assessed students, which partially led to scarce awareness of the necessity to use non-traditional assessment to evaluate students’ achievement (Jia et al., 2006; Sultana, 2019). This concern was confirmed by Portelli and O’Sullivan (2016) that the unbalanced focus of assessment practices in classroom was likely to be constrained by the external policy and testing agenda at the national level, consequently narrowing LAL among language teachers.

At the meso level, EFL teachers from public schools were reported to be more

knowledgeable in language assessment than their peers working in private schools where teachers were seldomly allowed to develop assessments by themselves and where teachers' professional knowledge in assessment was not valued highly compared with their English proficiency (Farhady & Tavassoli, 2018; Tavassoli & Farhady, 2018). Besides, schools advocating collaboration among teachers benefited teachers considerably in their LAL improvement because consultation with colleagues functioned as a primary source of teachers' subjective theory related to assessment (Jia, et al., 2006; Seden & Svaricek, 2018).

At the micro level, Han and Kaya (2014) indicated that class size and teaching hours had a significant impact upon teachers' preferences towards language assessment. Teachers with more than 25 students in the classroom were more likely to utilize assessment for instruction and informing than teachers in a smaller size of the class. Teachers who taught over 25 hours in a week favored assessment to inform than those who had fewer teaching hours. In general, teachers who worked longer hours with a larger size of the class were more preferred to the communicative function of assessment.

Besides, other stakeholders (e.g., parents and students) also functioned as a mediating factor in this process. With a focus on EFL teachers' LAL in reading assessment, Jia et al. (2006) indicated that parents indirectly influenced EFL teachers' assessment process in school and that students' widely varying reading ability in a class

complicated the assessment tasks because assessment practices appropriate for high-achievers might not work for slow learners. Rogers et al. (2007) confirmed that students' English proficiency not only influenced assessment procedures their teachers chose, but also the assessment conceptions embraced by teachers.

In summary, though they were identified in different categories, these mediating factors could not be separated from each other, nor did they influence independently on teachers' LAL in a vacuum. The contextual factors were inter-related and interactive with LAL, just as what Edwards (2017) stated that although some elements of LAL were generalizable to all contexts, some components were context-dependent. Teachers constantly made a variety of "compromises" between LAL and the context in the negotiation among various tensions (Y. Xu & Brown, 2016, p. 157). Language teachers' LAL was shaped and even constrained by the socio-cultural contexts where they were (Inbar-Lourie, 2012, 2017; Stabler-Havener, 2018). However, the contextual factors did not directly function on individuals, they were mediated through experiential factors (X. Yan et al., 2018).

2.5.2.2 Experiential Factors

The experiential factors sorted out from the available literature is also ordered from AL first and then followed by LAL. Similar with the contextual factors, the identified experiential factors in AL are outnumbered in LAL.

2.5.2.2.1 In the field of AL

A diversity of experiential factors has been identified in the literature to mediate the evolvement of teachers' AL to a varying degree, for instance, educational level, professional learning experience, being assessed experience, career experience, and conceptions towards assessment.

In terms of educational level, teachers with a Master degree were indicated to be more assessment literate than Bachelor degree holders for potential extended exposure to academic learning of assessment-related components. Master degree holders were more likely to outperform their Bachelor counterparts on every measured dimension and they were also reported to be more knowledgeable in authenticity in assessment (R. Huang & Jiang, 2020; Prizovskaya, 2018). The reason may be that a Master's degree enabled them to be exposed to more advanced educational concepts compared with those who had not experienced postgraduate studies (Napanoy & Peckley, 2020).

Professional learning is widely acknowledged as a crucial element in AL development. Whether teachers got access to any forms of training on assessment seemed to be a notable variable in their varied assessment practices (Cheng & Sun, 2015; Coombs et al., 2018). Generally, teachers who attended assessment training courses, programmes, or workshops appeared to be more proficient than those who did not (Alkharusi et al., 2012; Mertler, 2009; Plake & Impara, 1993). Especially the sustainable, on-going, and

high-quality trainings were more profitable than the ad-hoc, short-term, and ineffective programmes (Alkharusi et al., 2011; Conca et al., 2004; Daniel & King, 1998; DeLuca et al., 2015; Koh et al., 2012; Pedder & James, 2012).

Well-trained teachers demonstrated a better AL than those with limited training in assessment (Levy-Vered & Alhija, 2015). A huge improvement was observed in various aspects in participating teachers' assessment practices, including understanding assessment (Coombe et al., 2020; Koh, 2011), designing sound assessment tools and tasks (Koh, 2011; Koh & Velayutham, 2009; Stiggins et al., 1989), using diverse assessment strategies (Cheng et al., 2008; Stewart & Houchens, 2014), interpreting and communicating assessment results (Mahapatra, 2016; Zhang & Burry-Stock, 2003), and involving students in the whole assessment process (Chappuis & Stiggins, 2002).

Moreover, teachers' self-efficacy in assessment played a critical role in fostering AL. Self-efficacy in assessment and assessment knowledge were moderately or even highly correlated, which indicated that the higher self-efficacy towards assessment, the more knowledgeable a teacher was in assessment, or vice versa (Gotch & McLean, 2019; Huai et al., 2006). Another important facilitative factor in AL evolvement was the reflection on their own assessment conceptions or practices, either independently or collaboratively. Teachers who were self-reflective in assessment tended to be more assessment literate than those who seldom or less frequently reflected on their

assessment practices (Bijsterbosch et al., 2019; Howley et al., 2013).

The prior assessment experience of teachers also played a decisive role in their conceptions towards assessment (Azis, 2012; Wyatt-Smith et al., 2010). Their personal experience of being assessed as students in schools largely determined their conceptions of assessment and influenced their preferences of assessment procedures when conducting assessment activities (Jane, 2012; Rogers et al., 2007). Teachers tended to test as they had been tested in their schooling time (Firoozi et al., 2019). Teachers with a comparatively positive experience in assessment during their schooling days were more likely to hold a positive attitude towards assessment whereas negative experiences with assessment might lead to a negative perception about assessment (Quilter, 2000). Thus, the teacher's current attitudes to assessment possibly resulted from a combination of affective and cognitive factors, with a higher proportion of affective variables (Laren & James, 2008; Quilter & Gallini, 2000).

The studies on whether years of teaching experience accounted for the variation in teacher assessment performance yielded competing results. Some studies argued it was a variable in that more experienced teachers differed significantly than less experienced teachers in assessment practices (e.g., Alkharusi et al., 2012; Seden & Svaricek, 2018; Suah & Ong, 2012). The longer years in career enabled teachers to possess greater skills in various aspects of assessment activities, from designing to implementing and using the results to provide feedback (DeLuca et al., 2018). But

there was also evidence favoring less experienced teachers for their more frequent use of alternative assessment methods than experienced teachers with more than 30 years of teaching (Mertler, 1998). Other studies revealed no relationship between years of teaching and assessment conceptions and their assessment performance (Calveric, 2010; Jawhar & Subahi, 2020; Napanoy & Peckley, 2020; Prizovskaya, 2018).

Teachers' assessment-related conceptions might affect their approaches to assessment and choice of assessment strategies (Levy-Vered & Alhija, 2015). For instance, their conceptions of teaching and learning (Djoub, 2017; Muñoz et al., 2012), the conceptions of learners (Thomas, 2012), the conceptions of their identity of assessors (Looney et al., 2018), and more importantly, the conceptions of assessment (Abrams et al., 2016; Brown et al., 2011; Calveric, 2010; Harris & Brown, 2009). Changes in teachers' assessment practices needed a change in the way they conceived new forms of assessment in advance (Koh, 2011). Firoozi et al. (2019) reconfirmed the crucial influence of conceptions by arguing that it would be unrealistic to change teachers' assessment practices without any changes in how teachers conceived assessment.

Another factor that needed to mention here was gender. Inconsistent findings were reached whether female teachers perform differently from male teachers. According to Jawhar and Subahi (2020), gender was not a significant variable to account for the divergence among the participants whereas Alkharusi et al. (2012) found that female teachers tended to possess more assessment knowledge, perceive themselves to be

more competent, and use non-achievement elements in grading more frequently than males.

2.5.2.2.2 In the Field of LAL

In the field of experiential factors mediating EFL teachers' LAL, similar categories and research findings are identified. Research findings in this theme are to be presented following the same sequence, education level, professional learning experience, being assessed experience, career experience, and finally conceptions of assessment.

There was contradictory evidence on whether education level impacted LAL among EFL teachers. Tavassoli and Farhady (2018) supported that educational level was a key factor to shape EFL teachers' perceptions and priorities of the language assessment knowledge needs. Master teachers were reported to perceive themselves as more knowledgeable in language assessment and their priorities of needs for improvement in language assessment varied from Bachelor teachers' priorities. However, Büyükkarcı (2016) indicated the increase in education level from Bachelor, through Master to PhD did not add on their LAL, which was not influenced by the level of education they received.

Professional learning remained a facilitator in fostering LAL. Continuous professional learning enabled teachers to be knowledgeable and skilled in language assessment (Davidson & Coombe, 2019; Hakim, 2015; Sultana, 2019). Saputra et al. (2020)

provided more recent evidence on the effectiveness of the professional learning community on improving EFL teachers' LAL. The findings confirmed that the professional learning community enhanced teachers' understanding of LAL regarding the shift from AoL to AfL and AaL. More than changes in assessment conceptions, participants also mastered the skills of how to apply these understandings into their assessment practices to motivate students to perform better. Moreover, another study by Baker and Riches (2017) reported that besides the attainment in assessment-related knowledge and skills, teachers achieved more development in issues around morality and ethics in assessment, which suggested their direction towards student-centeredness. Also, reflection, as a critical constituent of teachers' professional learning, assisted teachers in promoting language assessment practices (Ashraf & Zolfaghari, 2018; Babaii & Asadnia, 2019; X. Yan & Fan, 2020). However, ineffective assessment training might be the same as no training. Or even in some cases, ineffective training programmes had potentially dire consequences on teachers' assessment performance (X. Yan et al., 2018).

In terms of learning experiences, teachers who had learned how to teach EFL were more assessment competent than those without such a background (Watmani et al., 2020). EFL teachers' language assessment practices were deeply rooted in what they experienced in classroom in previous time (Berry et al., 2017; Newfields, 2007). Their negative views to assessment often derived from past unpleasant personal assessment experiences (López & Bernal, 2009). Therefore, past schooling experience was a key

element that needed to be taken into account to explore teachers' LAL deeply and comprehensively.

With regard to career experience, there were generally two voices on whether years of teaching experience made a difference on teachers' LAL proficiency level. One side upheld experienced teacher demonstrated a better understanding of LAL, manifested more awareness of the assessment criteria, and applied language assessment methods in a more productive way than novice teachers (Hakim, 2015; Tajeddin et al., 2018). The other voice, however, argued that teachers' LAL did not increase with the accumulation of years of teaching (Büyükkarcı, 2016; Lin & Su, 2015). Teachers with much longer years of teaching did not differ with less experienced teachers in perception and knowledge of assessment (Jannati, 2015; Ölmezer-Öztürk & Aydın, 2019; Öz & Atay, 2017; Pishghadam & Shayesteh, 2012).

Similar to findings in AL, the conceptions of language assessment among EFL teachers directed their assessment practices and purposes to a large extent (Han & Kaya, 2014). Chan (2008) showed a significantly positive relationship between assessment conception and practices in multiple assessments. He illustrated that the frequency of using multiple assessments in classroom activities depended on to what extent teachers believed in it. The stronger teachers believed in multiple assessments, the more frequent use of multiple assessment in their teaching activities.

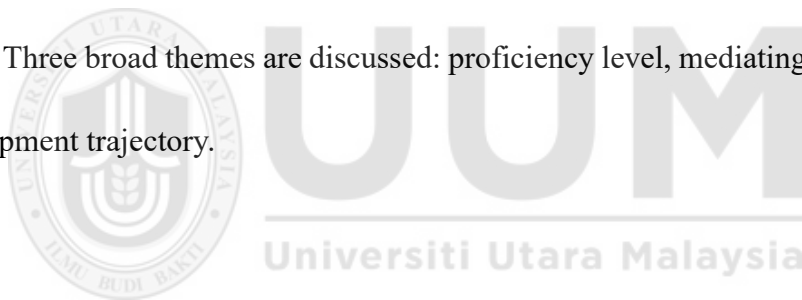
Another factor was gender. Inconsistent findings were reached on whether it was a factor contributing to the different assessment conceptions of teachers. Gender was reported to be a great variable in different assessment conceptions among EFL teachers in the research by Yetkin and Özer (2020). Male EFL teachers in their study more preferred to consider assessment for the purpose of school accountability than female EFL teachers. However, no relationship was detected between gender and the conceptions of assessment among participating teachers in the research by Pishghadam and Shayesteh (2012).

In short, there is a necessity to reconceptualize AL as a dynamic development process, which is mediated by a network of diverse correlated factors, including (1) the contextual factors: at the macro level (e.g., social-cultural environment, national policies or standards), at the meso level (e.g., school contexts, school culture, school system), and at the micro level (e.g., classroom contexts, workload, other stakeholders); (2) the experiential factors: educational background (e.g., educational level, being assessment experience), on-the-job experience (e.g., professional learning experience, career experience), and conceptions of assessment; and (3) the personal attributes (e.g., gender) (DeLuca et al., 2016a; Edwards, 2017; Luthfiyyah et al., 2020; Y. Xu & Brown, 2016; X. Yan et al., 2018; Zolfaghari & Ahmadi, 2016). According to Inbar-Lourie and Donitsa-Schmidt (2009), teachers' assessment activity is not just the reflection of an assessment method, but also the reflection of "a social and educational paradigm encompassing micro constraints (technological), macro influences (political),

ideologies and commonly-held beliefs (cultural) as well as evidence of critical pedagogy” (p. 185). In other words, the improvement of EFL teachers’ LAL needs to obtain support from all stakeholders and to take educational contexts at varying levels into account (Coombe et al., 2020).

2.6 Empirical Studies on AL and LAL among Pre-service Teachers

Compared with the large proportion of research on in-service teachers, the studies focused on AL or LAL among pre-service teachers are not only few in number but also limited in scope (Giraldo & Murcia, 2019; Sevimeel-Sahin & Subasi, 2019; Ukrayinska, 2018). Three broad themes are discussed: proficiency level, mediating factors, and the development trajectory.



2.6.1 AL and LAL Proficiency level

Primarily relying on self-report questionnaires as the tool to collect pre-service teachers’ performance on AL or LAL, numerous studies have all along warned the less satisfactory level among them across nations (Smith et al., 2014; Stobaugh et al., 2010; Volante & Fazio, 2007; B. Yan et al., 2017). Overall, the teacher candidates were found to demonstrate a low or intermediate level of AL (Lian & Yew, 2016; Volante & Fazio, 2007). This was also the case in China where pre-service teachers only answered less than half of the items correctly (B. Yan et al., 2017).

A closer look into the research findings revealed that an overwhelming majority of prospective teachers held a strongly negative attitude towards educational measurement and assessment (Childs & Lawson, 2003; Deneen & Brown, 2011; Gok et al., 2012). They tended to view assessment and teaching as discrete concepts, which they felt hard to directly link together (Campbell & Evans, 2000). Most pre-service teachers usually associated assessment with traditional summative purposes (AoL), although only a few manifested AfL notion (Smith et al., 2014; Volante & Fazio, 2007).

Pre-service teachers seemed to have great difficulty in transferring what they had been taught in assessment into practices, for they presented no sign of using assessment soundly (Campbell & Evans, 2000; Odo, 2016). Further, although they could articulate the theories related to AaL, some of the pre-service teachers were unable to associate basic practices with AaL, which suggested they would be less possible to conduct AaL after entering the career (DeLuca & Lam, 2014). The insufficient application indicated their inability to completely internalize what they had learned (Ogan-Bekiroglu & Suzuk, 2014).

A few studies provided encouraging discoveries. Prospective teachers began to enhance their AaL skills since they increasingly realized AL not only enabled to assess students' achievement productively, but also an indispensable constituent of the teaching profession (Bozkurt, 2020; Clark, 2015; M. Hill et al., 2017). This awareness of viewing AL as an integral part of identity construction towards assessors was also

upheld by AL framework developed by Y. Xu and Brown (2016).

Different from the research aiming to evaluate AL proficiency level by academic professionals, other studies were interested in self-evaluated proficiency level by pre-service teachers themselves. The findings were not consistent. They seemed to self-evaluate themselves as inadequately prepared for the expected assessment responsibilities (Kavaklı & Arslan, 2019; Volante & Fazio, 2007). Pre-service teachers were indicated to be less confident to deal with the assessment issues in the later teaching career (Volante & Fazio, 2007). However, others drew an opposite conclusion that participants generally overestimated their competence in assessment (Grainger & Adie, 2014; Kruse et al., 2020; Ogan-Bekiroglu, 2009; Sahinkarakas, 2012). This overestimation of AL could be the reflection of their unrealistic optimism (Weinstein, 1988) or meta-ignorance (i.e., being ignorant of their ignorance) (Dunning, 2011), which were also prevalent in assessment-related issues (Kruse et al., 2020). Yet, others argued that such an optimistic evaluation was not unrealistic but realistic towards assessment responsibilities they would be engaged in their future teaching career (DeLuca et al., 2013). DeLuca et al. (2013) revealed although pre-service teachers were optimistic towards assessment, they also acknowledged the realistic challenges or difficulties in applying assessment methods to inform teaching and learning. Therefore, the optimism was realistic not unrealistic towards assessment.

In the field of LAL among pre-service EFL teachers, deficient proficiency was also

extensively reported in the literature. Research showed that pre-service EFL teachers held a narrowed understanding of assessment and treated assessment equivalent to traditional tests, which were usually in the form of multiple choice and short answers to check students' mechanic memorization of facts and details (Graham, 2005). They also gained a limited scope of knowledge in both general testing principles and English language testing practice (Hatipoğlu, 2015). Though most of them grew to recognize alternative assessment as a valuable means to facilitate teaching and learning as advocated in assessment policies, pre-service EFL teachers were indicated to use portfolio assessment least frequently in language assessment practices (Graham, 2005; Kavaklı & Arslan, 2019).

In a word, although there are some desirable improvements in several aspects of AL or LAL observed in recent studies, AL and LAL proficiency is reported to be insufficient among pre-service teachers. Besides, as to the AL proficiency level self-evaluated by pre-service teachers, there are contradictory results. Therefore, one of the objectives of the current study is to provide evidence in LAL field to this issue from the context of China.

2.6.2 Factors Mediating AL and LAL Evolvment

Given the insufficient LAL proficiency among pre-service teachers, studies intend to explore the facilitating or hindering factors impacting on AL. Thus an overwhelming body of literature deals with the effectiveness of assessment-related courses and

practices in pre-service teacher education programmes.

There seemed a consensus among scholars that the assessment course was of paramount significance in initial teacher education to equip pre-service teachers with threshold assessment competence (Deneen & Brown, 2016; Huai et al., 2006; Kruse et al., 2020; Kyttälä et al., 2021; Lian & Yew, 2016; McGee & Colby, 2014; O'Sullivan & Johnson, 1993; Wise & Lukin, 1993). The assessment course helped to build a foundation for teacher candidates to better understand how to assess learners' achievement and to be more familiar with practical skills for implementing assessment activities in future teaching tasks (Deneen & Brown, 2011; McGee & Colby, 2014). Besides, DeLuca et al. (2013) added pre-service teachers were more confident in theories of assessment after the training in the course. What is more important, the purpose of this course was also to foster their long-lasting interest in further exploring issues associated with assessment in the later professional careers (DeLuca et al., 2010). Thus, explicit explanation of assessment might substantially assist them in AL improvement (Kruse et al., 2020; B. Yan et al., 2017).

However, the assessment course was not paid due attention to, nor was it delivered as effectively as expected (Hussain, 2017). Many pre-service programmes failed to offer adequate treatment of assessment (Gok et al., 2012) and remained less responsive to the latest trend of AfL to renew the teaching content (Volante & Fazio, 2007). In the survey by Volante and Fazio (2007), prospective teachers were reported to need more

trainings in newly-emerging alternative assessment instead of traditional assessment methods. Moreover, the delivery of professional knowledge about assessment was found to be disconnected from the real assessment practices and isolated from the complicated contexts in the classroom (Shulha, 1999). Hence, the assessment courses were strongly recommended to improve in multiple aspects. It was suggested to address the assessment preconceptions of prospective teachers (Brookhart, 2011; Izci & Siegel, 2018; Shepard, 2006), or else, their assessment practices might be hard to be accordingly improved (Deneen & Brown, 2016). Pre-service teachers' practical assessment-related experiences should also be involved as a basis for learning (Coombs et al., 2018; DeLuca et al., 2010). Besides, the problem-based learning approach was indicated to be adopted in assessment courses because it was productive to enhance pre-service teachers' AL through solving problems in the practical contexts (Koh & Tan, 2016).

Teacher educators in charge of other subjects in the pre-service teacher education programme were also supposed to provide a model by utilizing a wide range of assessment tools within their own coursework in order to enable prospective teachers to have a deeper understanding of the usage of diverse assessment methods from the perspective as students (Allen & Flippo, 2002; Izci & Siegel, 2018). Active and continuous participation in peer-assessment practices helped pre-service teachers to be more literate assessors (Grainger & Adie, 2014). As evidenced by Reynolds-Keefer (2010), pre-service teachers who had the experience of using rubrics as students might

be more possible to utilize rubrics to their own students when they became teachers. That means modeling the use of assessment tools could increase the likelihood of them being used in pre-service teachers' future teaching careers. Unfortunately, however, Hussain (2017) indicated a large majority of teacher educators attained a mediocre AL proficiency level, practiced AoL by overly relying on traditional tests, and knew little about how to develop and select appropriate assessments for multiple purposes.

The challengeable situation was also what EFL teacher education in the pre-service stage faced. The language testing and assessment trainings needed to get the attention it deserved in the initial teacher education programmes (Giraldo & Murcia, 2018, 2019; Hatipoğlu, 2015). According to Giraldo and Murcia (2019), they indicated such assessment trainings improved participants' quality of assessment design in theoretical, technical, and practical dimensions. Besides, a radical shift in assessment conception was revealed from treating assessment as grading or scoring to a broader view of treating assessment for multiple instructional purposes.

However, the quality of assessment trainings needed much improvement. In terms of teaching content in the courses, the language assessment-related knowledge was a notoriously hard domain for pre-service teachers because of the abstract terminologies in developing and using assessment tools (O'Loughlin, 2006). The dominant coverage of theoretical knowledge with little mentioning of its practical application in the contexts was less successful to help teacher candidates internalize the knowledge

(Gebril, 2017). Therefore, there was a call for integrating theoretical knowledge with the practice of language testing in current situations (Giraldo & Murcia, 2018). On the other hand, Jeong (2013) revealed the teaching content of the courses was significantly determined by the instructors' background. He pointed out that teachers without language testing background tended to concentrate more on classroom assessment and be less confident in technical issues in testing than those with language testing background.

As to the examination coverage of the language testing course, an overwhelming proportion of the questions in the examinations were occupied by the lower thinking skills like remembering and understanding the technical terms while an extremely limited range of questions dealt with measuring higher-thinking skills of evaluating or creating (Mohammadi et al., 2015). This unbalanced coverage of items in examinations might be more likely to direct pre-service teachers to mechanically memorizing what they had learned from the course and might be less likely to direct them to master the competence of language assessment in complicated contexts.

Within China, the situation was no better than other countries. According to Jin's survey (2010), the language testing courses were optional for 60% of programmes and compulsory for the rest 40%. A majority of the courses did not cover the new assessment paradigm and did not pay adequate attention to alternative assessment approaches (Jin, 2010). In the context of Hong Kong, the language assessment training

was still deficient and language assessment courses surveyed were demonstrated to be less successful in linking theory with the practice against the background of assessment reform (Lam, 2015).

Another vital variable identified in the literature was practicum, but there existed competing results. Practicum referred to the period of time which pre-service teachers spend in observing and engaging in the authentic teaching and learning setting in the field schools (Buckworth, 2017). The practicum was intended to enable pre-service teachers to become acquainted with the requirements and practices of their future careers, learn about teachers' daily routines, and gain experience in schools (Buckworth, 2017). Practicum was considered as an inhibitor by Lam (2015), who criticized that it was demanding for pre-service teachers to improve LAL through the period of practicum, as their attempts to implement assessment activities might be confined by the school assessment culture and the evaluation system by the school-based mentors. They had few opportunities and little autonomy to utilize innovative assessment methods which they learned from assessment courses in university. The pre-service teachers were likely to have a limited application of knowledge and skills related to assessment in the school-based contexts. Another criticism stemmed from the evaluation system of the practicum, which focused more on pedagogical skills than assessment competence. Since the practicum supervisor assessed the pre-service teachers on the quality of lesson delivery rather than the assessment activities, the pre-service teachers might have little possibility to manifest language assessment

strategies during the whole period of supervised teaching practices in the classroom.

Nonetheless, the positive evidence from the case study by Y. Xu and He (2019), describing practicum as a facilitator in both fostering pre-service teachers' conceptions of assessment and their identity construction as an assessor. The researchers indicated that after practicum participants changed their understandings of assessment from superficially taking the assessment for measuring students' achievement to a more profound understanding of assessment for multiple purposes. Meanwhile, the pre-service teachers had a more realistic understanding of assessment criteria in schools and began to realize fairness in assessment. The research findings also revealed that during the practicum the pre-service teachers had more opportunities to be immersed in the complex teaching reality and to construct their identity as an assessor. The assessor-identity construction was seen as the ultimate goal of LAL development in Y. Xu and Brown's (2016) model.

Regarding the experiential factors mediating pre-service teachers' AL or pre-service EFL teachers' LAL, prior school-based experience of being assessed as students was reported to be the dominant factor (Y. Xu & He, 2019). The previous assessment experiences of pre-service teachers might influence various aspects of their LAL, including the needs analysis of LAL improvement (Hatipoğlu, 2015), the conceptions of assessment (Bolívar, 2020; Brunker et al., 2019; Crossman, 2004; Deneen & Brown, 2011; Kvasova, 2022; Kyttälä et al., 2021; Smith et al., 2014), and willingness to grasp

the new language assessment knowledge or skills (O'Loughlin, 2006).

To summarise, various contextual and experiential factors mediating AL or LAL evolvement among pre-service teachers have been figured out in the existing literature. It appears to be that the quality of pre-service teacher education around the issue of assessment is not as desirable as expected. There is large room for such trainings to be improved to meet the needs of pre-service teachers and other stakeholders. The prior assessment experience and opportunities for professional development also exert impact on LAL evolvement. However, the existing findings focus on identifying what the factors are on AL or LAL evolvement, scant attention has been paid to how the mediating factors interact with AL or LAL evolvement. Hence, more explorations of this topic with evidence from different contexts are welcomed.

2.6.3 The Development Trajectory of AL and LAL

Up to date, only a few studies available have explored the evolvement of pre-service teachers' AL. Among them, the changes in the conceptions of assessment before and after the initial teacher education project are primarily focused on.

Taking about 1,000 teacher candidates in a 3-year initial teacher education programme for preparing teachers in primary schools in New Zealand as participants, Smith et al. (2014) implemented mixed research to investigate changes in the conception of assessment. They argued that initially the participants' conception seemed to be

derived from personal experiences of being assessed in schooling time and mostly seemed to be restricted to AoL, although they had some notions of AfL. After the 3-year programme, the pre-service teachers broadened their views towards assessment by embracing assessment as a means of learning improvement and by developing an emerging awareness of students' role in self-assessment. More positively, there was a noticeable shift in viewing assessment from the stance of students to the stance of teachers.

A survey-based study by DeLuca et al. (2013) in the USA reconfirmed that pre-service teachers' views towards assessment expanded from sole purpose for testing to multiple assessments for varying purposes. The participants were reported to develop from a one-dimensional conception of assessment based on testing to a multi-dimensional understanding of assessment in relation to the teaching and learning process. Besides, the pre-service teachers seemed to feel increasingly confident in understanding a wide range of assessment methods.

Also, in the United States of America, another supporting evidence came from a qualitative study by Wallace and White (2014), who confined the scope to mathematics pre-service teachers only. The authors classified the evolvement of perspectives and practice of assessment among the participants into three distinctive stages: from test-oriented, through task-oriented to tool-oriented. Likewise, in the early stage, the notion of assessment as tests for grading was dominant, then developed into a broader view

of assessment as tasks for informing student learning, and evolved to the last stage of viewing assessment as tools for enhancing student learning.

A more recent study in Columbia by Bolívar (2020), analyzing the journals written by pre-service EFL teachers to track the development of their LAL through an assessment course. Consistent with previous results, these prospective teachers also demonstrated a change in the conception of assessment from limited assessment purpose for testing to an expanded understanding of serving multiple purposes, especially for student learning. Besides, advancements were observed in other dimensions in LAL, such as a deeper understanding of the roles of students in language assessment activities, more clarity about the strategies of assessment, clearer distinguishment between ethical and unethical assessment, and more awareness of implications of language assessment.

In summary, from the above quantitative and qualitative studies conducted in different contexts, it seems that regardless of variations in subjects and teacher education programmes, prospective teachers' conception of assessment is likely to experience a desirable change from AoL to AfL. However, other dimensions of LAL are largely not mentioned in the literature, for example, assessment knowledge, assessment practices, and identity construction as an assessor. In addition, a few studies have been conducted in western countries (e.g., USA, Columbia, and New Zealand), little is known from the context of China. Thus, what is still unclear is the LAL evolvement trajectory among pre-service EFL teachers, let alone the voice from China.

2.7 Summary

In this chapter, the literature related to the concept AL and LAL are reviewed. The conceptualizations of AL and LAL are clarified based on its evolvement from the traditional understanding as a combination of assessment-related knowledge plus skills to a dynamic, multi-layered and complex notion of a process in which AL is not a static concept but is constantly embedded in the negotiation with the local contexts. Against the backdrop of assessment paradigm shift from traditional psychometric to social-cultural interpretative, the assessment is reoriented from AoL to AfL or even AaL. It poses unprecedentedly greater challenges for teachers to become highly assessment literate to fulfill the responsibility mandated in reform policies in modern times, not only for the benefits of teachers and students, but also for successful implementation of curriculum and school renewal.

However, numerous studies have consistently proven the deficient AL among both pre-service and in-service teachers across the world. They tend to lack sufficient AL to engage in assessment-related activities for enhancing student learning in the classroom. Various factors are reported to account for the disturbing situation, including contextual factors (e.g., national policy, school culture, class size) and experiential factors (e.g., learning experience, conception of assessment), but whether the list of these identified mediating factors is comprehensive and how the mediating factors interact with LAL evolvement are still unclear. Generally, the pre-service teachers are

reported to experience professional development in the conception, knowledge, and skills in assessment in the initial teacher education programme, although to a varying degree. However, the research on LAL evolvement trajectory is still lacking.

To sum up, much have been done on the theme of AL and LAL in the aforementioned literature, however, what is still unclear is pre-service EFL teachers' conceptualization of LAL, their self-evaluated LAL proficiency level, their LAL evolvement trajectory, and the interaction between the mediating factors and LAL evolvement. To address these research gaps, the current study is designed. The research methodology is discussed in detail in the next chapter.



CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes rationales for using narrative inquiry, details of the research scene and participants, data collection procedures, and data analysis methods. The role of the researcher and the trustworthiness of the study are also discussed to present a comprehensive understanding of the entire procedures.

The goal of the current study is to explore pre-service EFL teachers' LAL conceptualization, proficiency level, evolvement trajectory, and interaction with the identified mediating factors by targeting at pre-service EFL teachers in X Normal University in China. To this end, the study is qualitative in nature, using narrative inquiry to answer the following questions.

- (1) What is the conceptualization of LAL by the participants?
- (2) What is the LAL proficiency level self-evaluated by the participants?
- (3) How do the identified mediating factors interact with LAL evolvement among the participants?
- (4) How does LAL evolve among the participants?

3.2 Research Design

Aimed to answer the research questions, the present study adopts narrative inquiry in

qualitative research, since the research purpose is in alignment with the intent of qualitative research, which was defined by Creswell (2007) as

it begins with assumptions, a worldview, the possible use of a theoretical lens and the study research problems inquiring into the meaning individuals or groups ascribe to a social or human problem. To study this problem, qualitative researchers use an emerging qualitative approach to inquiry, the collection of data in a natural setting sensitive to the people and places under study. The final written report or presentation includes the voices of participants, the reflexivity of the researcher, and a complex description and interpretation of the problem, and it extends the literature or signal a call for action. (p. 37)

Qualitative research is suitable to obtain a complicated and detailed understanding of the issue, to empower participants to voice their opinions, to share their experiences, and to understand the contexts or settings where participants resolve the issue (Creswell, 2007). In other words, qualitative research intends to understand or explore the meanings and capture some aspects of the real world (Braun & Clarke, 2013). It fits in well with the intention of the current study, which is aimed to hear the voice of the participants on LAL and to understand their assessment experiences in the contexts where their LAL evolves. Furthermore, as the qualitative research is not rooted in a positivist paradigm, the current study is not aimed to seek an ultimate truth, but rather to discover meanings described from the participants' perspectives and to understand how the participants approach LAL.

Narrative inquiry, as a specific type of qualitative research, focuses on gathering information through collecting stories, reporting personal experiences, and discussing the meaning of those experiences upon an individual (Creswell, 2012). It is a research

approach for exploring the individual's lived and told stories to understand human experience (Clandinin & Johnson, 2014; Josselson, 2010). The narrative inquiry paradigm, different from the hypothesis-testing model, is aimed to understand and describe rather than measuring and predicting, to focus on meanings and interpretations rather than causation and statistical analysis, to acknowledge the significance of discourse rather than the numerical representation, to interest in contexts rather than the context-independent elements, and to attach critical importance to subjectivity rather than probing for objectivity (Josselson, 2010).

According to Webster and Mertova (2007), narrative inquiry is not a process of reconstructing life objectively, but “a rendition of how life is perceived”, which cannot be measured by statistical techniques (p. 3). They also argued that narrative inquiry is more appropriate to deal with complexity and human-centeredness that are not easily captured by traditional approaches. The narrative approach advocates pluralism, relativism, and subjectivity (Lieblich et al., 1998). Bell (2002) further pointed that narrative is better to understand the individual's experience, to collect information which the participants might not know consciously, and to explore the notion of experience and understanding of events changes.

In teaching and learning contexts, narrative inquiry is also well suited to addressing the complexity of human-centered experience (Webster & Mertova, 2007). It serves as a powerful tool to explore teacher professional growth and teacher preparation (Conle,

2000; Contreras & Gerardo, 2000). Additionally, narrative inquiry is therapeutic in nature and beneficial for the narrators who are provided with an opportunity to express their stories (Murray, 2018). Mishler (1986) further contends that the narrators are not only empowered to articulate their experiences, but they may move beyond the text to potential actions and changes through the narratives. Thus, it makes sense that the pre-service EFL teachers are provided with an opportunity to reflect on their professional development through sharing the narrative account of their assessment experiences, which may facilitate their further concern and subsequent actions on LAL.

Considering that the research objective of the current study is to explore the inner world of pre-service EFL teachers towards LAL conceptualization, LAL proficiency level, LAL evolvement trajectory, and the interaction between mediating factors and LAL, narrative inquiry is proper to understand human complexity, particularly in the cases in which the influencing variables cannot be controlled (Josselson, 2010). Narrative inquiry can provide in-depth evidence of the development trajectory of LAL over time and reveal the interaction between mediating factors and LAL growth (Harding & Brunfaut, 2020). Therefore, in view of all these advantages of narrative inquiry and its alignment with the research purpose, the current study takes narrative inquiry as the research method to address the research questions.

3.3 Research Settings

The present narrative inquiry is an endeavor to explore LAL among pre-service EFL

teachers for primary and secondary schools in China. The study took place in a pre-service EFL teacher education programme offered by X Normal University in China. X Normal University was chosen for purposive sampling. The researcher has been working there for over 14 years as an EFL teacher educator. The familiarity with the setting in the study makes it easier for the researcher to understand the participants' experience in pre-service teacher education programmes and more helpful to establish trust between researcher and participants in researcher-as-instrument qualitative research (Miles et al., 2014).

3.3.1 Background Information of X Normal University

The X Normal University selected in the study is located in the western region, a less developed area of China. Founded in 1978, it is a province-owned university featuring undergraduate teacher training. The university covers an area of more than 680,000 square meters of land and more than 16,000 students reside on the campus. The university is made up of 19 schools or departments and 28 research institutions. There are currently 54 undergraduate majors, of which 19 majors are teacher-training ones and 35 non-teacher-training ones.

The number of staff is 1,100, of whom 698 are full-time teachers in 2021. The number of professionals including professors and associate professors is 233. For the past 40 years, the university abides by the fine tradition of normal education and sticks to the university-running philosophy of serving the locals and leading the development of

society. The university-running orientation is the establishment of an applied mode university featuring teacher-training, with multiple disciplines and levels for serving the economic development of the society.

3.3.2 Background Information of Pre-service EFL Teacher Education Programme

With regard to the pre-service EFL teacher education programme in X Normal University, the programme provides pre-service teachers four years of tertiary education, including the coursework and teaching practicum, cited from the talent training scheme (See Table 3.1). Every year nearly 180 pre-service EFL teachers are enrolled and placed in six classes with about 30 students in each class.

Table 3.1

Outline of Courses Offered in the Pre-service EFL Education

Type	Term	Courses	Percentage of the total credits (total)
General courses	1-7	Computer, College Chinese	29.83% (181)
Professional courses	1-7	Intensive English, Listening, Translation, Academic Paper Writing	44.2% (181)
Teacher education courses	2-7	Educational Psychology, Class Management, English Pedagogy	7.74% (181)
Teaching practicum and training	1-6	Teaching Skills Training, English Lesson Design Training	15.47% (181)
Innovation and entrepreneurship courses	1-8	No specific course, just participating activities or competitions	2.76% (181)

The general courses are a must to be attended by all majors in X Normal University to meet the basic requirement of undergraduates in China. A series of elementary language proficiency courses are offered to improve pre-service teachers' English, for instance, Oral English, Intensive Reading, English Listening, and English Writing. They account for the most credits in the talent training scheme, revealing that the English proficiency is placed in the priority. The third type is the teacher education courses, which are arranged from the first year to the last year, including English Pedagogy, Educational Psychology, and English Teaching Activity Design. Language Testing and Assessment is a selective course of two credits while English Pedagogy is a compulsory course with two credits for all pre-service teachers.

The teaching practicum in field school lasting for 18 weeks is compulsory at the second term of the third year. Before starting to teach in field schools, all pre-service teachers must attend the intensive teaching training lasting for two weeks offered by the university to strengthen their teaching skills under the guidance of the mentor. Then they are allocated to the field school according to the arrangement of the university. Each pre-service teacher is assigned with two mentors, one from university and the other from field school, who are collaboratively in charge of the pre-service teacher in practicum. The mentor in the field school is responsible for guiding, supervising, and evaluating the pre-service teacher's work in practicum while the university mentor regularly visits the school for providing necessary support. The last type of courses are innovation and entrepreneurship courses, which needs pre-service teachers to

participate in activities or English competitions to obtain the corresponding credits.

3.4 Participants' Selection

The participants selected in the study came from X Normal University in China where the researcher has been working for nearly 14 years. The familiarity with the contexts makes it more possible to establish a greater degree of rapport and trust between participants and researcher. People are usually reluctant to do self-exploration of their experience and feelings to strangers, however, participants are more open to sharing if they trust the interviewer in most cases (Polkinghorne, 2007). The greater degree of trust established between researcher and participants, the greater degree of self-revealing from participants, the greater degree of richness of information (Josselson, 2007).

As to the number of participants, qualitative research does not have any specific rules for a study (deMarrais, 2004). The qualitative research tends to explore depth and details, usually focusing on a few participants is more proper (Fink, 2000; Miles & Huberman, 1994). Since the intent of qualitative research is not to generalize the research results from a sample to a population but to present an in-depth understanding, a larger number of cases may result in superficial analysis (Creswell, 2012). The deep understanding, analysis, and representation matter more in qualitative research. Thus, the in-depth exploration may provide rich information to the research questions so that the number of participants in narrative inquiry is restricted to a few individuals because

more participants may lead to shallow analysis and may hinder the access of rich and free-ranging discourse (Josselson, 2010). J. Kim (2016) also contends that the sample size tends to be smaller in narrative inquiry if the focus is to collect the life stories of the interviewees as the interviewing may be lengthy. Beitin (2012) suggests six to twelve participants is a proper sample size because there may be thematic redundancy. Bearing this in mind, my study selects six participants to explore their LAL.

In this narrative study, purposeful sampling is utilized. Purposeful sampling is defined as “researchers intentionally select individuals and sites to learn or understand the central phenomenon” (Creswell, 2012, p. 206). The rationale for using this selection method is that it helps to select participants who could detail their assessment experiences. Maximal variation sampling, also known as heterogeneous sampling, is a purposeful sampling method. The participants are sampled on different dimensions of characteristics or traits (Creswell, 2012). The assumption underlying this sampling method is to view from all potential angles, thereby involving participants selected across a wide spectrum to reach a greater understanding (Etikan et al., 2016). Rather than attempting to be generalizable to other contexts, the maximal variation sampling method intends to elicit multiple perspectives, therefore participants are purposefully as different from each other as possible for a diversity of views (Creswell, 2007, 2012). This sampling method is appropriate for the study because it enables to draw diverse perspectives and experiences from participants around the issue of LAL.

To be specific, one participant in pilot study and six participants in the formal study were selected. They were pre-service EFL teachers at the final year from X Normal University in the western region of China. The name list including some background information was suggested by the counselors who were in charge of the senior students. Among the 220 pre-service EFL teachers who would graduate in 2022, the participant, Gina, in pilot study, was suggested by the counselor according to the researcher's criterion: willing and interested to participate, cooperative and communicative.

In the formal study, among the potential participants, the researcher purposefully chose six participants with varying background information, such as gender, class number, practicum field school, and English proficiency. English majors were more female than male with the ratio of approximately 6:1 in X Normal University. The small proportion of male pre-service EFL teachers means it is comparatively difficult to get access to the same number of male participants with the female ones. After the researcher's explanation and invitation for several potential participants to join in the research, a few of them declined politely for various reasons: busy with finding a job, preparing for post-graduate entrance examination, or not interested in the research at all. At last, only one male responded to the invitation with explicit willingness to participate. So due to these practical limitations, the sample included only one voluntary male participant and the rest were female.

Different class number meant that the same course might be taught by different

teachers to elicit more information about the pre-service teacher education programme. Diverse practicum schools meant various teaching practice experience, so it was a factor to be considered. The last one is English proficiency. Participants with different proficiency level might provide more perspectives and hold different attitude towards assessment. English proficiency was judged both by their academic performance in each term in the university and the score attained in a national Test for English Majors (TEM-band 4) in 2021.

The researcher contacted the potential participants individually through text-messages for inviting to take part in the research voluntarily. Then the researcher organized a meeting at their convenient time to introduce the details of the research and get the consent form signed (See Figure 3.1).

