

Ling Gan
Ricky Lam

Language Assessment Literacy Development

An Ethnographic Case Study
on University English Instructors
in China

 Springer

Ling Gan and Ricky Lam

**Language Assessment Literacy
Development**
**An Ethnographic Case Study on University
English Instructors in China**



Ling Gan

School of Languages and Communication, Beijing Technology and
Business University, Beijing, China

Ricky Lam

Department of Education and Psychology, Hong Kong Baptist
University, Hong Kong, China

ISBN 978-981-95-2324-5 e-ISBN 978-981-95-2325-2

<https://doi.org/10.1007/978-981-95-2325-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive
license to Springer Nature Singapore Pte Ltd. 2026

This work is subject to copyright. All rights are solely and exclusively
licensed by the Publisher, whether the whole or part of the material is
concerned, specifically the rights of translation, reprinting, reuse of
illustrations, recitation, broadcasting, reproduction on microfilms or in
any other physical way, and transmission or information storage and
retrieval, electronic adaptation, computer software, or by similar or
dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks,
service marks, etc. in this publication does not imply, even in the
absence of a specific statement, that such names are exempt from the
relevant protective laws and regulations and therefore free for general
use.

The publisher, the authors and the editors are safe to assume that the
advice and information in this book are believed to be true and accurate
at the date of publication. Neither the publisher nor the authors or the
editors give a warranty, expressed or implied, with respect to the
material contained herein or for any errors or omissions that may have

been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Acknowledgements

The publication of this book will not have happened without the support and encouragement from many people. First and foremost, we are deeply thankful to the four university English instructors who joined this study for their openness, patience, and generosity in sharing their experiences. They have become role models for us with their quality instructions. We will continue to be enlightened by their reflections of personal and professional stories in language assessment. We express sincere thanks to Dr. Changsheng Jiang for his consistent support in data analysis and Dr. Xun Yan for his insightful comments on the study. We also extend our thanks to the editors and reviewers at Springer for their constructive feedback and suggestions.

The first author would like to thank Hong Kong Baptist University for providing her with the valuable opportunity to complete her PhD study in Hong Kong. This experience is the most challenging yet most rewarding in her academic career, laying a solid foundation for her research in language assessment. Great gratitude also goes to the staff and friends in the Faculty of Arts and Social Sciences, Hong Kong Baptist University, whose endless support and encouragement are crucial for this study.

Finally, we owe many thanks to our families. The first author's special thanks go to her son, Tony Zhan, who has encouraged her to keep going with his warm words: "Mom, you are the brightest star shining upon me", to her husband, Xinyu Zhan, who has shouldered all family responsibilities so that she could focus on the study, to her parents, who have been always supportive of and self-sacrificing for whatever she does. The second author also extends his gratitude to his family who shows constant support throughout his academic career.

Contents

1 Introduction

1.1 Background of the Study

1.2 The Research Problem

1.3 Purposes and Significance of the Study

1.4 Overview of Research Methodology

1.5 Layout of the Book

References

2 Literature Review

2.1 Overview

2.2 Definition of LAL

2.3 LAL Frameworks

2.3.1 LAL Component Frameworks

2.3.2 LAL Continuum Frameworks

2.4 Research on Teachers' LAL

2.4.1 Teachers' LAL Level and Needs

2.4.2 Teachers' LAL Development

2.4.3 Teacher Assessment Identity in LAL Development

2.4.4 Factors that Influence Teachers' LAL Development

2.5 Theoretical Framework of the Study and Research Questions

2.6 Summary

References

3 Research Methodology

3.1 Overview

3.2 Research Context

[3.3 An Ethnographic Case Study](#)

[3.4 Selection of Participants](#)

[3.5 Data Collection](#)

[3.5.1 Narrative Frames](#)

[3.5.2 Interviews](#)

[3.5.3 Field Observations and Post-Observation Conversations](#)

[3.5.4 Documents](#)

[3.5.5 Summary of Data Collection](#)

[3.6 Data Analysis](#)

[3.6.1 Analysis within Data Sets](#)

[3.6.2 Cross-Cases Analysis](#)

[3.7 Trustworthiness, Ethical Issues and Roles of the Researcher](#)

[3.8 Summary](#)

[References](#)

[4 The Case of Emma](#)

[4.1 Overview](#)

[4.2 Background Information](#)

[4.3 Prior Assessment Experiences](#)

[4.4 An Active Assessment Practitioner](#)

[4.4.1 Understanding Assessment Requirements](#)

[4.4.2 Integrating Peer Assessment into Speaking](#)

[4.4.3 Conducting Portfolio Assessment](#)

[4.4.4 Grading](#)

[4.4.5 Preparing Students for CET4](#)

[4.5 An Earnest Assessment Researcher](#)

[4.6 Perceived Assessment Identity](#)

[4.6.1 “I Am not an Assessor but an Explorer”](#)

[4.6.2 “I Am an Assessment Facilitator”](#)

[4.6.3 “Diminish My Identity as an Assessor”](#)

[4.7 Emma’s Case in Summary](#)

[4.7.1 LAL Profile](#)

[4.7.2 LAL Developmental Trajectory](#)

[4.7.3 Assessment Identity](#)

[4.7.4 Challenges and Opportunities for LAL Development](#)

[References](#)

[5 The Case of Rachel](#)

[5.1 Overview](#)

[5.2 Background Information](#)

[5.3 Prior Assessment Experiences](#)

[5.4 An Unswerving Assessment Practitioner](#)

[5.4.1 Understanding Assessment Requirements](#)

[5.4.2 Utilising Traditional Vocabulary Tests](#)

[5.4.3 Giving Oral Feedback](#)

[5.4.4 Grading](#)

[5.5 An Active Assessment Explorer](#)

[5.5.1 Conducting the First Peer Assessment on Students’ Writing](#)

[5.5.2 Improving Speaking Tests](#)

[5.6 Participating in a Teaching Contest](#)

[5.7 Perceived Assessment Identity](#)

[5.7.1 “I Am not an Assessor but a Guide”](#)

[5.7.2 “I am a Stumbling Novice Assessor Who Fumbles in the Assessment”](#)

5.8 Rachel's Case in Summary

5.8.1 LAL Profile

5.8.2 LAL Developmental Trajectory

5.8.3 Assessment Identity

5.8.4 Challenges and Opportunities for LAL Development

References

6 The Case of Ryan

6.1 Overview

6.2 Background Information

6.3 Prior Assessment Experiences

6.4 An Experienced Assessment Practitioner

6.4.1 Understanding Assessment Requirements

6.4.2 Optimising the Use of Quizzes

6.4.3 Conducting Group Peer Assessment on Writing Essays

6.4.4 Giving Feedback

6.4.5 Grading

6.4.6 Preparing Students for CET 4/6

6.5 Participating in a Teaching Contest

6.6 Perceived Assessment Identity

6.6.1 "I Don't Have a Strong Identity as an Assessor"

6.6.2 "I Am a Designer and a Facilitator As Well"

6.7 Ryan's Case in Summary

6.7.1 LAL Profile

6.7.2 LAL Developmental Trajectory

6.7.3 Assessment Identity

6.7.4 Challenges and Opportunities for LAL Development

References

7 The Case of Tim

7.1 Overview

7.2 Background Information

7.3 Prior Assessment Experiences

7.4 A Confident Test-Orientated Practitioner

7.4.1 Understanding Assessment Requirements

7.4.2 CET4/6-Orientated Practices

7.4.3 Using Self- and Peer Assessment in Writing Classes

7.4.4 Grading

7.4.5 Giving Individualised Feedback

7.5 Participating in Teaching Contests

7.6 Perceived Assessment Identity

7.6.1 “Am I an Assessor?”

7.6.2 A Life Mentor

7.7 Tim’s Case in Summary

7.7.1 LAL Profile

7.7.2 LAL Developmental Trajectory

7.7.3 Assessment Identity

7.7.4 Challenges and Opportunities for LAL Development

References

8 Discussion

8.1 Overview

8.2 LAL Needs for University English Instructors

8.2.1 Knowledge of Language Assessment

8.2.2 Language Assessment Skills

8.2.3 Principles of Language Assessment

8.2.4 Assessment Identity

8.2.5 An LAL Profile for University English Instructors

8.3 LAL Developmental Trajectories

8.4 Construction of Assessment Identity

8.5 Challenges and Opportunities for LAL Development

8.5.1 Institutional and Classroom Contexts

8.5.2 Assessment Experiences

8.5.3 Teachers' Conceptions of Language Assessment

8.5.4 A Recap of Challenges and Opportunities for LAL Development

8.6 Summary

References

9 Conclusion

9.1 Overview

9.2 Key Findings

9.3 Implications of the Study

9.3.1 Implication for University English Instructors

9.3.2 Implications for University/Department Administrators

9.3.3 Implications for Assessment Trainers/Training Providers

9.4 Limitations of the Study and Suggestions for Future Research

9.5 Conclusion

References

Appendix 1: Narrative Frame 1

Appendix 2: Narrative Frames 2-4

[Appendix 3: Interview Protocol](#)

[Appendix 4: Observation Protocol](#)

List of Figures

[Fig. 2.1 Language Assessment Literacy: An Expanded Definition](#)

[Fig. 2.2 Taylor's LAL Profile](#)

[Fig. 2.3 Contextual Levels Impacting on Teacher Beliefs of LAL Training Needs](#)

[Fig. 2.4 An Adapted LAL Framework for Language Teachers](#)

[Fig. 4.1 Students' Peer Assessment Sheets on Speaking](#)

[Fig. 4.2 A Record of Peer Assessment with a New App](#)

[Fig. 4.3 A Snapshot of Students' Notebook Portfolio](#)

[Fig. 4.4 A Grading Sheet of an Oral Test](#)

[Fig. 4.5 Emma's LAL Profile](#)

[Fig. 4.6 A Trajectory of Emma's LAL Development](#)

[Fig. 5.1 A Snapshot of a Vocabulary Test Used by Rachel](#)

[Fig. 5.2 A Screenshot of an Overview of Rachel's Grading](#)

[Fig. 5.3 A Sheet of Rachel's Grading in a Speaking Test](#)

[Fig. 5.4 Rachel's LAL Profile](#)

[Fig. 5.5 A Trajectory of Rachel's LAL Development](#)

[Fig. 6.1 A Screenshot of Ryan's Narrative Writing Assessment Rubric for Narrative Writing](#)

[Fig. 6.2 A Sample of Peer Feedback Sheet](#)

[Fig. 6.3 A Sample of Ryan's Written Feedback](#)

[Fig. 6.4 A Record Sheet of Students' Performance on Assessment Tasks](#)

[Fig. 6.5 A Screenshot of Ryan's Reflective Note of Teaching Design](#)

[Fig. 6.6 Ryan's LAL Profile](#)

[Fig. 6.7 A Trajectory of Ryan's LAL Development](#)

[Fig. 7.1 Tim's LAL Profile](#)

[Fig. 7.2 A Trajectory of Tim's LAL Development](#)

[Fig. 8.1 An LAL Profile for University English Instructors in the Chinese Context](#)

[Fig. 8.2 A Trajectory of LAL Development for University English Instructors](#)

[Fig. 8.3 Four Quadrants of Instructors' Assessment Identity and LAL Development](#)

[Fig. 8.4 A Trajectory of Instructors' Construction of Assessment Identity.](#)

List of Tables

[Table 2.1 Operationalisation of the Eight Components in Taylor's LAL Profile](#)

[Table 2.2 Operationalisation of the Five Levels in Taylor's LAL Profile](#)

[Table 3.1 Biographic Information of the Participants](#)

[Table 3.2 A Summary of Collected Data](#)

[Table 3.3 Assessment Scenes from Classroom Observations](#)

[Table 4.1 Assessment Requirements for College English by Emma's Department](#)

[Table 4.2 Components of "Assignments"](#)

[Table 4.3 Criteria for Peer Assessment on Speaking](#)

[Table 4.4 The Frequency of Using MCQs in the Observed Class Sessions](#)

[Table 5.1 Assessment Requirements for College English by Rachel's Department](#)

[Table 5.2 Rachel's Peer Assessment Sheet for Presentations](#)

Table 6.1 Assessment Requirements for College English by Ryan's Department

Table 7.1 Assessment Requirements for College English by Tim's Department

Table 7.2 A Criteria Sheet for Writing

Table 8.1 LAL Needs of University English Instructors in the Chinese Context

About the Authors

Ling Gan has a Ph.D. in Education from Hong Kong Baptist University. She is now an Associate Professor at the School of Languages and Communication, Beijing Technology and Business University in China. Her research interests include language assessment literacy, classroom-based assessment and teacher professional development. She has published articles in *RELC Journal*, *Education Sciences*, *Chinese Journal of Applied Linguistics*, *Language Assessment Quarterly* and *Language Testing in Asia*, and reviewed SSCI and SCOPUS-indexed journals including *Teaching and Teacher Education*, *Language Assessment Quarterly*, *System*, *Language Testing in Asia* and *Journal of Language, Identity and Education*.

Ricky Lam is an Associate Professor and Director of Master of Education Programme in the Department of Education and Psychology at Hong Kong Baptist University. He is an associate editor of *Frontiers in Psychology* and has served on several international editorial boards, such as *Language Assessment Quarterly*, *Assessing Writing*, *RELC Journal*, and *Journal of Asia TEFL*. His research interests include second language writing assessment, digital portfolios, and language assessment literacy. His publications have appeared in *Language Assessment Quarterly*, *System*, *Assessing Writing*, *Language Testing*, *TESOL Quarterly*, *ELT Journal*, and other SSCI- and SCOPUS-indexed journals. He is also the author of *'Integrating E-Portfolios into L2 Classrooms: Education for Future'* (Multilingual Matters, 2024).

1. Introduction

Ling Gan¹✉ and Ricky Lam²✉

- (1) School of Languages and Communication, Beijing Technology and Business University, Beijing, China
- (2) Department of Education and Psychology, Hong Kong Baptist University, Hong Kong, China

✉ Ling Gan

Email: linggan@btbu.edu.cn

✉ Ricky Lam (Corresponding author)

Email: rickylam@hkbu.edu.hk

1.1 Background of the Study

Language testing and assessment have received more attention than ever in language education, particularly in the past decades. They are used to establish a connection between language teaching and learning, as they can tell whether teachers have taught what is expected students to learn and whether students have learned what has been taught (Stiggins, [1995](#)). This connection can be illustrated in three ways. The first way is associated with the impact of language testing on instruction and learning, namely, washback (Alderson & Wall, [1993](#)). Washback shows itself mainly in testing culture where high scores are the goal for teaching. The second approach of linking assessment with teaching and learning finds its way into the improvement of curricular-driven assessment tasks, which are used to understand how well the objectives of a curriculum have been met (Purpura, [2016](#); Rea-Dickins, [2004](#)). The third approach to connecting assessment with teaching is to embed assessment throughout the instructional activities and

performances. This can be realised by classroom teachers who conduct assessment practices with the purpose of fine-tuning instructions to meet students' learning needs (Hidri, [2021](#)).