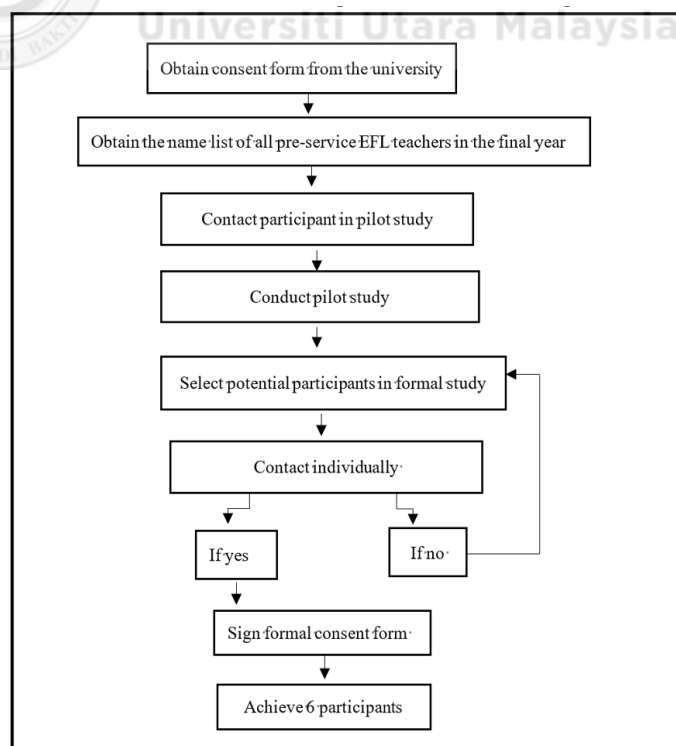


Figure 3.1 Participant selection procedures

After identifying the six participants, the profiles of each participant with pseudonyms from A-E according to English proficiency are listed in Table 3.2. All participants (five female and one male) were around 21 years old, among whom two came from village, two from county, one from town, and one from city. Four of them passed TEM-4 with two obtaining more than 70 in the examination. All of them had teaching experience; some tutored and some taught during the pre-service teacher education period. All participants attended practicum except Daisy. Due to some personal factors, Daisy missed the opportunity to complete the practicum. But she was a qualified potential participant according to the maximal variation principle in purposive sampling and she was strongly recommended by the coordinator for her cooperative attitude and talkative personality. Therefore, she was still selected for inclusion of various perspectives with different experiences.

Table 3.2

Profiles of the Participants

No.	Pseudo -nym	Gender	Age	Place of birth	Scores in TEM4	Teaching experience	Type of practicum school
P1	Amy	Female	22	Village	74	Tutor	City school
P2	Betty	Female	20	County	72	Tutor	City school
P3	Carol	Female	20	Village	63	Teacher	County school
P4	Daisy	Female	21	Town	60	Tutor	No participation
P5	Edwin	Male	22	City	52	Assistant, Tutor	City school
P6	Flora	Female	22	County	50	Teacher	City school

3.5 Data Collection

In qualitative research, researchers need to collect multiple sources of data to answer the research questions from different perspectives rather than relying on a single data form (Creswell, 2014). In this study, data were gathered from various sources: the semi-structured interview, documents (talent training scheme, syllabus of English Pedagogy course, and syllabus of practicum), artifacts (lesson plans and any assessment artifacts designed by the participants), and journals (from participants and the researcher). The data inventory is listed in Table 3.3 for an overview of all the data sources. More detailed description will be included in the following sub-section.

Table 3.3

Data Inventory

Data type	Purpose	Sources	Details	Total
Main data				
Interview	To understand LAL of pre-service EFL teachers	Amy	3.7 hours (61,413 words)	19 hours (282,624 words)
		Betty	2.7 hours (36,818 words)	
		Carol	3.4 hours (49,411 words)	
		Daisy	2.5 hours (35,057 words)	
		Edwin	3.4 hours (53,994 words)	
		Flora	3.3 hours (45,931 words)	
Supplementary data				
Document	To understand arrangement	Talent training scheme	1 file (6,007 words)	3 files (21,445 words)
	To understand the content	Syllabus of English Pedagogy	1 file (6,615 words)	
	To understand the practicum arrangement	Syllabus of Practicum	1 file (8,823 words)	

Table 3.3 continued

Artifacts	To understand assessment design	Assessment artifacts	1 Teaching practicum handbook from each participant (64 pages)	6 handbooks (384 pages in A4 size)
Journals	To reflect on LAL	Reflective journals	3 journals from 3 participants	3 journals (16 pages in A5 size)
	To self-reflect critically	Reflexive journals	1 journal from the researcher	1 journal (4 pages in A5 size)

3.5.1 Semi-structured Interview

The interview is well suited to concentrate directly on issues and to provide insightful perspectives related to the central phenomenon (Yin, 2018). Fontana and Frey (2008) state that the interview is one of the powerful means to understand humans. Interview with individuals is the most commonly used method in qualitative research and foremost in narrative inquiry (J. Kim, 2016). As the most widespread form in narrative inquiry, the interview-based design places interviewees at the heart of the study, intends to gather participants' own stories, and helps researchers to understand participants' lived experiences and behaviors (Josselson, 2010; Murray, 2018).

There are three types of interviews: structured, semi-structured, and unstructured interviews (Schwandt, 2007). Structured interviews tend to favor closed-ended questions pre-determined by the researchers, in which the questions generally remain the same for all interviewees who are restricted to a limited set of responses (Yin, 2015). Unstructured interviews are non-directive in nature, in which the questions are not generated beforehand but on the spot by the researchers, much like an informal

conversation in daily life (Glesne, 2011).

The semi-structured interview with a general structure set beforehand covers the main questions to be asked and the detailed questions are left to be worked out during the flow of interviewing (Drever, 1995). That is, the same topics serve as the basis for questioning, yet the sequencing of probing is participant-led (Roulston & Choi, 2018). This gives the interviewee a large degree of freedom in deciding what and how much to talk about, therefore it is very flexible for this study. Further, Fontana and Frey (2008) suggest building a collaborative relationship between the researcher and the interviewees, who are supposed to work collaboratively to generate a narrative interview. In narrative inquiry, responses that contribute to eliciting lived experience and stories are the desired outcome of the interview. Therefore, the semi-structured interview is appropriate to elicit responses of the participants towards LAL, thus it is adopted as a primary tool to collect data.

3.5.1.1 Development of the Interview Protocol

In order to stay focused on the issue of the research, this narrative inquiry used the interview protocol designed by the researcher based on research objectives, assessment practices in China, and prior studies adapted from Bolívar (2020), Giraldo and Murcia (2019), Harding and Brunfaut (2020), Newfields (2007), Prasetyo (2018), Tsagari and Vogt (2017) and Yan and Fan (2020) (See Table 3.4).

Table 3.4

Structure of the Interview Protocol

Interview	RQ addressed	Purpose
Round 1	RQ 1	To collect conceptualization of LAL
Round 2	RQ 3, 4	To gather past experience of being assessed in primary and middle school
Round 3	RQ 3, 4	To gather past experience of being assessed in university
Round 4	RQ 2, 3, 4	To gather past experience in assessment practices and evaluation of LAL

The first round aimed to understand the participant's background information and LAL conceptualization to seek the answer for Question 1. The second and third interview was carried out to collect the participant's prior experience of being assessed in primary, middle school, and university. The inclusion of the prior experience in primary and middle school is justified by the existing evidence which has demonstrated that personal history and prior assessment experience have considerable influence on the assessment conceptions and practices of pre-service teachers (Deneen & Brown, 2011; Reynolds-Keefer, 2010; Smith et al., 2014). Thus, a better understanding of the pre-service teacher's prior educational background information concerning language assessment experience is critical to explore the mediating factors in LAL evolvement.

The last round was interested in their critical experience of assessment practice in teaching practicum in field school and LAL proficiency self-evaluated by themselves after the pre-service education. The last three rounds of interview intended to collect data on LAL self-evaluation, mediating factors, and evolvement trajectory during the

process for addressing Question 2, 3 and 4. For more details about the interview protocol see Appendix C.

3.5.1.2 Refinement of the Interview Protocol

The interview protocol was refined through four stages suggested by Castillo-Montoya (2016). The first stage was to ensure the questions asked in the interview were in alignment with the research questions. Phase two required constructing an inquiry-based conversation. The protocol in the interview needed to be reviewed by expert for feedback in Phase 3. The last phase entailed validating it in a pilot study. Followed by the procedures, the interview protocol used in the study was refined through the four stages as indicated (See Table 3.5).

Table 3.5

Refinement of the Interview Protocol

Phase	Refiner	Method	Content focused	Content Revised
1	Researcher	Re-examining	Alignment	/
2	Researcher	Re-examining	Format	Adjust the sequence
3	Expert	Reviewing	Structure	Add questions
4	Researcher	Pilot study	Comprehensibility	Add and change several questions

Firstly, it was re-examined by the researcher to check its alignment with the research questions. The questions in the interview were not conversed directly from the research questions for the purpose of being understandable and comprehensible to the participants. Moreover, the questions were checked whether they were organized in a

conversational order. Some of them were re-sequenced and re-classified by the researcher in a more logic order.

Thirdly, the protocol was closely read by an EFL associate professor to provide feedback on its structure, length, writing style, and clarity. The EFL expert was selected purposefully for her expertise in EFL teacher education and qualitative research. She had rich teaching experience and had taught pre-service EFL teachers for 16 years. Besides, she had been the leading researcher of four province-funded projects focusing on EFL teacher education qualitatively. She commented positively that the whole interview questions were closely related to the research questions and were arranged in a reasonable order. She also provided a suggestion for adding a question at the end of the interview: Do you think this interview has any changes on your assessment understanding and has any influences on your future assessment-related activities? Her advice was adopted in the final version to understand whether this interview had any influence, positive or negative.

Then, a pilot study was conducted to try out the protocol for the researcher to get a realistic picture of the length of the interview and whether the questions were answerable for the participant. Before the pilot study, the interview protocol had been translated into Chinese, the mother tongue of the participants and the researcher, for an accurate understanding and expression of experiences. The translation process followed Brislin's (1986) model (See Figure 3.2).

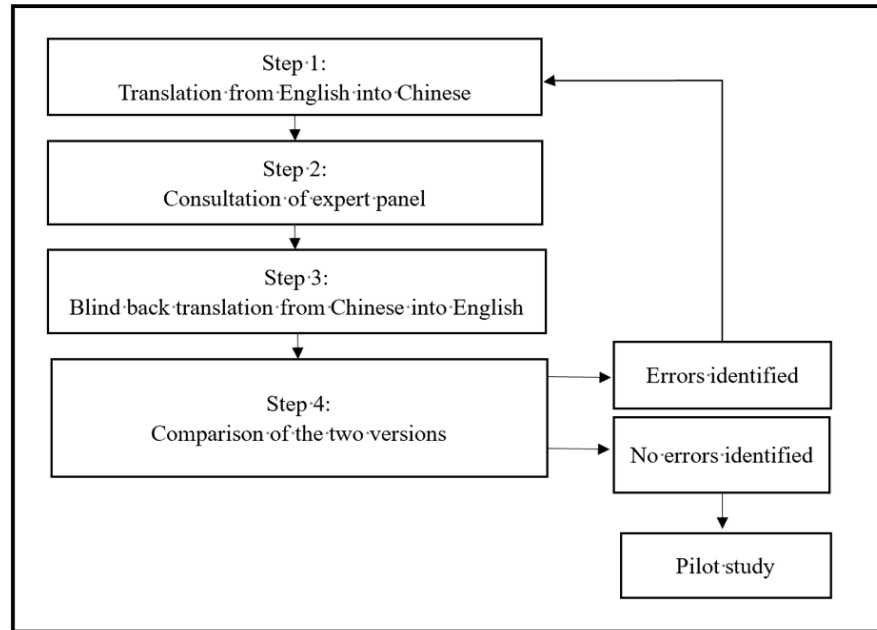


Figure 3.2 Translation process of interview protocol

The first step was translation of the English version into Chinese by the researcher. Two associate professors who had engaged in translation for more than 10 years in EFL education in X Normal University were invited as the expert panel to examine the two versions for any potential misunderstandings. The third step was to translate backward from the Chinese version into English blindly by one of my colleagues, who was also an experienced EFL teacher. Finally, comparison between the original English version and the backward translated English version of the interview protocol was made by the researcher from linguistic and cultural equivalence. If any errors were identified, then went back to the first step in translation for revision. If no errors were identified, then the translation of the interview protocol from English to Chinese was completed. After such a validation process, the Chinese version was implemented in pilot study.

In the pilot study of Gina, the Chinese interview protocol was revised a little: add and change some questions. Add the questions about the experience of intensive training of teaching skills in university before practicum for a complete understanding of practicum by including pre-practicum activities. Besides, it was confirmed that the selective course Language Testing and Assessment, which was chosen by fewer than 20 students, was not offered according to the regulations of the university. Therefore, some questions concerning this course had to be changed into inquiry about the reasons of not selecting and the challenge of learning assessment.

3.5.1.3 Implementation of the Interview Protocol

Undergone the stages of development, refinement, translation, and validation, the interview protocol was adopted in the formal study. The interviews with each participant were carried out in four rounds. The interview duration of the six participants was 19 hours in total, amounting to over 282,000 words after transcribing. The longest duration was Amy with over three and a half hours while the shortest was Daisy with only two hours and a half (See Table 3.6).

Table 3.6

Time Duration of Each Interview Round

Participants	Round 1	Round 2	Round 3	Round 4	Total
Amy	48 mins	46 mins	53 mins	57 mins	3hrs 42 mins
Betty	20 mins	44 mins	43 mins	38 mins	2hrs 40mins
Carol	20 mins	54 mins	58 mins	1hr 3 mins	3hrs 25 mins
Daisy	20 mins	30 mins	49 mins	38 mins	2hrs 30mins
Edwin	31 mins	41 mins	1 hr	1hr 2 mins	3hrs 24 mins
Flora	39 mins	43mins	45 mins	1 hr 5 mins	3hrs 20 mins

3.5.2 Documents

The document is valuable in the qualitative study and represents another form of primary evidence (Yin, 2015). The document is a way to understand and to make sense of social and institute practices (Coffey, 2014). Thus, the document collected includes: the talent training scheme (1file), syllabus of English Pedagogy course (1file), and syllabus of teaching practicum of the pre-service EFL education programme in X Normal University (1 file).

The talent training scheme included the training orientation and comprehensive arrangement plan through the 4-year initial teacher education. It helped the researcher to obtain a general outline of the training project, especially the arrangement of the courses and practices offered for the participants. The teaching syllabus of English Pedagogy was gathered to reveal the teaching content concerning language assessment in the coursework. In addition, the syllabus of teaching practicum was also helpful to understand the practicum plan and schedules. Thus, one talent training scheme, one teaching syllabus of the course, and one syllabus of practicum were gathered.

3.5.3 Artifacts

Artifacts can also be revealing in data collection in a qualitative study (Yin, 2015). In this study, the teaching practicum handbook (64 pages in A4 size) from each participant was copied as the artifacts including: eight lesson evaluation records, eight teaching plans, three thematic class activity records, three headteacher work logs, one

teaching investigation report, and one practicum summary. Such kind of information could help to illustrate how the participants designed language assessment activities in the lesson plan and how they developed language assessment.

3.5.4 Journals

Journals provide the advantage of being expressed in the language and own words of the participants and can reveal a rich source of information (Creswell, 2012). For participants, they were suggested to keep a reflective journal to jot down anything related to the issues of LAL in the present study. However, only three of the six participants, who were Amy (9 pages), Carol (1page), and Flora (6 pages) kept the journals in A5 size. The other three said they had expressed all their opinions in the interview and had nothing else to record in the journal.

The researcher is strongly advised to keep a reflexive journal during the research process. In the qualitative study, such a journal recording the introspections into the reaction and feelings about the fieldwork may later provide insights into the research perspectives and any possible undesired biases (Yin, 2015). Reflexivity is more than reflection because reflection is to take one step back from the phenomena under study whereas reflexivity is to take one more step back from reflection (J. Kim, 2016). In other words, reflexivity involves reflecting on the reflection (Jenkins, 1992). It offers an opportunity for the researcher to conduct self-reflection or self-evaluation critically as it pertains to the study by examining how the researcher's position or interest affects

the process in the research (Primeau, 2003). The reflexive practice will not end with the completion of interviews, but still keeps on to the data analysis stage and drafting of the findings (Roulston, 2014).

Followed by the suggestions mentioned above, the researcher kept a reflexive journal from the very beginning of the research, through data collection and data analysis to the completion of the data presentation. In total, the journals were kept in 4 pages in A4 size (nearly 3,000 words). Since the study could not be separated from my subjectivity, reflexive journals permit me to reflect critically on the role of my values, assumptions, and beliefs upon the research process, which may add rigor and trustworthiness of the research.

3.6 Research Procedures

After getting the consent from X Normal University, the researcher conducted a pilot study before the main study to validate the interview protocol as well as enhance the interviewing skills. The pilot study offers another opportunity to practice for it helps to refine the main study from several aspects, such as design, data collection instruments, and data analysis plans (Yin, 2015). Malmqvist et al. (2019) also advocate that the pilot study may better inform and prepare the researcher to face the challenges possibly arising in the main study and increase the researcher's confidence in the instrument in gathering data. The weakness of the research may be identified through properly analyzing the steps and results from the pilot study; thus, the quality of the

research will be improved by a well-organized and well-managed pilot study.

The pilot study was conducted in Chinese with one pre-service EFL teacher, Gina, in X Normal University (See Figure 3.3). Gina was suggested by the counselor according to the researcher's criteria: communicative, having completed practicum and the required courses, interested and having free time to participate. After the collection of the consent form, the assessment artifacts were also gathered. Then, the face-to-face interview in four rounds was audio-taped to make sure the correct use of the device. Later the interview was transcribed and analyzed accordingly. At last, the comments from the interviewee and the result of data analysis might be useful to refine the protocol and the interview implementation.

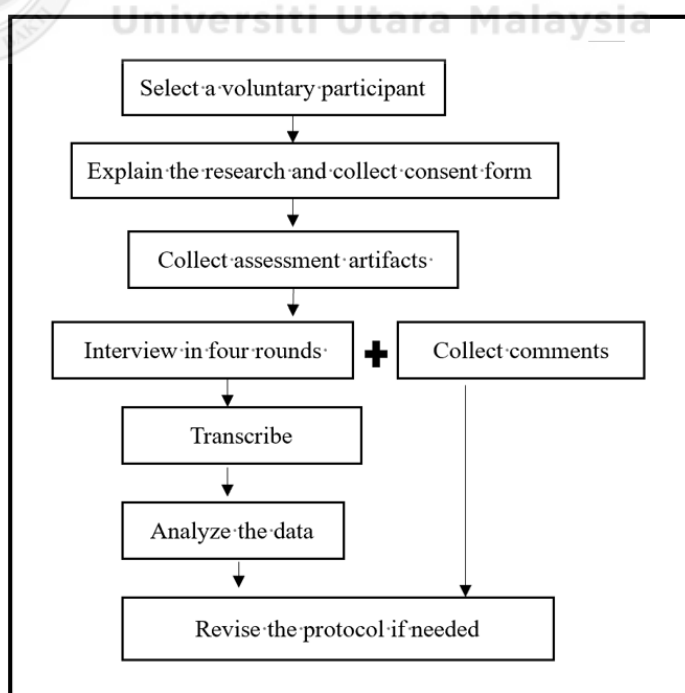


Figure 3.3 Procedures of the pilot study

As to the main study, it was proceeded through four stages: preparation, implementation, analysis, and conclusion (See Figure 3.4). Just to clarify each procedure in the research plan, the research was not linear but iterative in nature. The details and selection of six voluntary participants were mentioned in section 3.4 and the refinement of the interview protocol was introduced in section 3.5.1.

After collection of all the written documents, files, and artifacts mentioned in the previous section (See 3.5), the researcher conducted four rounds of semi-structured interviews with six participants at intervals during a month. Seidman (1991) suggests that at least three times of interviews over time can help the participants to gain trust in the interviewer and overcome hesitancy to reveal themselves in a single interview. Kvale (1996) also holds that in narrative inquiry a minimum of three rounds of interviews focusing on gathering open and in-depth stories is necessary.

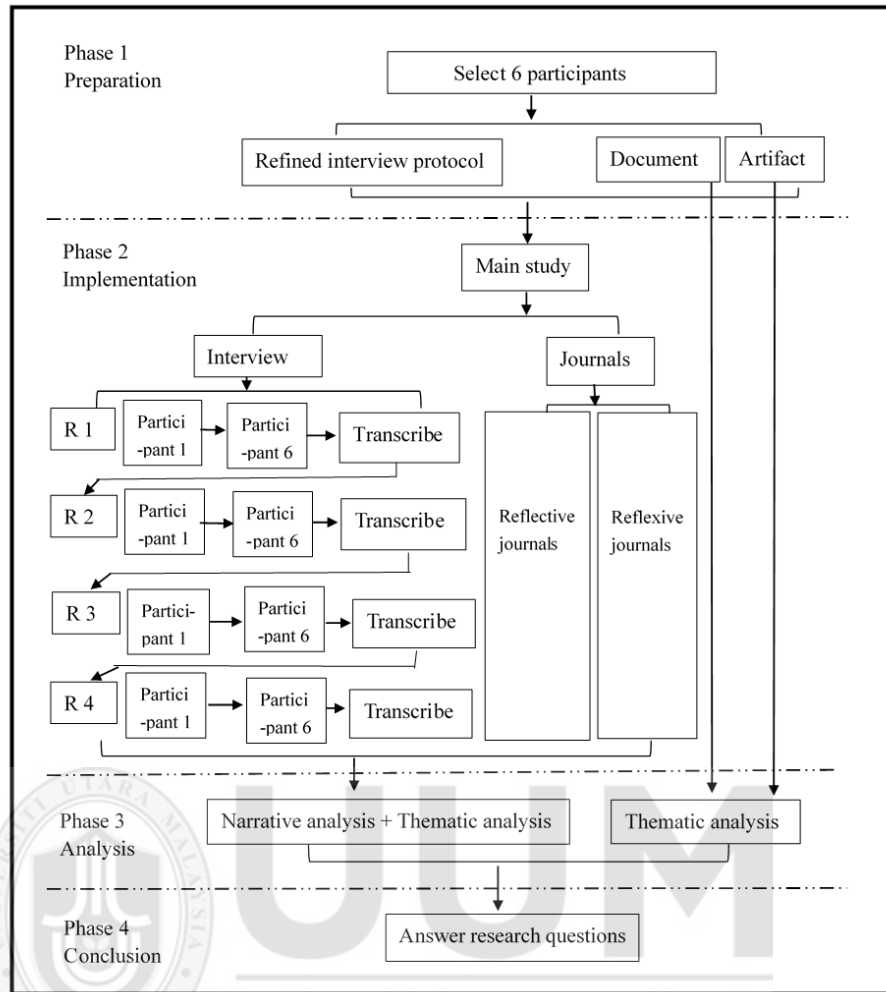


Figure 3.4 Procedures of the main study

The site for interviewing is important because a proper setting may make the interviewees feel comfortable to recount details of their experiences (Murray, 2018). All of the interviews were arranged in a small meeting room at the available time of the participants. The interviews were audio-taped under the consent of the participants. After the completion of each round of the interview, the researcher transcribed the audio recordings verbatim in Chinese. The original interview transcripts were recorded and analyzed in Chinese in order not to lose meaning and avoid any misunderstandings because “some nuances of one language may never be adequately translated into another” (Esin et al., 2014, p. 208). For example, the word “pi ping” in Chinese can be

translated into “criticise” or “comment” in English. The researcher initially translated the word “pi ping” in Chinese to “criticise” in English in Edwin’s interview transcripts. But after re-examining the data and reconfirming with Edwin, the researcher realized that “criticise” was not appropriate for it refers to pointing out the faults of someone. While in the interview, Edwin intended to convey “comment” in that story, so the researcher decided to translate “pi ping” into “communicating feedback” after clarifying the subtle difference. The translated excerpts in English were presented in the data citations, as recommended by Van Nes et al. (2010) to stick to original language as long and as much as possible to avoid potential limitations of translation from source language to target language in qualitative data analysis. Besides, the reflective journals from the participants and the reflexive journals kept by the researcher were also collected to be incorporated into the data analysis. The specific data analysis methods and procedures will be introduced in the next section.

3.7 Data Analysis Methods

Data gathering and analyzing in qualitative research is not a linear but an iterative process. The process weaves back and forth between data collection and analysis. In qualitative research, the researcher makes subjective judgement on whether the state of saturation has been reached: when new data will not add any new information for generating categories (Creswell, 2012).

Formal and systematic data analysis is done continuously as soon as the data are

collected. It begins during the process of data collection and interpretation (Roulston, 2014). The process of data analysis is to piece together data, to make the implicit clues explicit, to make decisions on whether it is significant or not, and to connect seemingly unrelated experiences (Josselson, 2010).

To be specific, narrative analysis and thematic analysis were utilized in this study to interpret the data (See Figure 3.5). The four-round interviews were transcribed verbatim and analyzed narratively and thematically. All the documents (talent training scheme, syllabus, and documents about practicum), artifacts (lesson plans) and journals (from participants and the researcher) were analyzed along with the interview transcripts through thematic analysis.

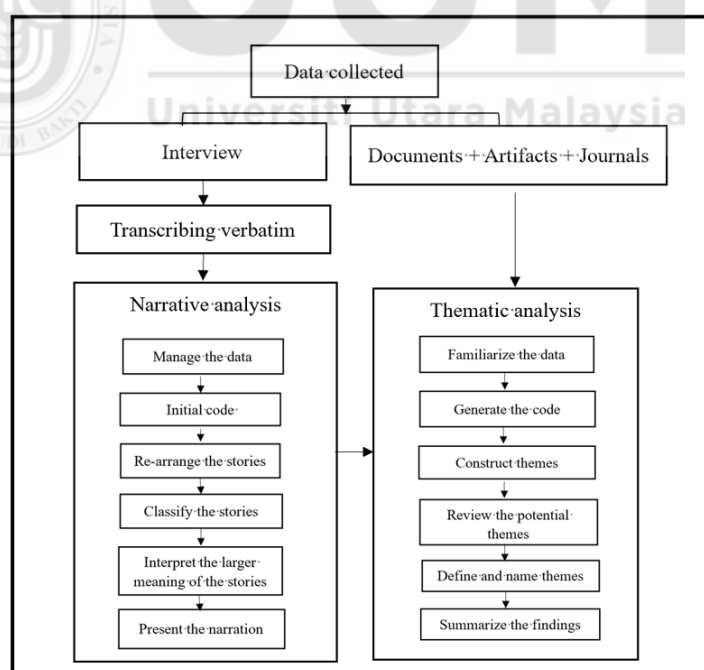


Figure 3.5 Procedures of data analysis

Narrative analysis is useful to develop an in-depth perspective of participants'

experiences related to a particular issue or phenomenon (Riessman, 2001). The narrative accounts collected in the study need to be analyzed for a chronology of unfolding events and the turning points (Creswell, 2007). Creswell (2007) proposed six-step narrative analysis procedures, which was described in a detailed way and easy to follow to handle the data for novice researchers. Hence, it was adopted to analyze the narrative data in the study. As is indicated by Creswell (2007), the analysis procedures cover six steps. (1) The first step is data managing to create files for data. (2) The second procedure involves reading and memoing the texts for initial codes. The interview transcripts will be closely read and re-read by the researcher to be immersed in materials. (3) Then the stories or experiences from the participants will be re-storied and re-arranged according to the chronological order. (4) The fourth step classifies the stories and the contextual factors involved. (5) The fifth deals with the interpretation of the larger meaning of the story elicited from the participants. (6) The final one is to present the narration focusing on issues involved in the research questions.

After the initial analysis of narrative accounts from each participant, the cross-cases analysis was performed to draw similar patterns or themes across the individual narrative account or probe into the differences among the narrative experiences (Josselson, 2010). To this end, the thematic analysis was conducted subsequently. Thematic analysis is an essential data analysis method in qualitative research, primarily analyzing the patterns and themes repeatedly emerging in the textual data

(Braun & Clarke, 2012). It is “a data reduction and analysis strategy by which data are segmented, categorized, summarised, and reconstructed in a way that captures the important concepts within a data set” (Ayres, 2008, p. 867). That is to say, it is mainly a descriptive strategy in search for patterns of experience in the qualitative data.

Thematic analysis is helpful to theorize across a set of cases and to find thematic elements across participants and the events elicited from the participants (Riessman, 2005; Vaismoradi et al., 2013). Thematic analysis best fits into elucidating the specific nature of the participant’s conceptualization of the phenomenon being studied (Joffe, 2011). It enables the researcher to state the main topics that the interview transcripts turn out to be about. Thus, it is appropriate to analyze the participants’ conceptualization of LAL and the emerging themes in LAL evolvement trajectory across cases. Furthermore, it is suggested to be a flexible research tool to reveal a rich and detailed account of data and to categorize the themes occurring in the documents and artifacts related to the research topic in textual data (Braun & Clarke, 2006).

After its rationale was explained, the thematic analysis was implemented in six steps suggested by Braun and Clarke (2006, 2012) because their method provides a clear and concise step-by-step guideline for the researcher to follow. The thematic analysis, like most qualitative analysis methods, is a non-linear process but iterative and recursive: the researcher moves back and forth among the different phases (Braun & Clarke, 2006). (1) The first phase is to familiarize with the data set, often beginning

during data collection. The researcher actively engaged in the data by attentively reading the textual data to generate very early and provisional analytic ideas. (2) The second phase aims to generate codes. Coding means identifying relevant data and labeling them with a few words that cover the meaning of the segment. Coding helps the researcher develop an insightful understanding of the data and provides a thorough basis for the analysis. The researcher embraced an open and flexible attitude to coding, refining, and revising codes throughout the entire research process.

(3) The third procedure involves constructing themes. Guided by the research questions, the researcher examined, combined, and clustered the codes together into a possible theme. (4) The fourth procedure is to review the potential themes respectively. After all candidate themes were developed, the researcher reviewed the network of themes to further shape, clarify, or even reject the theme. (5) The fifth step is to define and name themes. Defining a theme means providing a short summary of the core concept and abstract of each theme. It is concerned about guaranteeing the clarity, cohesion, precision, and quality of developing thematic analysis. (6) The final phase is writing the report to summarise the findings.

3.8 Researcher's Role

The role of the researcher should be taken into account in narrative inquiry (Murray, 2018). One of the key features of narrative inquiry is the collaborative relationship between the researcher and the participants (Connelly & Clandinin, 1990). The

interviewer is not simply a passive recipient (Gubrium & Holstein, 1998). The interview is co-created through the interaction between the interviewer and the interviewee (J. Kim, 2016; Polkinghorne, 2007; Savin-Baden & Niekerk, 2007). For this reason, the researcher is supposed to reflect on contribution to the shaping of the narration (Murray, 2018).

The interviewer should guard against simply eliciting the responses they have expected and should take an open listening attitude, attach importance to the unexpected responses from participants, and empower the interviewees by recognizing they are the only ones to get access to their experiences (Polkinghorne, 2007). The researcher should be a good listener not falling into a trap of existing preconceptions, remain adaptive, and have a firm mastery of issues under study from the professional stance (Yin, 2018).

The interviewer must make non-judgment about the participants' life and equality between the researcher and participants is critically important in narrative research (Josselson, 2007). The ideal state between the researcher and participants is to reach a joint inter-subjective understanding throughout the process in the research (Connelly & Clandinin, 1990). Additionally, the interviewer should better take an empathic attitude to the interviewees and try to understand their experiences from their perspectives (Josselson, 2010).

Bearing these criteria in mind, the researcher approached the participants with the identity of a researcher not an authority, trying to build an equal relationship with the participants. Being an attentive listener during the interview and keep a dialogical partnership with the participants, the researcher also kept a sensitive attitude to the unexpected responses or elicitation from the participants. During the interview, the researcher used a range of paralinguistics (such as nodding, eye contact) and short phrases to provide a supportive response and to encourage the interviewees to go on with their stories.

However, no matter how much the researcher puts aside the biases or associations to the interview, the interview content is indeed influenced by the researcher who must acknowledge and reflect on such an impact (Josselson, 2010). Therefore, reflexivity of the researcher is invaluable since it takes honesty to the fore, acknowledging a variety of factors influencing the data collection and interpretation (Kemmis & McTaggart, 2003). The researcher's reflexivity involves researchers in a critical and conscientious reflection of their subjectivity in relation to the participants and events being studied (Bishop & Shepherd, 2011; Primeau, 2003). Likewise, May and Perry (2014) stated that "Reflexivity involves turning back on oneself in order that processes of knowledge production become the subject of investigation" (p.109). This process enhances the accuracy of the study by bringing researchers to be aware of the influence of their preconceptions or biases that might remain hidden in the entire research process, ranging from situating the study, getting access, positioning of self, staying in

the field to telling the narrative account (Bishop & Shepherd, 2011; Josselson, 2007; Primeau, 2003).

With regard to the reflexivity, it is in the form of a constant account of the researcher's self-critique and self-appraisal to reveal what is occurring (Koch & Harrington, 1998). During the entire process of the narrative inquiry, the researcher kept a reflexive journal to write down the immediate thoughts about the participants' overall observation of their appearance, behavior, and narrative styles as soon as possible after the completion of each interview. Deliberately examining my experience with the participants, the researcher paid special attention to my emotions, thinking, and writings about the participants. The researcher reflected on how to get access to the participants, how to interact, and how to collect, analyze, and interpret data in the research, which would orient the direction of the research. The reflexive journal was incorporated with the narrative account of the interviewees by bringing the interviewer's reflexion into the analysis (Goldstein, 2017; Koch & Harrington, 1998).

3.9 Ethical Considerations

Protecting human rights through applying ethical guidance is of great importance in all kinds of research (Arifin, 2018). In the qualitative research, the ethical consideration becomes salient owing to the in-depth exploration and highly personal materials usually elicited during the research process (Josselson, 2010). In narrative inquiry, the ethical considerations need to be attended to throughout the complete

research process (Clandinin & Connelly, 2000).

The study followed three ethical principles highlighted in the literature. Firstly, informed consent should be obtained from university and participants (Braun & Clarke, 2013; Creswell, 2012). The approval from the university and the participants' consent form were obtained by researcher's detailed explanation of the research purpose and process (see Appendix A & B). The researcher ensured that all of the participants involved in this study had agreed to participate voluntarily and signed written informed consent. The potential participants were accessed individually and provided with a thorough explanation. They were given enough time to ask questions and to address any puzzles or concerns. The participants in the research were also informed of their rights to withdraw anytime during or after the research (Braun & Clarke, 2013). Their refusal to participate or withdraw from the research at any point would not affect their subsequent learning and graduation in any way.

The second principle was to maintain participants' privacy, confidentiality, and anonymity. Participants' anonymity and confidentiality should be preserved by not disclosing their identity in the process of gathering, analyzing, and reporting in the study (Arifin, 2018; Braun & Clarke, 2013; Creswell, 2012; Josselson, 2010). All the participants were referred to by their pseudonym names and no identifying information was revealed in the study.

The last ethical principle was to impose no harm towards the participants (Qu & Dumay, 2011). The researcher should attempt to establish reciprocity with the participants, sometimes in the form of a small reward for their participants (Creswell, 2012). In this study, the researcher prepared some small gifts for the participants involved in the study for their time and cooperation.

3.10 Trustworthiness of the Study

Trustworthiness or rigor of a qualitative study is defined as the degree of confidence in data, interpretation, and methods adopted to ensure the quality of the study (Polit & Beck, 2014). Lincoln and Guba (1985) outlined four criteria of trustworthiness which included credibility, dependability, confirmability, and transferability. To be specific, triangulation, member checking, prolonged engagement, peer review or debriefing, and thick description were used in this study to ensure the trustworthiness of the research.

Triangulation has a long history in qualitative research (Flick, 2018). Triangulation is the process by which more than one method of data collection or sources of data are utilized to explore the issue, aiming to approach the truth from diverse perspectives (Braun & Clarke, 2013; Creswell, 2012). In triangulation, the researcher corroborates evidence from multiple sources to reduce ambiguity, to identify inconsistencies, and to build a coherent rationale for the theme (Creswell, 2007, 2014). By exploring the possible varying interpretations, the researcher attempts to understand the varied

nature of experience rather than to seek the ultimate truth behind the narrative accounts (Murray, 2018). In this study, the findings of the research were triangulated by different sources of data to maintain the rigor of the study. Multiple sources of data were collected to validate the findings, including face-to-face interviews, documents, artifacts, and journals.

Another way to make the research trustworthy is member checking. This technique is considered to be the most critical technique to establish credibility (Lincoln & Guba, 1985). Member checking or member validation refers to check the accuracy of the transcription, interpretation, and explanations with the participants to determine the results' credibility and dependability from the perspective of participants (Creswell, 2007; Miles et al., 2014; Moen, 2006). It typically involves giving the participants the written draft of the findings and asking them to comment on the accuracy of the account (Braun & Clarke, 2013). Participants may be asked about various aspects of the research, including whether the description is complete, whether the interpretation is fair and representative, and whether the themes are accurate (Creswell, 2007). In this narrative inquiry, the interview transcripts and written drafts of the research findings were shared with the participants for them to check any misunderstandings of their narrative accounts to validate the findings.

The third technique employed to verify the research is prolonged engagement in the field. Prolonged engagement includes establishing trust with participants, learning the

culture, and checking for any possible misinformation or misunderstandings (Creswell, 2007; Moen, 2006). Spending prolonged time in the field enables the researcher to develop an in-depth understanding of the issue in the study. The longer experience that the researcher has with the participants in the settings, the more valid the findings will be (Creswell, 2014). The collection of data for this study lasted for about three months, followed by other several months to analyze the data, conduct member checking, peer-reviewing, and write the draft. During the research process, the researcher kept in contact with participants for clarifying any misunderstandings.

Peer review or debriefing is also an important technique to ensure the quality of the findings. It provides an external examination of the research process (Creswell, 2007). This strategy involves a peer debriefer to review the qualitative study and to ask questions so that another interpretation beyond the researcher is invited to add the validity of the results (Creswell, 2014). The peer in this study was an EFL teacher educator, associate professor in X Normal University. The external voice provided additional perspectives and reviews towards the study, which would add rigor to the study.

Another criterion to ensure trustworthiness is to make a rich and thick description, which allows readers to decide to what extent the study can be transferred (Creswell, 2007). Creswell (2014) suggested that this description might provide readers with great details about the settings and might give the discussion a shared experience. The more

details were described in the settings or themes, the more realistic and richer the results might become. In my study, a thick description of the participants and the setting (e.g., assessment contexts in China, X Normal University, the pre-service EFL programme, the profiles of the participants) were presented from various aspects for readers to have a comprehensive understanding. The details of the data collection and data analysis procedures were described in this chapter for readers to determine the transferability of the research.

3.11 Summary

This chapter describes the methodology. It first presents the rationale of adopting narrative inquiry in qualitative research and explains the suitability to the current study. After a brief introduction of X Normal University and the pre-service EFL teacher education programme where the participants come from, detailed information about the participants selection and data collection methods are presented. The six participants are selected purposefully by the researcher according to maximal sampling method. The sources of collected data include semi-structured interviews, documents, artifacts, and journals. As to the data analysis methods, the specific procedures of narrative analysis and thematic analysis are discussed for interpreting the data.

In order to guarantee the rigor and trustworthiness of this qualitative study, several methods are adopted, including a pilot study, triangulation, member checking,

prolonged engagement, peer debriefing, and thick description. The role of the researcher is also paid attention to during the whole process of research. Another important aspect in narrative inquiry is the ethical considerations. The study follows three ethical principles: obtaining informed consent form from the institute and participants, maintaining participants' privacy, and imposing no harm to the participants.



CHAPTER FOUR

RESEARCH FINDINGS

4.1 Introduction

In this section, the findings of data analysis and interpretations of the results are demonstrated. The data are collected from six pre-service EFL teachers in X Normal University in China by narrative inquiry, which is suitable to elicit their perspectives and experiences of LAL. The findings of the study are presented based on the sequence of research questions for a clear understanding. The research questions proposed in the study are:

- (1) What is the conceptualization of LAL by the participants?
- (2) What is the LAL proficiency level self-evaluated by the participants?
- (3) How do the identified mediating factors interact with LAL evolvement among the participants?
- (4) How does LAL evolve among the participants?

4.2 Findings for Conceptualization of LAL

The first research question is “What is the conceptualization of LAL by the participants?” Concerning the conceptualization of LAL, the interview data from the six participants were analyzed narratively and thematically. The investigation of the response yielded that LAL was composed of eight dimensions with varying importance

owing to the shaping effect of the social cultural context. Meanwhile, each participant's LAL conceptualization profile was unique from others.

The eight dimensions in LAL were listed as followed:

- 1) Disciplinary & cross-disciplinary competence: means knowledge of language, English curriculum, and other related disciplines, as well as a command of English.
- 2) Pedagogical content knowledge: means knowledge of how to teach the curriculum-based content to learners and the competence to sustain learning.
- 3) Personal conception & attribute: means knowledge of how one's own preconceptions, understandings and opinions may inform one's conceptualizations, interpretations, judgments and decisions in assessment, and the individual's personality.
- 4) Assessment principle & ethics: means knowledge of why to assess (i.e., formative, summative) and understanding legal and ethical responsibilities concerning the design, use, storage, and dissemination of assessment.
- 5) Assessment methods & implementation: means knowledge of a wide range of assessment strategies and competence of applying and carrying out them for the target learners.
- 6) Assessment content & criterion: means knowledge of how to assess the learning goals and specific content being learned (academic achievement or affective performance), and knowledge of rationale for grading or rubrics.
- 7) Assessment washback: means knowledge of potential influence of assessment,

whether beneficial or damaging, on teaching and learning.

8) Assessment interpretation & communication: means knowledge of ways of interpreting evidence generated from assessment, and ways of communicating assessment results to stakeholders such as students, parents, managers/administrators, and the general public. (See Table 4.1).

Table 4.1

Eight Dimensions of LAL Emerging from the Categories

Dimension	Category	Source					
		P1	P2	P3	P4	P5	P6
1) Disciplinary & cross-disciplinary competence	English proficiency		√		√	√	√
	Linguistic knowledge				√	√	√
	Other disciplinary knowledge					√	
	English curriculum	√					
2) Pedagogical content knowledge	Educational psychology	√					
	Competence to sustain learning	√					
3) Personal conception & attribute	Patient	√	√				
	Responsible	√					
	Adaptable	√					
	Attentive		√				
	Positive					√	
	Not overwhelmed by exam scores				√		
4) Assessment principle & ethics	Assessment purpose & principle	√			√	√	
	Assessment ethics	√		√			√
5) Assessment methods & implementation	Various & Individualized assessment	√	√		√	√	
	Competence to organize & implement assessment			√			√
6) Assessment content & criterion	Cognitive dimension assessment	√	√	√	√	√	√
	Affective dimension assessment				√	√	
	Assessment criterion						√
7) Assessment washback	Washback on students			√			
	Washback on course			√			
8) Assessment interpretation & communication	Assessment result interpretation		√				
	Feedback communication	√				√	√

4.2.1 Conceptualization of LAL from Participant 1

From Amy's definition, she considered LAL as a comprehensive term covering all the assessment-related tasks conducted on four skills: listening, speaking, reading, and writing in English. She also provided a list of constituents: familiarity with English curriculum, competence of sustaining English learning, knowledge of Educational Psychology, character of teachers (patient, responsible, and adaptable), knowledge of assessment purpose & ethics, mastery of individualized assessment methods, and competence of communicating feedback to students. Among the list, knowledge of Educational Psychology was given the greatest priority as teachers were believed to know students and know how they learn first, which could be seen from the excerpt of the interview data.

[EFL teachers] should be familiar with English curriculum they are teaching, for example if they teach in primary school, they must know the entire framework from grade three to grade six...they should learn Educational Psychology, the most important among all the components, to better understand the mechanism of student learning and then they can design individualized assessment for student benefit as well as for students learning...they are supposed to be patient and responsible during the assessment process, adaptable to students' response...besides, communication of feedback is also important to convince student into believing their assessment results. (Amy, Interview Round 1)

Another dimension is to encourage students to sustain their passion and enthusiasm for English learning. (Amy, Interview Round 3)

For a more vivid presentation, Amy's conceptualization of LAL was depicted in the spider diagram, with the most important dimension labelled with 2 and other dimensions labelled with 1 to demonstrate the scaled importance while 0 stood for

dimensions without any reference (See Figure 4.1). Based on Amy's perspective, seven dimensions were involved in total, among which PCK was more critical than others whereas assessment washback was not explicitly mentioned.