The close connection of language testing and assessment with teaching and learning has brought into the forefront stakeholders' professional knowledge of language testing and assessment. Among various stakeholders, it is teachers who use their assessment professional knowledge to play an important role in classroom assessment (Koh et al., [2019](#)). Language teachers possibly spend from one-third to half of their time on assessment tasks (Cheng, [2001](#); Coombe et al., [2012](#)). These activities include the design and administration of language assessment, the use of results derived from assessment to make decisions of learners for diverse purposes, and provision of feedback based on assessment data. To ensure the effective implementation and the high quality of these tasks, teachers should have sufficient language assessment literacy (LAL). LAL, a concept originating from assessment literacy in general education, usually cover knowledge, skills, and principles various stakeholders obtain to implement assessment-related activities (Fulcher, [2012](#); Inbar-Lourie, [2017](#)). Thus, so far, there has been no consensus on its definition. Given the burgeoning assessment responsibilities and accountabilities imposed on teachers (Koh et al., [2018](#)), LAL is urgently required among language teachers.

LAL is necessary and significant for language teachers. They need to become assessment literate, as they often have been on the central stage of connecting curriculum, instruction, and assessment (Gareis & Grant, [2015](#)). Sufficient LAL enables teachers to diagnose students' weaknesses and strengths in learning, provide accurate inferences to students and other stakeholders, and adjust their teaching accordingly. In contrast, insufficient LAL will often, if not always, reduce the reliability and validity of language assessment and, thus, negatively impact instructions and education (Xu & Brown, [2017](#)). Therefore, LAL should be put at the core of successful language teaching and teacher professional development (Cheng & Ma, [2015](#)). In this way, teachers' LAL can benefit both students' learning and teachers' long-term professional well-beings.

However, extensive LAL literature has pointed out that, owing to the lack of assessment training/learning opportunities, most language teachers do not have sufficient LAL to prepare themselves for the aforementioned assessment practices (Fulcher, [2020](#)). LAL inadequacy among language teachers would stand as a stumbling block of a successful language teaching and learning programme. It is this reason that highlights the significance of addressing language teachers' LAL development.

In the Chinese context, university English instructors, who teach English as a foreign language to the largest population of adult learners in the world (Y. Xu, [2017](#)), are also required to acquire adequate LAL to qualify themselves for various assessment-related issues. However, they have a low level of LAL in general (Jiang, [2020](#); Xu & Brown, [2017](#)), and particularly need to develop specific assessment skills, such as choosing classroom-based assessment methods, developing assessment criteria, and utilising assessment results (Jiang, [2020](#)). The paucity of LAL has become a barricade to the language assessment reform initiated in 2007 by the China Ministry of Education through issuing the *College English Curriculum Requirements*. This document has proposed the parallel implementation of formative assessment and summative assessment into College English classes. Yet the implementation was not conducted smoothly (Liu & Xu, [2017](#)). Taken together, LAL development of university English instructors has been badly needed.

1.2 The Research Problem

The due importance attached to teachers' LAL in recent years has sparked a rapid growth of research on this topic. As the starting point of LAL research, teachers' LAL status quo and their LAL training needs have been widely examined (e.g., Crusan et al., [2016](#); Fulcher, [2012](#); Vogt & Tsagari, [2014](#)). These studies revealed that language teachers' lack of assessment training led to a scarcity of LAL among them. As such, lack of assessment training and dissatisfactory status quo of teachers' LAL has inspired researchers to explore how to develop teachers' LAL efficiently (Deygers & Malone, [2019](#); Levi & Inbar-Lourie, [2020](#)).

Formally, university-based assessment training programmes/courses are one of the widely-used methods to provide language teachers with assessment knowledge, skills, principles, and theories in a professional way (Malone, [2017](#)). Assessment training programmes/courses have been found to facilitate teachers' LAL development in three aspects. First, instructors gain a better understanding of the relationship between language teaching, learning, and assessment, and between the assessment process and consequences of assessment (Baker & Riches, [2018](#); Lee, [2019](#)). Second, they have acquired and improved their assessment knowledge, skills and the know-how of how to integrate assessment into teaching and learning (Baker & Riches, [2018](#); Lee, [2019](#); Levi & Inbar-Lourie, [2020](#)). Third, they have increased self-efficacy and confidence in conducting assessment activities consistent with teaching and learning objectives (Lee, [2019](#); Levi & Inbar-Lourie, [2020](#)).

Informally, scholars have attempted to enhance teachers' LAL through alternative methods other than training programmes/courses, including teachers' involvement in online tutorials (Malone, [2017](#)), assessment task designs (Koh et al., [2018](#)), professional learning community (Ho & Yan, [2021](#); Saputra et al., [2020](#)), and test design and development (Cui et al., [2022](#); Vogt & Tsagari, [2022](#); Walters, [2010](#)). These alternatives aimed to empower language teachers with assessment skills and theories by engaging them in specific assessment practices where through the engagement, teachers' central role in assessment can be foregrounded (Xerri & Briffa, [2018](#)).

Both assessment training and alternative methods attempt to meliorate teachers' LAL by means of external forces. With their help, teachers' LAL can be enhanced promptly and palpably (Gan & Lam, [2022](#)). However, these methods have been overemphasised to enhance teachers' LAL in a top-down way. That is, they are used to impart assessment knowledge and skills to teachers who are viewed as an inactive agent in constructing their LAL. As LAL is a dynamic process-orientated construct (Inbar-Lourie, [2017](#); Levi & Inbar-Lourie, [2020](#)), its development is also an intricate process where teachers need to actively or passively interact with multiple social, historical, cultural, educational, and experiential factors (Scarino, [2013](#); Yan et al., [2018](#)). With the interactions, language teachers will gain their gradual LAL

growth from initial LAL derived from their prior assessment experiences (Yan & Fan, [2021](#)). Understanding how language teachers develop their LAL lends itself to the better delivery of assessment training to disseminate LAL. The problem is, however, as Yan and Fan ([2021](#)) noted, that few empirical studies have examined the LAL developmental process of language teachers in their teaching context, a natural milieu.

In China, developing university English instructors' LAL is particularly necessary and pressing. First, although the assessment reform has been implemented in College English for many years, there is a lacuna between the assessment tasks instructors are expected to conduct and their competencies to meet the reform requirements. For instance, their formative assessment practices are still used for summative purposes (Liu & Xu, [2017](#)) and they have difficulty with applying assessment theories into practices (Zhang et al., [2024](#)).

Second, few university English instructors are given opportunities to take assessment training to enhance their LAL (Gan & Lam, [2020](#)). In China, not all university English instructors have received the same level/hours of assessment training programme as those in primary or secondary schools who are required to complete assessment courses before the commencement of their teaching. Instead, university English instructors are usually offered an assessment-irrelevant short-term orientation training before they start to teach. In addition, the lack of institutional support on language assessment also discourages university English instructors from taking assessment training provided by educational institutions or language textbook publishers (Gan & Lam, [2020](#); Jiang, [2020](#)).