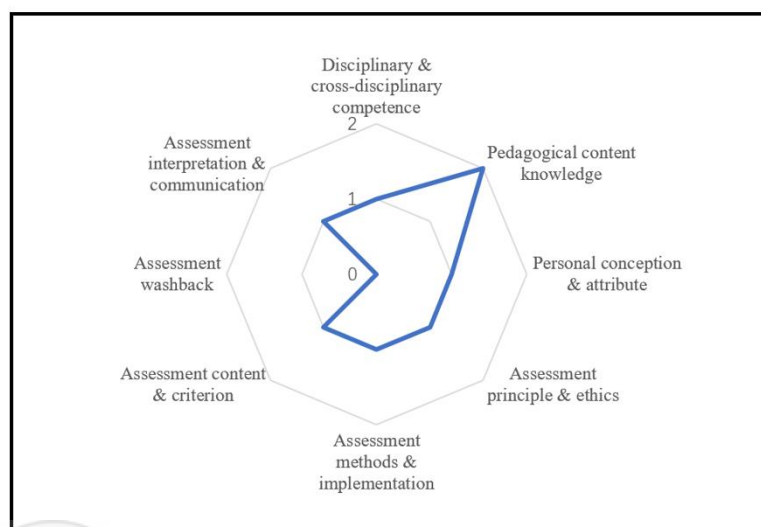


Figure 4.1 Conceptualization of LAL from P1 (Amy)

Amy lay great emphasis on Educational Psychology, which falls into the dimension of PCK. She believed Educational Psychology knowledge was even more important than disciplinary knowledge for EFL teachers because the teachers were supposed to know students well and the mechanism of their learning first. Then, such knowledge enabled the teachers to set personalized learning or assessment plans for students to facilitate learning. Her priority of this may derived from the great merits she obtained after learning the course Educational Psychology in university, which she strongly recommended to be taught at middle school for students to possess a better mastery of the learning methods based on the learning theories to enhance learning.

4.2.2 Conceptualization of LAL from Participant 2

In a more focused way, Betty, the second participant, defined LAL as the literacy on assessing language solely from linguistic aspects, such as grammar, pronunciation.

From literal interpretation, I feel that language assessment literacy should focus on [assessing] language. For instance, assessing the students' English competence from the accuracy of grammar, sentences or pronunciation, just within the scope of language. (Betty, Interview Round 1)

She also depicted the profile of a highly literate teacher in language assessment: equipped with high English proficiency, patient with assessing a large number of students, attentive to select a proper assessment method for each student's individualized needs, and able to interpret the assessment results for subsequent analysis, as revealed in the interview excerpt.

At first, the EFL teacher is obliged to be proficient in English, or else, she/he can't figure out the mistakes made by the students. Secondly, the teacher should be patient to conduct assessment for so many students in the class. Also, the teacher is attentive to each student's needs to select a proper assessment for them...being attentive is more important and is the basis. If teachers are careless to students' needs, it [assessment] is still fruitless even though the teacher is so competent...at last, the ability of summarizing and interpreting assessment results is so important to diagnose students' language deficit. (Betty, Interview Round 1)

They should know how to conduct exams or quizzes weekly, monthly, or at the end of the term to test whether students have mastered the knowledge. (Betty, Interview Round 3)

Compared with the rest of the four dimensions, personal conception & attribute (being attentive) was of greater importance, serving as the basis or pre-requisite in LAL from her point of view (See Figure 4.2). Three were not touched on: PCK, assessment principle & ethics, and assessment washback.

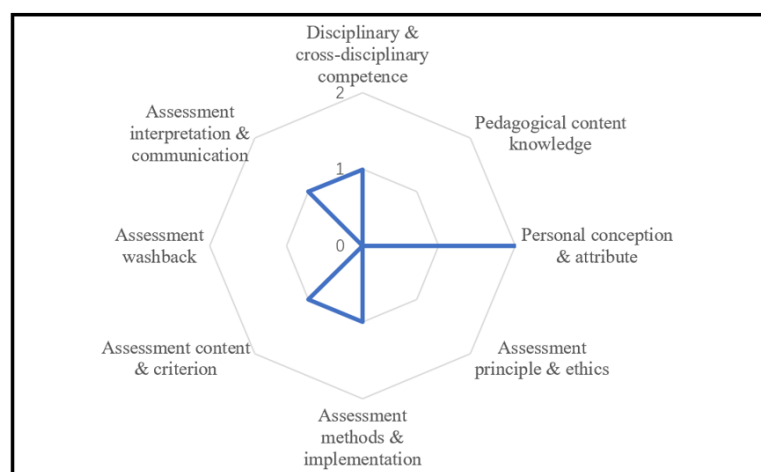


Figure 4.2 Conceptualization of LAL from P2 (Betty)

Unlike Amy who labeled knowledge with utmost importance, Betty considered the teacher's personal attribute to be of overwhelming significance. She laid great emphasis on the teacher's quality of being patient and attentive for they needed to be patient to deal with assessing so many students in a class in China's context and be attentive to all the details in assessment implementation and results interpretation. These professional qualities served as the pre-requisite for all the other components of LAL.

4.2.3 Conceptualization of LAL from Participant 3

Largely the same with Amy and Betty, Carol outlined LAL as the quality or competence to assess students' overall language proficiency. A qualified EFL teacher should know how to assess soundly rather than conducting assessment merely based on experience or instinct. Additionally, teachers should be aware of the potential effect of assessment imposed on students and courses. All these proposed components of

LAL were equally important for EFL teachers just like “Cask effect explanation”.

EFL teachers should know the possible effect of assessment on students and courses. If they are totally ignorant of the assessment-related knowledge and skills, they may either harm students or implement an unsound assessment unconsciously. Or maybe these behaviors are encouraged without any awareness...I think they [the components] are equally important without any discrimination, just like “cask effect”. All of these should be advanced side by side. (Carol, Interview Round 1)

From the description, it could be seen that Carol laid great emphasis on the assessment implementation, washback, ethics, and content, with no reference to other dimensions.

(See Figure 4.3)

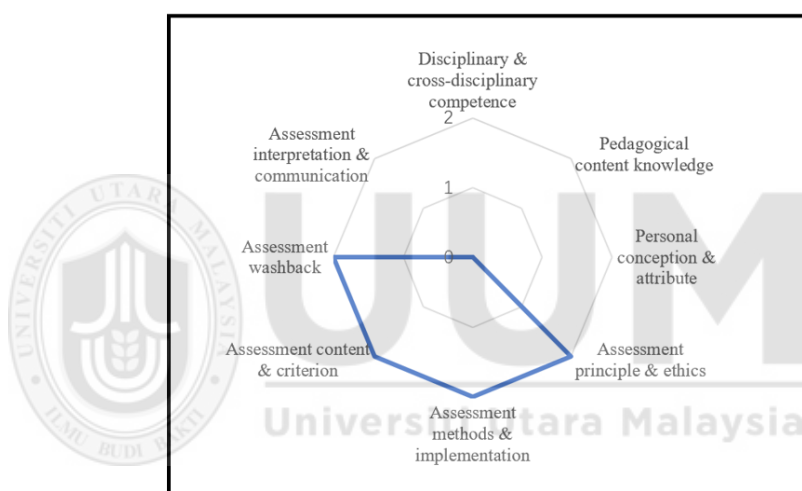


Figure 4.3 Conceptualization of LAL from P3 (Carol)

Carol was the only participant who noticed the assessment washback, that means, the possible impact on students imposed by the assessment, either positive or negative. This explicit notice may be derived from a shift in her conception in an occasional interaction with her classmate who was praised by teacher with “Excellent”. Her classmate told her in excited tone about the praise received from teacher. She suddenly realized that teachers’ assessment was taken seriously by her classmate. From then on, she did not take assessment as merely a formality but a penetration into students’ inner

heart. This incident impressed her a lot about the influence of assessment on students. That is why she stressed the washback effect in the interview with the vivid story.

4.2.4 Conceptualization of LAL from Participant 4

Daisy, the fourth participant, conceptualized LAL more from the benefit of students as “capability of knowing how to assess students accurately and beneficially to students’ English learning” (Daisy, Interview Round 1). The EFL teacher who possessed LAL should take the entire learning process and learning attitude into account rather than relying on the learning results (e.g., exam scores) exclusively. Also, the teacher should be competent in linguistic knowledge and skills apart from the ability of assessing students’ four language skills. Proficiency in English language should not lag behind. All these sub-categories were of equal importance but they were not in conflict.

My ideal EFL teacher should not be totally overwhelmed by exam scores, although in reality, this is the most common practice. Students should be assessed from the whole learning process and learning attitudes...teachers should design tasks in assessing students’ four skills to examine where the deficiency is. Moreover, the teachers’ language literacy, that is, linguistic knowledge and skills is also a must... I think they are equally important, but that does not mean they are conflictory with each other. All of them should be developed without a shortage in one aspect, only in this way, the one can become a professional EFL teacher. (Daisy, Interview Round 1)

EFL teachers are suggested to notice the assessment towards students’ emotions. That means to notice the subtle changes in their emotions. (Daisy, Interview Round 3)

Daisy mentioned five dimensions of LAL: disciplinary competence, personal conception, assessment principle, assessment methods, and assessment content. All

these elements mentioned were labelled with the same importance weight as displayed in Figure 4.4.

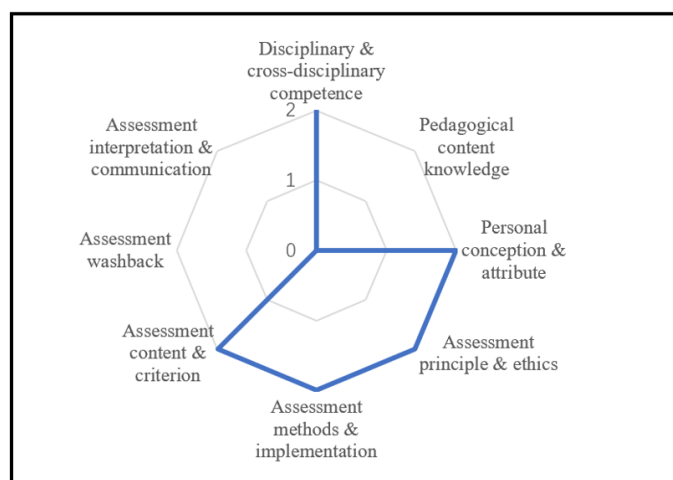


Figure 4.4 Conceptualization of LAL from P4 (Daisy)

There are two impressive highlights in Daisy's conceptualization. One is she emphasized the assessment literate teacher should not solely rely on examination score to assess students. Instead, she advocated multiple assessment methods to obtain a more comprehensive picture of the whole learning process. The other one is she was the first participant who stressed the affective dimension in assessment content, which means not only the academic performance, but also the students' attitudes and emotions should be incorporated into the assessment domain. In other words, she extended the coverage of the assessment domain to include the affective dimension. These two highlights may be closely related to her antipathy towards exam-oriented education. From her school-based assessment experience, she frequently complaint exam-oriented tendency in English teaching modes, which narrowed down the learning content to what might be covered in the high-stake exams and the serious

anxiety experienced during the exams. This unsatisfying experience seems to motivate her into thinking about the expected or ideal profile of assessment practices implemented by EFL teachers.

4.2.5 Conceptualization of LAL from Participant 5

The fifth participant, Edwin, a male pre-service EFL teacher, was applying for TESOL (Teach English for Speakers of Other Languages) in England. He impressed the researcher with more technical terms in assessment when answering the questions in the interview, like diagnostic assessment, criterion-referenced exam. He conceptualized LAL from the following aspects: knowledge of assessment principle, skills, and methods; the inclusion of affective dimensions in assessment content; attitude towards assessment; way of feedback; disciplinary literacy and broad knowledge base in other related disciplines.

To my knowledge, I think you [the EFL teacher] must know some assessment methods, like group assessment, diagnostic assessment, formative assessment. Also, you should master some assessment principles, such as whether the assessment involves all students, whether it assesses various aspects of the students, including the learning attitude towards English whether the students are positive or not on English learning...then, you should also hold a positive attitude to assessment...At last, your own professional literacy should achieve a qualified level, with proficient English level and broad knowledge base not only in the discipline, but also in the related area to broaden students' horizon and arouse their learning interests. (Edwin, Interview Round 1)

You [EFL teachers] need to pay attention to the way of feedback, like oral or written feedback, direct or indirect feedback, and the timing of feedback, when to correct the mistakes made by students. (Edwin, Interview Round 3)

When asked about the sequence of these elements based on their significance, he

provided a rather comprehensive opinion, “Ideally, these elements should be of the same significance, however, in reality, it is not the case” (Edwin, Interview Round 1). He was the first participant who had pointed out the context-sensitive nature of LAL. From his observation in school, he attached greater importance to assessment methods, as evidenced in the interview transcripts.

From my own limited experience in the school, I think the assessment methods are important than others because you [the EFL teacher] may encounter various students with different characters; if you assess them with a unified method, the results may not be beneficial to students’ learning. For the difference among students, you need to adopt a proper assessment method, which is more critical than others. (Edwin, Interview Round1)

As shown in Figure 4.5, his conceptualization covered six dimensions among which assessment methods & implementation was of utmost importance based on his own experience. All the dimensions identified by him were exposed to the social cultural context represented by the shadow in the diagram.

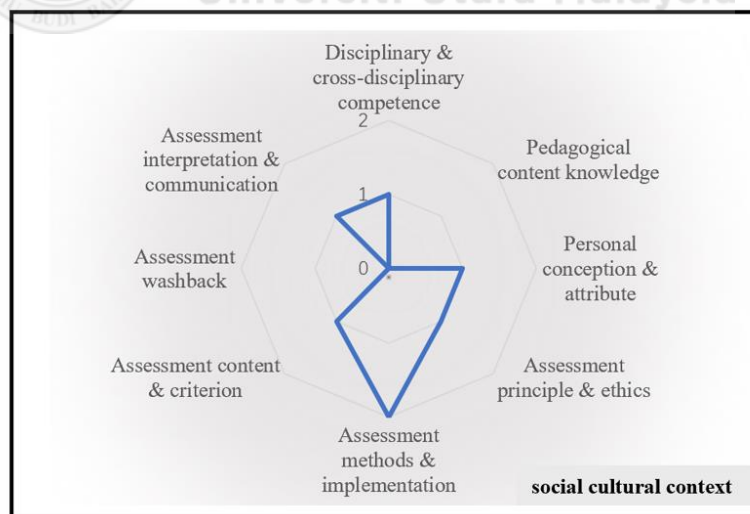


Figure 4.5 Conceptualization of LAL from P5 (Edwin)

In Edwin’s conceptualization of LAL, three salient features were obvious: the coverage of affective dimension in assessment content, the explicit notice of the context-

dependent nature of LAL, and the very emphasis on assessment methods. These features may be explained by his assessment experiences. His inclusion of learning attitude into the assessment domain was likely to be the result of his schooling experience. His academic performance in primary school was rather unstable, either the second or the second to last. With his efforts, he ranked in the middle of the class in middle school. So, he strongly suggested that not only the learning outcome but also the learning attitude should be a part of the assessment content.

Besides, from his practicum experience, his implementation of portfolio assessment did not proceed smoothly as a result of the little cooperation of some students. Thus, he experienced the constraints of local context on the assessment practices, which reinforced his awareness of the context-sensitive nature of LAL. At last, his emphasis on assessment methods may come from his accumulation of assessment knowledge in the preparation for TESOL, when he became increasingly familiar with assessment terminology, such as group assessment, formative assessment, summative assessment, and norm-referenced assessment.

4.2.6 Conceptualization of LAL from Participant 6

Likewise, Flora, perceived LAL as an integrated term encompassing everything related to language assessment. She also identified five dimensions of LAL with equal importance: English proficiency & linguistic competence, assessment implementation, assessment ethics, assessment criterion, and feedback communication.

At first, I feel like linguistic knowledge and related skills must be necessary and proficiency in English is also essential for the EFL teacher should be knowledgeable to teach...secondly, the teacher should be skilled in organizing and implementing assessment activities... at last, students' psychological state should also be given enough attention...meanwhile, the teacher is advised to deliver the assessment results or feedback in an appropriate way for students to better absorb. That is a big challenge for teacher to provide feedback in a more acceptable way for students. (Flora, Interview Round 1)

The EFL teacher is supposed to know the assessment criterion: how to distinguish the student with better academic achievement. (Flora, Interview Round 3)

Flora also emphasized LAL's context-dependent nature by explaining the shaping effect of high-stake examinations thrust on LAL structure. Ideally, an EFL teacher should be competent in assessing the four linguistic skills, nonetheless, the teacher's LAL on assessing oral English was deficient to a large extent in the real society due to the absence of testing spoken English in most large-scale exams in China. Thus, LAL's construct was largely shaped by the construct and the corresponding weight of influential exams.

To be frank, I think they [assessing the four skills] are equally important, but in the society, it is totally different... the skills like [assessing] reading, listening, writing, are more important than oral English, which is usually absent in the large-scale examination. (Flora, Interview Round 1)

As demonstrated in Figure 4.6, the background of LAL was shadowed to signify the shaping effect of the social cultural context. The five dimensions figured out by Flora were assigned with equivalent importance.

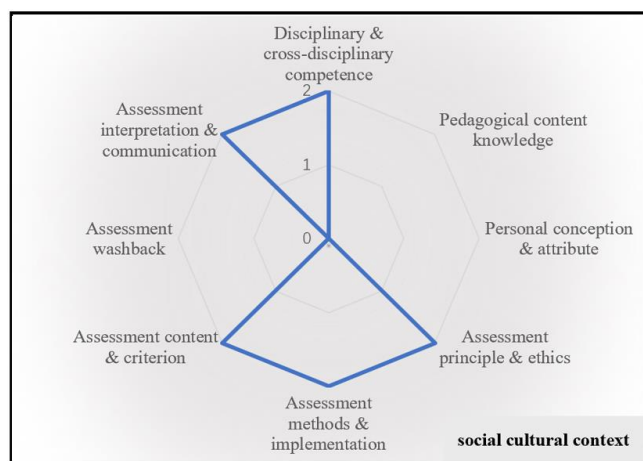


Figure 4.6 Conceptualization of LAL from P6 (Flora)

Flora's notice of the context-dependence of LAL possibly resulted from her practicum experience, in which the inadequate proficiency in oral English among the students in practicum school attracted her attention and she went further to select this as the topic of practicum investigation report. In the report, she attributed the scant attention on oral English primarily to its absence in high-stake examinations which in turn decreased its due importance in EFL teaching and learning. During the research process, she might come to realize teachers' assessment priority in real practices was dramatically shaped by the social cultural context.

4.2.7 Overall conceptualization of LAL

All of the six participants seemed to reach a consensus that LAL was a quality or competence of EFL teachers to assess students' English proficiency to facilitate learning by adopting various methods, though varying emphasis was laid on different dimensions in their references. The overall distribution summarised from the six

participants among the eight dimensions of LAL was displayed in Table 4.2 with numbers meaning different importance ranging from 2 for the most important, 1 for important, and 0 for not being mentioned in their perspectives.

Table 4.2

The Overall Distribution among the Eight Dimensions of LAL

Dimension	Scores on importance						Mean
	P1	P2	P3	P4	P5	P6	
Disciplinary & cross-disciplinary competence	1	1	0	2	1	2	1.67
Pedagogical content knowledge	2	0	0	0	0	0	0.33
Personal conception & attribute	1	2	0	2	1	0	1
Assessment principle & ethics	1	0	2	2	1	2	1.33
Assessment methods & implementation	1	1	2	2	2	2	1.67
Assessment content & criterion	1	1	2	2	1	2	1.5
Assessment washback	0	0	2	0	0	0	0.33
Assessment interpretation & communication	1	1	0	0	1	2	0.83

From Table 4.2, the eight dimensions are different in frequency of being mentioned by the participants. Pedagogical content knowledge and assessment washback are least frequently mentioned except by only one participant. It means the six participants least focus on these two dimensions. By contrast, another two dimensions are covered by six participants though with varying importance: assessment methods & implementation and assessment content & criterion, which are considered by all the participants to be the very core component of LAL.

For a more vivid presentation, the overall conceptualization of LAL from the six participants was summarised in Figure 4.7. As shown in Figure 4.7, the synthesized conceptualization of LAL from the six pre-service EFL teachers in China was characterized by four salient features: multi-dimension, graded importance, context-dependence and uniqueness.

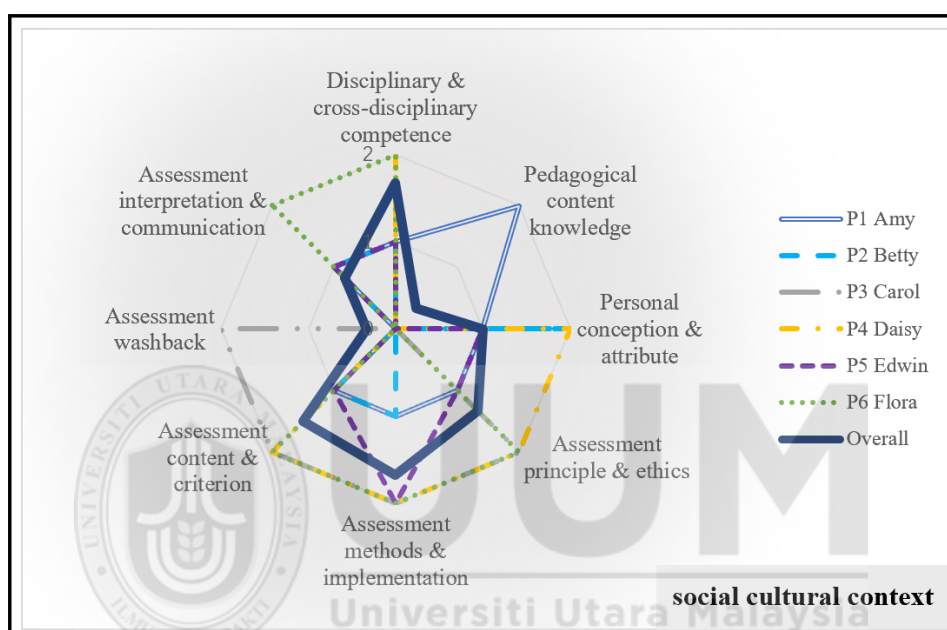


Figure 4.7 The combined overall conceptualization of LAL

(1) LAL was perceived as a multi-layered concept, encompassing eight dimensions, which could be roughly classified into three broad domains. Discipline-related competence (disciplinary & cross disciplinary competence, pedagogical content knowledge), person-related competence (personal conception & attribute), and assessment-related competence (assessment principle & ethics, methods & implementation, content & criterion, washback, interpretation & communication).

(2). LAL's constructs were not equally important. The scaled importance from the most

important to the least was: Disciplinary & cross-disciplinary competence and Assessment methods & implementation > Assessment content & criterion > Assessment principle & ethics > Personal conception & attribute > Assessment interpretation & communication > PCK and Assessment washback. The participants in the present study attached considerable importance to disciplinary competence and practical implementation of assessment methods while they tended to give scarce attention to PCK and the potential influence of assessment.

(3) LAL was dependent on the social cultural context. The constructs of LAL were shaped by the context where LAL was embedded. They were oriented explicitly or implicitly by the social cultural context, including the macro exam culture and the micro teaching context, which were highlighted by two of the participants (See Edwin and Flora). Thus, LAL was sensitive to the local context and was printed with the distinctive feature of the context.

(4) Each participant's spider diagram of LAL conceptualization was unique and different from others. None of them was identical.

In summary, the pre-service EFL teachers in China conceptualized LAL as a multi-layered concept in which eight dimensions were graded with varying importance due to the shaping function of the social cultural context where LAL was situated. Besides, the individual conceptualization of LAL from each participant was idiosyncratic and

unique from others. There were not two identical conceptualizations of LAL in the study.

4.3 Findings for Self-evaluated LAL Proficiency

The second research question is “What is the LAL proficiency level self-evaluated by the participants?”. The interview data showed a diversification in their self-evaluated proficiency and self-diagnosed improvement focus. Their self-evaluated proficiency was diversified into three dimensions: insufficiently qualified, marginally qualified, and satisfactorily qualified. Meanwhile, they expressed diverse improvement focus in LAL.

4.3.1 Self-evaluated LAL Proficiency

There was a wide range of LAL proficiency scores self-evaluated by the six participants, four of whom scored themselves as insufficiently qualified and the other two perceived themselves as qualified to different extent for various reasons. The scale ranged from less than 60 to 89 against the full mark of 100 (See Table 4.3).

Table 4.3

The LAL Proficiency Self-evaluated by the Participants

Participants	Self-evaluated proficiency scores	Classification
P1 (Amy)	89	Satisfactorily qualified
P2 (Betty)	65	Marginally qualified
P3(Carol), P4 (Daisy), P5 (Edwin), P6 (Flora)	Below 60	Insufficiently qualified

It was generally believed among the participants that they were still inadequately prepared although they already had some exposure to assessment learning and practices. Four participants (Carol, Daisy, Edwin, and Flora) classified themselves as insufficiently qualified for the forthcoming EFL teachers' assessment jobs due to various reasons: low willingness to join in the teaching profession (Carol), less satisfactory assessment preparation in teacher education programme (Daisy, Edwin, and Flora), and personal late start in learning (Edwin).

I evaluate myself to be inadequately prepared in LAL. Because I don't desire to become a teacher at the bottom of my heart, thus, I lack the autonomy to learn more [about the assessment] to approach the standard of teaching qualification. (Carol, Interview Round 4)

I feel I may get a failing grade. Firstly, I have an insufficient mastery of theoretical knowledge in assessment, just a superficial understanding about how to assess and how to feedback. Secondly, I am not familiar with other aspects of assessment. Only assessing students based on their exam scores, relying more on the results instead of the learning process or other aspects. Thus, I evaluate myself as unqualified. I still need to improve on assessment to enrich myself and to have a more comprehensive understanding, which I can apply to the teaching practices. (Daisy, Interview Round 4)

I think I may fail. I have not achieved...how to say, I have not achieved such a proficient level to stand on the podium. I can explain in two aspects. On the one hand, the curriculum design puzzles me a lot. Frankly speaking, I feel that some courses, such as Integrated English, Reading can be changed into other courses. On the other hand, my personal attitude. I have not realized to study until my third year in college. In the first two years, I did not treat learning very seriously, which leads to my unsolid foundation and less desirably professional competence. I haven't passed TEM-Band 4, of course, which cannot represent my real English proficiency, but I feel that I am not professional enough to teach students. (Edwin, Interview Round 4)

Maybe the score is 50, below the qualified level, because I think what I have learned in university is so limited... What I have substantially obtained is very limited and inadequate. (Flora, Interview Round 4)

Different from the majority, the other two participants, Amy and Betty, evaluated themselves as satisfactorily and marginally qualified respectively. As shown in the scripts, they expressed a comparatively optimistic evaluation towards LAL.

I believe I am capable, too general, if it can be scored, I will assign 89 to me because 90 represents excellence. A point less may keep me modest. (Amy, Interview Round 4)

I can say just so-so and the score is about 65, merely scrape through. I have not got enough exposure [to assessment] and even the course, English Pedagogy, lays greater emphasis on teaching design, covering too little on assessment. From my perspective, the marginal pass is not easy for me. (Betty, Interview Round 4)

The comparatively optimistic evaluation of LAL from Amy and Betty may be attributable to their excellent academic performance in university. They were top students in their respective classes and passed all final examinations with high marks. So, they tended to believe they were proficient in every aspect in pre-service stage, with no exception of LAL. Consequently, they showed great confidence in self-evaluated LAL proficiency.

To summarise, there was a diversification in their self-evaluated LAL proficiency level, ranging from insufficiently qualified through marginally qualified to satisfactorily qualified along the continuum (See Figure 4.8). Most of their self-evaluations fell into the first category as a result of mixed factors, including external (i.e., the limited amount of exposure to assessment) and internal factors (i.e., low willingness to become a teacher or to invest in learning).

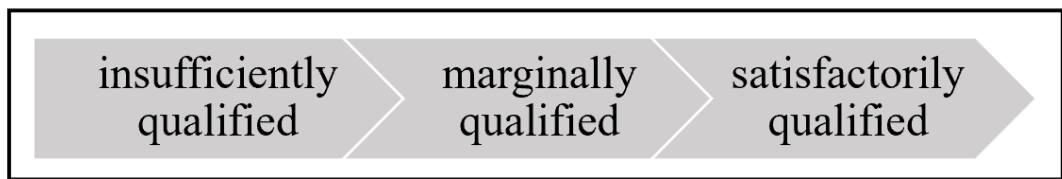


Figure 4.8 Self-evaluated LAL Proficiency Continuum

4.3.2 Self-diagnosed LAL Improvement

There were both similarities and discrepancies in their self-diagnosed LAL improvement area. What was similar was that nearly all of them expressed a strong desire to further improve LAL driven by the motivation to become an excellent and effective EFL teacher both from the interview data and reflective journals (See Table 4.4). In the interview data, the five participants highlighted the significance of LAL in teaching and they desired to become a good and helpful EFL teachers to facilitate teaching and learning. This motivated them to further enhance LAL. Besides, in the reflective journals kept by Flora, who confirmed the desire to become a (good) teacher motivated her to learn more to enrich her toolbox.

However, only one participant, Carol, expressed her low willingness to become a teacher in the future, therefore, comparatively, she was less motivated to improve LAL for professional development.

Table 4.4

The Self-diagnosed LAL Improvement among the Participants

Participant	LAL improvement focus	LAL improvement desire
P1 (Amy)	Way of feedback Humanistic care in assessment	To become a (good) teacher
P2 (Betty)	Assessment theory Assessment methods and implementation	To facilitate teaching and learning
P3 (Carol)	Assessment framework Assessment implementation assessment washback	Little desire to learn because of no desire to become a teacher
P4 (Daisy)	Way of feedback Theoretical knowledge	To facilitate teaching and learning
P5 (Edwin)	The dimensions of assessment, especially in reading and listening	The alignment with postgraduate specialization The importance of assessment in teaching and learning
P6 (Flora)	Assessment methods The result and washback of assessment	To become a good and helpful teacher

Yet, what was discrepant was their divergent LAL improvement focus. Four of them mentioned they wanted to learn more about feedback and washback of assessment (Amy, Carol, Daisy, and Flora); three needed to improve in theoretical learning in assessment (Betty, Carol, and Daisy); three wished to be more familiar with assessment methods & implementation (Betty, Carol, and Flora); humanistic care in assessment (Amy) and assessment dimensions (Edwin) were referred to only once by participants. Therefore, a wide range of aspects in assessment were underscored to be improved with the assessment theory, methods, implementation, results and washback mostly noted by the participants.

It was worth to point out that Carol was the only participant who expressed the

possibility of becoming a teacher is fifty. She desired to be a translator and applied for postgraduate study in translation too. So, it is no wonder she expressed a low willingness to improve LAL initiatives because becoming a teacher is not her first choice in the future career. But she still diagnosed her LAL and expressed improvement aspects in LAL.

To summarise, the self-diagnosed LAL improvement was identified to be focused on diverse aspects of assessment (from the most frequently mentioned to the least): feedback & washback, theoretical learning, assessment methods & implementation, humanistic care in assessment, and assessment dimensions. But they seemed to be driven by the identical power to become an effective and helpful teacher to facilitate teaching and learning in the future teaching profession.

4.4 Findings for Interaction between Mediating Factors and LAL Evolvment

Research question three is “How do the identified mediating factors interact with LAL evolvment among the participants?”. The third research question aims to find out the interaction between identified mediating factors and LAL evolvment. The first subsection presents the identified mediating factors in three categories: experiential, contextual, and personal. Then followed by the analysis of interaction between them.

4.4.1 The Identified Mediating Factors in LAL Evolvement

The identified factors mediating pre-service EFL teachers' LAL evolvement are classified into three dimensions: experiential, contextual, and personal. As demonstrated in Table 4.5, the experiential dimension includes positive or negative assessment experience, either being experienced as assessor, assessee, or assessor-to-be. The second contextual dimension contains three levels from macro through meso to micro level. At last, personal factors also play a role in mediating LAL evolvement.

With regard to the teacher preparation programme, it seems to play a mixed role in mediating LAL evolvement, participants' attendance of assessment courses or involvement in assessment practices are more likely to be the experiential factors while the assessment-related policy and the high-stake examination are better to be classified into contextual factors. Thus, it may be placed in an overlapped area between experiential and contextual factors.

A small part of the assessment-related experiences in programme is analyzed within the educational continuum from school to university in the experiential dimension while a systematic discussion of assessment-related policy, courses, practices, and high-stake examination are presented together for a holistic and in-depth exploration of the role of pre-service teacher preparation programme in LAL evolvement in order to avoid the scattered discussion.

Table 4.5

The Dimensions of Mediating Factors

Three dimensions	Sub-category		Source
Experiential	Positive assessment experience		P1- P6
	Positive interpretation of negative assessment experience		P2 & P3
	Negative assessment experience		P1- P6
	Negative interpretation of positive assessment experience		P3
Contextual	Macro-level	Social assessment culture	P4
	Meso-level	School policy & culture	P2
		Educational section	P2
	Micro-level	Identity of head teacher	P2
		Students	P1, P2, P3, P5
		Parents of students	P2
Experiential & contextual	Teacher preparation programme		P1-P6
Personal	Self-centeredness		P1
	Characteristics		P5
	Mental age & horizon		P5

4.4.1.1 Experiential Factors

Assessment experience was identified as an influential factor in mediating the participants' LAL involvement, whether being experienced as a student in the schooling time or teacher-to-be in the university, whether in the subject of English or other subjects, whether the experience is positive or negative. They acknowledged the explicit or implicit reference to the previous assessment experience when becoming an assessor in the future, such as Carol, Daisy, Edwin, and Flora.

I may treat my students in the way as I was treated by my teachers (Carol, Interview Round 2).

Teachers' performance in the class may exercise a subtle and formative

influence on my adoption of the procedures in assessment. (Daisy, Interview Round 4)

When facing the challenge of conducting a satisfying lecturing, my mind is full of the voices of my former teachers (Edwin, Interview Round 2) ...they come into my mind very naturally. (Edwin, Interview Round 4)

The assessment activities adopted by my teachers will more or less impact on my future decisions in assessment. I will follow their good practices and improve not so good practices in assessment to enhance my teaching. (Flora, Interview Round 3)

Assessment experience seemed to deeply influence the assessment preference or willingness to transfer in the future assessment practices. The positive assessment experience and positive interpretation of negative assessment experience tended to be positively transferred (+) in assessment preference while the negative assessment experience and negative interpretation of positive assessment experience usually negatively influenced (-) assessment preference in LAL involvement (See Figure 4.9). Thus, the willingness to transfer appeared to depend on whether the interpretation of the assessment experience is positive or not, instead of the assessment experience itself is positive or not.

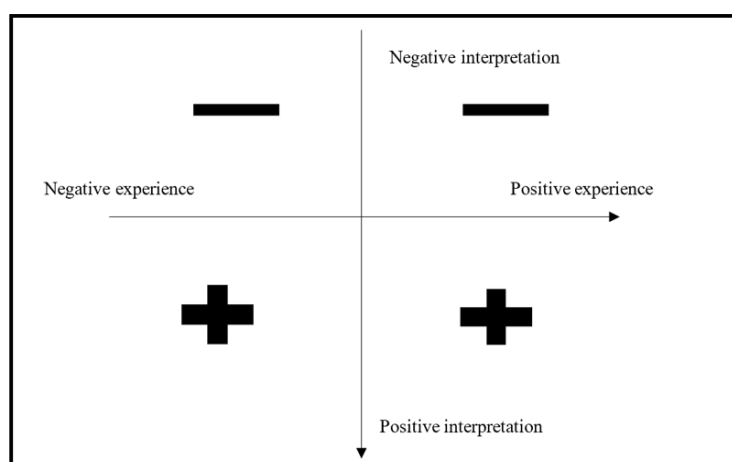


Figure 4.9 Assessment experience transfer conditions

Notes: “+” means positive transfer; “-” means negative transfer.

4.4.1.1.1 Positive Assessment Experience

Positive assessment experience tended to be positively transferred. All of the participants agreed that the positive assessment experience would be more likely to have a positive impact on their assessment decisions in future teaching career, especially in assessment tool, assessment methods, feedback, and assessment results application.

(1) Assessment tools. The participants who mentioned their favored assessment tools with pleased experience showed greater preference in utilizing it when becoming an EFL teacher. For instance, Carol, Daisy, and Edwin would advise their students to use online writing automatic feedback systems or English vocabulary assessment application programmes due to their recognition of the assessment tools introduced in their being assessed experience in teacher preparation programme.

[Assessee in university] Usually, I use more than one online writing automatic feedback platforms except pigai.com. Some useful alternative words or phrases will be given in Pigai.com, but I think the alternatives are a little mechanic. So, I will use youdao.com and another software, which also has the function of instant feedback to make comparison of their feedback of the same composition...I will recommend to my students because sometimes instant feedback is impossible if teachers are not available around you. Besides, it is difficult to find out your own mistakes by yourself alone. Although it is a little mechanic, it provides the alternative vocabulary and the suggested sentence patterns, which are quite beneficial. (Carol, Interview Round 1)

[Assessee in university] I have used the APPs for memorizing vocabularies, by the way to test my vocabulary for a reference. Then I would know well about my [English] level in my mind...I will suggest my students using these APPs to facilitate English learning, since the pronunciation and phonetic symbols are

marked in the applications, which I feel are suitable for students. (Daisy, Interview Round 1)

[Assessee in university] I have used European Dictionary APP, which I bought lifetime membership for 80 years at the cost of 198 RMB. I like it very much for its rich resources and English explanation, which is much better than Chinese explanation...I will advise students to use when I become a teacher for its advantages I mentioned just now. (Edwin, Interview Round 1)

(2) Assessment methods. Methods recognized by the participants when they were engaged in the assessment activities would be likely to be transferred in their future decision-making. The assessment methods like performance assessment, peer assessment, and parent-involved assessment were mentioned by the participants when they were in their schooling time, after-school training, or university period as an assessee or assessor. For instance, Amy, Edwin, and Flora favored performance assessment due to the pleasant experience during the after-school training or teacher education programme. Similarly, Carol and Flora preferred peer assessment experience obtained in elementary school or university. Additionally, Flora valued parent-involved assessment highly.

[Assessee in university] I prefer the practice implemented by my teacher who is in charge of Japanese, Ms. Li. She involves us in various interesting activities, like how to make Matcha and dress Kimono. She assigns the tasks in advance to give us enough time to prepare and collect information. She prepares the materials needed in making Matcha and then we make on the stage by introducing at the same time. After all the presentations, she provides her assessment and feedback. Similar procedures in making Sushi, I like this assessment very much because it involves us through the whole process and I feel so relaxed, thus the motivation is aroused ... I will follow Ms. Li [in designing assessment activities]. I will put forward an interesting topic, then advise them to search on the internet before class. In the class, they will be asked to show and present. (Amy, Interview Round 3)

[Assessee in after-school training] What impressed me most is my after-school

training during my elementary school period. In that institution, I was involved in diverse games or activities like building trains or Words Solitaire, which has exerted great influence on my present teaching. Sometimes, I referred to these activities during my practicum and I felt they were quite effective. (Edwin, Interview Round 2)

[Assessee in university] I remember that in the first or second year of the university, Ms. Yang, who lectured in Intensive English, assigned homework to play a drama based on the content from the textbook. We played in the classroom and the atmosphere was so funny. I felt so interested for this assessment activity, which had three advantages: deepen the feelings among the students, strengthen the understanding of the text, and add the amusing interaction. The audience applauded enthusiastically and spontaneously. (Flora, Interview Round 3)

From the above description of performance assessment in English learning, including games, presentation, and drama, the participants explicitly asserted that they would adopt these in their future assessment for their joyful feelings and effectiveness in assessing language. Additionally, the peer assessment was also positively transferred for the same reason, although some experience was gained from other disciplines (e.g., mathematics).

[Assessee in elementary school] In mathematics, the teacher asked us to solve the question in a group by discussion and then to exchange cross groups until the answer was found out. During the process, we would present our thinking and assess others... I think I will add this part in my teaching because sometimes discussion and assessment among students was necessary and more important. It provided us with opportunities to express our ideas and utter our own voices. (Flora, Interview Round 2)

[Assessor or assessee in university] In the course of English Pedagogy, when the student finished trial lecture, she/he would be assessed by the a few appointed audiences. When the appointed audience shared the similar comment with me, I would feel that I was recognized... another example was that after my trial lecture, other students also pointed out the shortcomings, such as the design was not complete; the assignment was not dealt with correctly. I felt their comment was pertinent...thus, I felt the peer assessment was quite useful and I would utilize it in the future. (Carol, Interview Round 3)

At last, so was the parent-involved assessment. Flora expressed her favor of this assessment method both in reflective journals and interview for its advocate of cooperation between parents and schools.

[Assessee in middle school] I might imitate the assessment methods adopted by my English teacher in middle school. For example, I would plan to involve parents in checking the homework like listening to the tape and vocabulary dictation. After the checking, the parent needed to sign their names on the checklist like a calendar. I like this method because cooperation between school and parents plays a monitoring role and parents should know what the kids have learned at school. (Flora, Interview Round 2)

Every day, the junior English teacher assigned the tasks, like listening to the tape and asked parents to check or monitor, which not only informs parents about their kids learning but also enhances students' learning autonomy. (Flora, Reflective Journal)

(3) Assessment feedback. The recognized way of assessment interpretation and feedback by the participant would be more likely to be imitated. Most of them felt encouraged by the positive praise from the teachers or rewards given for the progress in academic achievement. Such experiences were gained either from assessee in their schooling or after-school time (e.g., Amy, Betty, Edwin, and Flora) or from the practicum observation as assessor-to-be (e.g., Amy).

[Assessee in primary school] In primary school, I remembered that in an English composition, I used the word “wonderful”, which was praised by English teacher that the word was excellent. I felt so happy. (Amy, Interview Round 2)

In the future, I would certainly encourage my students and provide more positive feedback. (Amy, Interview Round 3)

I felt it [positive feedback] was necessary and might mean that the teacher paid close attention to me if I were the student. (Amy, Interview Round 4)

[Assessee in middle school] In middle school, I liked my Chinese teacher so much. She often provided positive feedback to me maybe because I was high-achiever in the class. She commented like you performed well recently, which

made me so happy for the recognition from the teacher. Since then, I was determined to study hard and treat the course seriously. I liked this teacher very much. We still kept in touch. She always encouraged me to follow in her footsteps to become a teacher. (Flora, Interview Round 2)

[Assessee in after-school training] What impressed me most was not the school teachers, but English teachers from out-of-school training institution. She meant a lot to me. She always encouraged me by saying you would be no problem in the future. (Edwin, Interview Round 2)

In the future, I would feedback positive comments first then followed by negative comments if any...I don't know why I chose this way to feedback maybe because my prior assessment experience, which came into my mind naturally. (Edwin, Interview Round 4)

[Assessee in after-school training] To maintain the customers, the after-school training institution during my primary schooling time usually rewarded us, including certificate of merit, or other small prizes. Such encouraging measures were highlighted, which left a deep impression on me...I felt if I were to teach primary school, I would reward my students for their academic improvement to some extent. (Betty, Interview Round 2)

[Assessor-to-be in practicum] During the teaching practicum, I found a very good practice for praise after student's correct answer. If they answered correctly, the students would be rewarded by a hand gesture made among the whole class, like lighting fireworks. They treated this hand gesture as a reward... It was very spectacular when the whole class made the same gesture. I would consider to use it when I became a teacher. (Amy, Interview Round 4)

(4) Assessment results application. The assessment results, usually the exam results tended to be applied to class management in China. The practice of arranging the seat based on the principle of slow learners learning from high-achievers was favored by one participant, Edwin, who was a slow learner in junior middle school, benefited from the seat arrangement by a desk mate good at learning English. He commented he learned a lot from the excellent desk mate and developed a good habit of note-taking, which was still kept until nowadays.

[Assessee in middle school] When I become a teacher, I will follow the practice in my junior middle school, where weaker students and better students seated as deskmate to study together. I felt this would be beneficial to learning and weaker students would be driven to learn, like me. I learned a lot from my excellent deskmate, who was one of the top students at that time. I imitated his note-taking style and knew how to learn. Since then, no matter how terrible my study was, my notes were always neat and orderly. Until nowadays, I still kept this good learning habit... therefore, I uphold mutual learning from each other when arranging the seat. (Edwin, Interview Round 2)

4.4.1.1.2 Positive Interpretation of Negative Assessment Experience

Another type of positive transfer was the positive interpretation of negative assessment experience. That is to say, the assessment-related activities experienced as negative at that moment by the participant might be re-interpreted as positive along with the shift of perspective or identity from being an assessee to assessor. In the interview, the way of communicating assessment (Parents Meeting) mentioned by Betty and assessment methods (performance assessment) mentioned by Carol were listed as examples. Initially, Betty and Carol had a comparatively negative experience in Parents Meeting and performance assessment respectively when they were assessed at school or in university. But along with the perspective shift to assessors, they gradually changed the negative interpretation to positive perspective, which contributed to the active utilization in their teaching career in future.