Third, university English instructors cannot turn to official assessment documents for help to learn how to conduct assessment for learning. Although there is a national document to postulate assessment policies for College English, no specific guidelines for assessment have been given to guide instructors on how to effectively implement assessment in their classrooms. Given the dissatisfactory LAL status quo, lack of external support and mandated guidelines, it is important for individual university English instructors to understand how to develop their LAL in their assessment practices when external support for LAL enhancement is constrained.

1.3 Purposes and Significance of the Study

The purposes of the study are twofold. The primary goal is to track the LAL development of university English instructors over time in the Chinese context. Specifically, the study tries to identify localised LAL needs for university English instructors in a natural teaching context. It also documents LAL developmental trajectories by tracking how university English instructors gain, enhance, and sustain their LAL through their engagement in assessment practices. Besides, the study investigates how university English instructors shape their identity as assessors in assessment practices along with their LAL development.

The other purpose is to explore how contextual, experiential, and personal factors influence university English instructors' LAL development. It intends to probe into challenges and opportunities university English instructors may encounter in LAL development mediated by contextual, experiential, and personal factors. It also explores how university English instructors mobilise these challenges and opportunities to promote their LAL development and construction of assessment identity.

In view of the research purposes, the study on LAL development of university English instructors in the Chinese context has theoretical, methodological, and practical significance to LAL scholarship. The theoretical significance lies in its addition to LAL literature by proposing a possible LAL profile, a context-dependent LAL development trajectory for university English instructors, and four quadrants of the relationship between LAL development and assessment identity. The LAL profile attempts to validate the application of Taylor's LAL profile for classroom teachers in a specific educational context. The LAL developmental trajectory moves our understanding of LAL from a static knowledge base or skill kit to a dynamic progressive process in which language teachers experience challenges and opportunities, negotiate, and compromise with institutional requirements. Meanwhile, the four quadrants of LAL development and assessment identity reconceptualise LAL beyond knowledge, skills, and practices to language assessment identities.

Methodologically, as the initial longitudinal qualitative research on university English instructors' LAL, this study provides valuable

information about the complexity of LAL by closely looking into the LAL development process over time. The longitudinal study has reported and monitored the wax and wane of LAL development, and identified LAL developmental patterns for language teachers in a nuanced way, filling a LAL literature gap of “how LAL develops within individuals” (Taylor, [2013](#), p. 407).

Practically, the study offers several implications for university English instructors, university/department administrators, and assessment trainers/training providers. Those English instructors eager to become more assessment literate may gain useful insights from this study into how to enhance their LAL by utilising available resources when external constraints are imposed. The study also provides a venue for university/department administrators to understand university English instructors’ challenges and opportunities for their professional development in language assessment. These challenges and opportunities are expected to draw more financial and emotional support at the institutional level for instructors’ LAL development. In the same vein, assessment trainers hope to obtain a nuanced understanding of university English instructors’ LAL needs and development patterns from this study. By being more mindful of these elements, they would provide customised assessment training programmes to meet instructors’ assessment needs in and out of classrooms.

1.4 Overview of Research Methodology

Given that the purpose of the study is to delve into the LAL developmental process, an ethnographic case study, a qualitative approach in nature, was adopted to track the changes, addressing the “how” and “why” questions regarding LAL development. The increase in qualitative approaches in LAL study is in line with the post-positivistic understanding that teachers’ LAL is dynamic and context-situated (Giraldo, [2020](#); Inbar-Lourie, [2017](#)). This perspective implies that LAL research should take a close look at language teachers’ interpretative “life-world” (Scarino, [2013](#), p. 316) in natural teaching contexts. This qualitative approach is meant to seek the LAL

development reality by providing a thick description of the participants' specific settings (Mackey & Gass, [2005](#)).

In addition, this ethnographic case study is longitudinal. The intensive long-term participant observation enables researchers not only to collect different data about the specific contexts but also to check and confirm researchers' inferences in the course of research (Maxwell, [2013](#)). Through a three-semester research engagement, we gained access to more complete data about university English instructors' LAL development. We also constantly constructed, revised, and revamped the interpretations about instructors' LAL development by comparing data in different semesters. This process, thus, has ensured the quality of the qualitative study. Furthermore, by comparing the results of each case, initial interpretations of the results could be tested and revised with those emerging from other individual cases, leading to either a literal replication or a theoretical replication (Yin, [2018](#)).

1.5 Layout of the Book

The book consists of nine chapters. Chapter 1 introduces the background of the study, identifying the research problem, and outlining the purposes and significance of the study. It also gives an overview of the research methodology and ends with a summary of the book organisation.

Chapter [2](#) reviews the relevant theorisation of and empirical studies on LAL. It begins with the evolution of LAL definitions, frameworks and models, and operationalises LAL for this study. The chapter then synthesises LAL empirical studies for language teachers. Insights from this review are generated to identify the research gaps, based on which the research questions are proposed. At the end of the chapter, an adapted framework is outlined to guide the present study.

Chapter [3](#) records the choices and procedures of research approaches and methods. It starts with an introduction to the research context, followed by a justification of the research design. The methodological details are subsequently specified, including participant selection, data collection, and analysis. Finally,

trustworthiness, ethical considerations, and reflections on the researcher's role are included.

Chapters [4](#) to [7](#) report on the findings of the study. The narrative of each case is portrayed respectively, focusing on participants' prior assessment experiences and their assessment practices where LAL developmental trajectories, assessment identity, and challenges and opportunities for LAL development are identified and presented.

Chapter [8](#) synthesises the key findings of the study and discusses them aligned with the research questions. In this chapter, a prospective LAL profile, a LAL developmental trajectory, and four quadrants of LAL development and assessment identity are proposed to advance our conceptual understanding of LAL development.

Chapter [9](#) summarises the findings and identifies the contributions and limitations of the study. It provides practical implications for university English instructors to engage in ongoing professional development and assessment trainers/training service providers to design appropriate coaching events. The final chapter also suggests some guidelines for future research on LAL.

References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129.

<https://doi.org/10.1093/applin/14.2.115>

[Crossref]

Baker, B. A., & Riches, C. (2018). 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>

[Crossref]

Cheng, L. (2001). An investigation of ESL/EFL teachers' classroom assessment practices. *Language Testing Update*, 29, 53–83.

Cheng, L., & Ma, J. (2015). "It puts us in our students' shoes": Listening to voices from teacher candidates on their test-taking experiences. *International Journal of Humanities and Social Science Research*, 1, 46–56. https://www.lifescienceglobal.com/media/zj_fileseller/files/IJHSSRV1A5-Cheng.pdf

Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges and recommendations. In P. D. C. Coombe, B. O'Sullivan, & S. Stoyloff (Eds.), *The Cambridge guide to second language assessment* (pp. 20–29). Cambridge University Press.