[Assessee in middle school] Previously in my middle school, I disliked the Parents Meeting, which made me nervous and worried. But nowadays, with the growing of ages, I have changed my perspective. I think it is very necessary. Because sometimes the communication through telephone is not so effective to interchange the deficits of the students. A fluent and good communication between parents and teachers is indispensable, especially the Parents Meeting, which gathers all the parents in a meeting. The teacher can convey the common problems among the students, and parents can learn more information about

other students to make a comparison. Certainly, after the meeting, the parents know more about their kid... maybe due to the increase of ages, some opinions are totally opposite to prior ones. Now, I think Parents Meeting is in fact essential. (Betty, Interview Round 2)

[Assessee in university] I remembered in a Reading class; the teacher assigned us a task to shoot a video based on the text. It took us a long time to make just a short video. I felt disgusting in my heart because I had to do some things... and I was not proficient in technology in making videos. I thought it was so time-consuming...but when I become a teacher, I will adopt this assessment method. I will not care about the time cost and I will design the assessment activities more from my stance of being a teacher. (Carol, Interview Round 3)

4.4.1.1.3 Negative Assessment Experience

Negative assessment experience seemed to be less likely transferred. The negative or unrecognized assessment experience which was gained throughout the participants' entire educational stages appeared not to be transferred in their future assessment decision-making. Their negative experience in assessment tools (Amy and Betty), way of feedback (Amy, Daisy, and Flora), and assessment results application (Betty, Carol, Edwin, and Flora) might decrease the willingness of utilizing it to their students when they become teachers for the unpleasant feelings.

(1) Assessment tools. The participants would not recommend the unrecognized assessment tools experienced as being assessed to their future students. As what Amy and Betty mentioned in the interview, they would not advise the assessment tools, like online writing feedback system (Pigai.com) and vocabulary testing APPs for the disapproval of the assessment feedback and results. Amy was suspicious of the validity of the feedback and Betty felt doubtful about the representativeness of the tested

sample vocabulary.

[Assessee in university] I have used the online writing assessment platform, like pigai.com for instant feedback, but I do not approve of its suggestions. Maybe due to my deficiency in English, I do not know whether its suggestions are more native-like or not...I will not advise my students to try these APPs, which I disapprove of. (Amy, Interview Round 1)

[Assessee in university] I have tested my vocabulary by some APPs. Although they are scientific to some extent, I still feel there exists randomness or chances. For example, sometimes I happen to know the words tested in the APP. However, in fact, I do not know enough in that field... thus, it is less possible for me to recommend it to learners. (Betty, Interview Round 1)

(2) Way of feedback. When they experienced negatively or sometimes felt harmed in teachers' feedbacks, the participants intended not to hurt their students in the same way. For instance, Amy and Daisy were discouraged by the improper feedback which hurt their confidence or self-esteem. Flora paid close attention to the appropriateness and acceptance of her feedback to students in the teaching practicum largely due to the disappointing comments received in subject of English and Mathematics in the schooling time. As an assessor-to-be in practicum, Amy observed an improper way of feedback conducted by the teacher to send the videoed misconduct of students to parents' chatting group, which she considered harmful to students and instead she would utilize other ways like home visit to protect students' privacy in communicating the results.

[Assessee in university] The teacher of English Pedagogy always discouraged us by saying that we did not live up to his expectation and we were born to be not so good enough. By the way, he belittled our parents by attributing the misbehavior (e.g., playing Tik Tok during the break) to our parents' addiction to these APPs...so he has severely blown my confidence inadvertently. (Amy, Interview Round 3)

[Assessee in primary school] I remembered that when we were self-studying in the classroom, the teacher was checking the homework on the platform. When the teacher found out some mistakes from the homework, he would call the student to the stage and criticize face-to-face. I think this could greatly hurt the student's self-esteem...I will never do like this. (Daisy, Interview Round 2)

[Assessee in middle school] My English teacher in senior middle school replied my inquiry of the exercise with she had taught the item before. I felt surprised and depressed that time, which impressed me most. I guessed maybe the teacher disliked me, which resulted in my subsequent negative attitude to English learning... if I were the teacher, I could explain the resolution and remind of the knowledge point again by resorting to the textbook rather than neglecting the student's emotion. (Flora, Interview Round 1)

My mathematics teacher sometimes criticized us with unacceptable words and tones, like we are too stupid to work out this problem. (Flora, Interview Round 2)

When I become a teacher, I will improve the assessment practices more from students' benefit by not using extreme or sensitive words...I have taken student's acceptance degree into account when designing assessment tasks in practicum. (Flora, Interview Round 4)

[Assessor-to-be in practicum] Some teachers in the practicum school took rather extreme reaction to student's violation of classroom discipline. The teachers videoed the clip and sent it to the parents' group instantly. Indeed, they did so. At that time, I was sitting at the back of the classroom. I totally disagreed with this way of communicating and it went too far. I was so puzzled... if I were the teacher, maybe I would take other alternative ways to communicate students' learning attitude to their parents, such as home visit, to protect their privacy instead of popularizing. (Amy, Interview Round 1)

(3) Assessment results application. The unrecognized application of the assessment results was to arrange the seat based on the students' exam scores, which was one of the unspoken common practices in China. The higher score the student obtained, the more priorities the student had to select the ideal seat in the classroom. Three participants who had such an experience were strongly opposed to this practice, as Betty, Edwin, and Flora noticed the potential consequence of unequalness and harm to

students who got less desirable results in the exam. They all consistently expressed their unwillingness to adopt this practice in their professional career.

[Assessee in middle school] Though the teacher did not say explicitly, our seats were arranged in accordance with our academic performance to a great extent. The student who behaved better in the exam sat at the front while those who got worse results sat at the back. That is to say, the exam results were represented implicitly in the seat arrangement... I may not do like this to my students. (Betty, Interview Round 2)

[Assessee in middle school] In my senior middle school, the teacher asked us to select seats based on the exam scores. I felt so ashamed and discouraged, especially when I was the last one. All the students having their own seats already watched me to sit down. Moreover, there was no other places for me to choose. (Edwin, Interview Round 2)

[Assessee in middle school] I had such an experience in middle school. We were ordered according to the exam score to choose the seat one by one. I completely disagree with this practice. Because only the top students can select the ideal seats, others who fail in the exam have no choice to sit where they like. Maybe they have to sit at the back or other unideal places, which is not beneficial to their further improvement and finally a vicious circle is the possible ultimate result...I will not follow this. (Flora, Interview Round 2)

4.4.1.1.4 Negative Interpretation of Positive Assessment Experience

The assessment experience labeled as positive might be re-interpreted from a negative perspective. The participant, Carol, provided an example from her own experience of seat arrangement. Initially, she preferred such a practice of seat arrangement based on the exam performance as being assessed in her schooling time. However, now she changed her ideas as she witnessed the adverse results in such a practice in practicum from the stance of an assessor. The acknowledgement of the neglected disadvantages of the once positive assessment experience prevented her from adopting it in subsequent assessment-related activities.

[Assessee in middle school] I felt it was acceptable for me. Because it broke the seat monopoly until the college entrance examination. The common practice before the seat arrangement based on the scores was that we could choose the seat by ourselves. Such practices led to seat monopoly that some students who chose seats earlier might reserve deskmates to their friends. When I intended to sit there, I was told in a low voice that it was reserved. So, the seat arrangement based on the scores broke the monopoly of the seat. I welcomed such a decision from teacher. (Carol, Interview Round 2)

When I become a teacher, I would not let them choose seats by themselves [based on the academic performance]. They usually chose to seat around their intimate friends if they had the choice. With time went on, they tended to violate classroom discipline. As what I experienced in practicum, the students might chat in class when they sat together for a long time. When I was lecturing at the stage, they were talking quietly in class, which was so unkindly and disgusting... although I have not worked out a right way to arrange the seat, I will certainly not follow this practice. (Carol, Interview Round 2)

4.4.1.2 Contextual Factors

The second type of mediating factors identified was the context. LAL does not evolve in a vacuum, but in the context where LAL is situated. The participants also noticed the influence of context at three levels from macro through meso to micro. They articulated more meso and micro factors than macro factors, maybe due to the more direct and visible impact of the first two types. Macro speaking, the social assessment culture influenced the assessment conception. At the meso level, the school policy & culture and education section they would teach influenced their assessment practices in LAL. Concerning the micro contextual factors, their future identity as head teacher and other stakeholder groups might also impact on their assessment preference and decisions. Finally, the facilitating role of pre-service teacher education programme in LAL evolvement was explained in a detailed way.

4.4.1.2.1 Macro level

At the macro level, the exam-oriented culture influenced the pre-service EFL teacher's assessment conceptions, where AoL and AfL coexisted. The social assessment culture, especially the exam culture in China dramatically oriented the school education to prepare students to perform better in high-stakes examinations. Immersed in such an exam-oriented assessment culture, the conception of AoL was dominant among the participants. For instance, Carol, Daisy, and Flora recognized that the exam was the most effective and direct way to assess students. Moreover, Carol's first choice was to utilize exercises to test students' mastery of knowledge points.

The exam-oriented education prioritized exam scores in large-scale or small-scale. However, Daisy opposed to focusing only on scores and argued the ideal assessment conception should not concentrate only on scores (learning results), but should take the learning process and attitude into account. She supported that the ultimate purpose of assessment was supposed to facilitate learning, thus, AfL conception was apparent in her expression.

[Assessee in middle school] In my senior middle school, I was tested frequently by kinds of examinations and then ordered according to the scores in the exam. And the teaching content was confined within the examination content. The teacher told us to do the model tests as many as possible to gain a better score in college entrance exam...I think this practice is unavoidable against such a social surrounding. However, maybe it could be improved along with the advancement of the society. (Daisy, Interview Round 2)

Ideally, my assessment conception is not focusing only on scores when

assessing students, although relying on scores is prevalent. (Daisy, Interview Round 1)

...when I assess students, I would like to use assessment methods varying from person to person...The biggest principle of assessment should facilitate learning not inhibiting learning. (Daisy, interview Round 4)

4.4.1.2.2 Meso level

At the meso level, overall, the policy and culture of the school and the educational section where they would teach seemed to shape their assessment decisions. As Betty illustrated, whether she would send the grade sheet to the parents' group depended on the school policy and culture. If the school did not mandate and other colleagues did not do so, she would choose not to communicate the assessment results in this way. In other words, the school policy and culture would press her into taking reluctant actions in ways of assessment results communication.

I feel that [sending the grade sheet to the parents' group] depends on my colleagues' behavior. If they choose to share in this way, I will certainly follow. Or else, the parents may feel I am not responsible...another is policy of the school. If the school is open and has no mandatory requirements, I will choose not to communicate in this way. But if parents strongly require me to do so, I may obey their requirements because keeping the job is more important. (Betty, Interview Round 2)

Furthermore, the education section where she would teach also mattered, especially on assessment methods selection. She mentioned peer assessment was more suitable to be adopted in middle school than in primary school. That is to say, the education section where she would teach might influence her assessment practices in LAL.

If peer assessment is adopted in primary school, I believe a chaotic scene may be the final result due to everyone wants to express their ideas. Moreover, I think it is meaningless and fruitless to utilize peer assessment in primary school,

for they are too young to acknowledge the problem and to assess others. Maybe it is more appropriate and meaningful to apply to junior or senior middle school students. (Betty, Interview Round 3)

4.4.1.2.3 Micro level

Concerning the micro-level, three factors were figured out in the present study that influenced their assessment practices: identity of head teacher, students, and parents.

If they were appointed to be the head teacher of the class, they would be more likely to adopt certain assessment method, like portfolio assessment, by taking the advantages of being a head teacher (Betty).

If I will be the head teacher of the class, I will [implement portfolio assessment]. If not, I think it is not convenient for me to implement if I am only the teacher of English in the class. Though it still can be carried out, I feel its quality is not maximized without the advantage of being a head teacher...the head teacher may have a strong controlling power over the whole class through regular contact with the students after class. By the way, the students are more collaborative and have deeper feeling with head teacher. (Betty, Interview Round 3)

Besides, the student-related information like their family background, characteristics, age, number, and academic performance would also be a part of their consideration when designing assessment activities or providing feedbacks (Amy, Betty, Carol, and Edwin). For instance, Amy would consider the number of students being assessed; Betty would provide attractive rewards to students based on their age; Carol would choose assessment methods according to students' academic performance and behaviors; the student's family background and characteristics would be placed as the first priority by Edwin when providing feedback.

The time for observation in assessment may be tight if there is a large number

of students in a class. There may not be enough time to observe and record their performance. So, I think these may be the possible reasons I will consider when trying out other new assessment methods. (Amy, Interview Round 3)

I remember that in practicum, I saw the teachers gave students who made progress in the exam or in the homework some rewards, like pens or candies. But the students told me that they did not like these...I think if the rewards are not attracting to students, the motivation of the rewards is lightened a lot. I think the rewards must be based on their age by investigating what interests them most, then the positive rewards may work and be meaningful. (Betty, Interview Round 4)

I like to choose encouraging methods to provide feedback, because the students are quite disciplined and behaved well in the class, with only a very few rebels... another is definitely academic performance. I will choose assessment methods according to their exam scores and daily performance. (Carol, Interview Round 4)

I learned a lesson from the teaching practicum where I provided feedback to a sensitive student who came from single-parent family. I criticized him in the class. He felt so sensitive to my comments and kept on asking me why. I did not know how to handle and asked my tutor for help. My tutor complained that I needed to know such information in advance. This impressed me a lot...so in the future, when I need to provide feedback to students, I will certainly collect information about their family background or characteristics before I give assessment feedback. (Edwin, Interview Round 1)

Parents, an important stakeholder group, potentially impacted on their way of communicating assessment results. Just as what Betty mentioned, parents would also become a pressure for her to choose sending the record sheet into the parents' group. Or else she would be considered as irresponsible by the parents. Thus, the pressure or requirements from parents might also have an influence on the assessment decisions.

4.4.1.2.4 Teacher Preparation Programme

Pre-service teacher education programme served as a huge facilitator to participants'

LAL evolvement and self-reflection. It was in the pre-service teacher education that they developed LAL theoretically and practically in a systematical way. Therefore, it plays a dominate role in participants' LAL evolvement.

In general, though to a varying degree, the pre-service teacher education programme improved pre-service EFL teachers' LAL in various dimensions as illustrated in Y. Xu and Brown's (2016) framework, including constructing the assessment knowledge, shaping assessment conceptions, enriching assessment practices, and facilitating identity construction as an assessor. For instance, Carol concluded that she accumulated disciplinary knowledge in LAL. Amy, Betty, and Flora summarised the programme strengthened their AfL conceptions.

Definitely helped. It enhanced my professional knowledge in English and enriched my assessment experience through formative assessment by teacher educators in the programme. (Carol, Interview Round 4)

Regarding the assessment conception, I think it [programme] just strengthened with little change. I always hold assessment for learning, for the benefit of students. (Amy, Interview Round 4)

...also, it strengthened my assessment conception of serving students and broadened my assessment practices observed in practicum. (Betty, Interview Round 4)

It deepened my consistent assessment understandings that when providing positive assessment feedback, I can communicate at the presence of the whole class. But when I need to provide negative feedback, I must try to avoid communicating publicly and instead privately. (Flora, Interview Round 4)

Their assessment practices were enriched by observing diverse useful assessment practices from the teacher educators in university or tutors in practicum school, which was referred to by all the participants.

I think it has changed. The biggest change is in lecturing, especially lighting fireworks by hand gestures in assessment practices. Another one is scoring according to the performance of the entire group, which may motivate them towards a shared learning objective and I think it is a good practice. (Amy, Interview Round 4)

It enhanced my understandings and related knowledge about assessment. Before the programme, I did not have a clear outline about it and considered it merely as feedback of the result after being assessed by teachers. After learning the related theories, I knew assessment served many purposes and had a variety of methods. By integrating the various assessment methods into teaching appropriately, the ultimate outcome of each class will be maximized. (Daisy, Interview Round 4)

It improved my LAL dramatically. These teachers make me know what is good assessment, which I will intake seriously. The university, especially teaching practicum changed me a lot. (Edwin, Interview Round 4)

It enriched my assessment tools. I knew more assessment methods, like oral assessment, and written assessment. (Flora, Interview Round 4)

The last dimension, identity construction as assessors, was also facilitated. As noted by Betty and Flora, their perspectives towards assessment from the identity of being an assessee (student) had been shifted along the process to be an assessor (teacher) after the teacher preparation programme.

Yes, it has changed. I perceived assessment from the identity of a student before. That is, I answered what teachers asked. When teacher praised me, I was so happy. When teacher criticized me, I was depressed, just from the perspective of a student. Now the teacher preparation programme enables me to perceive from the stance of teacher. How can I guide and facilitate students through assessment...So, it definitely helped to improve. (Betty, Interview Round 4)

Definitely it helped. Before university, I just received and accepted feedback from teachers. I seldomly reflected on assessment, for example why teacher assessed me like that? From what aspects? I did not think over about this. However, being an EFL teacher candidate in the university, I need to take it into account. I reflect on why teachers assessed me like that at that time and how I am going to assess my students when I become a teacher in the near future...I have to change the identity to think over assessment. (Flora,

Interview Round 4)

To be specific, the facilitating role of pre-service teacher education programme played in pre-service EFL teachers' LAL evolvement was elaborated from four aspects: the assessment-related policy (i.e. the formative and summative assessment implemented by teacher educators), the assessment-related courses (i.e. English Pedagogy), the assessment-related practices (i.e. intensive teaching training & teaching practicum), and the assessment-related high-stake exam (i.e. Teacher Certificate Qualification Exam).

(1) Assessment-related policy: formative and summative assessment. All the courses in pre-service teacher education programme were implemented with the policy of formative assessment plus summative assessment although with different proportions in the final integrated scores. The assessment policy indirectly shaped pre-service EFL teachers' LAL evolvement, to some extent, since it was implemented by teacher educators, who played the role of assessment modelling by involving pre-service teachers in to enrich their assessment experiences.

The participants experienced various assessment methods implemented by teacher educators in the programme, such as performance assessment, peer assessment. Amy mentioned the performance assessment, like dressing Kimono and making Matcha, conducted by teacher educator in second foreign language class was so interesting that she would adopt such a kind of assessment in future EFL teaching. Carol preferred the

peer assessment in English Pedagogy, which would be more likely to be transferred in her assessment practice. Flora was also fond of peer assessment for the benefit of students.

These assessment activities adopted in the university might exert influence on my future practices in assessment, more or less...for example, I will imitate some, like peer assessment. I will ask one student to answer the question, then require others to assess or comment on whether the answer is correct or not. If not, what is your opinion? If correct, why? Just in this way to enhance their attention to others' answers and their own thinking ability. (Flora, Interview Round 3)

(2) Assessment-related course: English Pedagogy. Unfortunately, the selective course which was directly dealt with language assessment, Language Assessment & Testing, had not been offered for fewer than 20 students to choose according to the university regulations. When being asked the reason why not select such a course, the participants provided extensive hindering obstacles: fear of failure (Carol, Daisy, and Edwin), no basis in assessment before (Amy, Betty, Daisy, Edwin, and Flora), no interests (Amy, Carol, and Flora), no close relationship with the major (Betty), and unfamiliarity with the teacher (Daisy). These reasons also reflected a fact that they appeared not to completely acknowledge the significance of assessment in EFL teaching. Therefore, it was only possible to find out from other related courses covering the topic of assessment. For the close inter-relatedness between assessment and teaching specially in AfL paradigm and also the inclusion of disciplinary knowledge & PCK in Y. Xu and Brown's (2016) model, the course English Pedagogy was focused as the target.

Generally speaking, English Pedagogy improved pre-service teachers' assessment

understandings, disciplinary knowledge & PCK, and reflective ability in LAL involvement through involvement in various assessment activities, like being assessed by themselves, peers, and teachers though limited and to a varying degree partly due to the variance in lecturers' capability and attention in assessment. Even the same lecturer facilitated pre-service teachers' LAL improvement to different levels.

There were three teachers, Ms. Gao (Betty), Ms. Li (Carol, Daisy, and Edwin), and Mr. Lei (Amy and Flora), who lectured this course to the participants in different classes in the present study. They all organized the assessment activities at three levels: self-assessment, peer assessment, and teacher assessment, which enriched the pre-service teachers' assessment experience as being assessed and assessing others and enhanced their reflective ability in assessment in LAL. More importantly, the disciplinary knowledge and PCK were dealt with great emphasis, which was an indispensable sub-category in knowledge base of LAL.

Ms. Gao presented the theories and terms in teaching EFL in the first half of the term and then assigned tasks to us to apply these theories in mock teaching presentation around 10 minutes. After each presentation, peer assessment and feedback would be proceeded...The classmates provided assessment from various aspects: teaching manner. Most of the students were nervous when it was the first time to stand on the platform. Second is teaching content. They might cover so many points without any emphasis. When we first touched on this course, we were not familiar with it. Most of us assessed from manner and the volume of teaching content. What we focused on might be superficial...then, the teacher gave feedback. First from a general impression, followed by the specific problems. Her feedback was more professional in content and provided improvement direction and suggestions...at the end of each class, we would fill in a self-assessment form: what I have mastered, what I have puzzled, what I want to test others. (Betty, Interview Round 3)

Ms. Li planned the class according to how to teach the four skills and how to

write lesson plan. Every week, we were asked to write lesson plans (including teaching reflection) and presented at the platform in turns, then assessed by other students and the teacher...the others often assessed from voice, teaching plan, atmosphere of the class, inter-action, teaching language in the lecture, such like that. The teacher assessed us from the advantages of lesson design or the shortcomings in the mock lecturing. With much emphasis on the lesson design and implementation, little attention had been paid to the element of assessment in such design. (Daisy, Interview Round 3)

Ms. Li guided us in how to assess others. We did not know how to assess at the beginning and could only reply with good, just like my senior middle school English teacher. I remembered when I was assessed by one of my classmates, he commented my lecture was vivid and interesting. Ms. Li further enlightened us with raising questions like: “Why good enough? In what aspects? Does he explain clearly?” (Edwin, Interview Round 3)

Mr. Lei imparted theoretical knowledge in language teaching in the first class, then organized us into lecturing in the second class. After our presentation of the mock lecturing, we were required to do a self-assessment first, like teaching reflection on the advantages and disadvantages. Then other students assessed based on their analysis of the mock lecturing, from teaching manner, blackboard writing, voice, dressing. At last, the teacher would summarise with an assessment from the teaching content to provide improvement. (Flora, Interview Round 3)

Though it included the content about assessment in teaching and enriched assessment experiences, the course was evaluated by the participants as limited improvement to a varying extent in LAL. Even the course taught by the same teacher (Ms. Li) was evaluated differently among the three participants (Carol, Daisy, and Edwin) from little to limited help. Three reasons were reported to contribute to this small amount of growth. The first reason was too much emphasis on how to teach than on how to assess (Carol, Daisy, and Edwin). The second contributor was the exposure to implicit assessment elements with little emphasis on explicit assessment explanation (Betty and Carol). Another reason was likely to be the absence of assessment scale in all the

three-level assessment activities. All participants mentioned the teachers never adopted any rubrics in assessment, which might result in assessing based on their own understandings without formal and systematic scaffoldings in assessment.

[Mr. Lei] Virtually no help in assessment learning. (Amy, Interview Round 3)

[Ms. Gao] Limited improvement. Just let us try to assess but gave us little help in what should be assessed. We had the initial impression about assessment, but had little chance to be deeply involved in assessment. (Betty, Interview Round 3)

[Ms. Li] Limited improvement. The teacher paid more attention to lesson design, teaching activities, and the teaching procedures. I remembered that the teacher just mentioned by the way, right, by the way, when writing the lesson plan, we should list assessment as a separate column at right for a clear presentation. But in reality, the teacher did not explicitly require us to do so and we seldom adhered to this rule in the following classes. (Carol, Interview Round 3)

[Ms. Li] A little. The main focus of the course lay in how to teach and how to write teaching plans with a few coverages on assessment. (Daisy, Interview Round 3)

[Ms. Li] Little help. The course seldomly covered assessment. So, what we focused on was how to make a good teaching plan rather than on assessment preparation. (Edwin, Interview Round 3)

[Mr. Lei] Some improvement. Especially we were involved in self-reflection about the teaching and assessment, then being assessed by peers and teachers. (Flora, Interview Round 3)

To summarise, owing to the over-reliance on teaching, implicit assessment explanation and scarce use of assessment scaffoldings, the participants in the present study claimed their marginal LAL evolvement in the course English Pedagogy. Although limited, the course played a facilitating role in LAL evolvement among the participants to a varying degree by exposure to disciplinary knowledge and PCK and engagement in three-level assessment activities which helped their assessment understandings

internalized and reflecting on their own or others' assessment in teaching activities.

(3) Assessment-related practices: intensive training and teaching practicum. Intensive training, aiming to sharpen the pre-service teachers' skills in teaching design and implementation, was arranged before the teaching practicum and lasting for two weeks under the guidance of an EFL teacher from the pre-service teacher education programme. The content was mainly composed of mock lecturing and presentation of lesson plans by every pre-service teacher for at least twice. Then the peer assessment and assessment feedback from teacher were followed to help hone their teaching ability.

Intensive training was evaluated by the participants as a trifle contributor to LAL involvement. The contribution was indicated to lead to their enhanced assessment awareness and increasingly honed teaching design ability. For instance, Edwin's assessment awareness was enhanced, and Flora appreciated more practices in EFL teaching design where assessment was embedded.

The teacher, Ms. Su, was so responsible. She impressed me with her rigorous attitude. When I finished my presentation, she asked me to assess by myself, then by peers. At last, she commented. It was at that time when I was carefully thinking over assessment feedbacks from teachers. I tended to totally ignore assessment before. That means, the session cultivated my awareness of assessment. (Edwin, Interview Round 4)

The intensive training session provided more opportunities for us to practice through trial lecture and sharing lesson plans. To some extent, it was a promotion to our teaching where assessment is embedded (Flora, Interview Round 4)

However, as a result of variance in teacher educators' capability and in the training

focus as well as the realistic factors (limited training duration), their LAL evolved to different degrees. Just as what Betty complained, the teacher in this session was not professional enough to provide useful and focused feedback. The training focused more on teaching than assessing, which was mostly handled with less attention in an implicit way, as figured out by Betty and Carol.

The intensive training helped a little in LAL. It paid much attention to teaching instead of assessing. That is little coverage of assessment. Moreover, the teacher in such a training was not as professional as Ms. Gao [the teacher of English Pedagogy]. She just provided positive feedback to us and did not point out our problems or shortcomings. Her comments merely encouraged us in our confidence and lack of focus. I felt in form it was no difference from the course English Pedagogy, but in quality, it was less satisfying for the teacher was not helpful enough. (Betty, Interview Round 4)

The intensive training session was focused on sharpening teaching skills rather than assessment. In other words, a little neglected or less attention. (Carol, Interview Round 4)

Moreover, the arrangement of the session was also complained by the participants (Amy, Daisy, and Flora). The limited time duration and inappropriate arrangement at the end of the term were suggested to be improved further. Therefore, these above-mentioned factors greatly influenced the effectiveness of the training and consequently, the benefits of facilitating LAL evolvement varied from person to person.

The time for intensive training was in conflict with preparation of the final exam, which made us in a dilemma. We intended to treat intensive training seriously, but the exam was more overwhelming. So maybe the outcomes of the training were not as desired as expected. (Amy, Interview Round 4)

However, owing to the limited time span and placement at the end of the term, when we are busy with preparing for final exam, some of its benefits were lost. (Flora, Interview Round 4)

I suggest the time duration in this session last for a longer time. In this way, it may be more effective. (Daisy, Interview Round 4)

The other type in assessment-related practices is teaching practicum, which promoted the participants' LAL greatly in various dimensions in Y. Xu and Brown's (2016) framework. For some realistic reasons, Daisy had not participated in teaching practicum. So, her data in this part was absent. In terms of the rest of participants, the involvement in real assessment practices during the teaching practicum broadened their assessment knowledge base in several sub-categories, strengthened their AfL assessment conception, enriched their assessment practices mainly through observation as an assessor-to-be, and facilitated the identity construction approaching an assessor through reflecting on the assessment activities more from the stance of an assessor or assessor-to-be. Although dramatic promotion in LAL evolvement was indicated, there still existed some inhibitors. The tutors in university and practicum school helped little when the participants encountered difficulties in assessment for the tutors were primarily concerned about security and teaching guidance respectively.

It would be better to analyze the promotion of LAL in practicum in great detail. The first dimension was broadening the assessment knowledge, especially in the knowledge of feedback communication. Edwin cited his experience as an example in which he had a deeper understanding of the potential influence of feedback for a disadvantaged student and he learned how to handle this kind of issues more carefully and appropriately by incorporating more background information of the student in advance when communicating feedbacks.

In the practicum, I did not know the student was so sensitive and grew up in

a single-parent family. I thought I communicated the feedback to him in a proper way, but he tended to enlarge these feedbacks and contributed to other factors. I was so puzzled and did not know how to face him. After this case, I knew when communicating assessment results or feedbacks to a special group of students, I needed to gather more information before hand, like the family background, their characteristics. (Edwin, Interview Round 1)

The participant, Edwin, knew more about how to communicate assessment with a special group of students in a suitable way after such an impressive experience in practicum. In other words, the practicum helped him to accumulate knowledge in assessment feedback communication, a sub-category in assessment knowledge base in LAL framework. Also, how to be ethical in assessment and protect students' benefit aroused his concern and focus on related knowledge.

Secondly, their assessment conceptions that assessment should serve students' benefit and enhance learning (AfL) was further strengthened. Betty, Carol, and Edwin asserted their human-oriented assessment conception was further consolidated for the positive example (Betty) or counter example (Edwin) experienced in practicum school. As explained by Betty, she witnessed the desired outcome of positive assessment and thus more insisted on her conception of assessment. By contrast, Edwin experienced a counter example through which he was more conscious of the dire influence of improper assessment on students, thus, he was more determined to adhere to the conception of serving students benefits.

For the assessment conception, I always think the students' mental and physical health should be set in priority in all assessment activities. The assessment methods were more important than the assessment content. For the terrible assessment may influence or even hurt students in an irreversible way, especially the sensitive student from single-parent family I have mentioned

before... I was really aware of the influential effect of assessment on a person [student]. (Edwin, Interview Round 4)

I consistently hold that assessment should exert positive influence on students. Even the most troublesome students, they will behave well when receiving positive feedback. So, the teaching practicum just strengthened my conception; especially these examples observed in school further proved my adherence. Another I learned during the practicum was that the assessment activities should be designed by taking the psychological character of students in different age stage into account. (Betty, Interview Round 4)

I always uphold the human-oriented conception in assessment. After the practicum, I have never changed this conception, just know more about how to assess for students' good. (Carol, Interview Round 4)

Thirdly, their assessment practices were enriched primarily through observation as an assessor-to-be (Amy, Betty, Carol, and Flora) or engagement as an assessor (Edwin). All the participants emphasized their assessment practices were enriched to a great extent by observing the model assessment activities designed by tutors in practicum school or their own engagement in conducting an assessment. For instance, Amy broadened her horizon in such innovative assessment practices in feedback communication.

I broadened my horizon. I saw some innovative assessment practices used by teachers in practicum school. The lighting fireworks by hand gestures, scoring based on their team performance, which could motivate students to learn towards a shared learning objective. I think it is an excellent practice. Another practice I appreciated a lot was feedback to students' English composition with expression in picture, not words, but simple line-drawings. Students were more motivated to English writing for collecting more expressions in picture from teachers. They treated English writing not as a burden. I think highly of this practice. The students' strong motivation was greatly due to the assessment practices conducted by teachers. (Amy, Interview Round 4)

Betty learned to provide positive feedback to retain students' motivation and reward students based on assessment results.

What I learned was providing positive feedback and comments when students behaved well to encourage them. In the school, the teacher gave some small rewards to those students who made progress or performed better in the homework, though sometimes, the rewards were not charming enough to students. When I become a teacher in the future, I will provide rewards according to their interests, maybe, to improve the effect of rewards. (Betty, Interview Round 4)

Carol benefited from the reward criteria implemented by the teachers: summative assessment (i.e., exam scores), value-added assessment (i.e., progress in learning), and peer assessment (i.e., team leaders' assessment).

From the tutor, I have learned more about the assessment practices which I have never experienced before. The teachers in practicum school rewarded students according to their progress or performance. For example, the top students in English or other subjects were rewarded. The students who made great progress in exam scores were rewarded. Or the students who behaved well recently judged by the team leaders were also rewarded by the teacher. (Carol, Interview Round 4)

Flora summarised the assessment practices should be assigned to assess students in alignment with their English proficiency in order not to embarrass them at the presence of the whole class. Besides, the equal opportunity of being involved in assessment activities was also noticed.

The teachers in practicum school were so experienced. They taught us to take care of students. For example, when questioning, we should not throw the difficult questions to those students who might be less proficient. Besides, we should not ask a student to answer questions repeatedly and frequently in a class. Instead, we should cover most students as wide as possible to involve them in the classroom activities. (Flora, Interview Round 4)

Apart from the assessment practices observed from the identity as an assessor-to-be, Edwin had the opportunity to participate in portfolio assessment as an assessor in practicum where he got the first-hand experience of portfolio assessment. Owing to

such a valuable experience, he commented that he was more familiar with the procedures, the possible challenges, and the shortcomings in the implementation process.

I helped the teacher in practicum school to implement portfolio assessment for nearly three months. We collected their scores in each exam or test to make a line chart. With the process going on, the workload was a little large because we did manually and some data from the students were missing. At the later stage, we made the chart roughly based on the performance in every two tests. I actually felt that it posed a great amount of workload on teachers. (Edwin, Interview Round 3)

Fourthly, the identity construction of an assessor was facilitated through reflecting on the assessment activities more from the stance of an assessor-to-be or an assessor. The above-mentioned reference of the assessment practices denoted that they consciously paid attention to assessment models set by the tutors from the identity of an assessor-to-be. Some of the models would be directly transferred to their future assessment practices while others would be adapted to enhance its effectiveness. Take Betty and Edwin as an example, when she witnessed that the rewards were not attractive to the primary students in practicum school, Betty intended to improve the attractiveness of the rewards in her future career to make it more fruitful after reflecting on such a practice from the identity of an assessor-to-be. Edwin went further in identity construction to being an assessor as he was a primary implementer in portfolio assessment, which enabled him to perceive and reflect from the stance of an assessor.

At last, although they reaped benefits from observing tutors' assessment models, the participants seemed to reach a consensus that they gained little help from tutors in

assessment, which was less likely to be the focus of the communication (Amy, Betty, Carol, Edwin, and Flora). The tutors in practicum school, primarily responsible for enhancing participants' teaching designs and classroom management, seldomly guided on assessment in an explicit way. What they focused on was how to teach and how to manage the class on most occasions.

The tutor paid much attention to the teaching design. He also emphasized the blackboard design, which should be vivid and clear for students to understand. Concerning the assessment, he mentioned little. (Carol, Interview Round 4)

They seldomly did. They were more focused on my performance in designing and implementing teaching activities. (Amy, Interview Round 4)

He [the tutor] did not teach me about how to design and implement an assessment plan. In reality, the assessment is not the main focus, so we seldomly pay attention to it. What we talked a lot was the classroom management. That is, how to discipline students in class. (Betty, Interview Round 4)

The tutor helped little in assessment. What I have learned depends on what I constantly see and hear in practicum school. How to say, by observation. (Edwin, Interview Round 4)

The tutor just cultivated me in teaching capability or teaching reflection, how to do better. However, he did not teach me how to assess students better. The assessment was not the main focus in the communication between us. The reason might be that we all tend to neglect the assessment. (Flora, Interview Round 4)

Thus, by combining the data from assessment practices and tutors' guidance reported by the participants, it could be concluded that tutors in practicum school implicitly influence pre-service EFL teachers' LAL through their assessment modelling in practices rather than verbal lecturing or any formal communication in guiding pre-service teachers in assessment learning.

(4) Assessment-related high-stake exam: The National Teacher Certificate Qualification Exam. The Qualification Exam is organized by the nation every year to issue teacher certificate to those who have passed, meaning they are qualified legally to teach in primary or middle schools. To the researcher's best knowledge, it is the only high-stake exam covering pre-service teachers' LAL at the national level. Thus, it was also considered as a possible factor which might have a washback on their LAL evolvement.

The participants varied divergently in the evaluation of the contributing role of such an exam in promoting their LAL evolvement. Two of them asserted significant promotion (Amy and Edwin); three considered limited promotion (Betty, Carol, and Flora); only one (Daisy) said little promotion. To be specific, Amy and Edwin argued that the high-stake exam oriented LAL evolvement significantly.

All reforms begin from exams. Once it appears in exams, the reform begins. So, I think it significantly orients the attention and focus of teacher candidates to the elements of assessment appeared in the exam. (Amy, Interview Round 4)

Absolutely. The exam was one of the exams which I prepared most carefully and diligently. I made full preparation to this exam. During the preparation, I knew the assessment in teaching procedures, the technical terms like nor-referenced tests, criterion-referenced test, diagnostic tests. All these words or phrases I learned in the preparation lingered on in my mind. I was so impressive even now when you were asking me. (Edwin, Interview Round 4)

Betty, Carol, and Flora shared the same opinion that the exam helped to improve LAL but to a limited extent because the exam merely covered a few items in assessment. The light weight of assessment-related elements in the exam might not matter much to pre-service teachers.

The exam covers some assessment items, which only occupy a small proportion. Or sometimes, it is impossible to notice there are items testing assessment-related knowledge. (Betty, Interview Round 4)

The exam is definitely beneficial to LAL improvement...However, the assessment occupies a small coverage in the exam. (Carol, Interview Round 4)

It helped to improve LAL definitely but limited to some extent for the testing of assessment knowledge merely in two or three items, not a high proportion in the score. (Flora, Interview Round 4)

Different from other participants who acknowledged the promoting role of the exam in LAL, Daisy believed that it did not help. Only rote remembering for the purpose of passing the exam helped little in promoting LAL. What had learned in the process of preparation was less possible to be transformed into practices or literacies for they were only temporary and mechanic in mind.

No. It is just an exam. We may review the assessment-related points, but we just try to rote remember the points for passing the exam without really exploring the meanings and its underlying rationales. So, it did not facilitate our assessment learning, in other words, helped little. (Daisy, Interview Round 4)

4.4.1.3 Personal Factors

The last type was personal factors. The participants articulated their personality also had an influence on their assessment preferences and practices (Amy and Edwin). Amy attributed her preference of self-assessment to her self-centeredness. Edwin explained the favor of the question-answering assessment method was mostly owing to his characteristics. Additionally, he also mentioned the mental age and horizon might be the influential factors in LAL evolvement.

I do not like peer assessment because I think no one can understand me. I prefer self-assessment maybe due to my self-centeredness to some extent. (Amy, Interview Round 3)

I like question-answering method because I am very enthusiastic when I am lecturing. I think asking questions will activate the atmosphere in the class...after the practicum, I made judgement about the assessment methods which were more suitable to my personality. I will filter these and apply to the future assessment practices if possible. (Edwin, Interview Round 4)

I feel that LAL is closely connected to their [teachers] horizon, experience, mental age etc. not only the recitation of the assessment books. (Edwin, Interview Round 4)

To summarise, the identified mediating factors spread in three categories: experiential, contextual, and personal. For a clear presentation, the network of mediating factors was listed as separate dimensions, but that does not mean the factors function separately. Instead, they were inter-related and inseparable from each other in the network and functioned collectively on LAL evolvement among the pre-service EFL teachers in the present study (See Figure 4.10).

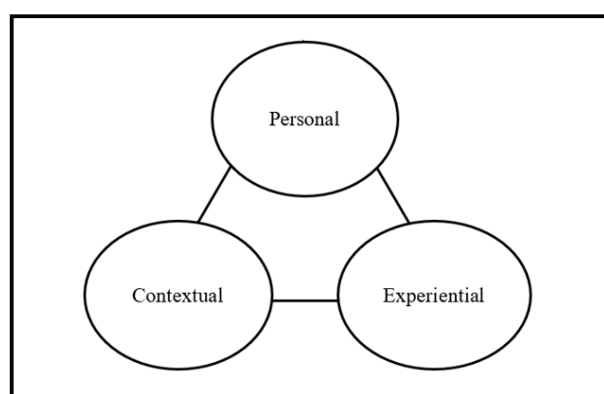


Figure 4.10 The inter-related network of the mediating factors

4.4.2 Interaction between the Mediating Factors and LAL Evolvement

The network of mediating factors together mediated LAL evolvement among pre-

service EFL teachers to different degrees in four dimensions encompassed in Y. Xu and Brown's (2016) framework: the knowledge base, assessment conceptions, assessment practices, and identity construction as an assessor. From bottom-up, the mediating factors participated in constructing pre-service EFL teachers' knowledge base, shaping their assessment conceptions, filtering the assessment practices, and facilitating the assessor identity construction. With the enhanced LAL, they were more attentive to reflect on the influence of these mediating factors. The interaction between the network of mediating factors and LAL evolvement was shown in Figure 4.11.

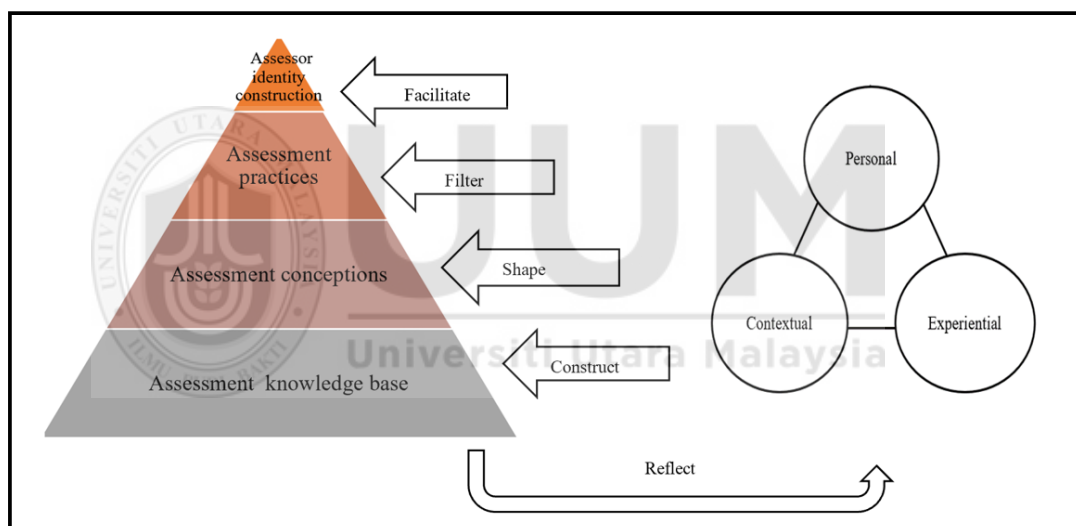


Figure 4.11 Interaction between mediating factors and LAL evolvement

(1) At the bottom level, the participants' knowledge base in LAL was constructed by the function of mediating factors, among which the contextual type contributed most. As illustrated in the contextual factors (See 4.4.1.2), the participants' assessment knowledge was primarily constructed in pre-service teacher education programme, which offered assessment-related courses to build disciplinary knowledge & PCK and deepen their understandings of assessment in an explicit and systematic way. Moreover,

the assessment-related practices section (intensive teaching training and teaching practicum) offered them to obtain assessment knowledge in practice, especially the knowledge of various assessment methods (Edwin), knowledge of feedback (Amy, Betty, Edwin, and Flora), knowledge of assessment results communication (Betty), knowledge of grading (Carol), as well as knowledge of ethics (Carol, Edwin, and Flora) either from the reflective observation of tutors' modelling in assessment practices or from the hands-on engagement in conducting assessment activities.

(2) At the second tier, the assessment conceptions were shaped under the influence of the mediating factors. The dominant conception of AoL among the participants might be shaped by the exam-oriented culture in China. For example, Flora believed that exam was the most effective way to assess students, which was likely to be shaped by the exam-oriented education she had been exposed to. But the limitedness of the exam-oriented education also figured out by Betty, who argued that exam scores should not become an overwhelming assessment standard. Besides, the participants' AfL assessment conceptions were further strengthened in practicum in teacher preparation programme (See 4.4.1.2.4).

Additionally, the experiential and personal factors also helped to shape their assessment conceptions, especially in affective dimensions. As elaborated in experiential factors (See 4.4.1.1), if the participants experienced an assessment positively (in assessment tools, methods, feedback, and results application), they

tended to hold a positive conception towards it, which was more likely to be applied into their assessment practices, although some of the positive assessment conceptions might be reshaped into negative along with the accumulated assessment experiences.

Apart from the experiential factors, the personal type played a shaping role in their assessment conceptions, too (See 4.4.1.3). As expressed by Amy, her comparatively negative conception towards peer assessment but positive towards self-assessment was mostly shaped by her self-centeredness. Another example came from Edwin, who attributed his positive conception of certain assessment methods (e.g., questioning) to his character of enthusiasm, to a greater extent.

(3) At the third tier, the assessment practices were filtered through the mediating factor network. The combination of the factors seemed to function as a filter in mediating the assessment practices among the participants. That means, the assessment practices which were in alignment with their characteristics, or the assessment context, or their experiences had more chances to be adopted by the pre-service EFL teachers. Or else, the practices appeared to be less utilized. For instance, the participants expressed their selection of the assessment activities depended on the personal factors (e.g., character from Edwin), the contextual factors (e.g., students' information from Betty, Carol, Daisy, and Flora), and experiential factors (e.g., performance assessment from Amy). Thus, the mediating network influenced the participants' assessment practices like a filter.

(4) The assessor identity construction was facilitated by exposure to the mediating factors, especially in teaching practicum. As illustrated, the teaching practicum played a significant role in identity shift from an assessee to assessor-to-be or further assessor. The pre-service teachers reported their perspectives and understandings of the assessment more from the identity of assessors-to-be (Amy, Carol, Betty, and Flora) or assessors (Edwin).

Take Carol as an example, the practicum experience facilitated her identity construction from being assessed to assessor-to-be, as revealed in her perspective towards seat arrangement based on exam scores. From the perspective of an assessee, she viewed such seat arrangement practice as positive, but during the practicum, her identity shifted to an assessor-to-be, which changed her positive perspective to negative interpretation.

In addition, Edwin went further in the identity construction process. He substantially participated in the portfolio assessment as a primary assessor in the practicum, which provided him opportunities to approach the stance of an assessor. Besides, his personal factors may also facilitate such an identity construction process. He was the only participant who applied for TESOL in postgraduate study, which means he was more focused on effective EFL teaching and more reflective on assessment for learning. His personal proactive attitude to effective teaching which cannot be isolated from

effective assessment seems to help facilitate the assessor identity construction. As a result, under the interplay of all these mediating factors, he was able to reflect on the assessment methods and implementation more from the identity of an assessor than an assessor-to-be. Thus, it could be seen that the shifting of the perspectives denoted the identity shift after practicum experience.

(5) The enhanced LAL enabled the participants to reflect on the influence of the mediating factor network to become a more self-directed assessor. The enhanced LAL might drive them to critically reflect on assessment conceptions and the assessment experiences mediated by the factors. For instance, with the growing LAL, Betty reflected on the shaping effect of the exam-oriented culture on her assessment conception: she completely disagreed with the sole reliance on exams to make assessment on students.

In addition, Betty was also more reflective on the assessment experience, like effectiveness of the assessment practices conducted by the tutors and intended to improve in her future professional career (e.g., the less interested rewards offered to students in practicum school). Equipped with the increased LAL, she was more likely to engage in self-reflection, including reviewing her assessment experiences in the local context and reconsidering her assessment-related conceptions and practices critically.

To sum up the findings for the research question three, the mediating factors identified

in the present study were classified into three types: experiential, contextual, and personal, which functioned together in mediating LAL evolvement among the participants to a varying degree. Furthermore, the factors functioned differently in each dimension in LAL proposed in Y. Xu and Brown's (2016) model, as depicted in Figure 4.11.

4.5 Findings for LAL Evolvement Trajectory

The last research question aimed to find out "How does LAL evolve among the participants?". Framed in Y. Xu and Brown's (2016) model, the LAL evolvement was analyzed within each level and across three mastery levels. As proposed by Y. Xu and Brown (2016), LAL evolved at three mastery levels: the first is the mastery of basic assessment knowledge; the internalized understanding and skills were the second mastery level; the ultimate goal of the assessor identity construction was placed at the ultimate goal. Overall, LAL evolvement among the participants had individualized but also shared features.

Within each level, each of the participants had their own distinct LAL evolvement trajectory resulting from a bulk of mediating factors mentioned in 4.4. They all revealed varying development in knowledge base, internalized understanding (assessment conceptions and assessment practices), and assessor identity construction. The six participants' LAL evolvement trajectories were analyzed in a sequence.

Amy constructed five sub-categories in assessment knowledge: disciplinary knowledge & PCK (from assessment-related course), knowledge of assessment purpose & methods (e.g., performance assessment methods), knowledge of ethics (avoiding hurting students), knowledge of assessment interpretation & communication (avoiding sharing the videos of misbehavior of students to parents chatting group), and knowledge of feedback (avoiding discouraging feedback). At the second mastery level, she expressed her assessment conceptions and practices of AfL, like formative assessment methods. She was also familiar with how to provide proper feedback and how to be ethical in assessment from the critical reflection on the assessment practices observed in pre-service teacher education programme from the identity of an assessor-to-be, which was the third level in LAL evolvement. Amy's LAL evolvement trajectory was illustrated in Figure 4.12.

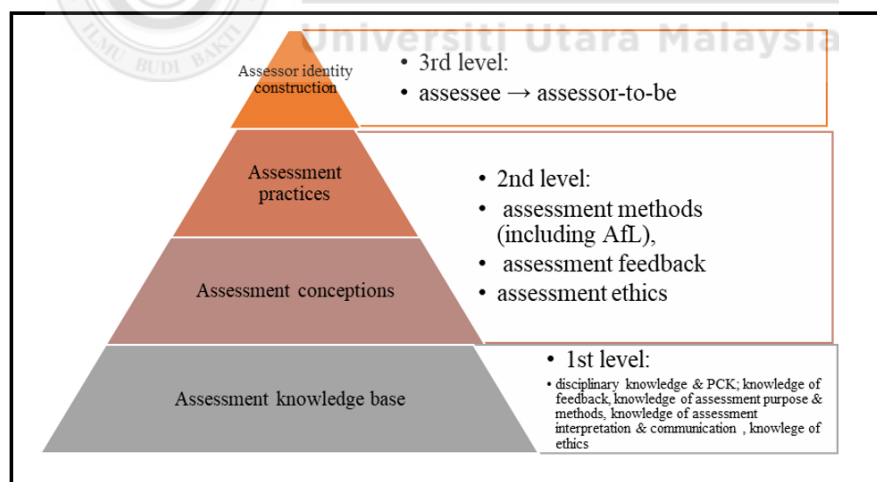


Figure 4.12 LAL evolvement trajectory of P1 (Amy)

Betty, the second participant, developed similar sub-categories in knowledge with Amy, in disciplinary knowledge and PCK, knowledge of assessment purpose and methods (assessment exit slip), knowledge of feedback (comprehensive and

encouraging feedback), knowledge of assessment interpretation and communication (parents meeting), and knowledge of assessment ethics (not giving up the low-achievers). These types of knowledge were also revealed in her assessment conceptions and practices. Besides, she was one of the participants who explicitly noticed the identity shift from being assessed (student) to an assessor (teacher).

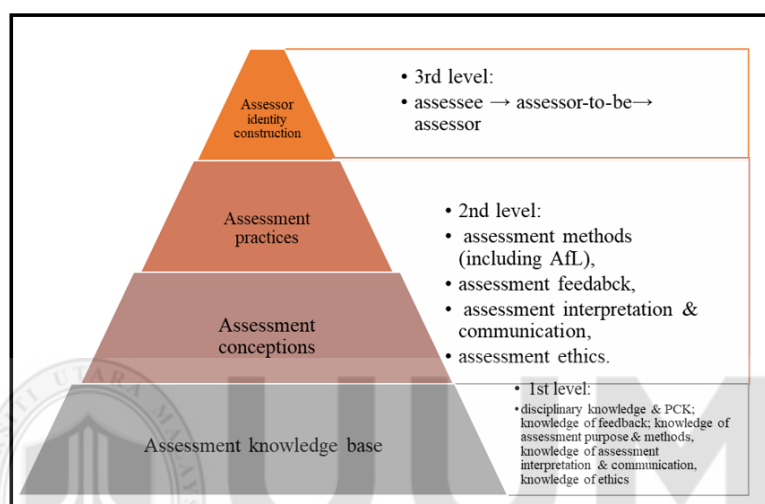


Figure 4.13 LAL evolution trajectory of P2 (Betty)

Carol reported development in four sub-categories in assessment knowledge: disciplinary knowledge and PCK, knowledge of assessment purpose and methods (online writing autonomous grading system), knowledge of grading (marking English tests), and knowledge of ethics (student-oriented). The assessment methods, especially the AfL methods (performance assessment) and student-oriented assessment conceptions were apparent in her preference of assessment selection. At last, her identity construction as an assessor-to-be was prominent in her reflective observation of the practices in practicum school.

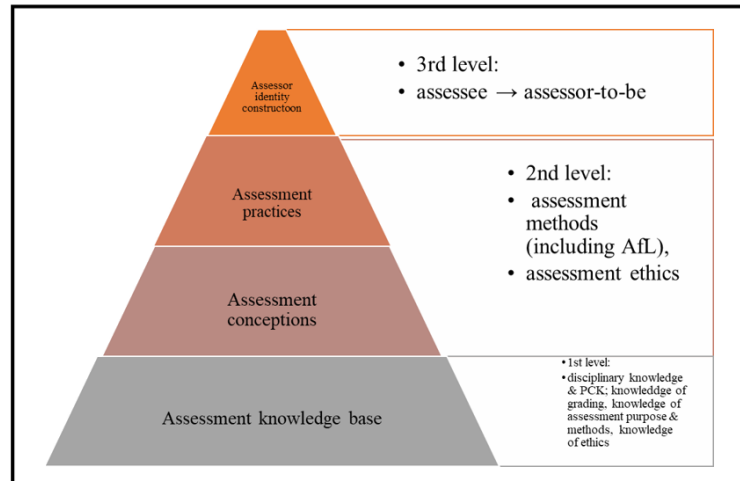


Figure 4.14 LAL evolvement trajectory of P3 (Carol)

Daisy's assessment knowledge was constructed in three sub-categories: disciplinary knowledge and PCK, knowledge of assessment purpose and methods (vocabulary testing APP), and knowledge of ethics (avoiding hurting students' self-esteem). Her assessment conceptions and practices were directed by AfL because she advocated that sole reliance on exam scores was improper to assess students. At the third level, her identity shifted from being an assessee to an assessor-to-be was also evident.

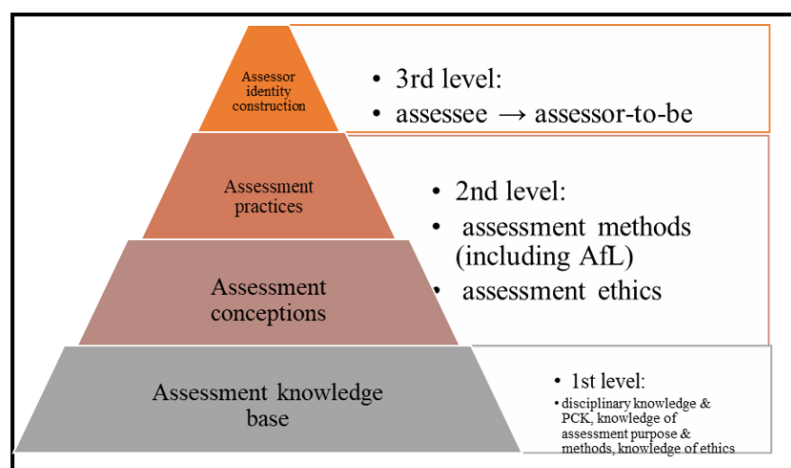


Figure 4.15 LAL evolvement trajectory of P4 (Daisy)

Edwin developed four sub-categories in assessment knowledge: disciplinary

knowledge and PCK, knowledge of assessment purpose and methods (vocabulary testing APP), knowledge of feedback (the improper feedback to a sensitive student), and knowledge of ethics (avoiding bad influence on students). He highlighted the purpose of assessment was to facilitate students' learning and utilized various assessment methods (e.g., questioning, quizzes, portfolio assessment) to implement the AfL conceptions. What's more, he was the only participant who was engaged in an assessment activity (portfolio assessment) as a primary agent, thus, his assessor identity construction became clearer by his critical reflection on the assessment practices from the stance of an assessor.

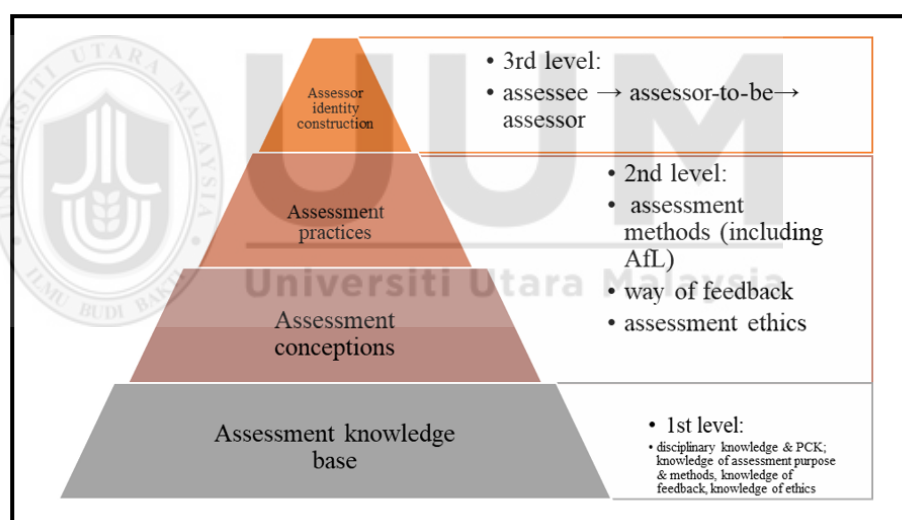


Figure 4.16 LAL evolution trajectory of P5 (Edwin)

Flora's assessment knowledge was developed in five sub-categories: disciplinary knowledge and PCK, knowledge of assessment purpose and methods (e.g., performance assessment), knowledge of feedback (encouraging and positive feedback), knowledge of assessment interpretation and communication (parents meeting), and knowledge of ethics (protecting their privacy and self-esteem). Her assessment

conceptions and practices seemed to be still dominant by AoL, but the emergence of AfL was also prominent. She considered the exam was the most effective and direct way to assess, but she also preferred to adopt performance assessment (e.g., drama playing) in the future assessment activities. At the last mastery level, her identity shifting to assessor-to-be was facilitated greatly in practicum by reflecting on the assessment practices of the tutors.

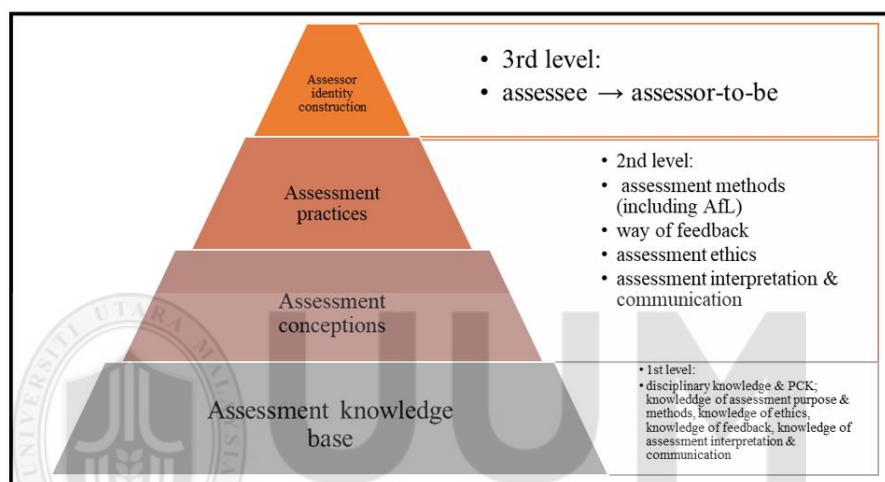


Figure 4.17 LAL evolution trajectory of P6 (Flora)

The six participants' LAL evolution shared some features in common. The LAL evolution trajectory among the overall participants were shown in Figure 4.18. Firstly, they all developed several sub-categories in assessment knowledge to different extent. The overlapped sub-categories among the participants were disciplinary knowledge and PCK, knowledge of assessment purpose and methods, and knowledge of ethics. By contrast, what they seldom mentioned was the knowledge of peer and self-assessment. Initially, the participants appeared to be less focused on assessment, thus their knowledge construction tended to be implicit and deficit. With the increasing exposure to implicit or explicit assessment education in pre-service education

programme, their knowledge in assessment experienced multi-dimensional development to a varying degree under the impact of the mediating factors. Then, they became clearer about the assessment knowledge, which presented the feature of explicit focus and self-directed growth.

Secondly, AoL and AfL coexisted in their assessment conceptions and practices, with AfL being further strengthened. Thirdly, their identity construction experienced from being an assessee through an assessor-to-be (Amy, Carol, Daisy, Flora) to assessor (Betty and Edwin). Lastly, all the participants went through the three mastery levels from the first knowledge base through internalized understanding to identity construction.

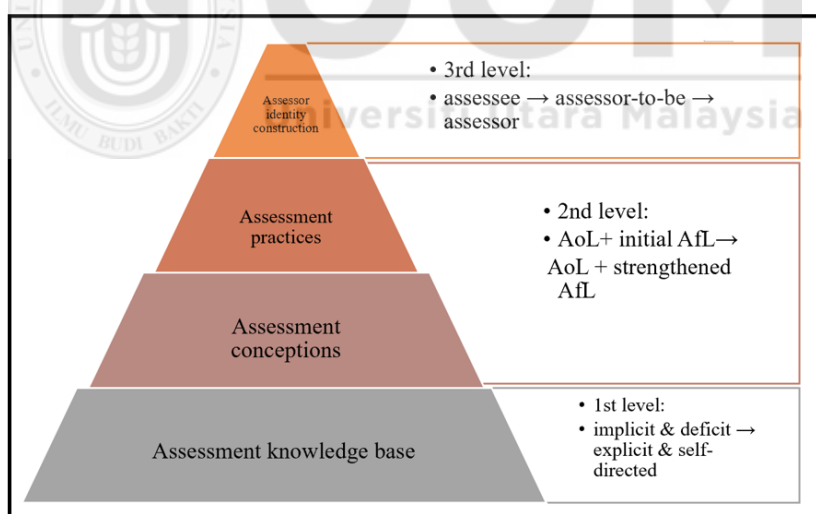


Figure 4.18 The overall LAL evolvement trajectory among the participants

4.6 Summary

In the section, the findings of the present study were summarised and presented based on the sequence of the research questions.

(1) LAL was perceived as a multi-layered concept, encompassing eight dimensions, which are not equally important. LAL was dependable on the social cultural context. Each participant's spider diagram of LAL conceptualization was unique and different from others. None of them was identical, and they turned out to be branded with distinctive personal characteristics.

(2) The second research question is concerned about the proficiency level self-evaluated by the participants. The findings showed that their self-evaluated proficiency was spread into three categories along a continuum: insufficiently qualified, marginally qualified, and satisfactorily qualified.

(3) The third research question is aimed to identify the mediating factors in LAL evolvement and its interaction. Three types of mediating factors were identified from the current study: experiential, contextual, and personal. On the one hand, the network of mediating factors together mediated LAL evolvement among EFL teacher candidates to different degrees in three mastery levels: the knowledge base, internalized assessment (assessment conceptions & assessment practices), and identity construction as an assessor. On the other hand, with the enhanced LAL, they were more attentive to reflect on the influence of these mediating factors.

(4) The last research question is concerned about how LAL evolve among the participants. Overall, LAL evolvement among the participants had individualized but also shared features. Within each level, each of the participants had their own distinct

LAL evolvement trajectory profile resulting from a bulk of mediating factors. All the participants went through the three mastery levels from the first knowledge base through internalized understanding to the ultimate mastery level of assessor identity construction.



CHAPTER FIVE

DISCUSSION AND IMPLICATIONS

5.1 Introduction

In this section, the findings of the present narrative inquiry are discussed by referring to the literature reviewed according to the sequence of the research questions for a more in-depth understanding of the research results. Additionally, implications for participants, other pre-service EFL teachers, teacher educators, and the teacher education programmes are provided based on the research findings.

5.2 Discussion of Findings for Conceptualization of LAL

RQ 1 is to elicit participants' conceptualization of LAL, which is characterized by eight graded dimensions (See Figure 4.7). As illustrated in the literature review, the conceptualization of LAL from pre-service EFL teachers is few, therefore, their conceptualizations of LAL are analyzed with reference to frameworks from the perspective of academic researchers or in-service teachers.

5.2.1 Eight Dimensions

The conceptualization of LAL from the participants' perspective overlapped greatly with the existing frameworks derived from other stakeholders. Nonetheless, a few

dimensions mentioned in the present study are seldom referred to in other models, and vice versa (See Figure 5.1).

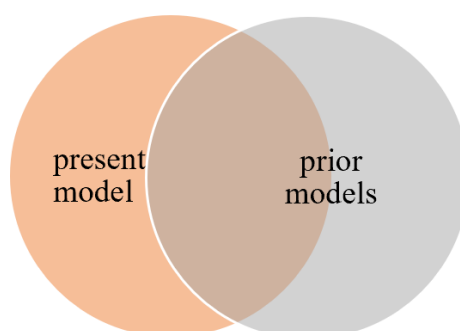


Figure 5.1 Comparison with other models in dimension classifications

The participants in the study perceive LAL as multi-dimensional which is widely supported by the literature (Baker & Riches, 2017; Bøhn & Tsagari, 2021; Davies, 2008; Fulcher, 2012; Kremmel & Harding, 2020; Taylor, 2013; Y. Xu & Brown, 2016). Most of the dimensions identified by the EFL teacher candidates are greatly overlapped with the previous models listed chronologically (See Table 5.1). They are roughly consistent with the classifications into three broad domains: discipline-related competence, person-related competence, and assessment-related competence.

Table 5.1

The Overlaps of Eight Dimensions in the Study with Previous Frameworks

Dimensions in this study	Dimensions in previous frameworks
Disciplinary & cross-disciplinary competence	Language pedagogy (Taylor, 2013)
	Disciplinary knowledge (Y. Xu & Brown, 2016)
	Language pedagogy (Baker & Riches, 2017)
	Language structure, use, and development (Kremmel & Harding, 2020)
	Disciplinary competence (Bøhn & Tsagari, 2021)

Table 5.1 continued

Pedagogical content knowledge	Language pedagogy (Taylor, 2013)
	PCK (Y. Xu & Brown, 2016)
	Language structure, use, and development (Kremmel & Harding, 2020)
	Language pedagogy (Bøhn & Tsagari, 2021)
Personal conception & attribute	Awareness of personal belief/attitude (Taylor, 2013)
	Conception of assessment (Y. Xu & Brown, 2016)
	Awareness of personal beliefs/attitudes (Baker & Riches, 2017)
	Personal beliefs & attitudes (Kremmel & Harding, 2020)
	Personal beliefs & attitudes (Bøhn & Tsagari, 2021)
Assessment principle & ethics	Principles (Davies, 2008)
	Why (Inbar-Lourie, 2008)
	Principles (Fulcher, 2012)
	Knowledge of assessment ethics (Y. Xu & Brown, 2016)
	Assessment principles & interpretation (Kremmel & Harding, 2020)
	Principles (Bøhn & Tsagari, 2021)
Assessment methods & implementation	Skills (Davies, 2008)
	How (Inbar-Lourie, 2008)
	Practice (Fulcher, 2012)
	Task performance (Taylor, 2013)
	Assessment purpose & content & methods (Y. Xu & Brown, 2016)
	Task performance (Baker & Riches, 2017)
	Developing & administering language assessment; Statistical & research methods (Kremmel & Harding, 2020)
	Technical skills (Bøhn & Tsagari, 2021)
Assessment content & criterion	Knowledge (Davies, 2008)
	What (Inbar-Lourie, 2008)
	Knowledge of grading (Y. Xu & Brown, 2016)
	Theoretical & conceptual knowledge (Baker & Riches, 2017)
	Scoring and rating (Kremmel & Harding, 2020)
	Knowledge of assessment theory; Scoring (Bøhn & Tsagari, 2021)
Assessment washback	Context (Fulcher, 2012)
	Washback and preparation (Kremmel & Harding, 2020)
Assessment interpretation & communication	Decision making (Taylor, 2013)
	Knowledge of feedback; Knowledge of assessment interpretation & communication (Y. Xu & Brown, 2016)
	Decision making (Baker & Riches, 2017)
	Assessment principle & interpretation (Kremmel & Harding, 2020)

However, the differences between the conceptualization of LAL from pre-service EFL teachers and others also deserve attention. On the one hand, what the pre-service EFL teachers has included but not mentioned in the prior models lies in two aspects: (1) cross-disciplinary competence (knowledge in other related discipline) and (2) personal attribute (patient, responsible, attentive). Few models in the literature seem to take the knowledge in related area and teacher's own characters into account. This is a new contribution to the existing literature through enriching the conceptualization of LAL from the neglected stakeholders, pre-service EFL teachers.

On the other hand, what the pre-service EFL teachers has not paid attention to but mentioned in the prior models exists in three aspects: (1) assessment policies and local practices (Baker & Riches, 2017; Bøhn & Tsagari, 2021; Kremmel & Harding, 2020; Taylor, 2013); (2) collaboration (Baker & Riches, 2017; Bøhn & Tsagari, 2021); and (3) teacher identity (re)construction as assessor (Y. Xu & Brown, 2016). One of the possible reasons maybe that these pre-service EFL teachers have few opportunities to implement or be involved in-depth in real assessment practices or activities in primary and middle school contexts even in teaching practicum, where they are just assigned the task of scoring the objective-tasks on most occasions. Therefore, they appear to be not so concerned about the practical factors in conducting an assessment such as assessment policies, local practices in schools, cooperation in the assessment, and identity construction as assessors.

5.2.2 Scaled Importance

The pre-service EFL teachers' conceptualization in the present study share something in common with other stakeholders' (language teachers, teacher educators) perception of the priority in LAL, but interestingly, they have their own extraordinary classification in nonpriority along the LAL priority continuum (See Figure 5.2).

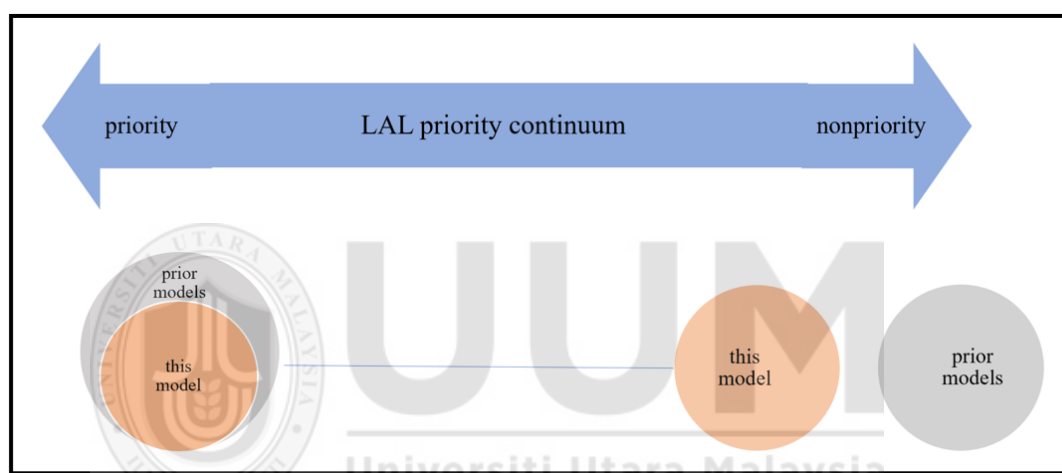


Figure 5.2 Comparison with other models in scaled importance

The eight dimensions showed in the study are scaled according to the participants' perception of each dimension's significance. The varying importance is in wide consistency with the existing literature (e.g., Baker & Riches, 2017; Bøhn & Tsagari, 2021; Kremmel & Harding, 2020; Taylor, 2013). The comparison of the significance continuum on two extremes from the most important to the least important with other spider diagrams is listed in Table 5.2.

Table 5.2

Dimensions on the Two Extremes of Importance Continuum

Source	The most important	The least important
Present study	Disciplinary & cross-disciplinary competence	PCK
	Assessment methods & implementation	Assessment washback
Taylor (2013)	Language pedagogy	Knowledge of theory
		Scores & decision-making
		Principle & concepts
Baker & Riches (2017)	Language pedagogy	Theoretical & conceptual knowledge
	Awareness of local practices	
	Awareness of personal beliefs/attitudes	
Kremmel & Harding (2020)	Assessment principles & interpretation	Statistical & research methods
	Language structure, use & development	
	Washback & preparation	
Bøhn & Tsagari (2021)	Assessment in language pedagogy	Technical skills
	Disciplinary competence	
	Principle	
	Language pedagogy	Collaboration competence

On the most important extreme, the present conclusion is mostly similar to Kremmel and Harding (2020) and Bøhn and Tsagari (2021), whose studies also highlight discipline-related competence. This competence is considered by the participants as the most important requirement for EFL teachers' LAL. Meanwhile, the mastery of assessment methods and implementation of the assessment are of the same importance. Likewise, assessment in language pedagogy is also labelled as utmost important in Kremmel and Harding (2020)'s research. That means, the pre-service EFL teachers in China share roughly the same understanding of priority in the LAL conceptualization with language teachers in UK (See Kremmel & Harding, 2020) and teacher educators in Norway (See Bøhn & Tsagari, 2021).

On the contrary, the other extreme presents a totally different picture. On the least important extreme, the PCK and assessment washback summarised from the study are seldom classified as the comparatively less important dimensions in other frameworks, which consistently take theoretical knowledge in assessment as the less demanding requirement for language teachers' LAL. This discrepant classification may be explained by the well-acknowledged divergence among different stakeholders' conceptualization of LAL and the variance in each research context which LAL depends on.

5.2.3 Context-dependence

The third prominent feature, context-dependence, means LAL is mediated by the contexts where it is rooted. This dependence has been agreed by a wide range of studies in the literature (Baker & Riches, 2017; Crusan et al., 2016; Davidson & Coombe, 2019; Farhady & Tavassoli, 2018; Han & Kaya, 2014; Jia et al., 2006; Portelli & O'Sullivan, 2016; Rogers et al., 2007; Sultana, 2019; Y. Xu & Brown, 2016). Instead of treating LAL in a vacuum, the pre-service EFL teachers in China has noticed the influential impact imposed by the social cultural contexts on LAL constructs.

5.2.4 Idiosyncrasy

The very core meaning of LAL conceptualization among the participants is rather

similar whereas the peripheral aspects of the conceptualization tend to be more characterized by their own individual experiences (See Figure 5.3).

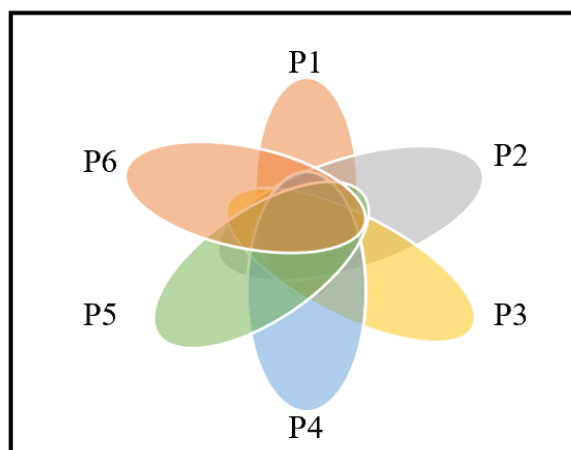


Figure 5.3 Comparison of LAL conceptualization among the participants

What they have reached a consensus is that LAL is concerned about assessing language and is a multi-layered concept. They all agree the very core meaning of LAL but differ in the peripheral meanings. None of these conceptualizations of LAL is identical as displayed in Figure 4.7, which exhibits unique spider diagrams from each participant in the current study. This personal idiosyncrasy may result from experiential factors, including the unique assessment experience during the entire educational stage and the different assessment conceptions held by the participants.

5.3 Discussion of Findings for Self-evaluated LAL Proficiency

The second question is concerned about LAL proficiency self-evaluated by the participants. The findings reveal three types in a continuum: insufficiently qualified, marginally qualified, and satisfactorily qualified. The participants also express their

desire to improve LAL focusing on diverse aspects but motivated by the same objective to be an effective teacher to enhance teaching and learning. These findings are discussed in relation to the literature in a sequence in the following sub-sections.

5.3.1 Self-evaluated LAL Proficiency

The present study classifies LAL proficiency self-evaluated by the prospective EFL teachers into three types along a continuum: insufficiently qualified, marginally qualified, and satisfactorily qualified (See Figure 4.8). The three types may co-exist simultaneously without any conflict among the pre-service EFL teachers.

Most of them perceive themselves as insufficiently qualified in the present study, which is consistent with Kavakli and Arslan's (2019) findings that pre-service EFL teachers tend to evaluate themselves as inadequately prepared to conduct assessment-related tasks in the future teaching profession. Besides, the finding is in alignment with pre-service teachers in other disciplines who also seem to consider themselves as unqualified, as evidenced from Volante and Fazio (2007). This self-perceived inadequacy in LAL may be largely attributable to the unsatisfying teacher preparation in assessment (Kavakli & Arslan, 2019; Volante & Fazio, 2007). Apart from the external factor of the limited teacher preparation in assessment as identified in the literature, the present study also adds the internal factors, such as the low willingness to become a teacher and to engage in learning, which is found to be contributing to this self-described inadequacy. This finding complements the ever known external

factors in literature and provides a rather comprehensive picture of the possible causes from both external and internal perspectives.

The second type is labelled as marginally qualified in the current study, which is a new contribution to LAL proficiency self-evaluated among the pre-service EFL teachers. Up to date, few literature has figured out this category. The existing classification is three types listed in the problem statement: insufficiently qualified, unrealistically optimistic, and realistically optimistic (DeLuca et al., 2013; Kavakli & Arslan, 2019; Kruse et al., 2020; Sahinkarakas, 2012). These categories may not explicitly reveal the marginally qualified as a distinctive type, which is worth to be marked as a separate section in the LAL proficiency continuum. Hence, the present study innovatively lists it as a separate type with the same status of the other two counterparts.

Furthermore, the third type, satisfactory qualification, is also emerged in the present study. Only one participant conveys a realistically optimistic evaluation of LAL: the participant is optimistic about her LAL meanwhile she is aware of the potential difficulties in the forthcoming assessment tasks, which is consistent with what DeLuca et al. (2013) labelled as realistic optimism among pre-service teachers in AL. Thus, realistic optimism in AL is upheld by the supporting evidence from LAL field that it can also be applied to pre-service EFL teachers. But for the purpose of establishing a continuum of LAL proficiency self-evaluated by the pre-service EFL teachers, the third type is relabeled as satisfactorily qualified to keep in form with the other two

types.

At last, a point needs to be pointed out is that the unrealistic optimism revealed in other studies has not been supported in present study (e.g., Ogan-Bekiroglu, 2009; Kruse et al., 2020; Sahinkarakas, 2009). That is to say, the pre-service EFL teachers in current study have not demonstrated unrealistic optimism, which is defined as over-estimation with little awareness of the complexity in assessment-related tasks (Kruse et al., 2020).

Thus, it can be concluded that the findings from the present study uphold two voices in the literature that the pre-service EFL teachers tend to evaluate themselves as insufficiently qualified or satisfactorily qualified (realistic optimism), with little demonstration of unrealistic optimism. Apart from this clarification, the study newly adds another category to build a LAL self-evaluated proficiency continuum by labelling marginally qualified as separate to explicitly acknowledge the significance of the firstly mentioned category from pre-service EFL teachers in China.

5.3.2 Self-diagnosed LAL Improvement

The participants convey diverse LAL improvement focus nearly covering all stages in carrying out an assessment: theoretical knowledge in assessment, assessment methods & implementation, assessment ethics, assessment dimensions, and feedback & washback in assessment. But they seem to be driven by the similar objective of being a helpful teacher to enhance teaching and learning.

The LAL improvement focus identified by the six pre-service EFL teachers in China not only bears a few similarities with their peers in other country, but also has overlaps with that of in-service teachers. To be specific, the participants in present study in China share something in common with their counterparts in Turkey where the assessment methods, approaches, and utilization are also prioritized in their self-described LAL improvement desires (Kavakli & Arslan, 2019).

Additionally, the improvement focus from pre-service teachers are partially overlapped with in-service teachers' concerns. For example, assessment knowledge and skills are also highlighted universally by teachers in their improvement interests (Firoozi et al., 2019; Matovu & Zubairi, 2014; Mede & Atay, 2017; Prizovskaya, 2018; Vogt & Tsagari, 2014). Assessment practice and various aspects in language assessment are figured out in several studies as well (Chan & Luo, 2020; Tavassoli & Farhady, 2018; Yan et al., 2017). At last, the potential ethical issues in assessment should be given equal emphasis in LAL improvement among in-service teachers (Chan & Luo, 2020). Thus, regardless of the pre-service or in-service teachers, they almost express their LAL improvement desires in various aspects in language assessment, with assessment methods and implementation remaining at the very center of improvement concern.

Moreover, all of the participants (except Carol) overwhelmingly explain the

motivation to improve LAL comes from the aim to be a good teacher in helping students to learn effectively. The identity recognition of becoming a good teacher can be further elaborated into the identity construction as a good assessor, because the participants have acknowledged to facilitate student learning through designing, implementing, and utilizing a quality assessment. This echoes the role of identity construction as assessor in Y. Xu and Brown's (2016) AL framework, which asserts that assessor identity construction facilitates teacher learning autonomy and self-directed awareness of reflection in assessment practices.

In a word, the majority of the participants reveal a strong desire to improve LAL in diverse aspects, which are partially alike with pre-service and in-service teachers in other countries. In addition, the driving power of the LAL improvement seem to be the assessor identity construction as illustrated by Y. Xu and Brown (2016).

5.4 Discussion of Findings for Interaction between Mediating Factors and LAL Evolvement

The third research question intends to find out the interaction between identified mediating factors and LAL evolvement. The proposed interaction model in Figure 4.11 will be discussed with literature.

5.4.1 The Mediating Factors Identified in the Study

Prior studies indicated that LAL was mediated by two broad types of factors: contextual and experiential factors (e.g., Crusan et al., 2016; Vogt & Tsagari, 2014). Personal factors were added by Y. Xu and He (2019). Similarly, the three types of mediating factors are identified in the present study: experiential, contextual, and personal.

5.4.1.1 The Experiential Factors

The experiential factors are considered to be the dominant factors in mediating LAL involvement among pre-service EFL teachers, similar with previous studies (e.g., Graham, 2005; Hatipoğlu, 2015; O'Loughlin, 2006; Y. Xu & He, 2019). The prior assessment experience gained either as being assessed as a student in schooling time or observing as an assessor-to-be in practicum or designing as an assessor in the assessment activity, has played a decisive role in LAL involvement. Regardless of the identity in the assessment experience, all these have been integrated into their experience and become a part of their lifeworld. As Giraldo and Murcia (2019) claimed that experience, a component of the lifeworld, is fundamental in developing LAL.

Besides, consistent with the results from in-service teachers, the prior assessment experience of being assessed as students in schooling time dramatically determines their preferences in assessment practices (Berry et al., 2017; Firoozi et al., 2019; Jane,

2012; Newfields, 2007; Rogers et al., 2007). In-service teachers have such a tendency to assess their students as they were assessed in their schooling time. Their assessment experiences seem to be transferred in their assessment practices, implicitly or explicitly.

The present study broadens the existing knowledge about the transfer conditions. In the literature, it is concluded whether the experiences will be transferred in their teaching career depends on the experience positive or not. Evidence from in-service teacher shows that teachers with positive experience when being assessed as students tended to hold a positive attitude and prefer to adopt the favorable assessment in their practices while the negative experience is less likely to be transferred (Jane, 2012; Quilter, 2000). This claim is upheld by the present study which provides supporting evidence in LAL field from pre-service EFL teachers who also demonstrate such a transfer tendency.

Furthermore, apart from what is known in literature, the current study also discovers another two transfer conditions: the positive interpretation of the once negative assessment experience may be transferred. On the contrary, the negative interpretation of the once positive assessment experience may not be transferred. In other words, the positive interpretation of the positive or negative assessment experience will result in active transfer while the negative interpretation will lead to less willingness to transfer when they become teachers. Thus, it is clear that the transfer conditions are extended to include interpretation as the determinant.

Finally, the present study specifies the transfer content in LAL, which used to be neglected in the literature. Due to the limited results in LAL among pre-service teachers, the findings from in-service teachers are referred to. The findings from present study have identified various aspects in transfer content, including assessment tool selection, assessment methods preference, way of feedback, assessment results communication and application. Among them, only assessment methods preference has been mentioned to be transferred among the in-service teachers (Berry et al., 2017; Firoozi et al., 2019; Newfields, 2007; Rogers et al., 2017). Thus, the findings contribute to the topic by elaborating transfer content in LAL among the pre-service EFL teachers.

To sum up, the present study confirms that experiential factors play a decisive role in pre-service teachers' LAL involvement. It also broadens the existing knowledge of transfer condition to set interpretation of the assessment experience as a criterion. The prior assessment experience, even if once negative, seems to be transferred into their future assessment practices if they perceive it as positive. Moreover, a variety of transfer aspects are also elaborated to supplement the existing dimensions figured out in the prior studies.

5.4.1.2 The Contextual Factors

Concerning the contextual factors mediating AL or LAL among the pre-service

teachers, the majority of the literature is focused on the quality of assessment-related courses or practices in teacher preparation programme. So, the discussion begins from the effectiveness of the pre-service teacher education, then followed by other contextual factors.

(1) Assessment-related courses. As investigated by Jin (2010), who revealed that the course was selective in 60% of teacher preparation programmes in China, Language Testing and Assessment was optional in teacher preparation programme in current study. What is more disturbing is that participants tend not to choose this course for various reasons. This implies that the LAL cultivation has not received its due attention, both from the programme designer and pre-service teachers, though more than ten years have passed. So, the researcher, in alignment with the scholars, advocate to lay more emphasis on language testing and assessment training in pre-service teacher education (Giraldo & Murcia, 2019; Hatipoglu, 2015; Jin, 2010).

The assessment course, Language Testing, has been confirmed to be encouraging in LAL improvement in various dimensions by a range of studies (e.g., Gebril, 2017; Giraldo & Murcia, 2019; Jeong, 2013). Nonetheless, the assessment-related course, English Pedagogy, has received little attention. The present study extends the scope to assessment-related courses, covering English Pedagogy as a source of LAL for participants. The course is proven to be a facilitator in LAL evolvement among the pre-service EFL teachers, especially in the most fundamental assessment knowledge base and reflective ability of assessment.

The biggest reason appears to be evident that the course is mainly focused on the disciplinary knowledge and PCK, a sub-category of assessment knowledge in LAL. The direct exposure to the lecture contributes to the growth in pre-service EFL teachers' assessment understandings. Another reason may lie in pre-service teachers' involvement in a variety of assessment activities designed by the teacher educators of the course. Such an involvement as assessee may offer them students' perspectives if the assessment methods are adopted when they become teachers. The deep involvement enhances their assessment reflective ability to some extent. Thus, the focus on PCK and the diverse assessment activities of the course enhance the efficiency in cultivating LAL among the pre-service EFL teachers in China.