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>
[Crossref]

Cui, Y., Liu, Y., Yu, H., & Gao, Y. (2022). Developing English teachers' language assessment literacy in an EAP reform context through test design: A case study. *System*, 109, 1–12. <https://doi.org/10.1016/j.system.2022.102866>

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>
[Crossref]

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>

Fulcher, G. (2020). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
[Crossref]

Gan, L., & Lam, R. (2020). Understanding university English instructors' assessment training needs in the Chinese context. *Language Testing in Asia*, 10(11), 1–18. <https://doi.org/10.1186/s40468-020-00109-y>
[Crossref]

Gan, L., & Lam, R. (2022). A review on language assessment literacy: Trends, foci and contributions. *Language Assessment Quarterly*, 19(5), 503–525. <https://doi.org/10.1080/15434303.2022.2128802>
[Crossref]

Gareis, C. R., & Grant, L. W. (2015). *Teacher-made assessments: How to connect curriculum, instruction, and student learning*. Routledge.
[Crossref]

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>

Hidri, S. (2021). Language assessment literacy: Where to go? In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for improved student learning* (pp. 3–12). Routledge.

Ho, E. C., & Yan, X. (2021). Using community of practice to characterize collaborative essay prompt writing and its role in developing language assessment literacy for pre-service language teachers. *System*, 101. <https://doi.org/10.1016/j.system.2021.102569>

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 257–270). Springer International Publishing AG.

Jiang, J. (2020). Gaoxiao yingyu jiaoshi ceping suyang xianzhuang jiqi yingxiang yinsu yanjiu [Study of college English teachers' assessment literacy and its influencing factors]. *Foreign Language World*, 6, 18–26.

Koh, K., Burke, L., Luke, A., Gong, W. G., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 22(3), 264–288. <https://doi.org/10.1177/1362168816684366>
[Crossref]

Koh, K., DePass, C., & Steel, S. (2019). *Developing teachers assessment literacy: A Tapestry of ideas and inquiries*. Brill Sense.

Lee, J. (2019). A training project to develop teachers' assessment literacy. In E. White & T. Delaney (Eds.), *Handbook of research on assessment literacy and teacher-made testing in the language classroom* (pp. 58–80). IGI Global.
[Crossref]

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>
[Crossref]

Liu, J., & Xu, Y. (2017). Assessment for learning in English language classrooms in China: Contexts, problems and solutions. In H. Reinders, D. Nunan, & B. Zou (Eds.), *Innovation in language learning and teaching: The case of China* (pp. 17–38). Palgrave Macmillan.
[Crossref]

Mackey, A., & Gass, S. M. (2005). *Second language research methodology and design*. Lawrence Erlbaum Associates.

Malone, M. E. (2017). Training in language assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment, encyclopaedia of language and education* (pp. 225–239). Springer International Publishing AG.
[Crossref]

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). SAGE.

Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(S1), 190–208. <https://doi.org/10.1111/modl.12308>
[Crossref]

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249–258. <https://doi.org/10.1191/0265532204lt283ed>
[Crossref]

Saputra, E. R., Hamied, F. 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>
[Crossref]

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>
[Crossref]

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238–245.

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>

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>
[Crossref]

Vogt, K., & Tsagari, D. (2022). Language assessment literacy and test validation: Highlighting the role of language teachers. *Studies in Language Assessment*, 11(1), 1–15. https://arts.unimelb.edu.au/_data/assets/pdf_file/0006/4282557/SiLA-11.1-Vogt_and_Tsagari.pdf

Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly*, 7(4), 317–342. <https://doi.org/10.1080/15434303.2010.516042>
[Crossref]

Xerri, D., & Briffa, P. V. (2018). Introduction. In D. Xerri & P. V. Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 1–7). Springer International Publishing AG.
[Crossref]

Xu, Y. T. (2017). *Assessment literacy of university English instructors in China: A mixed-methods study* (Doctoral dissertation), The University of Hong Kong. The HKU Scholars Hub.

Xu, Y. T., & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133–158. https://www.altanz.org/uploads/5/9/0/8/5908292/9.si7xubrown_final_formatted_proofed.pdf

Yan, X., & Fan, J. (2021). “Am I qualified to be a language tester?”: Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246. <https://doi.org/10.1177/0265532220929924>
[Crossref]

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>
[Crossref]

Yin, R. Y. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.

Zhang, L., Zheng, Y., & Yang, D. (2024). Exploring Chinese university English teachers’ language assessment literacy: A mixed-method study. *Asia Pacific Journal of Education*, 1–19. <https://doi.org/10.1080/02188791.2024.2354684>

2. Literature Review

Ling Gan¹✉ and Ricky Lam²✉

- (1) School of Languages and Communication, Beijing Technology and Business University, Beijing, China
- (2) Department of Education and Psychology, Hong Kong Baptist University, Hong Kong, China

✉ **Ling Gan (Corresponding author)**

Email: linggan@btbu.edu.cn

✉ **Ricky Lam**

Email: rickylam@hkbu.edu.hk

2.1 Overview

LAL was originally derived from assessment literacy in general education. Coined by Stiggins ([1991](#)), assessment literacy as a term refers to the knowledge and skills an assessment-literate person (teachers, school administrators, test users, etc.) needs to construct, administer, or complete designated assessment tasks. Since then, its scope has been broadened. Drawing on assessment literacy, the term “language assessment literacy”, used by Inbar-Lourie ([2008](#)) for the first time, shares the general qualities with assessment literacy but keeps its unique feature. In other words, LAL embeds language construct into knowledge, skills, and practices in testing and assessment (Harding & Kremmel, [2016](#); Inbar-Lourie, [2008](#)). This chapter reviews LAL definitions, frameworks and existing studies related to language teachers’ LAL status quo and development. Teachers’ assessment identity along with their LAL development is also reviewed and discussed. Based on the review, a theoretical framework

and two main research questions are proposed to inform and guide the current study.

2.2 Definition of LAL

Subsumed into assessment literacy, LAL has been defined differently. Taylor ([2009](#)) used “assessment literacy” to describe LAL as “the level of knowledge, skills, and understanding of assessment principles and practice that is increasingly required by other test stakeholder groups, depending on their needs and context” (p. 24). O'Loughlin ([2013](#)) expanded the scope of LAL beyond the test score uses and advocated a critical perspective by defining LAL as “the acquisition of a range of skills related to test production, test score interpretation and use, and test evaluation in conjunction with the development of a critical understanding about the roles and functions of assessment within education and society” (p. 363). This socially embedded view of LAL coincides with Fulcher's ([2012](#)) definition of LAL. Fulcher ([2012](#)) elucidated that LAL entailed not only knowledge, skills, and abilities to cope with large-scale tests and classroom-based assessment but also understanding of assessment theoretical underpinnings and “the role and impact of testing on society, institutions and individuals” (p. 125).

The three scholars conceptualise LAL as assessment-relevant knowledge, skills, or abilities, highlighting the interaction between assessment activities and the social context. However, all of them barely emphasise understanding of language, language teaching or learning. That is to say, they fail to include *language* in their definitions, a crucial construct that differentiates LAL from assessment literacy in general education (Kremmel et al., [2017](#)).

To address this gap, LAL scholars tend to underline the role of language when defining LAL. Inbar-Lourie ([2008](#)) set a good example. She conceptualised LAL as a “language assessment knowledge base” (p. 390), consisting of multilayers of assessment skills incorporated with specific language competencies. The base enabled stakeholders to “ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the results” (p. 389). The term

“language assessment literacy” was used for the first time from then on in language testing and assessment.