Although efficient in LAL evolvement, the course still needs to be improved in several aspects, such as more explicit explanation on assessment and assessment scaffoldings. It is understandable that the course is primarily concentrated on how to teach, but it is feasible to integrate how to teach with how to assess, especially against the universal trend of paradigm shift to AfL, where assessment and teaching cannot be separated. The explicit emphasis on assessment within the relationship with teaching could make the course more facilitative by raising their awareness and attention to assessment. Another is suggested to provide available scaffoldings in assessment, like assessment criteria and standards in peer or self-assessment. Evidence shows that the pre-service teachers who experience assessment rubrics are more likely to utilize them than those

who do not have such an experience (Reynolds-Keefer, 2010). In short, the effectiveness of the assessment-related courses in LAL evolvement could be more significant if the assessment explanations were delivered in an explicit way and assessment scaffoldings were offered for reference.

(2) Assessment-related practices (Intensive teaching training & Practicum). Serving as a pre-practicum activity to hone their teaching abilities, intensive teaching training is rather a local university-based practice in the current study and is acknowledged as a marginal facilitator in LAL evolvement by offering additional opportunities to design teaching activities where assessment is embedded. The more opportunities to conduct assessment practices, the more possibilities in assessment competence growth (DeLuca et al., 2019). But the implicit and little coverage of assessment in intensive teaching training seems to restrict LAL improvement.

Concerning the teaching practicum, there are two voices in the literature. Lam (2015) argues the practicum as an inhibitor while Y. Xu and He (2019) perceive it as a facilitator in pre-service EFL teachers' LAL evolvement. The present study reveals that it greatly enhances pre-service teachers' LAL, mostly consistent with Y. Xu and He's (2019) study, which also considers practicum as a critical facilitator to LAL improvement in various aspects: accumulated knowledge and a broader understanding of feedback and assessment results application, enriched assessment practices by reflective observation of the school-based tutors, and shifted identity construction as

an assessor. However, what is inconsistent is that the participants' assessment conception is broadened to multiple purposes in Y. Xu and He's (2019) study while in the present study, the participants' assessment conception appears to be further strengthened towards AfL during the practicum.

But the findings are not totally against Lam's (2015) study. Lam (2015) claims that the practicum school-based tutors focus more on cultivating pre-service teachers' pedagogical skills than on assessment competence, which is considered as an inhibitor in effectiveness of practicum in improving LAL. This is also the case in the current study. The participants argue that their school-based mentors supervise little on how to assess while the mentors' attention is attracted to the teaching skills. In other words, their supervision does not instruct assessment explicitly, let alone systematically.

(3) Teacher educators. Teacher educators in pre-service teacher education programme are evidenced to play a modelling role in assessment to demonstrate and implement the assessment methods by involving the pre-service teachers in (Hill et al., 2017; Ölmezer-Öztürk & Aydın, 2019). Likewise, teacher educators in present study show various formative assessment methods to enrich pre-service teachers' assessment experience, some of which may be transferred into their future assessment practices. That means, teacher educators' AL or LAL matter dramatically in cultivating pre-service teachers' AL or LAL.

(4) The National Teacher Qualification Certificate Exam. As an entrance examination, this certificate exam has not put forward any requirements in teachers' assessment capability yet (Jin, 2018). It only tests a very few assessment-related terminologies in the form of blanks-filling items or multiple choices, which can be almost negligible for occupying less than 5% of the total score. So, there is no wonder that a total of four participants look upon it as a limited facilitator or even no effect in enhancing LAL.

Nonetheless, the other two participants (Amy and Edwin) argue that the limited inclusion of the assessment in the exam has still aroused their attention and focus to assessment learning, which is ultimately beneficial to LAL evolvement. Taking themselves as an example, they reveal that they have accumulated some assessment knowledge in the process of preparing for the exam. Thus, they recognize the strong driving power from the exam to push them into assessment accumulation.

(5) Other contextual factors (Macro, Meso, and Micro). Generally speaking, these contextual mediating factors at three levels are overwhelmingly overlapped with the results from in-service teachers in the literature, with a few newly added according to the findings of the current study.

What is already known is that the macro factor, exam-oriented culture and policy, is recorded to shape in-service language teachers' assessment conceptions and practices (Jia et al., 2006; Portelli & O' Sullivan, 2016; Sultana, 2019). At the meso level, school policies and culture play a role in mediating LAL (Farhady & Tavassoli, 2018; Jia et

al., 2006; Seden & Svaricek, 2018). Moreover, the grade level has been identified to influence teachers' assessment methods selection (Cheng & Sun, 2015). At last, the micro factors, such as number of students, other stakeholders (i.e., students' parents) also impact on their choices of assessment activities (Alkharusi et al., 2012; Cheng & Sun, 2015; Han & Kaya, 2014; Jia et al., 2006). The above-mentioned mediating contextual factors seem to be extended to pre-service EFL teachers in China, as evidenced in current study.

What is newly contributed to the existing knowledge is at the micro level. The pre-service EFL teacher would like to take concurrent post into consideration in their future assessment decisions. For instance, Betty points out whether portfolio assessment method would be adopted depends on her identity of head teacher or not. To implement portfolio assessment, the identity of head teacher may provide more convenience to collect comprehensive information from students through frequent contacts and students may be more cooperative with head teacher, who is responsible for nearly all affairs of students in the class in China.

5.4.1.3 The Personal Factors

Apart from the well-acknowledged experiential and contextual factors, the personal factors also become salient in mediating LAL among pre-service EFL teachers. Y. Xu and He (2019) identifies pre-service EFL teachers' agency in assessment as the personal mediating factor. The more reflective they are in assessment learning and

practices, the more possible to accumulate proficiency in AL (Ashraf & Zolfaghari, 2018; Yan & Fan, 2020). By contrast, the present study finds out three potential personal factors: characteristics, self-centeredness, and mental age & horizon, which seem to echo the claim that AL/LAL is partly a dispositional trait associated with teachers' psychology (DeLuca et al., 2019; Looney et al., 2018).

To sum up, the pre-service teacher education programme (i.e., assessment-related policies, courses, practices, and high-stakes exam) enhances participants' LAL to a varying degree in diverse aspects, ranging from assessment knowledge base to assessor identity construction. The majority of other contextual factors drawn from in-service teachers can also be applied to pre-service EFL teachers in China as well. What is newly contributed to the existing literature from the current findings is that the transfer conditions from the experiential factors are broadened to interpretation of the assessment experience. Besides, concerning the personal type, three factors closely related to pre-service teachers' psychology are originally identified.

5.4.2 The Interaction between the Mediating Factors and LAL Evolvement

The network of factors collectively but differentially mediates the pre-service EFL teachers' LAL evolvement, which is also responsive to the mediating factors (Y. Xu & He, 2019). The interaction model depicted in Figure 4.11 is supported by prior findings in the literature.

(1) The assessment knowledge base is constructed mainly through contextual factors, including the assessment-related courses and teaching practicum in pre-service teacher education programme. Compared with the extensive records that the assessment courses in AL or LAL field greatly help pre-service teachers to construct the knowledge in assessment (e.g., Deneen & Brown, 2016; Gebril, 2017; Giraldo & Murcia, 2018; Kruse et al., 2020), the present study indicates the neglected assessment-related course (i.e., English Pedagogy) also contributes to the knowledge construction among the pre-service EFL teachers. Besides, the role of teaching practicum in knowledge construction, especially in feedback, assessment results communication, and ethics echoes the findings in Y. Xu and He's (2019) study, which suggests the participants deepens their understandings of assessment in various aspects during the practicum.

(2) The shaped assessment conceptions are an outcome of the collective functions of these mediating factors. This study corroborates that the contextual and experiential factors which have shaped in-service teachers' assessment conceptions also played a shaping role in pre-service EFL teachers' assessment conceptions. For example, the social-cultural context (Black & Wiliam, 2005; Brown, 2011; Brown et al., 2011; Koh & DePass, 2019) and the schooling-based assessment experiences (Jane, 2012; Quilter, 2000; Rogers et al., 2007). Furthermore, the present study further supplements the personal factors identified by Y. Xu and He (2019). What is newly added is the sub-categories in personal type (i.e., characteristics, self-centeredness, and mental age &

horizon).

(3) Assessment practices are filtered through the network of mediating factors. A large number of studies demonstrate that preference of certain assessment methods or procedures in in-service teachers' assessment practice tend to be rooted in their prior assessment experience as being students in schools (Berry et al., 2017; Firoozi et al., 2019; Newfields, 2007; Rogers et al., 2007). The findings originated from pre-service EFL teachers also show such a tendency. Besides, consistent with study by Y. Xu and He (2019), the contextual factors, such as classroom realities are proven to be considered in designing the assessment practices. Thus, it becomes apparent that the assessment practices, which are less likely to be adopted seem to be filtered out through the network of mediating factors.

(4) Assessment identity construction as assessors are facilitated. The present study testifies Y. Xu and He's (2019) findings that pre-service EFL teachers' identity constructs from being assessed to assessors mainly during the practicum. Pre-service teachers undergo a shift in identity construction from being assessed as students into assessors who initiate the assessment, by negotiating the assessment identity within the learning community (Willis et al., 2013; Y. Xu & Brown, 2016).

(5) In the interaction model, the bidirectional relationship between LAL and the mediating factors is prioritized, the importance of which has also been highlighted in

prior studies (e.g., Giraldo, 2020; Mellati & Khademi, 2018). The present findings further clarify the interaction by offering original insights that LAL is not passively mediated but also empowers the pre-service teachers to actively reflects on the roles of mediating factors. For example, the participant (Betty) with enhanced LAL is more reflective on the shaping effect of contextual factors on assessment conceptions and more critical on the assessment practices observed in the practicum, some of which would be adapted by herself to effectively address the assessment issues situated in the local assessment contexts. Thus, the growing LAL enables the pre-service teachers to reflect on the potential mediating factors to become more self-directed assessors.

5.5 Discussion of Findings for LAL Evolvment Trajectory

The last research question focuses on LAL evolvment trajectory among the participants in China. The present study provides empirical supporting evidence that Y. Xu and Brown's (2016) model is suitable to trace pre-service EFL teachers' LAL evolvment trajectory. Although Y. Xu and Brown (2016) argue the rationale of application of the model to pre-service teachers to describe LAL evolvment trajectory, they merely conclude from speculative and experiential perspective. The current study qualitatively testifies the feasibility of framework application, and further clarifies LAL evolvment trajectory among the pre-service EFL teachers, who are long-neglected stakeholders in LAL literature.

Firstly, it reveals that the pre-service EFL teachers' LAL evolves multi-dimensionally,

including assessment knowledge, assessment conceptions & practices, and assessor identity construction. Consistent with previous studies primarily generated from pre-service teachers in general education, suggesting AL evolves in accumulated assessment knowledge (Bolivar, 2020; Smith et al., 2014), broadened assessment conceptions and practices (DeLuca et al., 2013; Smith et al., 2014; Wallace & White, 2014), and constructed assessor identity (Smith et al., 2014; Y. Xu & He, 2019), the present study has provided rigorous evidence that LAL evolves in multiple dimensions among pre-service EFL teachers in China, too.

Secondly, LAL evolvement trajectory of each participant is described to be distinct but have shared features. This trend may be due to the complex attribute of LAL, which is extensively acknowledged to be dependent on the local contexts (e.g., Fulcher, 2012; Giraldo, 2019; Kremmel & Harding, 2020; Taylor 2013) and individual interpretative inner world (e.g., Scarino, 2013). The shared features in LAL evolvement may be largely determined by the common macro social-cultural contexts where LAL evolves. Whereas, the distinct individualized characteristics engraved on LAL evolvement appear to result from unique personal character and experiences in assessment.

Thirdly, from bottom up, the findings confirm that the assessment knowledge base construction among the participants experiences differentiated development with some sub-categories more focused on while others less concerned. This partly accords with Bolivar's (2020) study, which shows that after the assessment-related course, the pre-

service EFL teachers construct assessment knowledge noticeably in assessment methods and assessment ethics. In contrast to the studies which highlight knowledge accumulation in students' agent role in assessment, like self- or peer-assessment (e.g., Bolivar, 2020; Smith et al., 2014), however, it does not appear to be the case in the present study arguing little knowledge construction in self- and peer-assessment among the participants. A possible reason for such neglect may be derived from their prevailing belief that teachers are the dominate agent in assessment while the students just act as a minor supplement. Therefore, their focus on assessment knowledge tends to be less attracted to knowledge in self- or peer- assessment, which emphasizing the agent role of students in assessment.

What's more, during the differentiated evolvement process, the assessment knowledge evolves from implicit and deficit to explicit and self-directed state. The observed increase could be mostly attribute to the overall satisfying assessment education in pre-service teacher preparation programme, which helps them to construct knowledge from the identity of being an assessor-to-be or an assessor.

Fourthly, it indicates that the internalized assessment conceptions and practices among the participants in China evolve from initial AfL to strengthened AfL. The result corroborates the findings of a great deal of the previous work that suggests such a trend in conceptions and practices among the pre-service teachers in other contexts (Bolivar, 2020; DeLuca et al., 2013; Smith et al., 2014; Wallace & White, 2014). Initially,

although still restricted to AoL, the pre-service EFL teachers have some notions in AfL. After the intervention of the pre-service teacher preparation programme, AfL notion is further strengthened among the pre-service teachers who are increasingly embracing assessment for students learning. Oriented by AfL conceptions, their preference of assessment practices also develops in the same direction.

Lastly, being the top mastery level in LAL evolvement, the assessor identity is shown to be constructed from being assessed through assessor-to-be to assessor in current study. Although pre-service teachers' identity shift from being assessed to an assessor has been consistently emphasized in prior studies (e.g., Smith et al., 2014; Y. Xu & He, 2019), the present findings further clarify the process by identifying an intermediate stage as assessor-to-be. As illustrated in contextual and experiential factor section, the participants (Amy, Carol, Daisy, and Flora) are actively engaged in critical reflection on the assessment models set by teacher educators in university or mentors in practicum school more from the identity of an assessor-to-be than being assessed, for they tend to relate such assessment models with their future assessment responsibilities.

5.6 Implications of the Findings

The study explores pre-service EFL teachers' LAL conceptualizations, self-evaluated proficiency level, the mediating factors, and LAL evolvement trajectories through a narrative inquiry conducted in China. The findings indicated from the present study have significant implications for the six participants, other pre-service EFL teachers,

teacher educators, and teacher preparation programmes in the similar context.

5.6.1 Implications for Participants

The implications of the findings for each participant are presented in the section. A brief summary of each participant's profile is analyzed and then the corresponding implications are elaborated.

Amy is a participant who covers the most dimensions (seven out of eight) in LAL conceptualization, self-evaluates LAL as satisfactorily qualified, constructs five sub-categories of assessment knowledge, reinforces AfL conceptions, and constructs her identity from assessee to assessor-to-be. Although she self-evaluates herself as satisfactorily qualified, there still are some improvement suggestions.

The finding reveals that she seems to neglect the assessment washback, which she seldomly mentions in LAL conceptualization as well as in the LAL improvement focus. In the accumulation of assessment knowledge base, two dimensions are rarely referred to in her interview: knowledge of grading and knowledge of peer & self-assessment. Therefore, it is strongly recommended that she needs to pay attention to the possible impact of assessment on learners, acquire knowledge of how to score and how to use rubrics, and makes efforts in empowering learners to engage in assessment practices, like peer or self-assessment. Additionally, the identity construction towards assessor may enable her to enhance LAL in a more self-directed way.

Betty conceptualizes LAL in five dimensions, with personal attribute as the most important component. She self-evaluates herself as marginally qualified, but still needs to improve in assessment theory and assessment methods to become a more effective EFL teacher. Her assessment knowledge is constructed in similar dimensions with Amy and lacks the knowledge of grading and knowledge of peer and self-assessment. Although she has some initial AfL conceptions and practices, AoL still dominates, for example she mentions the assessment may occupy 20% or 30% amount of teacher's time and a good teaching should be ended with a good assessment. She tends to separate assessment with teaching, and the function of assessment is to check the result of teaching and learning. Moreover, reflective learning seems to facilitate her identity construction towards an assessor.

According to the findings of the present study, apart from her self-diagnosed improvement focus in LAL, she is suggested not neglecting the pedagogical content knowledge, assessment principles and assessment washback. AfL conception needs to be reinforced by acknowledging the inseparable relationship between assessment and teaching and recognizing the agent role of students in assessment, which is also what the English Curriculum Reform in China advocates. In order to implement AfL well, she also needs to concentrate on the knowledge of grading and knowledge of conducting peer or self-assessment among learners to facilitate learning.

Compared with other participants, Carol's conceptualization of LAL covers the fewest dimensions (four out of eight). Though limited, she is the only participant who notices the washback of the assessment on courses and learners. She appears to focus on the more technical aspects of LAL, like assessment principle & criterion, assessment methods & implementation, assessment content & criterion, and washback which are also the improvement aspects self-diagnosed by herself. Nonetheless, she lacks the desire to further improve LAL because she is less motivated to become a teacher in the future. Largely due to the low willingness, she said she seldomly actively develops the necessary qualities of an effective EFL teacher, just passively follows all the arrangement designed by the university. Thus, she demonstrates a slow and passive identity construction towards assessor-to-be.

For Carol, except assessment-related competence, she is indicated to pay due attention to person-related and discipline-related competence in LAL to enlarge her existing conceptualization, which will surely benefit for her LAL improvement. Secondly, similar to Amy and Betty, Carol needs to construct assessment knowledge base especially in knowledge of how to provide feedback, how to interpret and communicate assessment results, and how to conduct peer or self-assessment among learners. Lastly, though she has little desire to become a teacher, as an undergraduate from Normal University, she is still advised to endeavor to be competent in teaching especially in assessment.

Daisy's conceptualization of LAL is featured by inclusion of affective dimension in assessment content, that is, not only the learning results but also the learning attitudes or learning process need to be included in teacher's assessment domains. She evaluates herself to be insufficiently qualified and wishes to get improvement in assessment theories and way of feedback. Besides, she is also inadequately prepared in knowledge of grading, assessment interpretation and communication, and involving learners in peer or self-assessment. In her assessment conception, AfL is evident since she encourages the multiple methods in assessing learning rather than the mere reliance on summative assessment (i.e., examination). Meanwhile, her identity is constructed gradually from being assessed to assessor-to-be during the pre-service stage. Yet, what is a regret for her is that she is absent from practicum.

The findings are enlightening for Daisy in several aspects. To begin with, her conceptualization of LAL fails to cover three dimensions: pedagogical content knowledge, assessment washback, and assessment interpretation and communication, which need to be incorporated into her partial coverage of LAL in order not to gain an unbalanced LAL. Secondly, she has AfL conception to some extent, but she still needs to learn more on how to implement these conceptions, such as how to use rubrics to engage learners in assessment, how to feedback, how to interpret and communicate the assessment results to stakeholders. All these assessment practices are critically necessary for her, who has not experienced practicum to hone her assessment skills in the real teaching contexts. Lastly, she is advised to reflect on the assessment-related

theories and practices more from the stance of a teacher, or more specifically an assessor, to become a reflective practitioner during LAL evolvement process.

Edwin, the only male participant, perceives the influence of social cultural context on EFL teachers' LAL. He conceptualizes LAL as a dynamic multi-dimensional concept, with the most importance in assessment methods & implementation. He seems to pay little attention to the pedagogical content knowledge and assessment washback. Motivated to be a TESOL postgraduate candidate, he desires to learn more in how to assess reading and listening, the receptive skills in English, compared with the productive skills, like writing and speaking. He often cites the assessment scales in productive skills (writing and speaking) from ELTS (English Language Testing Service) as examples whereas he mentions the assessment scales in receptive skills (reading and writing) is not that clear and explicit. With regard to his LAL evolvement trajectory, the knowledge in English discipline, assessment purpose & methods, feedback as well as the assessment ethics is constructed obviously. Meanwhile, AfL conception is reinforced by his agent role in assessment practices in practicum school, which dramatically facilitates his identity construction from assessor-to-be towards assessor.

Though he has obtained desirable LAL in a variety of aspects, the findings are still helpful for him. Like others, he is indicated to pay attention to the possible assessment washback, which he tends to neglect in LAL conceptualization. He seems not so

confident in LAL and still needs further improvement in knowledge of how to assess receptive skills, which he believes is more implicit to be assessed, knowledge of how to score and use rubrics, knowledge of how to implement peer or self-assessment, as well as knowledge of how to interpret and communicate assessment results with learners or other stakeholders. Deep involvement in teaching practicum plays a huge facilitative role in his LAL evolvement, especially in strengthening AfL conception and identity construction from assessor-to-be to assessor.

Similar to other participants, the last one, Flora covers part of the eight dimensions (five out of eight) of LAL. All the mentioned dimensions are labelled without discriminative importance. Besides, she also perceives the shaping effect of the social cultural context on LAL. Maybe due to the insufficient English proficiency, she self-evaluates herself as insufficiently qualified and intends to further improve in assessment methods and assessment result communication. Concerning LAL evolvement trajectory, she constructs the majority of knowledge base, strengthens the AfL conception for the benefit of learners, and constructs identity of being an assessor-to-be from an assessee in the teacher education programme.

The findings from the present study implicates that Flora is recommended to enlarge LAL conceptualization by adding pedagogical content knowledge, assessment washback, and personal conception in order to conceptualize LAL in an all-round way. Her motivation to improve the insufficient LAL comes from the wish of becoming an

effective and facilitative teacher to enhance learners' English proficiency. As to the LAL involvement, she seems not fully develop the knowledge of grading and knowledge of peer or self-assessment. Her assessment conception has the initial AfL trend although still dominated by AoL. She needs to develop a more balanced assessment conception between AoL and AfL, as advocated in English Curriculum. Her identity construction towards assessor-to-be is greatly facilitated by her self-reflection on the assessment practices in practicum in pre-service teacher education.

5.6.2 Implications for Pre-service EFL Teachers

For the pre-service EFL teachers in similar context, three implications can be drawn from the present study. Firstly, most of whom tend to self-evaluate themselves as inadequately prepared in LAL, they attribute such an insufficiency partly to internal factors: low willingness to become a teacher and late start in EFL learning. Thus, the pre-service EFL teachers are suggested to actively construct identity recognition towards teachers, or more specifically assessors. The assessor identity construction empowers them with enhanced agency to engage in sustainable assessment learning, either on language assessment or assessment in general education, which may lead to improved LAL.

Another implication for pre-service EFL teachers is to engage in active reflection on their own LAL. The reflection on LAL is more likely to direct them to acknowledge the influence of mediating factors and self-diagnose their need analysis on assessment

improvement. Consequently, to become a reflective practitioner in LAL enables the pre-service teachers to consolidate awareness that LAL is an indispensable part of effective teachers and highlight the significance of assessment in their professional development in order to attract their focus to assessment, which is long neglected or implicitly taught.

The last implication generated from the present study is that pre-service EFL teachers need to recognize the agent role of students in assessment by paying more attention to the knowledge deficit in peer assessment or self-assessment, especially against the assessment paradigm transition towards AfL. Besides, more accumulation in AfL methods, strategies, or implementation is also suggested based on the results of the current research.

5.6.3 Implications for Pre-service EFL Teacher Educators

As an important stakeholder in cultivating pre-service EFL teachers, teacher educators also need to be taken into account when addressing the issue of LAL improvement among the pre-service teachers. The teacher educators, especially those who are in charge of assessment-related course, are recommended to explicitly show assessment elements in lectures. For example, the adoption of assessment scaffoldings, like rubrics or exit slip, is encouraged to make the assessment criteria clearer and is also beneficial to cultivate the agent role of teacher candidates in assessment.

Additionally, such assessment practices adopted by teacher educators also play a modelling role in pre-service teachers' LAL. By being engaged in various assessment activities implemented by teacher educators, pre-service EFL teachers may tend to transfer the desirable assessment practices they have experienced to their future assessment tasks. In other words, the broader range of assessment practices teacher educators have implemented, the broader range of assessment practices teacher candidates will conduct when they become EFL teachers. Thus, teacher educators' LAL also needs to come into the researchers' horizon and to be improved so as to enrich pre-service teachers' assessment experiences and understanding of assessment in a range of assessment activities.

5.6.4 Implications for Pre-service EFL Teacher Preparation Programme

Pre-service EFL teacher preparation programme is indicated to be an influential contextual factor in mediating pre-service teachers' LAL evolvement. The inclusion of assessment-related courses and practices as well as the implementation of formative assessment policy and high-stake assessment exam in current study are revealed to be a facilitator in pre-service teachers' LAL evolvement. Although the programme sampled in the present study is evaluated as satisfying by participants, the pre-service teacher education programme still has much to be improved in developing pre-service teachers' LAL.

The first implication is concerned with emphasizing the construction of assessment-

related courses. The assessment course has not obtained its due attention yet and still deficiently offered in the programme. In the sampled X Normal University, the course on language testing and assessment is an optional course rather than a compulsory one for all pre-service EFL teachers. What's worse, due to various reasons, the course has not been available for graduates at that year. Thus, it is strongly recommended to be listed in the training scheme as a compulsory course for all the candidates to systematically construct assessment knowledge and explicitly understand assessment. Besides, other courses covering assessment, like English Pedagogy, should also rethink and redesign the teaching content to explicitly show the relationship with assessment within the AfL paradigm.

The second implication is pertaining to the effectiveness of assessment-related assessment practices (i.e., intensive teaching training and teaching practicum), which should be enhanced from several aspects. To begin with, the tutors in assessment-related practices are supposed to be not only skilled in EFL teaching skills but also literate in language assessment to provide professional assessment suggestions in guidance. Secondly, to guarantee the quality of intensive teaching training, it is suggested not to be arranged at the end of the term, when teacher candidates are busy with preparing for the final term exam. As a result, the conflict with exam preparation may distract pre-service teachers from attentive devotion to teaching training. Thirdly, the duration of such practices is expected to be prolonged, as indicated by the participants in present study. Finally, if possible, the teacher preparation programme is

strongly recommended to grant another opportunity for those who are absent from practicum like Daisy within the deadline of flexible educational system. Daisy in the present study missed the chance to participate in practicum, but she expressed her regret, and meanwhile admired other peers who had such an experience in practicum school by sharing with the researcher some interesting stories heard from her classmates. In addition, practicum is reported to be facilitating LAL evolvement greatly, especially in assessor identity construction by direct exposure to the complex teaching contexts where assessment is embedded. Thus, the participation of practicum needs to be given utmost importance in the training scheme in teacher education programmes.

The last implication generated from present study is that the weight of assessment should be correspondingly represented in National Teacher Qualification Certificate Exam. As a high-stake national exam covering assessment, it is recommended to test assessment-related knowledge in a larger proportion according to the critical position of LAL in EFL teacher professional development and in English Curriculum Standard issued in China. Moreover, the exam needs to be reformed from the existing test of rote memorizing the terminologies in a few multiple-choice items into integrated test of internalized assessment knowledge. Another improvement is welcomed to include assessment in other sections of the exam (e.g., the face-to-face interview) in order to direct the teacher candidates' internalized understanding of assessment.

5.7 Contribution to the Body of LAL Research

LAL research is still in its infantile stage and issues in LAL have not been completely resolved yet (Giraldo, 2018b; Jin, 2018; Kim et al., 2020; Yan & Fan, 2020). In this way, the current study contributes to the existing literature of LAL theoretically, methodologically, and pedagogically.

Theoretically, the voices of pre-service EFL teachers have been presented to supplement the inadequate focus on this critical stakeholder in LAL research scope, especially in the context of China, where LAL research is comparatively limited (Gan & Jiang, 2020). The pre-service EFL teachers' conceptualizations of LAL and self-evaluated proficiency level have been explored to provide a holistic picture of LAL from the long-neglected group.

What's more, it further improves Y. Xu and Brown's (2016) model by clarifying LAL evolvement trajectory among the pre-service teachers. Their framework merely points out the pre-service teachers go through three mastery levels from knowledge base through internalized assessment to assessor identity construction. However, how does LAL evolve in each level is not clear. This study provides original evidence that LAL evolvement in each level undergoes characterized development.

Another theoretical contribution is that it elucidates the interaction between LAL evolvement and mediating factors. The interaction relationship has been widely

acknowledged in the literature (E. Huang & Yang, 2019; Y. Xu & Brown, 2016; Xu & He, 2019), however, how do they interact is less researched. The findings from the current study have bettered the interaction relationship to address the paucity in LAL field.

Methodologically, it has enriched the research methods in LAL literature by offering empirical evidence that narrative inquiry, less commonly used in LAL field, is feasible to explore pre-service EFL teachers' assessment experiences. Narrative inquiry empowers researchers to explore participants' inner world in-depth, which seems to be irreplaceable by other qualitative methods.

Pedagogically, the findings provide pedagogical suggestions on how to improve LAL among pre-service EFL teachers in the assessment context like China. To maximize the efficiency in enhancing LAL, not only pre-service EFL teachers themselves, but also the parties involved need to collaborate together to underscore the significance of LAL and make the assessment education more effective.

5.8 Limitations of the Research

The current study has offered valid and trustworthy answers to the four research questions proposed and contributed to the existing body of LAL literature, particularly from pre-service EFL teachers in the context of China. However, there are three limitations which need to be pointed out.

The first limitation comes from the recruitment of only one male participant in this study. As revealed by previous studies (e.g., Alkharusi et al., 2012), gender may be a factor influencing the assessment knowledge and self-perception of assessment competence. However, the gender imbalance is widespread in English major, and the sampled X Normal University is no exception, which is overwhelmingly favored by female rather than by male students. In addition, based on the principle of voluntary participation, the number of male participants cannot be easily guaranteed. Therefore, it would be much better to recruit the same number of male participants as female ones for not over-representation of female voices.

The second limitation is related to the data collection period. The data were collected among the participants at the first term of their final year, when they had completed the large majority of the courses and practices, but were still having a very few disciplinary courses, like Advanced English. It would be better to collect the data at the last term of the final year when they had completed all the required courses offered in pre-service teacher education programme. The researcher believes if the data were collected at the time nearer to their graduation time, their evaluation of the effectiveness of the teacher preparation programme would be more accurate and comprehensive.

The last limitation is derived from the supplementary data collection that only three of

the six participants kept and submitted the reflective logs during the research process. Regrettably, due to the pressure of landing a job or preparing for the National Postgraduate Entrance Examination at December, the other three participants failed to keep the reflective logs. Though they argued they had expressed everything they intended to convey and had nothing to add, the reflective logs might help them to reflect on their assessment experiences and the questions put forwards in the interview with more details. Furthermore, owing to the participants' restricted involvement in initiating an assessment activity, the number of assessment artifacts designed by the participants was correspondingly limited, which might result in less evidence of how they conceptualized and implemented assessment.

5.9 Recommendations for the Future Research

Future research in LAL is still warranted in the following five aspects: extending research contexts, broadening the participants, adopting longitudinal research methods, collecting other sources of assessment performance, and clarifying the relationship among the mediating factors.

First of all, the present study was conducted in a pre-service teacher education programme situated in the western part of China, which was recognized as a less developed area with restricted assessment resources. The settings sampled in the current study may differ dramatically from teacher education programmes in other contexts, like other regions of China or other parts of the world. Thus, concerning the

LAL context-dependent nature, further studies are suggested to be carried out in the rest regions of China or outside China for a more comprehensive understanding of LAL in various assessment contexts.

Secondly, further studies are expected to broaden the research participants to focus on and compare LAL evolvement trajectories among novice teachers, proficient teachers, or expert teachers. Pre-service EFL teachers who are in the initial professional development stage is the focus of the current study. But, teachers in other professional development stages also need attention. For example, to what extent do they share similarities or to what extent do they differ in LAL evolvement is still unknown in the existing literature.

The third suggestion for future research originates from the research methods. The present study adopted retrospective narrative inquiry to explore the participants' assessment experiences and understandings. Besides, the longitudinal research method is also welcomed to collect data in exploration of LAL evolvement. Different research methods may supplement each other's shortcomings in uncovering an overall picture of LAL evolvement.

Fourthly, the studies conducted by incorporating a range of assessment artifacts are encouraged. A range of assessment artifacts designed by the participants can be drawn on to reveal their internalized assessment understandings or skills under the

negotiation among various contextual tensions.

Lastly, what is the relationship among the three mediating factors identified in the present study is still under-researched. The present study concludes three mediating factors from personal, experiential, and contextual dimensions. However, how do they interact with each other still needs more research to clarify the relationship among them.

5.10 Summary

In this chapter, the findings are discussed in relationship with the existing literature. Also, the implications originated from the current study are analyzed to maximize the efficiency in cultivating LAL among teacher candidates in the contexts like China. Besides, the contributions to the existing knowledge body of LAL are discussed from theoretical, methodological, and pedagogical dimensions.

Although it is enlightening in its original contributions, the study has a few limitations related to data collection. It ends with recommendations for directing future research in LAL from five topics: broaden the research contexts to other parts of China or outside China, extend the research participants to in-service teachers in various professional development stages, adopt longitudinal method, collect various forms of assessment artifacts, and clarify the interaction among the identified mediating factors.

RERERENCES

- Abrams, L., Varier, D., & Jackson, L. (2016). Unpacking instructional alignment: The influence of teachers' use of assessment data on instruction. *Perspectives in Education*, 34(4), 15-28. <https://doi.org/10.18820/2519593X/pie.v34i4.2>
- Alkharusi, H. A. (2011). An analysis of the internal and external structure of the teacher assessment literacy questionnaire. *International Journal of Learning*, 18(1), 515-528.
- Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2012). Educational assessment attitudes, competence, knowledge, and practices: An exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning*, 1(2), 217-232. <http://dx.doi.org/10.5539/jel.v1n2p217>
- Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and in-service teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*, 39(2), 113-123. <https://doi.org/10.1080/1359866X.2011.560649>
- Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, 20(1), 20-34. <https://doi.org/10.1080/0969594X.2012.736364>
- Allen, D. D., & Flippo, R. F. (2002). Alternative assessment in the preparation of literacy educators: Responses from students. *Reading Psychology*, 23(1), 15-26. <https://doi.org/10.1080/027027102317345385>
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association (AFT, NCME, & NEA). (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32. <https://doi.org/10.1111/j.1745-3992.1990.tb00391.x>
- Amirian, S. M. R., Pourfarhad, M., & Nafchi, A. M. (2016). Teachers' assessment literacy and its correlation with IELTS students' achievement in writing skill. *Theory and Practice in Language Studies*, 6(5), 994-999. <http://dx.doi.org/10.17507/tpls.0605.12>
- Andrews, M., Brown, R., & Mesher, L. (2018). Engaging students with assessment and feedback: improving assessment for learning with students as partners. *Practitioner Research in Higher Education*, 11(1), 32-46. <https://ojs.cumbria.ac.uk/index.php/prhe/article/view/460>

- Aria, D., Sukyadi, D., & Kurniawan, E. (2021). Teacher assessment literacy: Indonesian EFL secondary teachers' self-perceived on classroom-based assessment practice. *English Review: Journal of English Education*, 10(1), 15-26. <https://doi.org/10.25134/erjee.v10i1.5349>
- Arifin, S. R. M. (2018). Ethical considerations in qualitative study. *International Journal of Care Scholars*, 1(2), 30-33. <https://journals.iium.edu.my/ijcs/index.php/ijcs/article/view/82>
- Armstrong, M. C. F., Marín, M. I. R., & Matarroyo, L. M. (2004). Analysis of assessment instruments used in foreign language teaching. *Íkala*, 9(1), 171-201. <https://revistas.udea.edu.co/index.php/ikala/article/view/3146>
- Ashraf, H., & Zolfaghari, S. (2018). EFL Teachers' assessment literacy and their reflective teaching. *International Journal of Instruction*, 11(1), 425-436. <https://doi.org/10.12973/iji.2018.11129a>
- Ayres, L. (2008). Thematic coding and analysis. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 867-878). Sage. <https://dx.doi.org/10.4135/9781412963909.n451>
- Azis, A. (2012). Teachers' conceptions and use of assessment in student learning. *Indonesian Journal of Applied Linguistics*, 2, 40-52. <https://doi.org/10.17509/ijal.v2i1.72>
- Babaii, E., & Asadnia, F. (2019). A long walk to language assessment literacy: EFL teachers' reflection on language assessment research and practice. *Reflective Practice*, 20(6), 745-760. <https://doi.org/10.1080/14623943.2019.1688779>
- Bachor, D. G., & Anderson, J. O. (1994). Elementary teachers' assessment practices as observed in the province of British Columbia. *Assessment in Education: Principles, Policy & Practice*, 1(1), 63-94. <https://doi.org/10.1080/0969594940010105>
- Bachor, D. G., & Baer, M. R. (2001). An examination of preservice teachers' simulated classroom assessment practices. *Alberta Journal of Educational Research*, 47(3), 244-258. <https://journalhosting.ucalgary.ca/index.php/ajer/article/view/54877>
- Baker, B. A., & Riches, C. (2017). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 35(4), 557-581. <https://doi.org/10.1177/0265532217716732>
- Barkhuizen, G. (2016). A short story approach to analysing teacher (imagined) identities over time. *TESOL Quarterly*, 50(3), 655-683. <https://doi.org/10.1002/>

- Barthes, R. (1977). *Image-music-text* (S. Heath, Trans.). Glasgow: William Collins.
- Beitin, B. (2012). Interview and sampling. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (pp. 243-253). Sage. <https://dx.doi.org/10.4135/9781452218403.n17>
- Bell, J. S. (2002). Narrative inquiry: More than just telling stories. *TESOL quarterly*, 36(2), 207-213. <https://doi.org/10.2307/3588331>
- Berger, A. (2012). Creating language-assessment literacy: a model for teacher education. In J. Huttner, B. Mehlmauer-Larcher, S. Reichl, & B. Schiftner (Eds.), *Theory and practice in EFL teacher education: Bridging the gap* (pp. 57-82). Blue Ridge. <https://doi.org/10.21832/9781847695260-007>
- Berry, V., Sheehan, S., & Munro, S. (2017, October 28–29). *Beyond surveys: An approach to understanding effective classroom assessment practices* [Paper presentation]. IATEFL TEASIG/CRELLA Seminar, Luton, UK.
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers?. *ELT Journal*, 73(2), 113-123. <https://doi.org/10.1093/elt/ccy055>
- Bijsterbosch, E., Béneker, T., Kuiper, W., & van der Schee, J. (2019). Teacher professional growth on assessment literacy: A case study of prevocational geography education in the Netherlands. *The Teacher Educator*, 54(4), 420-445. <https://doi.org/10.1080/08878730.2019.1606373>
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L., & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117-140. <https://doi.org/10.1177/1478210314566733>
- Bishop, E. C., & Shepherd, M. L. (2011). Ethical reflections: Examining reflexivity through the narrative paradigm. *Qualitative Health Research*, 21(9), 1283-1294. <https://doi.org/10.1177/1049732311405800>
- Black, P. J. (1993). Formative and summative assessment by teachers. *Studies in Science Education*, 21, 49-97. <https://doi.org/10.1080/03057269308560014>
- 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), 9-

21. <https://doi.org/10.1177/003172170408600105>

Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *Curriculum Journal*, 16(2), 249-261. <https://doi.org/10.1080/09585170500136218>

Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research*, 12(2), 222-233. <http://dx.doi.org/10.17507/jltr.1202.02>

Bol, L., Stephenson, P. L., O'connell, A. A., & Nunnery, J. A. (1998). Influence of experience, grade level, and subject area on teachers' assessment practices. *The Journal of Educational Research*, 91(6), 323-330. <https://doi.org/10.1080/00220679809597562>

Bolívar, E. M. R. (2020). Monitoring preservice teachers' language assessment literacy development through journal writing. *Malaysian Journal of ELT Research*, 17(1), 38-52.

Boud, D. (2006). Foreword: Innovative assessment in higher education. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. xvii-xix). Routledge. <https://doi.org/10.4324/9780203969670>

Bozkurt, F. (2020). Teacher candidates' views on self and peer assessment as a tool for student development. *Australian Journal of Teacher Education*, 45(1), 47-60. <http://dx.doi.org/10.14221/ajte.2020v45n1.4>

Bracey, G. W. (2000). Thinking about tests and testing: A short primer in "assessment literacy". American Youth Policy Forum.

Braney, B. T. (2011). *An examination of fourth grade teachers' assessment literacy and its relationship to students' reading achievement* [Doctoral dissertation, University of Hartford]. ProQuest Dissertations and Theses Global.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper (Ed.), *APA handbook of research methods in psychology* (Vol. 2, pp. 57-71). American Psychological Association. <https://doi.org/10.1037/13620-004>

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Sage.

- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126-136). Cambridge University Press.
- Brislin, R. W. (1986). The wording and translation of research instruments. In W. L. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp.137-164). Sage.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy & Practice*, 11(1), 7-26. <https://doi.org/10.1080/0969594042000208976>
- Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles: research-based principles to guide classroom practice*. Assessment Reform Group.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Brown, G. T. (2011). Teachers' conceptions of assessment: comparing primary and secondary teachers in New Zealand. *Assessment Matters*, 3, 45-71. <https://doi.org/10.18296/am.0097>
- Brown, G. T., Harris, L. R., & Harnett, J. (2012). Teacher beliefs about feedback within an assessment for learning environment: Endorsement of improved learning over student well-being. *Teaching and Teacher Education*, 28(7), 968-978. <https://doi.org/10.1016/j.tate.2012.05.003>
- Brown, G. T., Hui, S. K., Flora, W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, 5(50), 307-320. <https://doi.org/10.1016/j.ijer.2011.10.003>
- Brown, G. T., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice*, 16(3), 347-363. <https://doi.org/10.1080/09695940903319737>
- Brunker, N., Spandagou, I., & Grice, C. (2019). Assessment for learning while learning to assess: Assessment in initial teacher education through the eyes of pre-service teachers and teacher educators. *Australian Journal of Teacher Education*, 44 (9),

89-109. <http://dx.doi.org/10.14221/ajte.2019v44n9.6>

Buckworth, J. (2017). Issues in the teaching practicum. In G. Geng, P. Smith, & P. Black (Eds.), *The challenge of teaching* (pp. 9-17). Springer.

Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, 38(3), 429-455. <https://doi.org/10.1177/0265532221992274>

Büyükkarcı, K. (2014). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction*, 7(1), 107-120.

Büyükkarcı, K. (2016). Identifying the areas for English language teacher development: A study of assessment literacy. *Pegem Eğitim ve Öğretim Dergisi*, 6(3), 333-346. <http://dx.doi.org/10.14527/pegegog.2016.017>

Cachia, R., Ferrari, A., Ala-Mutka, K., & Punie, Y. (2010). Creative learning and innovative teaching: *Final report on the study on creativity and innovation in education in EU member states*. Publications Office of the European Union.

Calveric, S. (2010). *Elementary teachers' assessment beliefs and practices* [Doctoral dissertation, Virginia Commonwealth University]. VCU Scholars Compass Open Access. <https://scholarscompass.vcu.edu/etd/2332/?mode=simple>

Campbell, C., & Collins, V. L. (2007). Identifying essential topics in general and special education introductory assessment textbooks. *Educational Measurement: Issues and Practice*, 26(1), 9-18. <https://doi.org/10.1111/j.1745-3992.2007.00084.x>

Campbell, C., & Evans, J. A. (2000). Investigation of preservice teachers' classroom assessment practices during student teaching. *The Journal of Educational Research*, 93(6), 350-55. <https://doi.org/10.1080/00220670009598729>

Castillo-Montoya, M. (2016). Preparing for interview research: The interview protocol refinement framework. *Qualitative Report*, 21(5), 811-831. <http://nsuworks.nova.edu/tqr/vol21/iss5/2>

Chai, Y. (2020). Jiyu fangtande xiaoxue xinshou jiaoshi ketang pingjia suyang yanjiu [Interview-based classroom assessment literacy research on novice teachers in primary schools]. *Shanghai Jiaoyu Pinggu Yanjiu*, 9(4), 63-67.

Chan, Y. C. (2008). Elementary school EFL teachers' beliefs and practices of multiple assessments. *Reflections on English Language Teaching*, 7(1), 37-62. <http://www.nus.sg/celec/research/books/relt/vol7/no1/37-62chan.pdf>

- Chan, C. K., & Luo, J. (2020). An exploratory study on teacher assessment literacy: do novice university teachers know how to assess students' written reflection. *Teachers and Teaching*, 1-15. <https://doi.org/10.1080/13540602.2020.1787375>
- Chappuis, S., & Stiggins, R. (2002). Classroom assessment for learning. *Educational Leadership*, 60(1), 40-44. <http://www.edtechpolicy.org/CourseInfo/edhd485/ClassroomAssessmentforLearning.pdf>
- Chappuis, S., & Stiggins, R. (2008). Finding balance: Assessment in the middle school classroom. *Middle Ground*, 12(2), 12-15.
- Chen, J., & Brown, G. T. (2013). High-stakes examination preparation that controls teaching: Chinese prospective teachers' conceptions of excellent teaching and assessment. *Journal of Education for Teaching*, 39(5), 541-556. <https://doi.org/10.1080/02607476.2013.836338>
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360-389. <https://doi.org/10.1191/0265532204lt288oa>
- Cheng, L., Rogers, W. T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, 33(1), 9-32. <https://doi.org/10.1080/02602930601122555>
- Cheng, L., & Sun, Y. (2015). Teachers' grading decision making: Multiple influencing factors and methods. *Language Assessment Quarterly*, 12(2), 213-233. <https://doi.org/10.1080/15434303.2015.1010726>
- Chew, A., & Lee, I. (2013, October 20-25). *Teachers' beliefs and practices of classroom assessment in Republic Polytechnic, Singapore* [Paper presentation]. The 39th Annual Conference of the International Association of Educational Assessment (IAEA) on Educational Assessment 2.0: Technology in Educational Assessment, Tel Aviv, Israel.
- Childs, R. A., & Lawson, A. (2003). What do teacher candidates know about large-scale assessments? What should they know?. *Alberta Journal of Educational Research*, 49(4), 354-367.
- Ch'ng, L. C., & Rethinasamy, S. (2013). English language assessment in Malaysia: teachers' practices in test preparation. *Issues in Language Studies*, 2(2), 24-39. <https://doi.org/10.33736/ils.1669.2013>
- Choi, J. (2017). Understanding elementary teachers' different responses to reform: The case of implementation of an assessment reform in South Korea. *International*

Electronic Journal of Elementary Education, 9(3), 581-598. <https://www.iejee.com/index.php/IEJEE/article/view/177>

Clandinin, D. J., & Connelly, F. M. (2000). *Narrative inquiry: Experience and story in qualitative research*. Jossey-Bass.

Clandinin, D. J., & Johnson, R. B. (2014). Narrative inquiry and case study research. In R. B. Johnson & L. Christensen (Eds.), *Educational research: Quantitative, qualitative, and mixed approaches* (pp. 417-441). Sage.

Clark, J. S. (2015). "My assessment didn't seem real": The influence of field experiences on preservice teachers' agency and assessment literacy. *Journal of Social Studies Education Research*, 6(2), 91-111. <https://dergipark.org.tr/en/pub/jsser/issue/19105/202755>

Coffey, A. (2014). Analysing documents. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 367-380). Sage. <https://dx.doi.org/10.4135/9781446282243.n25>

Conca, L. M., Schechter, C. P., & Castle, S. (2004). Challenges teachers face as they work to connect assessment and instruction. *Teachers and Teaching*, 10(1), 59-75. <https://doi.org/10.1080/13540600320000170927>

Conle, C. (2000). Narrative inquiry: Research tool and medium for professional development. *European Journal of Teacher Education*, 23(1), 49-63. <https://doi.org/10.1080/713667262>

Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher*, 19(5), 2-14. <https://doi.org/10.3102/0013189X019005002>

Contreras, E., & Gerardo, A. (2000). Self-storying, self-understanding: Toward a narrative approach to EFL teacher education. *TESOL Journal*, 9(3), 24-27.

Coombe, C. (2010). Assessing foreign/second language writing ability. *Education, Business and Society: Contemporary Middle Eastern Issues*, 3(3), 178-187. <https://doi.org/10.1108/17537981011070091>

Coombe, C., Troudi, S. & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyloff (Eds.). *The Cambridge guide to second language assessment* (1st ed., pp. 20-29). Cambridge University Press.

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: what do we need to learn, unlearn, and relearn?. *Language Testing in Asia*, 10(1), 1-16.

<https://doi.org/10.1186/s40468-020-00101-6>

- Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, 71, 134-144. <https://doi.org/10.1016/j.tate.2017.12.010>
- Cowie, B., Cooper, B., & Ussher, B. (2014). Developing an identity as a teacher-assessor: Three student teacher case studies. *Assessment Matters*, 7(6), 64-89. <https://doi.org/10.18296/am.0128>
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Sage.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods, approaches* (4th ed.). Sage.
- Crossman, J. (2004). Factors influencing the assessment perceptions of training teachers. *International Educational Journal*, 5(4), 582-590.
- Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28, 43-56. <https://doi.org/10.1016/j.asw.2016.03.001>
- Cumming, A. (2004). Broadening, deepening, and consolidating. *Language Assessment Quarterly*, 1(1), 5-18. https://doi.org/10.1207/s15434311laq0101_2
- Daniel, L. G., & King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *The Journal of Educational Research*, 91(6), 331-344. <https://doi.org/10.1080/00220679809597563>
- Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, 21(2), 149-166. <https://doi.org/10.1080/0969594X.2014.898128>
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8(1), 6-13. <https://doi.org/10.14507/epaa.v8n1.2000>
- Davidson, P., & Coombe, C. (2019). Language assessment literacy in the Middle East and North Africa (MENA) region. *Arab Journal of Applied Linguistics*, 4(2), 1-23.

- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-348. <https://doi.org/10.1177/0265532208090156>
- DeLuca, C., Chavez, T., & Cao, C. (2013). Establishing a foundation for valid teacher judgement on student learning: The role of pre-service assessment education. *Assessment in Education: Principles, Policy & Practice*, 20(1), 107-126. <https://doi.org/10.1080/0969594X.2012.668870>
- DeLuca, C., & Klinger, D. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438. <https://doi.org/10.1080/0969594X.2010.516643>
- DeLuca, C., Klinger, D., Pyper, J., & Woods, J. (2015). Instructional rounds as a professional learning model for systemic implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 22(1), 122-139. <https://doi.org/10.1080/0969594X.2014.967168>
- DeLuca, C., Klinger, D., Searle, M., & Schula, L. M. (2010). Developing a curriculum for assessment education. *Assessment Matters*, 2, 133-156. <https://doi.org/10.18296/am.0080>
- DeLuca, C., & Lam, C. Y. (2014). Preparing teachers for assessment within diverse classrooms: An analysis of teacher candidates' conceptualizations. *Teacher Education Quarterly*, 41(3), 3-24. <http://www.jstor.org/stable/teaceducquar.41.3.3>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248-266. <https://doi.org/10.1080/10627197.2016.1236677>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251-272. <https://doi.org/10.1007/s11092-015-9233-6>
- DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, 25(4), 355-375. <https://doi.org/10.1080/0969594X.2016.1244514>
- DeLuca, C., Willis, J., Cowie, B., Harrison, C., Coombs, A., Gibson, A., & Trask, S. (2019). Policies, programs, and practices: Exploring the complex dynamics of assessment education in teacher education across four countries. *Frontiers in*

- deMarrais, K. (2004). Qualitative interview studies: Learning through experience. In K. deMarrais & S. D. Lapan (Eds.), *Foundations for research: Methods of inquiry in education and the social sciences* (pp. 51-68). Lawrence Erlbaum Associates.
- Deneen, C. C., & Brown, G. T. (2011, November 23 – 28). *The persistence of vision: An analysis of continuity and change in conceptions of assessment within a teacher education program* [Paper presentation]. The 37th Annual Conference of the International Association for Educational Assessment (IAEA), Philippines, United States.
- Deneen, C. C., & Brown, G. T. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, 3(1), 1-14. <https://doi.org/10.1080/2331186X.2016.1225380>
- Deygers, B., & Malone, M. E. (2019). Language assessment literacy in university admission policies, or the dialogue that isn't. *Language Testing*, 36(3), 347-368. <https://doi.org/10.1177/0265532219826390>
- Díaz, C., Alarcón, P., & Ortiz, M. (2012). The English teacher: His beliefs about English language assessment at primary, secondary and tertiary levels. *Revista de lenguaje y cultura*, 17(1), 15-26.
- Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, 27(4), 307-329. [https://doi.org/10.1016/S0191-491X\(01\)00032-3](https://doi.org/10.1016/S0191-491X(01)00032-3)
- Ding, G., Li, M., Sun, M., Li, Y., Chen, J., Yang, F., Ju, Y., Zhuang, Y., Wang, J., Han, Y., Zhang, D., Chen, X., & Xun, Y. (2013). *Zhongguo gaodeng shifan yuanxiao shifansheng peiyang zhuangkuang diaocha yu fenxi baogao* [National survey and policy analysis on the pre-service education of student teachers in Normal universities and colleges]. East China Normal University Press.
- Djoub, Z. (2017). Assessment literacy: Beyond teacher practice. In R. Al-Mahrooqi, C. Coombe, F. Al-Maamari & V. Thakur (Eds.), *Revisiting EFL assessment* (pp. 9-27). Springer. https://doi.org/10.1007/978-3-319-32601-6_2
- Drever, E. (1995). *Using semi-structured interviews in small-scale research: A teacher's guide*. Scottish Council for Research in Education.
- Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 247-296). Elsevier. <https://doi.org/10.1016/B978-0-12->

- Dysthe, O. (2008). The challenges of assessment in a new learning culture. In A. Havnes & L. McDonald (Eds.), *Balancing dilemmas in assessment and learning in contemporary education* (pp. 15-32). Routledge.
- Earl, K., & Giles, D. (2011). An-other look at assessment: Assessment in learning. *New Zealand Journal of Teachers' Work*, 8(1), 11-20. <https://hdl.handle.net/10289/5766>
- Earl, L. (2003). *Assessment as learning*. Corwin.
- Earl, L. (2007). Assessment as learning. In W. D. Hawley (Ed.), *The keys to effective schools: Educational reform as continuous improvement* (2nd ed., pp. 85-98). Corwin.
- Earl, L. (2013). Assessment for learning; assessment as learning: Changing practices means changing beliefs. In Education Bureau (Ed.), *Assessment and learning* (pp. 1-5). The Hong Kong Government Printer.
- Earl, L., & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Western and Northern Canadian Protocol.
- East, M. (2008). *Dictionary use in foreign language writing exams: Impact and implications*. John Benjamins.
- Edwards, F. (2017). A rubric to track the development of secondary pre-service and novice teachers' summative assessment literacy. *Assessment in Education: Principles, Policy & Practice*, 24(2), 205-227. <https://doi.org/10.1080/0969594X.2016.1245651>
- Esfandiari, R., & Nouri, R. (2016). A mixed-methods, cross-sectional study of assessment literacy of Iranian university instructors: Implications for teachers' professional development. *Iranian Journal of Applied Linguistics*, 19(2), 115-154. <http://ijal.khu.ac.ir/article-1-2730-en.html>
- Esin, C., Fathi, M., & Squire, C. (2014). Narrative analysis: The constructionist approach. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 203-216). Sage. <https://dx.doi.org/10.4135/9781446282243.n14>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1-4.

- Fan, Y. C., Wang, T. H., & Wang, K. H. (2011). A web-based model for developing assessment literacy of secondary in-service teachers. *Computers & Education*, 57(2), 1727-1740. <https://doi.org/10.1016/j.compedu.2011.03.006>
- Farhady, H. (2019). A cross-contextual perspective on EFL teachers' assessment knowledge. *Yeditepe Üniversitesi Eğitim Fakültesi Dergisi*, 8(10), 1-19. <https://dergipark.org.tr/en/pub/edu7/issue/51553/665542>
- Farhady, H., & Tavassoli, K. (2018). Developing a language assessment knowledge test for EFL teachers: A data-driven approach. *Iranian Journal of Language Teaching Research*, 6(3), 79-94.
- Fink, A. S. (2000). The role of the researcher in the qualitative research process: A potential barrier to archiving qualitative data. Forum: *Qualitative Social Research*, 1(3), Article 4. <http://nbn-resolving.de/urn:nbn:de:0114-fqs000344>
- Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9(1), 1-14. <https://doi.org/10.1186/s40468-019-0078-7>
- Flick, U. (2018). Triangulation in data collection. In U. Flick (Ed.), *The SAGE handbook of qualitative data collection* (pp. 527-544). Sage. <https://dx.doi.org/10.4135/9781526416070.n34>
- Fontana, A., & Frey, J. H. (2008). The interview: From neutral stance to political involvement. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (3rd ed., pp. 115-160). Sage.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <https://doi.org/10.1080/15434303.2011.642041>
- Gan, L., & Jiang, C. (2020). Jiaoshi ceping suyang guowai yanjiu shuping [Literature review of research on teacher assessment literacy abroad]. *Zhonggugo kaoshi*, (1), 72-78.
- Gareis, C. R. (2007). Reclaiming an important teacher competency: The lost art of formative assessment. *Journal of Personnel Evaluation in Education*, 20, 17-20. <https://doi.org/10.1007/s11092-007-9044-5>
- Gareis, C. R., & Grant, L. W. (2015). Assessment literacy for teacher candidates: A focused approach. *Teacher Educators' Journal*, 4-21.
- Gebril, A. (2017). Language teachers' conceptions of assessment: An Egyptian

- perspective. *Teacher Development*, 21(1), 81-100. <https://doi.org/10.1080/13664530.2016.1218364>
- Gebril, A., & Brown, G. T. (2014). The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice*, 21(1), 16-33. <https://doi.org/10.1080/0969594X.2013.831030>
- Gee, J. P. (2003). Opportunity to learn: A language-based perspective on assessment. *Assessment in Education: Principles, Policy & Practice*, 10(1), 27-46. <https://doi.org/10.1080/09695940301696>
- Gielen, S., Dochy, F., & Dierick, S. (2003) Evaluating the consequential validity of new modes of assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards*. Kluwer Academic Publishers. https://doi.org/10.1007/0-306-48125-1_3
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. Falmer Press.
- Gipps, C. V. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24(1), 355–392. <https://doi.org/10.3102/0091732X024001355>
- Gipps, C. V., & Stobart, G. (2003). Alternative assessment. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation: Practice* (pp. 549–575). Kluwer. https://doi.org/10.1007/978-94-010-0309-4_33
- Giraldo, F. (2018a). A Diagnostic study on teachers' beliefs and practices in foreign language assessment. *Íkala, revista de lenguaje y cultura*, 23(1), 25-44. <https://orcid.org/0000-0001-5221-8245>
- Giraldo, F. (2018b). Language assessment literacy: Implications for language teachers. *Profile Issues in Teachers Professional Development*, 20(1), 179-195. <https://dx.doi.org/10.15446/profile.v20n1.62089>
- Giraldo, F. (2019). Language assessment practices and beliefs: Implications for language assessment literacy. *How*, 26(1), 35-61. <https://doi.org/10.19183/how.26.1.481>
- Giraldo, F. (2020). A post-positivist and interpretive approach to researching teachers' language assessment literacy. *Profile: Issues in Teachers' Professional Development*, 22(1), 189-200. <https://doi.org/10.15446/profile.v22n1.78188>

- Giraldo, F., & Murcia, D. (2018). Language assessment literacy for pre-service teachers: Course expectations from different stakeholders. *GiST Education and Learning Research Journal*, 16, 56-77. <https://doi.org/10.26817/16925777.425>
- Giraldo, F., & Murcia, D. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal*, 21(2), 243-259. <https://doi.org/10.14483/22487085.14514>
- Glesne, C. (2011). *Becoming qualitative researchers: An introduction* (4th ed.). Pearson.
- Gok, B., Erdogan, O., Altinkaynak, S. O., & Erdogan, T. (2012). Investigation of pre-service teachers' perceptions about concept of measurement and assessment through metaphor analysis. *Procedia-Social and Behavioral Sciences*, 46, 1997-2003. <https://doi.org/10.1016/j.sbspro.2012.05.417>
- Goldstein, S. E. (2017). Reflexivity in narrative research: Accessing meaning through the participant-researcher relationship. *Qualitative Psychology*, 4(2), 149-164. <https://doi.org/10.1037/qup0000035>
- Gonen, K., & Akbarov, A. (2017). Instructors' principles and practices in the implementation stage of classroom-based language assessment in higher education in Turkey. *Journal of European Education*, 5(3), 35-45. <http://www.eu-journal.org/index.php/JEE/article/view/29>
- Gonzales, R., & Aliponga, J. (2012). Classroom assessment preferences of Japanese language teachers in the Philippines and English language teachers in Japan. *MEXTESOL Journal*, 36 (1), 1-19. <http://dx.doi.org/10.2139/ssrn.2746287>
- Gotch, C. M., & McLean, C. (2019). Teacher outcomes from a statewide initiative to build assessment literacy. *Studies in Educational Evaluation*, 62, 30-36. <https://doi.org/10.1016/j.stueduc.2019.04.003>
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21(6), 607-621. <https://doi.org/10.1016/j.tate.2005.05.001>
- Grainger, P. R., & Adie, L. (2014). How do pre-service teacher education students move from novice to expert assessors? *Australian Journal of Teacher Education*, 39(7), 89-105. <https://doi.org/10.14221/ajte.2014v39n7.9>
- Gu, P. Y. (2014). The unbearable lightness of the curriculum: what drives the assessment practices of a teacher of English as a foreign language in a Chinese secondary school?. *Assessment in Education: Principles, Policy & Practice*, 21(3),

286-305. <https://doi.org/10.1080/0969594X.2013.836076>

Gubrium, J. F., & Holstein, J. A. (1998). Narrative practice and the coherence of personal stories. *Sociological quarterly*, 39(1), 163-187. <https://doi.org/10.1111/j.1533-8525.1998.tb02354.x>

Hakim, B. (2015). English language teachers' ideology of ELT assessment literacy. *International Journal of Education and Literacy Studies*, 3(4), 42-48. <http://dx.doi.org/10.7575/aiac.ijels.v.3n.4p.42>

Hamidi, E. (2010). Fundamental issues in L2 classroom assessment practices. *Academic Leadership*, 8(2), 1-17. <https://scholars.fhsu.edu/alj/vol8/iss2/21>

Han, T., & Kaya, H. I. (2014). Turkish EFL teachers' assessment preferences and practices in the context of constructivist instruction. *Journal of Studies in Education*, 4(1), 77-93. <https://doi.org/10.5296/jse.v4i1.4873>

Harding, L., & Brunfaut, T. (2020). Trajectories of language assessment literacy in a teacher-researcher partnership: Locating elements of praxis through narrative inquiry. In M. E. Poehner & O. Inbar-Lourie (Eds.), *Toward a reconceptualization of second language classroom assessment* (pp. 61-81). Springer. https://doi.org/10.1007/978-3-030-35081-9_4

Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 413-429). De Gruyter Mouton.

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education*, 35(2), 213-224. <https://doi.org/10.1080/03057640500146880>

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>

Hassan, M., & Coombe, C. (2020). Preface: language assessment literacy. *Language Testing in Asia*, 10(1). <https://doi.org/10.1186/s40468-020-00103-4>

Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: Expectations and needs of pre-service English language teachers. *ELT Research Journal*, 4(2), 111-128. <https://dergipark.org.tr/en/pub/eltrj/issue/28780/308006>

- Hildén, R., & Fröjdendahl, B. (2018). The dawn of assessment literacy—exploring the conceptions of Finnish student teachers in foreign languages. *Language*, 12(1), 1-24. <https://doi.org/10.17011/apples/urn.201802201542>
- Hill, K. (2017). Understanding classroom-based assessment practices: a precondition for teacher assessment literacy. *Papers in Language Testing and Assessment*, 6(1), 1-17.
- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395-420. <https://doi.org/10.1177/0265532211428317>
- Hill, M. F., Ell, F. R., & Evers, G. (2017). Assessment capability and student self-regulation: The challenge of preparing teachers. *Frontiers in Education*, 2(21). <https://doi.org/10.3389/feduc.2017.00021>
- Howley, M. D., Howley, A., Henning, J. E., Gilla, M. B., & Weade, G. (2013). Intersecting domains of assessment knowledge: School typologies based on interviews with secondary teachers. *Educational Assessment*, 18(1), 26-48. <https://doi.org/10.1080/10627197.2013.761527>
- Huai, N., Braden, J. P., White, J. L., & Elliott, S. N. (2006). Effect of an Internet-based professional development program on teachers' assessment literacy for all students. *Teacher Education and Special Education*, 29(4), 244-260. <https://doi.org/10.1177/088840640602900405>
- Huang, E., & Yang, M. (2019). Qingjing xuexi yu Shijian gongtongti kuangjiaxia de pingjia suyang fazhan yanjiu — “Xingping xinshou” daxue yingyu jiaoshi weili [Assessment literacy development research in the framework of contextual learning and practice community— Taking formative assessment novice university English teachers as an example]. *Waiyujie*, (6), 27-35.
- Huang, R., & Jiang, L. (2020). Authentic assessment in Chinese secondary English classrooms: teachers' perception and practice. *Educational Studies*, 1-14. <https://doi.org/10.1080/03055698.2020.1719387>
- Hudson, M. E. (2017). *Examining the relationship between selected grade 3-12 teachers' perceived assessment literacy and their classroom assessment practices* [Doctoral dissertation, Old Dominion University]. ODU Digital Commons Open Access. https://digitalcommons.odu.edu/cgi/viewcontent.cgi?article=1039&context=efl_etds
- Hughes, G. (2014) *Ipsative Assessment: Motivation through marking progress*. Palgrave Macmillan.

- Hussain, S. (2017). *Relationship of teacher educators' assessment literacy and classroom assessment practices with their students' academic achievement* [Doctoral dissertation, International Islamic University]. PASTIC Dspace Repository Open Access. <http://pr.hec.gov.pk/jspui/bitstream/123456789/9939/1/Assessment%20Literacy%20and%20practices%20of%20teacher%20educators%20PhD%20dissertation.pdf>
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402. <https://doi.org/10.1177/0265532208090158>
- Inbar-Lourie, O. (2012). Language assessment literacy. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-9). John Wiley & Sons.
- Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., pp. 257-270). Springer. https://doi.org/10.1007/978-3-319-02261-1_19
- Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy & Practice*, 16(2), 185-204. <https://doi.org/10.1080/09695940903075958>
- Inbar-Lourie, O., & Levi, T. (2020). Assessment literacy as praxis: Mediating teacher knowledge of assessment-for-learning practices. In M. E. Poehner & O. Inbar-Lourie (Eds.), *Toward a reconceptualization of second language classroom assessment* (pp. 241-259). Springer. https://doi.org/10.1007/978-3-030-35081-9_12
- Islam, M. S., & Stapa, M. (2019). Investigating the potentials of classroom assessment: A critical discussion. *International Journal of Engineering and Advanced Technology*, 8(5), 1172-1179. <http://www.dx.doi.org/10.35940/ijeat.E1165.0585C19>
- Izci, K., & Siegel, M. A. (2018). Investigation of an alternatively certified new high school chemistry teacher's assessment literacy. *International Journal of Education in Mathematics Science and Technology*, 7(1), 1-19. <https://doi.org/10.18404/ijemst.473605>
- James, M., McCormick, R., Black, P., Carmichael, P., Drummond, M J., Fox, A., MacBeath, J., Marshall, B., Pedder, D., Procter, R., Swaffield, S., Swann, J., & Wiliam, D. (2007). *Improving learning how to learn, in classrooms, schools and networks* (1st ed.). Routledge.

- Jane, S. M. (2012). Primary school teachers' conceptions of classroom assessment: A qualitative study. *International Journal of Cross-Disciplinary Subjects in Education*, 3(2), 699-706.
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teacher Education*, 6(2), 26-37.
- Jawhar, S. S., & Subahi, A. M. (2020). The impact of specialty, sex, qualification, and experience on teachers' assessment literacy at Saudi Higher Education. *International Journal of Learning, Teaching and Educational Research*, 19(5), 200-216. <https://doi.org/10.26803/ijlter.19.5.12>
- Jenkins, R. (1992). *Pierre bourdieu*. Routledge.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers?. *Language Testing*, 30(3), 345-362. <https://doi.org/10.1177/0265532213480334>
- Jia, Y., Eslami, Z. R., & Burlbaw, L. M. (2006). ESL teachers' perceptions and factors influencing their use of classroom-based reading assessment. *Bilingual Research Journal*, 30(2), 407-430. <https://doi.org/10.1080/15235882.2006.10162883>
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555-584. <https://doi.org/10.1177/0265532209351431>
- Jin, Y. (2018). Waiyu jiaoshi pingjia suyang fazhan: lilun kuangjia he lujing tansuo [Foreign language teachers' assessment literacy development: Theoretical framework and routine exploration]. *Waiyu jiaoyu yanjiu qianyan*, (2), 65-72.
- Jin, Y., & Yan, M. (2017). Computer literacy and the construct validity of a high-stakes computer-based writing assessment. *Language Assessment Quarterly*, 14(2), 101-119. <https://doi.org/10.1080/15434303.2016.1261293>
- Joachim, C., Hammann, M., Carstensen, C. H., & Bögeholz, S. (2020). Modeling and measuring pre-service teachers' assessment literacy regarding experimentation competences in Biology. *Education Sciences*, 10(5), 1-27. <https://doi.org/10.3390/educsci10050140>
- Joffe, H. (2011). Thematic analysis. In D. Harper & A. R. Thompson (Eds.), *Qualitative methods in mental health and psychotherapy: A guide for students and practitioners* (pp. 209-223). Wiley. <http://dx.doi.org/10.1002/9781119973249.ch15>
- Josselson, R. (2007). The ethical attitude in narrative research: Principles and

- practicalities. In D. J. Clandinin (Ed.), *Handbook of narrative inquiry: Mapping a methodology* (pp. 537-566). Sage. <https://doi.org/10.4135/9781452226552.n21>
- Josselson, R. (2010). Narrative research. In N. Salkind (Ed.), *Encyclopedia of research design* (pp. 869-875). Sage. <http://dx.doi.org/10.4135/9781412961288>
- Kahl, S., Hofman, P., & Bryant, S. (2013). *Assessment literacy standards and performance measures for teacher candidates and practicing teachers*. Measured Progress.
- Kalajahi, S. A. R., & Abdullah, A. N. (2016). Assessing assessment literacy and practices among lecturers. *Pedagogika*, 124(4), 232-248. <http://dx.doi.org/10.15823/p.2016.65>
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Praeger.
- Kanjee, A., & Mthembu, J. (2015). Assessment literacy of foundation phase teachers: An exploratory study. *South African Journal of Childhood Education*, 5(1), 142-168. <https://doi.org/10.4102/sajce.v5i1.354>
- Kaur, A., Noman, M., & Awang-Hashim, R. (2018). Exploring and evaluating differentiated assessment practices of in-service teachers for components of differentiation. *Teaching Education*, 30(2), 160-176. <https://doi.org/10.1080/10476210.2018.1455084>
- Kavaklı, N., & Arslan, S. (2019). Towards a continuum from know-how to show-how for developing EFL student-teachers' assessment literacy. *International Online Journal of Education and Teaching*, 6(1), 223-232. <http://www.iojet.org/index.php/IOJET/article/view/593>
- Kelly, M. P., Feistman, R., Dodge, E., Rose, A. S., & Littenberg-Tobias, J. (2020). Exploring the dimensionality of self-perceived performance assessment literacy (PAL). *Educational Assessment, Evaluation and Accountability*, 499-517. <https://doi.org/10.1007/s11092-020-09343-7>
- Kemmis, S., & McTaggart, R. (2003). Participatory action research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 336-396). Sage.
- Kılıçkaya, F. (2016, October 6). *EFL teachers' formal assessment practices based on exam papers* [Paper presentation]. The 2nd International Conference on Language Research and Teaching in the Global Economic Environment, Kyiv, Ukraine.
- Kim, J. H. (2016). *Understanding narrative inquiry: The crafting and analysis of*

stories as research. Sage.

- Kim, A. A., Chapman, M., Kondo, A., & Wilmes, C. (2020). Examining the assessment literacy required for interpreting score reports: A focus on educators of K–12 English learners. *Language Testing*, 37(1):54-75. <https://doi.org/10.1177/0265532219859881>
- Kiomrs, R., Abdolmehdi, R., & Rashidi, N. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Teaching*, 4(1), 156-161.
- Kirkgoz, Y., Babanoglu, M. P., & Agçam, R. (2017). Turkish EFL teachers' perceptions and practices of foreign language assessment in primary education. *Journal of Education and E-Learning Research*, 4(4), 163-170. <https://doi.org/10.20448/journal.509.2017.44.163.170>
- Kizlik, B. (2012) Measurement, Assessment and Evaluation in Education. Retrieved from July, 20, 2020, from <http://www.adprima.com/measurement.htm>
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, 16(3), 263-268.
- Klinger, D. A., Volante, L., & DeLuca, C. (2012). Building teacher capacity within the evolving assessment culture in Canadian education. *Policy Futures in Education*, 10(4), 447-460. <https://doi.org/10.2304/pfie.2012.10.4.447>
- Koch, T., & Harrington, A. (1998). Reconceptualizing rigour: The case for reflexivity. *Journal of Advanced Nursing*, 28, 882–890. <https://doi.org/10.1046/j.1365-2648.1998.00725.x>
- Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22(3), 255-276. <https://doi.org/10.1080/10476210.2011.593164>
- Koh, K., & DePass, C. (2019). Developing teachers' assessment literacy: Multiple perspectives in action. In K. Koh, C. DePass, & S. Steel (Eds.), *Developing teachers' assessment literacy: A tapestry of ideas and inquiries* (pp. 1-6). Brill Sense. https://doi.org/10.1163/9789004385672_001
- Koh, K., & Luke, A. (2009). Authentic and conventional assessment in Singapore schools: An empirical study of teacher assignments and student work. *Assessment in Education: Principles, Policy & Practice*, 16(3), 291-318. <https://doi.org/10.1080/09695940903319703>

- Koh, K., & Tan, C. (2016). Promoting reflection in pre-service teachers through problem-based learning: An example from Canada. *Reflective Practice*, 17(3), 347-356. <https://doi.org/10.1080/14623943.2016.1164683>
- Koh, K. H., Tan, C., & Ng, P. T. (2012). Creating thinking schools through authentic assessment: The case in Singapore. *Educational Assessment, Evaluation and Accountability*, 24(2), 135-149. <https://doi.org/10.1007/s11092-011-9138-y>
- Koh, K. H., & Velayutham, R. L. (2009). Improving teachers' assessment literacy in Singapore schools: An analysis of teachers' assessment tasks and student work. *Research Brief*, 9(2), 1-4. <http://hdl.handle.net/10497/4433>
- Kremmel, B., & Harding, L. (2019, March 9-12). *Exploring the language assessment literacy of SLA researchers* [Paper presentation]. The 2019 Conference of the American Association for Applied Linguistics (AAAL), Atlanta, United States.
- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100-120. <https://doi.org/10.1080/15434303.2019.1674855>
- Kruse, L., Impellizeri, W., Witherel, C. E., & Sondergeld, T. A. (2020). Evaluating the impact of an assessment course on preservice teachers' classroom assessment literacy and self-Efficacy. *Mid-Western Educational Researcher*, 32(2), 107-132.
- Kuze, M. W., & Shumba, A. (2011). An investigation into formative assessment practices of teachers in selected schools in Fort Beaufort in South Africa. *Journal of Social Sciences*, 29(2), 159-170. <https://doi.org/10.1080/09718923.2011.11892966>
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Sage.
- Kvasova O. (2022). The Impact of Student Teachers' Pre-Existing Conceptions of Assessment on the Development of Language Assessment Literacy within an LTA Course. *Languages*, 7(1):62-84. <https://doi.org/10.3390/languages7010062>
- Kyttälä, M., Björn, P. M., Rantamäki, M., Närhi, V., & Aro, M. (2021). Assessment conception patterns of Finnish pre-service special needs teachers: The contribution of prior studies and teaching experience. *European Journal of Special Needs Education*, 1-15. <https://doi.org/10.1080/08856257.2020.1853972>
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197. <https://doi.org/10.1177/0265532214554321>

- 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>
- Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, 61, 112-122. <https://doi.org/10.1016/j.stueduc.2019.03.003>
- Laren, L. V., & James, A. (2008). Selected teachers' understanding of assessment four years after the implementation of the new assessment policy. *Africa Education Review*, 5(2), 288-303. <https://doi.org/10.1080/18146620802449944>
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, 63(3), 19-24.
- Lee, J. E., & Son, J. W. (2015). Two teacher educators' approaches to developing preservice elementary teachers' mathematics assessment literacy: Intentions, outcomes, and new learning. *Teaching and Learning Inquiry*, 3(1), 47-62. <https://doi.org/10.2979/teachlearningqu.3.1.47>
- Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168-182. <https://doi.org/10.1080/15434303.2019.1692347>
- Levy-Vered, A., & Alhija, F. (2015). Modelling beginning teachers' assessment literacy: The contribution of training, self-efficacy, and conceptions of assessment. *Educational Research and Evaluation*, 21, 378-406. <https://doi.org/10.1080/13803611.2015.1117980>
- Lian, L. H., & Yew, W. T. (2016). A framework for examining assessment literacy of preservice teachers. *US-China Education Review*, 6(5), 294-300.
- Lian, L. H., Yew, W. T., & Meng, C. C. (2014). Enhancing Malaysian teachers' assessment literacy. *International Education Studies*, 7(10), 74-81. <http://dx.doi.org/10.5539/ies.v7n10p74>
- Lieblich, A., Tuval-Mashiach, R., & Zilber, T. (1998). *Narrative research: Reading, analysis, and interpretation*. Sage.
- Lin, D. (2019). *Zhongxiaoxue yingyu jiaoshi yuyan pingjiasuyang cankao kuangjia* [A reference framework for language assessment literacy of primary and secondary English teachers]. Foreign Language Teaching and Research Press.

- Lin, D., & Su, Y. (2015). An investigation of Chinese middle school in-service English teachers' assessment literacy. *Indonesian EFL Journal*, 1(1), 1-10. <https://doi.org/10.25134/ieflj.v1i1.607>
- Lin, D., & Wu, Z. (2014). Guowai yuyan pingjia suyang yanjiude zuixin jinzhan [The latest development of foreign language assessment literacy research]. *Xiandai waiyu*, 37 (5), 711-720 +731.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Sage.
- Liu J., & Xu, Y. (2017). Assessment for learning in English language classrooms in China: Contexts, problems, and solutions. In H. Reinder, D. Nunan, & Z. Bin. (Eds.), *Innovation in language learning and teaching* (pp. 17-37). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-60092-9_2
- Looney, A., Cumming, J., Van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442-467. <https://doi.org/10.1080/0969594X.2016.1268090>
- López, A., & Bernal, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile Issues in Teachers Professional Development*, 11(2), 55-70.
- Luthfiyyah, R., Basyari, I. W., & Dwiniasih, D. (2020). EFL secondary teachers' assessment literacy: Assessment conceptions and practices. *Journal on English as a Foreign Language*, 10 (2), 402-421. <https://doi.org/10.23971/jefl.v10i2.2101>
- Maclellan, E. (2001). Assessment for learning: the differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, 26(4), 307-318. <https://doi.org/10.1080/02602930120063466>
- Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education*, 20(5), 523-535. <https://doi.org/10.1016/j.tate.2004.04.008>
- Mahapatra, S. K. (2016). Tracking the development of teachers' language assessment literacy: A case study. In G. Pickering & P. Gunashekar (Eds.), *Ensuring quality in English language teacher education* (pp. 106-114). British Council.
- Malmqvist, J., Hellberg, K., Möllås, G., Rose, R., & Shevlin, M. (2019). Conducting the pilot study: A neglected part of the research process? Methodological findings supporting the importance of piloting in qualitative research studies. *International Journal of Qualitative Methods*, 18, 1-11. <https://doi.org/10.1177/160940691987834>

- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between test-ers and users. *Language Testing*, 30(3), 329-344. <https://doi.org/10.1177/0265532213480129>
- Marshall, B., & Drummond, J. M. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education*, 21(2), 133-149. <https://doi.org/10.1080/02671520600615638>
- Matovu, M., & Madziah Zubairi, A. (2014). Self-perceived assessment competencies and practices among university lecturers. *Journal of Applied Research in Higher Education*, 6(2), 269-284. <https://doi.org/10.1108/JARHE-04-2013-0020>
- May, T., & Perry, B. (2014). Reflexivity and the practice of qualitative research. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 109-122). Sage.
- McGee, J., & Colby, S. (2014). Impact of an assessment course on teacher candidates' assessment literacy. *Action in Teacher Education*, 36, 522-532. <https://doi.org/10.1080/01626620.2014.977753>
- McKellar, E. J. K. (2002, August 28-30). *Change our assessment practices? Why should we? The theory behind assessment practices* [Paper presentation]. The Learning Communities and Assessment Cultures Conference, Newcastle, U.K.
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, 87(1), 40-49.
- McMillan, J. H., & Nash, S. (2000, April 25-27). *Teachers' classroom assessment and grading decision making* [Paper presentation]. The Annual Meeting of the National Council of Measurement in Education, New Orleans, United States.
- Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi*, 168 (1), 43-60. <https://www.ceeol.com/search/article-detail?id=548906>
- Mellati, M., & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education*, 43(6), 1-19. <http://dx.doi.org/10.14221/ajte.2018v43n6.1>
- Mertler, C. A. (1998, October 14-17) *Classroom assessment practices of Ohio teachers* [Paper presentation]. The Annual Meeting of the Mid-Western Educational Research Association, Chicago, United States.