Taking the language construct into consideration, LAL has been understood from the perspectives of both teaching and non-teaching professionals. For example, Malone ([2013](#)) saw LAL as “language instructors’ familiarity with testing definitions and the application of this knowledge to classroom practices in general and specifically to issues related to assessing language” (p. 329). This view of LAL from language teachers’ perspectives was echoed in the studies by Vogt and Tsagari ([2014](#)) and Lam ([2015](#)). Vogt and Tsagari ([2014](#)) defined LAL as language teachers’ ability to “design, develop, and critically evaluate tests and other assessment procedures, as well as the ability to monitor, evaluate, grade and score assessments on the basis of theoretical knowledge” (p. 377). Similarly but differently, Lam ([2015](#)) understood LAL as teachers’ mastery of language assessment knowledge, skills, and procedures and their understanding of these elements of students’ learning within a sociocultural context. These scholars underscore the central role of LAL for language teachers in applying their LAL to classroom-based assessment practices. Arguing that LAL cannot confine itself to language teachers, Pill and Harding ([2013](#)) took LAL as “a repertoire of competencies that enable an individual to understand, evaluate and, in some cases, create language tests and analyse test data” (p. 382). They believed that “an individual” could be anyone who is involved in language testing and assessment activities. This perspective is consistent with Inbar-Lourie’s latest LAL definition (2017) that involves various assessment-related stakeholders.

The much-cited LAL definitions scholars have provided, though with various foci, have shown that LAL is multi-layered (Inbar-Lourie, [2013](#)). It can be an ability (Malone, [2013](#)), a context-dependent knowledge base (Fulcher, [2012](#); Inbar-Lourie, [2008](#)), and competence (Baker, [2024](#); Baker & Riches, [2018](#)) that empowers stakeholders with theoretical and practical language testing and assessment knowledge and skills. LAL involves not just knowledge and skills but also ethical practices in specific social contexts where an “assessment culture”, the assessment approach that an institution or a school chooses to take, needs to be promoted (Inbar-Lourie, [2008](#), p. 387). Based on the review of definitions, LAL in this study has been operationalised as language

teachers' competence in acquiring and applying language assessment knowledge, skills, and principles into assessment practices, in which they make informed decisions on students' performance and evaluate the impact of these decisions within a sociocultural context.

2.3 LAL Frameworks

The review of LAL definitions above disclosed two major divergences in LAL, namely, what elements should be entailed in LAL and what level of LAL should be needed by stakeholders. Centring around these issues, theoretical frameworks have been mapped out to explore the nature of LAL.

2.3.1 LAL Component Frameworks

Brindley ([2001](#)) is the first scholar to discuss LAL components in language testing and assessment (Fulcher, [2012](#)). He proposed that professional assessment development programmes should include five components with two core units and three optional units:

1. The social context of assessment (core unit)
2. Defining and describing proficiency (core unit)
3. Constructing and evaluating language tests (optional unit)
4. Assessment in the language curriculum (optional unit)
5. Putting assessment into practice (optional unit)

The first core unit involved considerations of social, political, and educational elements of assessment in teaching situations and wider communities. For instance, teachers need to consider assessment culture in different countries when designing and implementing assessment activities. Ethical considerations such as fair test use, and result reports are also paramount in this core unit. The other core unit was related to the definition and description of proficiency such as decisions made on students' performances. In addition, teachers should

acquire theoretical knowledge of language testing and assessment, including concepts of validity and reliability.

Brindley viewed the remaining three units of the model as optional. One unit was pertinent to the development and evaluation of language tests by introducing statistical analysis. Another optional unit suggested assessment in the curriculum, associated with using assessment alternatives (e.g., observation schedules, portfolios, self-assessment, project work) to construct criterion references for assessing progress and achievement. The last unit involved specifying an action plan to explore issues raised in assessment courses or workshops. For instance, teachers could apply for assessment-relevant projects or conduct classroom-based assessment research.

Brindley's modular outline underlines different assessment needs for language teachers, providing a useful framework for the LAL components. His modular framework thereby is regarded as the first and important step to introduce LAL in language testing and assessment (Inbar-Lourie, [2017](#)). However, this framework has been criticised for its understated roles of the three optional components. These components should be mandatory for language teachers in LAL professional development (Harding & Kremmel, [2016](#); Inbar-Lourie, [2008](#)) and essential for language teachers who need to validate assessments, particularly formative assessment in classrooms (Gu, [2021](#)).

Drawing on an analysis of Brindley's model ([2001](#)), Inbar-Lourie ([2008](#)) outlined three components for constructing LAL knowledge base: the "why" (rationales for assessment, similar to Brindley's core unit 1), the "what" (a theoretical basis of language teaching, learning, and assessment, similar to Brindley's core unit 2), and the "how" (assessment skills and methods, similar to Brindley's optional units 1–3). She agreed with Brindley that the social context of assessment was mandatorily considered in professional development programmes. She also argued that assessment was an individualised and socially constructed practice. This perspective was aligned with the current epistemological importance attached to the role of language testing and assessment in society. Therefore, the "why" component, namely, the social consequences of language assessment, should be both a component of assessment culture and a focus for LAL.

Inbar-Lourie also elaborated on her “what” and “how” aspects of LAL by commenting on Brindley’s outline. When discussing the trait of the “what”, Inbar-Lourie ([2008](#)) affirmed Brindley’s argument that language teachers should be well versed with concepts of validity and reliability, knowledge of language, language use, and language teaching pedagogy, constructing their individualised LAL knowledge base. In the component of the “how”, different from Brindley, she maintained that the assessment methods should be mandatory for all classroom practitioners, not just for those professionals who are required to gain advanced knowledge for assessment research. Assessment methods may vary because of different requirements of knowledge and skills for classroom-based assessment and standardised tests.

In the same year when Inbar-Lourie ([2008](#)) proposed her Why-What-How component outline, Davies ([2008](#)) identified a trend of skills, knowledge, and principles embedded in language testing textbooks. These three aspects were viewed as LAL components acquired by language assessment literate professionals (Fulcher, [2012](#)). Skills provide education with methodological tools for item writing, test analysis, statistics, and software programmes associated with testing procedures. Knowledge offers background relevant to measurement and language description in context setting. Principles refer to the proper use of tests, concerned with test fairness, ethics, and impact. Davies’ LAL triplet indicates that language assessment professionals should base their assessment activities on the solid assessment knowledge and should be able to take a critical view towards their own assessment practices.

Davies’ skills, knowledge, and principles are almost equivalent to the “How”, “What” and “Why” outlined by Inbar-Lourie ([2008](#)), which suggests stakeholders’ competence of test design and practice analysis, the setting for measurement, language proficiency, and contexts, and the social dimension of language testing and assessment. In reality, more emphasis has been placed on skills and knowledge than principles, a trend that also can be found in the studies of language testing courses (e.g., Jeong, [2013](#); Jin, [2010](#)). Both Inbar-Lourie ([2008](#)) and Davies ([2008](#)) acknowledged LAL as a knowledge base or a toolkit, with ethical considerations, but they both failed to recognise the interactions among these components.

Fulcher's (2012) expanded LAL definition model filled the aforementioned gap by disclosing the connections among LAL components (see Fig. 2.1). His definition, informed by data collected from 278 international language instructors through a survey, has broadened the scope of Davies' skills-knowledge-principle model. Fulcher's model has three layers, with practices at the bottom layer, principles in the middle, and contexts on the top. Practices, as the foundation of LAL, entail knowledge, skills, and abilities involved in language testing. Principles work as the guidance for practice, such as the processes, principles, and concepts related to testing and assessment. Contexts, at the topmost level, refer to a historical, social, political, and philosophical framework where practices and principles are enacted. Within the contexts, stakeholders can understand the origins and impacts of as well as reasons for language testing and assessment.