- Mertler, C. A. (2003, October 15-18). *Preservice versus inservice teachers' assessment literacy: does classroom experience make a difference?* [Paper presentation]. The Annual Meeting of the Mid-Western Educational Research Association, Columbus, United States.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving schools*, 12(2), 101-113. <https://doi.org/10.1177/1365480209105575>
- Mertler, C. A., & Campbell, C. (2005, April 11-15). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory* [Paper presentation]. The Meeting of the American Educational Research Association, Montreal, Canada.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). MacMillan.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. (3rd ed.). Sage.
- Ministry of Education. (2011). *Jiaoshi jiaoyu kecheng biao zhun* [Teacher education curriculum standards (Trial)]. http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/s6049/201110/xxgk_125722.html
- Ministry of Education. (2012). *Zhongxiaoxue jiaoshi zhuan ye fazhan biao zhun ji zhidao—yingyu* [The standards and guidance for primary and secondary teacher professional development--English]. Beijing Normal University Press.
- Ministry of Education. (2018). *Putong gaozhong yingyu kecheng biao zhun* [Curriculum standards for the English course in middle schools]. Renmin Education Press.
- Mishler, E. G. (1986) *Research interviewing: Context and narrative*. Harvard University Press.
- Moen, T. (2006). Reflections on the narrative research approach. *International Journal of Qualitative Methods*, 5(4), 56-69. <https://doi.org/10.1177/160940690600500405>
- Mohammadi, E., Kiany, G. R., Samar, R. G., & Akbari, R. (2015). Appraising pre-

- service EFL teachers' assessment in language testing course using revised Bloom's taxonomy. *International Journal of Applied Linguistics and English Literature*, 4(4), 8-20. <http://dx.doi.org/10.7575/aiac.ijalel.v.4n.4p.8>
- Montee, M., Bach, A., Donovan, A., & Thompson, L. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of the National Council of Less Commonly Taught Languages*, 14, 1-32.
- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education*, 30(1), 109-162. <https://doi.org/10.3102/0091732X030001109>
- Munasinghe, H. K. (2020, March 5-6). *Language assessment literacy of teachers of English: implications for assessing speaking and listening skills of students in the classroom* [Paper presentation]. Proceedings of the Sri Lankan English Language Teacher Educator Conference (SLELTEC), MAS Athena, Sri Lanka.
- Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile Issues in Teachers Professional Development*, 14 (1), 143-158. <https://www.redalyc.org/articulo.oa?id=169224317010>
- Murray, M. (2018). Narrative data. In U. Flick (Ed.), *The SAGE handbook of qualitative data collection* (pp. 264-279). Sage. <https://dx.doi.org/10.4135/9781526416070.n17>
- Napanoy, J. B., & Peckley, M. K. (2020). Assessment literacy of public elementary school teachers in the indigenous communities in northern Philippines. *Universal Journal of Educational Research*, 8(11B), 5693-5703. <http://doi.org/10.13189/ujer.2020.082203>
- Newfields, T. (2006). Teacher development and assessment literacy. In T. Newfields, I. Gledall, M. Kawate-Mierzejewska, Y. Ishida, M. Chapman, & P. Ross (Eds.), *Authentic communication: Proceedings of the 5th Annual JALT Pan-SIG Conference* (pp. 48-73). Tokai University Press.
- Newfields, T. (2007, August 24-25). *Engendering assessment literacy: Narrowing the gap between teachers and testers* [Paper presentation]. The 4th International Conference of Korea English Language Testing Association, Seoul, Korea.
- Nimehchisalema, V., & Bhattib, N. (2019). A review of literature on language assessment literacy in last two decades (1999-2018). *Assessment*, 8(11), 44-59.
- Nimehchisalem, V., & Hussin, N. I. S. M. (2018). Postgraduate students' conception o

- f language assessment. *Language Testing in Asia*, 8(1), 1-14. <https://doi.org/10.1186/s40468-018-0066-3>
- Noonan, B., & Renihan, P. (2006). Demystifying assessment leadership. *Canadian Journal of Educational Administration and Policy*, (56), 1-21.
- Norris, J. M. (2006). The issue: The why (and how) of assessing student learning outcomes in college foreign language programs. *The Modern Language Journal*, 90(4), 576-583. https://doi.org/10.1111/j.1540-4781.2006.00466_2.x
- Odo, D. M. (2016). An investigation of the development of pre-service teacher assessment literacy through individualized tutoring and peer debriefing. *Journal of Inquiry and Action in Education*, 7(2), 31-61.
- Ogan-Bekiroglu, F. (2009). Assessing assessment: Examination of pre-service physics teachers' attitudes towards assessment and factors affecting their attitudes. *International Journal of Science Education*, 31(1), 1-39. <https://doi.org/10.1080/09500690701630448>
- Ogan-Bekiroglu, F., & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice. *Curriculum Journal*, 25(3), 344-371. <https://doi.org/10.1080/09585176.2014.899916>
- Ölmezer-Öztürk, E., & Aydın, B. (2019). Investigating language assessment knowledge of EFL teachers. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34(3), 602-620. <https://dx.doi.org/10.16986/HUJE.2018043465>
- O'Loughlin, K. (2006). Learning about second language assessment: Insights from a postgraduate student on-line subject forum. *University of Sydney Papers in TESOL*, 1, 71-85.
- O'Sullivan, R. G., & Johnson, R. L. (1993, April 12-16). *Using performance assessments to measure teachers' competence in classroom assessment* [Paper presentation]. The Annual Meeting of the American Educational Research Association, Atlanta, United States.
- Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128-138. <https://doi.org/10.1016/j.tate.2019.05.003>
- Paterno, J. (2001). Measuring success: A glossary of assessment terms. In Building

- cathedrals: Compassion for the 21st century. Retrieved from July 15, 2020, from <http://www.tcsdk12.org/math/Documents/measuringsuccess.html>
- Pedder, D., & James, M. (2012). Professional learning as a condition for assessment for learning. In J. Garnder (Ed.), *Assessment and learning* (pp. 33-48). Sage. <https://doi.org/10.13140/2.1.2600.5284>
- Pfeiffer-Hoens, M. (2017). *An investigation of charter schools' school leader and teacher level of assessment literacy* [Doctoral dissertation, Middle Tennessee State University]. MTSU Repository Open Access. <https://jewlscholar.mtsu.edu/handle/mtsu/5368>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381-402. <https://doi.org/10.1177/0265532213480337>
- Pishghadam, R., & Shayesteh, S. (2012). Conceptions of assessment among Iranian EFL Teachers. *The Iranian EFL Journal*, 8(3), 9-23. <https://profdoc.um.ac.ir/paper-abstract-1028212.html>
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-27.
- Plake, B. S., & Impara, J. C. (1993, April 13-15). *Teacher assessment literacy: Development of training modules* [Paper presentation]. The Annual Meeting of the National Council on Measurement in Education, Atlanta, United States.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12. <https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>
- Polit, D. F., & Beck, C. T. (2014). *Essentials of nursing research appraising evidence for nursing practice* (8th ed.). Wolters Kluwer.
- Polkinghorne, D. E. (2007). Validity issues in narrative research. *Qualitative Inquiry*, 13(4), 471-486. <https://doi.org/10.1177/1077800406297670>
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental?. *Theory into Practice*, 48(1), 4-11. <https://doi.org/10.1080/00405840802577536>
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273. <https://doi.org/10.1080/08878730.2011.605048>

- Portelli, L., & O'Sullivan, K. A. (2016). Teachers' perceptions of the influence of assessment on their teaching of year 9 English. *English in Australia*, 51(1), 71-80.
- Prasetyo, A. H. (2018). Language assessment literacy development: A student-teacher's experiences in teaching practice program. *LLT Journal: A Journal on Language and Language Teaching*, 21(2), 199-206.
- Primeau, L. (2003). Reflections on self in qualitative research: Stories of family. *American Journal of Occupational Therapy*, 57(1), 9-16. <https://doi.org/10.5014/ajot.57.1.9>
- Prizovskaya, V. (2018). Identifying New Jersey teachers' assessment literacy as precondition for implementing student growth objectives. *Journal of Education and Learning*, 7(1), 184-207. <https://doi.org/10.5539/jel.v7n1p184>
- Puspawati, I. (2019). Understanding teachers' knowledge, skills, and principles on language assessment: A survey on teachers' language assessment literacy. In A. Nurmandi, Y. Chen, N. Ismail, & Z. Rafique (Eds.) *The Third International Conference on Sustainable Innovation 2019–Humanity, Education and Social Sciences (IcoSIHESS 2019)* (pp. 70-75). Atlantis Press. <https://doi.org/10.2991/icosihess19.2019.11>
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142–173. <https://doi.org/10.1191/0265532205lt300oa>
- Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. *Qualitative Research in Accounting & Management*, 8(3), 238-264. <https://doi.org/10.1108/11766091111162070>
- Quilter, S. M. (2000). *Inservice teachers' assessment literacy and attitudes toward assessment* [Unpublished doctoral dissertation]. University of South Carolina.
- Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator*, 36 (2), 115-131. <https://doi.org/10.1080/08878730009555257>
- Ragchaa, J. (2019). English language teachers' view on assessment and their reflection on teaching practices: A Mongolian case study. *European Journal of Educational Sciences*, 6(2), 1-14. <http://dx.doi.org/10.19044/ejes.v6no2a1>
- Redecker, C., & Johannessen, O. (2013). Changing assessment—Towards a new assessment paradigm using ICT. *European Journal of Education*, 48(1), 79-96. <https://doi.org/10.1111/ejed.12018>

- Riessman, C. K. (2001). Analysis of personal narratives. In J. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 695-710). Sage. <https://dx.doi.org/10.4135/9781412973588.n40>
- Riessman, C. K. (2005) Narrative analysis. In N. Kelly, C. Horrocks, K. Milnes, B. Roberts, & D. Robinson (Eds.), *Narrative, memory and everyday life* (pp. 1-7). University of Huddersfield Press. <http://eprints.hud.ac.uk/id/eprint/4920>
- Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment, Research, and Evaluation*, 15(1), 1-10. <https://doi.org/10.7275/psk5-mf68>
- Rodríguez-Gómez, G., & Ibarra-Saiz M. S. (2015). Assessment as learning and empowerment: Towards sustainable learning in higher education. In M. Peris-Ortiz & J. M. Merigó Lindahl (Eds.), *Sustainable learning in higher education* (pp. 1-20). Springer. https://doi.org/10.1007/978-3-319-10804-9_1
- Roeber, E. (2016, May 5). *REL Mid-Atlantic 2016 educator effectiveness* [Webinar]. Regional Educational Laboratory Mid-Atlantic. <https://files.eric.ed.gov/fulltext/ED571502.pdf>
- Rogers, W. T. (1993). Principles for fair student assessment practices for education in Canada. *Canadian Journal of School Psychology*, 9(1), 110-127. <https://doi.org/10.1177/082957358500900111>
- Rogers, W. T., Cheng, L., & Hu, H. (2007). ESL/EFL instructors' beliefs about assessment and evaluation. *Comparative and International Education*, 36(1), 39-61. <https://doi.org/10.5206/cie-eci.v36i1.9088>
- Rogler, D. (2014). Assessment literacy: Building a base for better teaching and learning. *English Teaching Forum*, 52(3), 2-13.
- Roulston, K. (2014). Analysing interviews. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 297-312). Sage.
- Roulston, K., & Choi, M. (2018). Qualitative interviews. In U. Flick (Ed.), *The SAGE handbook of qualitative data collection* (pp. 233-249). Sage. <http://dx.doi.org/10.4135/9781526416070.n15>
- Sahinkarakas, S. (2012). The role of teaching experience on teachers' perceptions of language assessment. *Procedia-Social and Behavioral Sciences*, 47, 1787-1792. <https://doi.org/10.1016/j.sbspro.2012.06.901>

- Sanders, J. R., & Vogel, S. R. (1993). The development of standards for teacher competence in educational assessment of students. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp.41-62). Burros Institute of Mental Measurements.
- Saputra, R., Hamied, A., & Suherdi, D. (2020). The development of beliefs and practices of language assessment literacy: does a professional learning community help?. *Journal of Education for Teaching*, 46(3), 414-416. <https://doi.org/10.1080/02607476.2020.1761250>
- Savin-Baden, M., & Niekerk, L. V. (2007). Narrative inquiry: Theory and practice. *Journal of Geography in Higher Education*, 31(3), 459-472. <https://doi.org/10.1080/03098260601071324>
- Scarino, A. (2009). Assessing intercultural capability in learning languages: Some issues and considerations. *Language Teaching*, 42(1), 67-80. <https://doi.org/10.1017/S0261444808005417>
- Scarino, A. (2010). Assessing intercultural capability in learning languages: A renewed understanding of language, culture, learning, and the nature of assessment. *The Modern Language Journal*, 94(2), 324-329. <https://doi.org/10.1111/j.1540-4781.2010.01026.x>
- 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. <https://doi.org/10.1177/0265532213480128>
- Schwandt, T. A. (2007). *The SAGE dictionary of qualitative inquiry* (3rd ed.). Sage.
- Seden, K., & Svaricek, R. (2018). Teacher subjectivity regarding assessment: Exploring English as a foreign language teachers' conceptions of assessment theories that influence student learning. *Centre for Educational Policy Studies Journal*, 8(3), 119-139. <https://doi.org/10.26529/cepsj.500>
- Seidman, I. E. (1991). *Interviewing as qualitative research*. Teachers College Press.
- Serafini, F. (2000). Three paradigms of assessment: Measurement, procedure, and inquiry. *The Reading Teacher*, 54(4), 384-393.
- Sevimel-Sahin, A. & Subasi, G. (2019). An overview of language assessment literacy research within English language education context. *Kuramsal Eğitim Bilim Dergisi*, 12(4), 1340-1364. <https://doi.org/10.30831/akukeg.501817>
- Shahzamani, M., & Tahririan, M. H. (2021). Iranian medical ESP practitioners' reading

- comprehension assessment literacy: Perceptions and practices. *Iranian Journal of English for Academic Purposes*, 10(1), 1-15.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational researcher*, 29(7), 4-14. <https://doi.org/10.3102/0013189X029007004>
- Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 623–646). Praeger.
- Shepard, L., Hammerness, K., Darling-Hammond, L., Rust, F., Baratz Snowden, J., Gordon, E., Gutierrez, C., & Pacheco, A. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275-326). Jossey-Bass.
- Shin, L. W. (2015). Teachers' assessment literacies and practices: Developing a professional competency and learning framework. *Advances in Scholarship of Teaching and Learning*, 2(2), 1-20.
- Shulha, L. M. (1999). Understanding novice teachers thinking about student assessment. *Alberta Journal of Educational Research*, 45(3), 288-303. <http://136.159.200.199/index.php/ajer/article/view/54698>
- Sisman, E. P., & Buyukkarci, K. (2019). A review of foreign language teachers' assessment literacy. *Sakarya University Journal of Education*, 9(3), 628-650. <https://doi.org/10.19126/suje.621319>
- Smith, L. F., Hill, M. F., Cowie, B. & Gilmore, A. (2014). Preparing teachers to use the enabling power of assessment. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.). *Designing assessment for quality learning* (pp. 303–323). Springer. https://doi.org/10.1007/978-94-007-5902-2_19
- Soh, K. C., & Zhang, L. (2018). Teacher assessment literacy scale design and validation. In K. C. Soh. (Ed.) *Teaching Chinese language in Singapore: Efforts and possibilities* (pp. 179-191). Springer. https://doi.org/10.1007/978-981-10-8860-5_13
- Stabler-Havener, M. L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Working Papers in Applied Linguistics & TESOL*, 18(1), 1-22. <https://doi.org/10.7916/salt.v18i1.1195>
- Stewart, T. A., & Houchens, G. W. (2014). Deep impact: How a job-embedded formative assessment professional development model affected teacher practice. *Qualitative Research in Education*, 3(1), 51-82. <http://dx.doi.org/10.4471/qre.2014.36>

- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-39.
- Stiggins, R. J. (1993). Teacher training in assessment: Overcoming the neglect. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 27-40). Erlbaum.
- Stiggins, R. J. (1994). *Student-centered classroom assessment*. Merrill.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- Stiggins, R. J. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86(1), 22-27. <https://doi.org/10.1177/003172170408600106>
- Stiggins, R. J. (2007a). Assessment through the student's eyes. *Educational Leadership*, 64 (8), 22-26.
- Stiggins, R. J. (2007b). Five assessment myths and their consequences. *Education Week*, 27 (8), 28-29.
- Stiggins, R. J. (2014). Improve assessment literacy outside of schools too. *Phi Delta Kappan*, 96(2), 67-72. <https://doi.org/10.1177/0031721714553413>
- Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). Classroom assessment for student learning: Doing it right, using it well. Assessment Training Institute.
- Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22(4), 271-286. <https://doi.org/10.1111/j.1745-3984.1985.tb01064.x>
- Stiggins, R. J., & Conklin, N. F. (1988). *Teacher training in assessment*. Northwest Regional Education Laboratory.
- Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, 26(3), 233-246. <https://doi.org/10.1111/j.1745-3984.1989.tb00330.x>
- Stobaugh, R. R., Tassell, J. L., & Norman, A. D. (2010). Improving preservice teacher preparation through the Teacher Work Sample: Exploring assessment and analysis of student learning. *Action in Teacher Education*, 32(1), 39-53. <https://doi.org/10.1080/01626620.2010.10463541>
- Stoynoff, S. (2012). Looking backward and forward at classroom-based language

- assessment. *ELT Journal*, 66(4), 523–532. <https://doi.org/10.1093/elt/ccs041>
- Suah, S. L., & Ong, S. L. (2012). Investigating assessment practices of in-service teachers. *International Online Journal of Educational Sciences*, 4(1), 91-106.
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 1-14. <https://doi.org/10.1186/s40468-019-0077-8>
- Taber, K. S., Riga, F., Brindley, S., Winterbottom, M., Finney, J., & Fisher, L. G. (2011). Formative conceptions of assessment: trainee teachers' thinking about assessment issues in English secondary schools. *Teacher Development*, 15(2), 171-186. <https://doi.org/10.1080/13664530.2011.571500>
- Tajeddin, Z., Alemi, M., & Yasaei, H. (2018). Classroom assessment literacy for speaking: Exploring novice and experienced English language teachers' knowledge and practice. *Iranian Journal of Language Teaching Research*, 6(3), 57-77.
- Talib R., Ghafar M.N.A., Naim H.A. (2014) Assessment literacy: A catalyst to the success of school-based assessment in Malaysian schools. In A. Kasim, W. S. Atikah, N. Razak, N. Musa, R. Halim, & S. Mohamed (Eds.), *Proceedings of the International Conference on Science, Technology and Social Sciences (ICSTSS) 2012* (pp. 197-202). Springer.
- Tatto, M. T. (2006). Education reform and the global regulation of teachers' education, development and work: A cross-cultural analysis. *International Journal of Educational Research*, 45, 231-241. <https://doi.org/10.1016/j.ijer.2007.02.003>
- Tavassoli, K., & Farhady, H. (2018). Assessment knowledge needs of EFL teachers. *Teaching English Language*, 12(2), 45-65. http://www.teljournal.org/article_74116html
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. <https://doi.org/10.1177/0265532213480338>
- Thomas, M. (2012). Teachers' beliefs about classroom assessment and their selection of classroom assessment strategies. *Journal of Research & Reflections in Education*, 6 (2).104-115.
- Tomaszewski, L. E., Zarestky, J., & Gonzalez, E. (2020). Planning qualitative research: Design and decision making for new researchers. *International Journal of Qualitative Methods*, 19. <https://doi.org/10.1177/1609406920967174>

- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education*, 14(3), 281-294. <https://doi.org/10.1080/09695940701591867>
- Tracy, S. J. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. (1st ed.). John Wiley & Sons.
- Tran, T. H. (2012, October 19-20). *Second language assessment for classroom teachers* [Paper presentation]. MIDTESOL-2012, Iowa Rolla, United States.
- Tsagari, D. (2016). Assessment orientations of state primary EFL teachers in two Mediterranean countries. *Center for Educational Policy Studies Journal*, 6(1), 9-30. <https://www.cepsj.si/index.php/cepsj/article/view/102>
- 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.
- Ukrayinska, O. (2018). Developing student teachers' classroom assessment literacy: The Ukrainian context. In S. Hidri (Ed.), *Revisiting the assessment of second language abilities: From theory to practice* (pp. 351-371). Springer. https://doi.org/10.1007/978-3-319-62884-4_17
- Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, 15(3), 398-405. <https://doi.org/10.1111/nhs.12048>
- Van Nes, F., Abma, T., Jonsson, H., & Deeg, D. (2010). Language differences in qualitative research: is meaning lost in translation?. *European Journal of Ageing*, 7(4), 313-316. <https://doi.org/10.1007/s10433-010-0168-y>
- Vandeyar, S., & Killen, R. (2007). Educators' conceptions and practice of classroom assessments in post-apartheid South Africa. *South African Journal of Education*, 27(1), 101-115. <https://www.ajol.info/index.php/saje/article/view/25101>
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402. <https://doi.org/10.1080/15434303.2014.960046>
- Vogt, K., Tsagari, D., Csépes, I., Green, A., & Sifakis, N. (2020). Linking learners' perspectives on language assessment practices to teachers' assessment literacy enhancement (TALE): Insights from four European countries. *Language Assessment Quarterly*, 17(4), 410-433. <https://doi.org/10.1080/15434303.2020>

- Volante, L. (2010). Assessment of, for, and as learning within schools: Implications for transforming classroom practice. *Action in Teacher Education*, 31(4), 66-75. <https://doi.org/10.1080/01626620.2010.10463536>
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749-770. <https://journals.sfu.ca/cje/index.php/cjerce/article/view/2973>
- Volante, L., & Melahn, C. (2005). Promoting assessment literacy in teachers: Lessons from the Hawaii school assessment liaison program. *Pacific Educational Research Journal*, 13(1), 19-34.
- Wallace, M., & White, T. (2014). Secondary Mathematics preservice teachers' assessment perspectives and practices: An evolutionary portrait. *Mathematics Teacher Education and Development*, 16 (2), 25-45.
- Wang, J. R., Kao, H. L., & Lin, S. W. (2010). Preservice teachers' initial conceptions about assessment of science learning: The coherence with their views of learning science. *Teaching and Teacher Education*, 26 (3), 522-529. <https://doi.org/10.1016/j.tate.2009.06.014>
- Wang, T. H., Wang, K. H., & Huang, S. C. (2008). Designing a web-based assessment environment for improving pre-service teacher assessment literacy. *Computers & Education*, 51(1), 448-462. <https://doi.org/10.1016/j.compedu.2007.06.010>
- Watmani, R., Asadollahfam, H., & Behin, B. (2020). Demystifying language assessment literacy among high school teachers of English as a foreign language in Iran: Implications for teacher education reforms. *International Journal of Language Testing*, 10(2), 129-144. http://www.ijlt.ir/article_118025.html
- Webster, L., & Mertova, P. (2007). *Using narrative inquiry as a research method: An introduction to using critical event narrative analysis in research on learning and teaching*. Routledge.
- Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, 16(3), 194-209. <https://doi.org/10.1016/j.jslw.2007.07.004>
- Weinstein, C. S. (1988). Preservice teachers' expectations about the first year of teaching. *Teaching and Teacher Education*, 4(1), 31-40. [https://doi.org/10.1016/0742-051X\(88\)90022-4](https://doi.org/10.1016/0742-051X(88)90022-4)

- White, E. (2009). Are you assessment literate? Some fundamental questions regarding effective classroom-based assessment. *OnCUE Journal*, 3(1), 3-25.
- Wicking, P. (2017) The assessment beliefs and practices of English teachers in Japanese universities. *JLTA Journal*, 20, 76–89. <https://doi.org/10.20622/jltajournal.20.076>
- William, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment* (pp. 53-82). Routledge. <https://doi.org/10.4324/9781315086545-3>
- Williams, J. C. (2015). “Assessing Without Levels”: Preliminary research on assessment literacy in one primary school. *Educational Studies*, 41(3), 341-346. <https://doi.org/10.1080/03055698.2015.1007926>
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers’ assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241-256. <https://doi.org/10.1007/s13384-013-00899>
- Wilson, L. (1994, April 4-8). *A theoretical framework linking beliefs with assessment practices in school mathematics: Assessment reforms in search of a theory* [Paper presentation]. The Annual Meeting of the American Educational Research Association, New Orleans, United States.
- Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations* (Document R-01-3). University of Washington. <https://www.education.uw.edu/ctp/sites/default/files/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf>
- Wise, S. L., & Lukin, L. E. (1993). Measurement training in Nebraska teacher education programs. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp.187-202). Buros Institute of Mental Measurements.
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers’ judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, policy & practice*, 17(1), 59-75. <https://doi.org/10.1080/09695940903565610>
- Xu, H. (2017). Exploring novice EFL teachers’ classroom assessment literacy development: A three-year longitudinal study. *The Asia-Pacific Education Researcher*, 26 (3-4), 219-226. <https://doi.org/10.1007/s40299-017-0342-5>
- Xu, Y. (2013). Waiyu jiaoshi ketang pinggu suzhi yanjiu shuping [Reviews on foreign

- language teachers' classroom assessment literacy]. *Waiyu ceshi yu jiaoxue*, (4), 42-50.
- Xu, Y. (2019a). English language teacher assessment literacy in practice. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 517-539). Springer.
- Xu, Y. (2019b). How teacher conceptions of assessment mediate assessment literacy: A case study of a university English teacher in China. In K. Bailey & R. Damerow (Eds.), *Global perspectives on language assessment: Research, theory, and practice* (pp. 197-211). Routledge. <https://doi.org/10.4324/9780429437922-14>
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Xu, Y., & Brown, G. T. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133-158.
- Xu, Y & He, L (2019) How pre-service teachers' conceptions of assessment change over practicum: Implications for teacher assessment literacy. *Frontiers in Education*, 4. <https://doi.org/10.3389/educ.2019.00145>
- Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia-Social and Behavioral Sciences*, 116, 2998-3004. <https://doi.org/10.1016/j.sbspro.2014.01.696>
- Yan, B., Hackett, R. K., & Xu, J. (2017). The development of assessment literacy in Chinese pre-service primary teachers. In *the 2nd International Conference on Education, E-learning and Management Technology (EEMT 2017)* (pp. 44-48). DEStech Publications. <https://doi.org/10.12783/dtssehs/eemt2017/14400>
- Yan, X., & Fan, J. (2020). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 1-28. <https://doi.org/10.1177/0265532220929924>
- Yan, X., Zhang, C., & Fan, J. J. (2018). "Assessment knowledge is important, but...": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158-168. <https://doi.org/10.1016/j.system.2018.03.003>
- Yetkin, R., & Özer, Z. (2020). An investigation into in-service and pre-service English teachers' conception of assessment. *Journal of Theoretical Educational Science*, 13 (2), 384-396. <https://doi.org/10.30831/akukeg.582886>

- Yin, R. K. (2015). *Qualitative research from start to finish*. Guilford publications.
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.
- Young, J. E., & Jackman, M. G. A. (2014). Formative assessment in the Grenadian lower secondary school: Teachers' perceptions, attitudes and practices. *Assessment in Education: Principles, Policy & Practice*, 21(4), 398-411. <https://doi.org/10.1080/0969594X.2014.919248>
- Yu, G., & Jin, Y. (2014). English language assessment in China: Policies, practices and impacts. *Assessment in Education: Principles, Policy & Practice*, 21(3), 245–250. <https://doi.org/10.1080/0969594X.2014.937936>
- Zeng, Y. (2016). *Zhongguo jiaoshi jiaoyushi* [The History of Teacher Education in China]. The Commercial Press.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323-342. https://doi.org/10.1207/S15324818AME1604_4
- Zhang, C., & Yan, X. (2018). Assessment literacy of secondary EFL teachers: Evidence from a regional EFL test. *Chinese Journal of Applied Linguistics*, 41(1), 25-46. <https://doi.org/10.1515/cjal-2018-0002>
- Zhao, S. (2020). Xiaoxue jiaoshi ketang pingjia suyang de xianzhuang yanjiu-jiyu shanghai shi Y qu 1032 ming xiaoxue jiaoshi de diaocha yu fenxi [A research on the present situation of primary school teachers' classroom assessment literacy—Based on the survey of 1032 primary teachers in District Y in Shanghai City]. *Shanghai jiaoyu keyan*, (8), 53-59.
- Zheng, D. (2010). Zhongxiaoxue jiaoshi pingjia suyang zhuangkuang: laizi Z sheng de baogao [The situation of primary and middle school teachers' assessment literacy: A report from province Z]. *Quanqiu jiaoyu zhanwang*, 39(2), 31-36+42.
- Zolfaghari, F., & Ahmadi, A. (2016). Assessment literacy components across subject matters. *Cogent Education*, 3(1), 1-16. <https://doi.org/10.1080/2331186X.2016.1252561>
- Zulaiha, S., Mulyono, H., & Ambarsari, L. (2020). An Investigation into EFL Teachers' assessment literacy: Indonesian teachers' perceptions and classroom practice. *European Journal of Contemporary Education*, 9 (1), 189-201. <https://doi.org/10.13187/ejced.2020.1.189>

APPENDICES

Appendix A: Informed Consent for the Institute

I am Geng juanjuan, a PhD candidate in Universiti Utara Malaysia (UUM), who is currently working on my doctoral thesis titled “**Narrative inquiry of language assessment literacy development of pre-service EFL teachers in China**”. To achieve the objectives of the study, I am very grateful if I would be granted the opportunity to conduct the study in the institute.

Purpose of the research: The primary purpose of the study is to explore the language assessment literacy development among the pre-service EFL teachers in China by using the method of narrative inquiry, thus a few pre-service EFL teachers in the university will be selected purposefully by the researcher to participate in the research.

Nature of participation: This research will involve the interviews with the pre-service EFL teachers in the foreign language department by the researcher, collection of the documents (e.g., talent training scheme, teaching syllabus of the assessment-related courses, and documents related to the teaching practicum) and assessment artifacts (e.g., the teaching plans) designed by the participants.

Confidentiality: All the identifying information will be excluded from the research and the name of the university and the participants will be completely anonymous in data analysis and reporting.

Voluntary participation: The decision to participate in the study is completely voluntary.

Benefits: Through this study, the pre-service EFL teachers will be more reflective of the assessment-related issues.

Contact information: If you have any questions, you can contact me through any of the following:

Name of the researcher: Geng juanjuan

Address: Universiti Utara Malaysia

06010 UUM Sintok

Kedah Druai Aman, Malaysia

Tel: 0177266280 (Malaysia); 15877405611(China)

E-mail: mylovegjj123@gmail.com

QQ: 106304615

Signature of dean : _____ Date _____ (Day/month/year)

Official stamp: _____

Appendix B: Informed Consent for Participants

Informed Consent

I am Geng juanjuan, a PhD candidate in Universiti Utara Malaysia (UUM), who is currently working on my doctoral thesis titled “**Narrative inquiry of language assessment literacy development of pre-service EFL teachers in China**”. To achieve the objectives of the study, I am very grateful if you would participate in the research.

Purpose of the research: The primary purpose of the study is to explore the language assessment literacy development trajectory among the pre-service EFL teachers in China, thus your experience about the language assessment can contribute much to my understanding of language assessment literacy.

Nature of participation: This research will involve your participation in four times of interviews during the interval of approximately one week. Each of the interviews will take about 60 minutes at your convenient time.

Confidentiality: Your responses are completely anonymous. No personal identifying information will be used when discussing and reporting data.

Voluntary participation: Your decision to participate in the study is completely voluntary. If you decide not to participate in the study, it will not affect your study, benefits, or services on the campus. If you decide to participate in the study, you may withdraw from your participation at any time without penalty.

Benefits: Through this study, you will be more reflective in assessment-related issues and all the participants will receive a gift after the study for your time and energy.

If you have any questions, you can contact me through any of the following:

Name of the researcher: Geng juanjuan

Address: Universiti Utara Malaysia

06010 UUM Sintok

Kedah Dual Aman, Malaysia

Tel: 0177266280 (Malaysia); 15877405611(China)

E-mail: mylovegjj123@gmail.com QQ: 106304615

CONSENT

I have read and I understand the provided information and have had the opportunity to ask questions. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason and without cost. I understand that I will be given a copy of this consent form. I voluntarily agree to take part in this study.

Signature of Participant : _____ Date _____ (Day/month/year)

Signature of Researcher: _____ Date _____ (Day/month/year)

Appendix C: Interview Protocol

Interview Protocol

(Round 1)

The purpose of round 1 interview is to collect background information of the participants and their conceptions of assessment and LAL.

Part A: Ask questions to gather background information of the participants. (e.g., age, years of learning English, English proficiency level, experience of involvement in tutoring or teaching, the reason for becoming a teacher candidate, and future career choice).

Part B: Conception of assessment and LAL

1. What is your definition of assessment? What does assessment mean to you?
Can you use a metaphor to explain? Like assessment is _____.
What is a GOOD language test/assessment? What is an ineffective one? Why?
2. Is assessment important? Why?
What is the relationship between assessment and teaching & learning? How?
In what ways do you think assessment affects students?
What should be assessed in the English class?
What are the results from assessment useful for?
Who should make decisions regarding assessment in the English class?
What do you think of teacher's role in assessment?
3. Have you ever used the E-assessment tools as assessor or assessee? (e.g., automated online writing grading software, vocabulary volume testing APP).
If yes, how do you know these APPs? (from teachers, peers or books...?) What is your feeling and comment? Will you use for your students in your teaching in the future?
If no, will you like to learn more about them? Why and how?
4. Do you think it is necessary for pre-service teachers to learn about methods of assessing student's learning? Why or why not?
5. What kinds of skills and knowledge does an English teacher in primary and middle school need in order to carry out good assessments? Or what the EFL teachers should know and be able to do in assessing students? Or what quality should EFL teachers possess in conducting a good assessment?
As for the things you mentioned in your response to question 5, would you say that there are some skills or knowledge that are important than others, or are they all equally important?

6. What's your understanding of LAL?
7. In your opinion, what does LAL mean to a qualified EFL teacher? If a teacher is not good at designing, implementing and interpreting the assessment, do you think that really matters? Why?

Interview Protocol

(Round 2)

The purpose of the 2nd round interview is to gather their past experience of being assessed in primary and middle schools or training institutions if any.

1. Can you briefly introduce your primary school? And the English class and English teacher?

What was your experience of being assessed in the subject of English or other subjects in primary school?

(e.g., whether teachers publicized the test results and ranking, face-to-face homework checking and feedback, communicated with your parents about your learning)

Any impressive stories related to assessment design, assessment result communication and exchange, and parents' concern about the assessment?

How do you comment on your experience of being assessed? Are you satisfied with the assessment practice implemented in primary school?

Is there anything you want to add to your assessment experience in primary school?

2. Can you briefly introduce your junior middle school? And the English class and English teacher?

What was your experience of being assessed in the subject of English or other subjects junior middle school?

Any impressive stories related to assessment design, assessment result communication and exchange, and parents' concern about the assessment?

How do you comment on your experience of being assessed? Are you satisfied with the assessment practice implemented in junior middle school?

Is there anything you want to add to the assessment experience in junior middle school?

3. Can you briefly introduce your senior middle school? And the English class and English teacher?

What was your experience of being assessed in the subject of English or other subjects senior middle school?

Any impressive stories related to assessment design, assessment result communication and exchange, and parents' concern about the assessment?

How do you comment on your experience of being assessed? Are you satisfied with

the assessment practice implemented in senior middle school?

Is there anything you want to add to your assessment experience in senior middle school?

4. Do you have any pleasant or unpleasant experience in senior middle entrance examination and college entrance examination? What is your attitude to such a large-scale exam?

Interview Protocol

(Round 3)

The 3rd round interview is designed to understand their experience of being assessed in university (i.e., pre-service teacher education stage).

1. I have learned the course Language Assessment and Testing was not offered because of the limited number of students. So, why did you not select this course? How do you enhance your understanding of assessment? In what ways or from what resources?

Do you understand the latest assessment methods? (e.g., portfolio assessment). Have you ever tried before?

In your opinion, what should a qualified EFL teacher possess in assessment knowledge?

Do you think the course is necessary to be offered as a compulsory course?

2. In the course of English Pedagogy, do you feel it helpful for assessment? In what aspects?

Are you familiar with assessment policy and standards in basic education?

Are you familiar with ethical considerations in assessment? If yes, briefly express your understanding and possible ways. Do you think they should be in the list of required learning needs of pre-service teachers?

Are you satisfied with English Pedagogy course? Any suggestions for improvement?

In your opinion, what should be put in priority in assessment education?

3. In the formative assessment practices across all courses, which one is your favorite? Why do you favor this practice?

What other formative assessment practices have you experienced?

Have you ever been involved in peer-assessment or self-assessment activities in any form? If yes, do you like such experience and what is the benefit? If no, what maybe the possible challenge to carry out these forms of assessment?

Is the assessment policy, “formative assessment + summative assessment”, effective to enhance your English learning?

4. Are you satisfied with the assessment practices in the university?

Do you have any suggestions for improvement?
Anything you want to add to the assessment experience in university.

Interview Protocol (Round 4)

The last round of interview focuses on the experience of assessment practices and evaluation of LAL & evaluation of the pre-service teacher education program.

Part A: Experience of assessment practice in intensive teaching skills training and practicum.

1. Please make a brief introduction of your intensive teaching training session. (e.g., the mentor, the procedures)
What are your feelings or comments? Any suggestions for improvement?
What have you learned about assessment in this session?
Any involvement in peer assessment or self-assessment activities? What is the effect?
2. Briefly introduce your practicum. (e.g., types of school, the class size, duration of the practicum, your tutor in university and middle school, your task assigned by the school, any involvement of teaching practice, and students' homework checking)
3. What is your relationship with your supervisor in school?
What is your relationship with your students in practicum school? And students' parents?
Are you aware of the assessment policies and procedures in your practicum school?
If yes, can you give examples?
4. Which assessment methods have you used when you assess your students in your practicum?
What is your most important consideration in choosing a method for assessing student achievement?
When assessing language, I should...// I should not...
What is (or are) the more effective assessment method(s)? Why?
Based on what principles, beliefs, and values do you do the process of assessment?
In designing the teaching plans, do you consider the assessment? Why? How?
Do your lecturers and teacher-supervisor guide you in designing and implementing your assessment plan? How do they go about doing it?
What assessment methods will you plan to use in the future? Why/Why not?
5. Are there any assessment challenges that you face in your practicum?
What coping strategies do you utilize when faced with some of the challenges you mentioned?

Do you seek help the resources such as books, teachers, peers, online courses, or lectures? Is it effective? How?

6. What have you learned about assessment from your school practicum experiences /part-time teacher so far?

Please explain. In what ways, if any, have there been instances of conflict or confusion between what you have learned about assessment at university and what you have learned or observed about assessment on school practicum?

How did practicum change your perspectives, attitude, or practice of assessment if any?

Part B: Evaluation of LAL and the pre-service teacher education program

1. How do you evaluate your language assessment literacy (LAL)?

Do you feel you are ready to implement the assessment after graduation? Why or why not?

Do you have areas for improvement in LAL? What are they?

What prompted the decisions you made to learn more about assessment?

Do you think assessment will be frequently used in your future teaching? In what type? For what purpose?

2. How, if at all, do you think your pre-service education has changed your views on or practices in language testing and assessment?

In what ways, if any, have your beliefs and understandings about assessment changed since you began your teacher education program?

3. Do you think teacher education prepares students well for carrying out good assessment practices?

What aspects of your teacher education program have been particularly helpful for your learning about assessment? Any strengths in your pre-service program in improving your LAL?

What do you think should be changed in teacher education regarding the topic of assessment? And any shortcomings need to be improved?

4. What assessment-related questions were in the teacher certification examinations?

Do you think assessment-related content should be included? How?

Do you think the inclusion of assessment knowledge in certificate will enhance pre-service teacher's LAL? Any other ways to enhance LAL?

5. Is there anything else you would like to share regarding your experiences with assessment?

Do you have anything else you would like to share about your beliefs, understandings and practices of assessment? If so, please explain.

6. Do you have any other comments that have been missed about assessment and should be added to the current interview questions?

7. In your opinion, what is the significance of assessment in an EFL teacher's profession? Or what does the proportion of assessment occupy in EFL teacher's daily work?
8. During the interview, you have reviewed the assessment experience and reflected on the assessment understandings, do you think this interview helps you to focus more on assessment or LAL? Will the participation of this interview influence your future assessment activities?



Appendix D: Interview Transcripts (Excerpt Sample)

The First Round of Interview Translated Transcript (Excerpt Sample from Daisy)

I: As what I have introduced in the first meeting, this is the interview about language assessment. Are you ready to begin? Or shall we start?

Daisy (D): Yes.

I: At first, I'd like to collect some background information. How old are you? When do you begin to learn English?

D: I'm 21 years old. I started to learn English from the third grade of primary school.

I: What English proficiency tests have you taken and what are your scores?

D: I have passed College English Test-band 4 and College English Test-band 6, and then I have passed Test for English Majors-band 4 with the score of 60.

I: Have you ever participated in tutoring or had such a teaching experience in training institutions?

D: Well, kind of. I worked in the second semester of sophomore year to be in charge of tutoring primary school students' homework. The salary was very low. The institution was located on campus, so after I finished my class, I went there.

I: Which grade were the students in?

D: Fourth grade or fifth grade.

I: That is, you were tutoring primary school students in grade 4 or 5 for their English homework for nearly a term?

D: Not only English, but included all the disciplines and subjects in primary school.

I: OK, then why do you become a pre-service EFL teacher?

D: The real situation is that I am not as good at other subjects, and English has always been better. Moreover, for girls, it is better to become teachers, so taking all these into account, I choose English as my major.

I: The choice of becoming a pre-service EFL teacher is the result of your family's wishes or your own?

D: Both. I don't rule out them. I think becoming a teacher is suitable.

I: So, will you be a teacher after graduation?

D: Yes. In fact, I think teaching in high school or junior high school is good for me. Because I have got the teacher's qualification certificate.

I: OK, so much about the background information collection, and then, let's talk about some terms in assessment. We know that assessment cannot be ignored in English teaching and learning process. According to your understanding, what do you think of assessment? How to define it in your own words?

D: To me, the assessment is like giving feedback of this stage. That means, when the

teacher sees the students' growth or achievements or the shortcomings in all aspects, the teacher gives a result and direct feedback on students' efforts at this stage.

I: In other words, the assessment is more like a kind of feedback telling the present position where you are.

D: Yes.

I: Assessment can be divided into effective and ineffective, or good and bad. What do you think of a good language assessment? And a bad language assessment? What principles do you follow to distinguish effective from ineffective?

D: A good assessment means that the teacher should focus on more broad aspects, apart from the defects. The teacher may contribute the possible reason to the surrounding contexts rather than the personal inadequacy in proficiency. By contrast, the ineffective assessment means the teacher provides a rather general conclusion of your shortcomings without pointing out where the core problems are or what the solutions are. He/she just tells you that you are insufficient in this aspect.

I: Just to say where the problem lies, but without a specific improvement measure and method, his/her feedback is not comprehensive enough, and may not have a direct promotion effect. Therefore, a good assessment takes all factors into account and is more comprehensive and then gives students feedback that can be used to improve their learning. This is a good language assessment.

D: Yes.

I: Do you think assessment is important in teachers' teaching and students' learning? Or what role does assessment play?

D: I think it is important, because assessment provides feedback to teachers' teaching about the degree of students' mastery of the teaching objectives at this stage. Moreover, it is beneficial for students too. It makes student know what they need to improve. Thus, it is mutual for both teacher's teaching and students' learning.

I: How do you think assessment affects students?

D: Sometimes, the teacher talks with the students, and then he/she says something. If the students can also feel the assessment information, they may work hard in this aspect, and then make improvement. In fact, it has something to do with the student's self-control ability. Sometimes the students may not change anything after receiving the assessment results from teacher.

...

Appendix E: Participant's Journal (Excerpt Sample)

Translated Reflective Journal Sample (From Flora)

Language learning experience

English learning experience probably started at the beginning of the junior high school. At that time in our place, English was not taught in primary school, so my English learning started from the first grade of junior high school. Since then, I deeply realized that a good foundation of English really mattered. Otherwise, the gap in learning English between me and my peers who learned from primary school would be even greater.

The English teacher of junior high school was a female teacher. Her teaching style was relatively strict and capable. She assigned tasks such as memorizing words and listening to tapes every day, and required parents to supervise the completion of such tasks. This was a model of home school cooperation. Not only could parents better understand the situation of students, but also could promote students' consciousness. In addition, the teacher also adopted the group leader responsibility system, in which students good at English needed to assist two students with poor English. This not only improved the group leader's English performance and enhanced the sense of responsibility, but also motivated the progress of the poor students.

In a word, the teacher's teaching methods were very effective, and my English achievements had also laid a foundation and made continuous progress at this stage. If I could become an English teacher in the future, I would also carry out my teaching activities according to the methods adopted by the teacher and timely adjustments according to the situation of students.

Finally, my high school English teachers and college English teachers were not very strict with us, so in these two stages, self-consciousness and self-study were crucial. I completed most of my courses through self-study. Therefore, in my language learning process, teacher management was very important, and independent learning was also essential.

Language assessment experience

In my impression, before going to college, the main body of language assessment has always been conducted by teachers, who assess students in class, such as students' classroom performance, the question-answering in class, and so on; Or in a stage of learning, teachers assess and analyze students according to their recent performance

or the results of an exam. Therefore, in this period, it is difficult for me to really understand what assessment is; most of the assessments are done by teachers. I only know how I should do after the teacher's assessment and analysis.

After going to college, some teachers designed peer assessment among students. As an assessment agent, I also began to understand the language assessment. For example, the teacher of translation let students translate, and other students commented and modified by pointing out the highlights and shortcomings of the translated version. The teacher of listening asked another student to supplement and modify the answer provided by others. What impressed me most was that the language assessment conducted in the course English Pedagogy was diverse. First of all, I would reflect on my teaching after the trial teaching, that is, I would assess myself. Secondly, other students assessed my language in different aspects according to the different dimensions of my lectures. Finally, teachers would assess my language, which promoted my progress and growth.

During the practicum, as a teacher in the middle school, I also applied the language assessment in my own classroom, and assessed the students' classroom performance and knowledge mastery. If I go to the profession in the future and have the opportunity to become an English teacher, I will continue to learn from and consult excellent English teachers about language assessment and try to apply properly them in my classroom.

...

Appendix F: Assessment Artifacts (Sample of Lesson Plan)

A Lesson Plan in Teaching Practicum (From Betty)

School: 2 nd Feng Dong Middle School		Class: Class 4 of Grade 9	Discipline: English
Subject: What are the shirts made of?			
Duration: 1 period of class		Type: integrated lesson	
Teaching text: People’s Education Press			
Textbook Analysis	The topic of this lesson is to discuss what materials are used for daily necessities and where are they made. Students are required to learn and master the names of some daily necessities. At the same time, in language learning training, students will be exposed to and learn the passive voice.		
Teaching Objectives	Knowledge aims: Students should master some key words: chopstick, coin, fork and so on; some important phrases: be known for, no matter, and they should learn to use the passive voice with present tense. Ability aims: 1. Develop to use the passive voice with present tense to train the ability of listening, speaking, reading and writing. 2. Learn to talk about what things are made of and where they were made. Moral aims: Students can learn more about traditional Chinese culture.		
Key points and Solutions	Students should master the new words and important phrases in this lesson and they should be able to describe and ask about materials in English.		
Difficult points and Solutions	Students should understand the use of the passive voice and its sentence structure.		
Teaching Methods	The communicative approach, task-based language teaching.		
Teaching Procedures	Step 1: Lead in 1. first, the teacher will play a video about the papermaking process and let the students know about traditional Chinese inventions. And after the playing, the teacher will ask them several questions like: who invented paper first? What is paper made of now? In this way, students can be tested to see if they are paying attention. Step 2: New knowledge presentation 1. The teacher will present the sentence structure and use the pictures on the powerpoint. ---- What is the golden medal made of? ----- It’s made of gold.		

-----Is this table made of glass?
----No, it isn't. It's made of wood.
----Is butter made from meat?
-----No, it's made from cream.

Then the teacher will teach the usage of “be made of” and “be made from” and the difference of them.

2. Second, the teacher will show some pictures on the powerpoint and try to teach the new words using “be made of” structure.
e.g., This pair of chopsticks are made of bamboo.
Is this blouse made of cotton?
What is the fork made of?

Students will discuss these questions with their deskmate and try to learn the new words.

Step 3: Practice

1. The teacher will ask students to read the things and materials in 1a. Students should discuss with their partners and match them with the materials. More than one answer is possible. The teacher will give the instructions: What are these things usually made of?

1. chopsticks	a. wool
2. window	b. gold
3. coin	c. silver
4. stamp	d. paper
5. fork	e. silk
6. blouse	f. glass

Then the teacher will check the answers with the students.

Step 4: Pair work

1. The teacher will ask students to make a conversation using the information in 2a and 2b.
e.g., A: What did you see at the art and science fair?
B: I saw a model plane.
A: What is it made of?
B: It's made of steel, glass, and plastic.

2. Then, students should make conversation based on their own life and practice the conversation in pairs. Some pairs will be invited to show for the whole class.

Step 5: Listening

1. The teacher, will ask students to hear a conversation about some things and material. Students should listen and match the products with what they are made of and where they were made.

Things	made of	made in
shirts	cotton	Korea
chopsticks	silver	Thailand
ring	steel	America

Step 6: Summary and Homework

Summary: in this section, students have learned some names of daily necessities and what are they made of and where are they made in. They also have a basic understanding of the passive voice.

Homework:

1. Students should read aloud and practice the conversation in 2 d.

	<p>2. Students should translate the following sentences into English.</p> <p>(1) 这个戒指是银制的。</p> <p>(2) 这种纸是由树木制成的</p> <p>(3) 油漆是由什么制成的。</p> <p>(4) 杭州因其茶叶而为人知。</p> <p>(5) 据我所知，茶树被种植于山坡上。</p>
Reflection	<p>Most of the students could keep up with the pace, however, there are still a small number of slow learners who are absent-minded to some extent. I will focus more on these students by engaging them in interactions, like questioning and encouraging them into active learning. An effective class is the result of good cooperation of teacher and students, not permitting anyone to lag behind. I need to do more to involve them in English learning.</p>