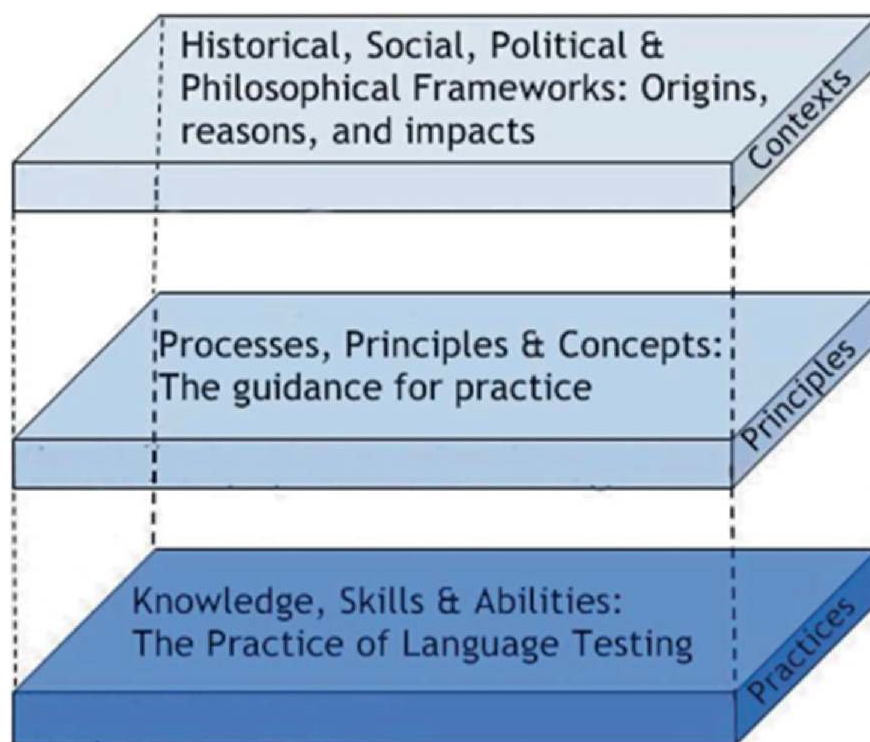


Fig. 2.1 Language Assessment Literacy: An Expanded Definition. *Note* Adapted from Fulcher (2012, p. 126)

Fulcher's LAL conceptualisation emphasises the interplay of the components at different levels and the view that language assessment

occurs within various contexts. His understanding of LAL indicates that stakeholders become assessment literate only when they fully understand how assessment practices are integrated into contexts. This conceptualisation is also geared to both large-scale and classroom-based tests and assessment, showing that language teachers' LAL is not confined to classroom assessments (Giraldo, [2018](#)). What is worth mentioning is that Fulcher ([2012](#)) argued not all elements were equally important for all stakeholders. Yet, he did not elucidate what components of LAL were needed for which group of stakeholders.

The aforementioned modular LAL frameworks have pinpointed that LAL is multidimensional and consists of diverse components. They also posit that LAL is a polarised concept, categorising stakeholders into being either literate or illiterate. However, this dichotomy seems to oversimplify stakeholders' LAL development and fails to capture the degree of the development.

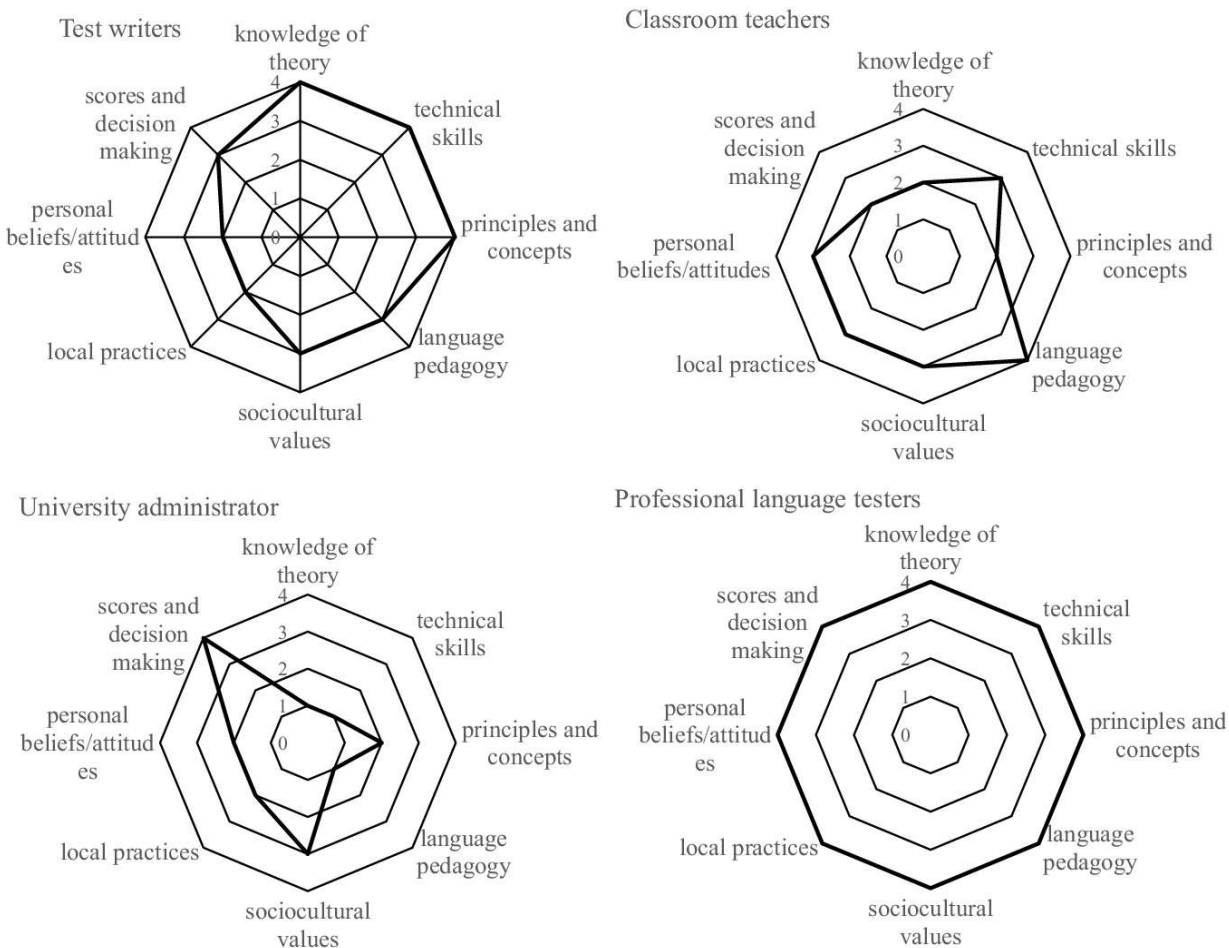
2.3.2 LAL Continuum Frameworks

A big leap in the development of LAL profiles is a shift from a componential view of LAL to a developmental perspective, namely, the LAL developmental progress (Kremmel & Harding, [2020](#)). Informed by a study on LAL among non-practitioners, Pill and Harding ([2013](#)) identified five proficiency levels of LAL: illiteracy, nominal literacy, functional literacy, procedural and conceptual literacy, and multidimensional or expert literacy. They argued that LAL could develop from being illiterate to expert literate in a continuum way. This continuum framework is valuable, as it accurately describes the developmental stages of LAL and showcases the dynamic nature of LAL. In addition, it innovatively combined mathematical literacy education with language research, highlighting LAL to be incremental proficiency.

However, this continuum framework is subjected to theoretical challenges. First, as Harding ([2016](#)) cautioned, this framework overemphasises the theoretical knowledge (i.e., concepts of assessment) and the descriptions at each level are inconsistent and ambiguous. Regardless of these deficiencies, the descriptions are still informative, as researchers can identify stakeholders' LAL levels through their own interpretations of findings. Second, the social-historical contextualisation of language testing and assessment in this

model is positioned on the top only for those who develop the advanced expertise of LAL. Yet, it has been widely recognised that LAL is context-specific for all assessment-literate professionals (Inbar-Lourie, [2017](#)). What matters most is that this continuum framework, similar to Fulcher's model, does not articulate what levels of LAL different stakeholders should achieve, as not all stakeholders need to develop their LAL into an expert level. Regarding this thorny problem, Taylor's ([2013](#)) LAL profile seems to provide a solution.

Taylor ([2013](#)) mapped eight LAL dimensions into Pill and Harding's ([2013](#)) five levels of LAL proficiency to construct a LAL profile for different stakeholders: test writers, classroom teachers, university administrators, and professional language testers (see Fig. [2.2](#)). These eight dimensions include knowledge of theory, technical skills, principles and concepts, language pedagogy, sociocultural values, local practices, personal beliefs/attitudes, and scores and decision-making. As Taylor ([2013](#)) implies LAL is not value-free, it embraces the perceptions and beliefs stakeholders bring to their assessment practices.



Note. Adapted from Taylor (2013, p. 410).

Fig. 2.2 Taylor's LAL Profile. Note Adapted from Taylor (2013, p. 410)

According to Taylor, different stakeholders should develop their LAL to an ideal level (0 to 4 stands for the incremental levels defined by Pill and Harding (2013) in the eight dimensions, depending on the specific assessment context and assessment needs. Figure 2.2 shows that the eight dimensions are placed in a spiderweb and that the more important the dimension is, the further it is placed in the web. For instance, the diagram of the LAL profile for classroom teachers suggests that the most important dimension for them is knowledge of language pedagogy, i.e., how to use assessment for pedagogical purposes (e.g., assessment for learning strategies). They should also master technical assessment skills, be familiar with local assessment practices and the social values shaping these practices and reflect on their attitudes towards assessment. Comparatively, principles and concepts of

assessment seem less important for classroom teachers. This spiderweb unfolds two essentials of LAL for stakeholders: components and levels.

However, there are challenges against the content of the eight dimensions and the exact levels stakeholders should achieve (e.g., Tsagari, [2021](#)). As Taylor's profile is based on researchers' assumptions instead of empirical data, the eight dimensions are more speculative than data-informed. They are ambiguous and sometimes overlap each other. For instance, "concepts" in "principles and concepts" may refer to such terms as validity, reliability, and so forth, overlapping with "knowledge of theory" to some extent. In addition, the ideal levels in Taylor's LAL profile are not always aligned with the realities. For example, classroom teachers were reported to require more advanced literacy in local practices and decision-making (Baker & Riches, [2018](#)) or less literacy in technical skills and language pedagogy (Lan & Fan, [2019](#)) than Taylor proposed. Nevertheless, Taylor's LAL conceptualisation has provided a useful tool for researchers to identify different stakeholders' LAL needs in specific educational settings. A lack of definitions may complicate the operationalisation of the components, but these components can be reinterpreted according to researchers' purposes and thus the interpretation becomes a starting point for empirical investigation.

After Taylor's LAL profile, Baker and Riches ([2018](#)) took the initiative to attest this profile for language teachers and assessment professionals by eliciting empirical data from a training programme. They synergised the eight components of Taylor's LAL profile into seven and relabelled them. These researchers also placed "sociocultural value" outside the spiderweb to highlight its impact on stakeholders' LAL. Although the level of each component achieved by the two groups of participants is experiential rather than data-driven, this study chartered a new venue for researchers to explore stakeholders' LAL needs in specific educational contexts.

Following Baker and Riches' study (Baker & Riches, [2018](#)), other scholars provided a sounder empirical basis by adopting a view of multidimensional models with LAL levels for various stakeholders. For instance, Kremmel and Harding ([2020](#)) identified nine LAL components with a large-scale survey on different stakeholders' LAL against the

eight components in Taylor's LAL profile. Like previous researchers, Kremmel and Harding (2020) did not define these components either. Moreover, these components were too complex, posing challenges to the operationalisation in empirical studies. Meeting these challenges, Bøhn and Tsagari (2021) built up a teacher LAL profile from a teacher educators' perspective. They operationalised each component informed by previous studies and validated it in the Norwegian educational context. This study provides essential evidence to show how Taylor's LAL profile can be further developed.

In sum, the conceptualisation of LAL's components and the discussion of LAL's scope and boundaries are still ongoing. It will be continuous to develop context-specific, pluralistic but differentiated frameworks or models that can address various needs of stakeholders (Vogt et al., 2024). LAL conceptualisation is an indicator of researchers' endless efforts to understand what LAL should entail and "who can be qualified to disseminate it and how" (Elder, 2017, p. 81). It is also implied that LAL is a concept that cannot be defined in limited terms (Giraldo, 2018). Focusing on different LAL competencies, all LAL frameworks underscore stakeholders' ability to describe language proficiency, understand assessment theories and methods, know how to conduct assessment practices and use assessment results within appropriate socio-cultural contexts.

2.4 Research on Teachers' LAL

While emerging LAL frameworks have informed and advocated LAL studies among various stakeholders, language teachers continue to be the central stakeholders and are expected to develop their LAL foremost (Giraldo, 2018). This can also be manifested by the fact that a majority of LAL studies have focused on teachers (Harding & Kremmel, 2016), a situation in which most of the LAL definitions are mainly drawn from teachers' data (e.g., Lam, 2015; Malone, 2013). The following sections will detail these studies on teachers.

2.4.1 Teachers' LAL Level and Needs

Investigating the current level of teachers' LAL is a starting point for LAL scholarship (Xu & Brown, 2017). Language teachers' lack of LAL

has been well documented in both English-speaking and traditionally categorised English-speaking ‘outer circle’ countries (e.g., Kunnan et al., [2021](#); Sayyadi, [2022](#); Sun & Zhang, [2022](#)), and this LAL scarcity has been mainly ascribed to insufficient assessment training among teachers (Berry et al., [2019](#)).

Despite scarce assessment training opportunities, language teachers expressed their assessment training needs and prioritised their needs in specific aspects (Gan & Lam, [2020](#)). For instance, Hasselgreen et al. ([2004](#)) found that language teachers voiced training needs in LAL, particularly training in test designs and alternative assessments such as portfolio and self-/peer assessment. This finding was also echoed in the studies by Vogt and Tsagari’s research team (2014, 2017, 2020). Their studies focused on assessment training among 853 language teachers across seven European countries. They confirmed that language teachers lacked assessment training, hence, a deficiency of LAL and that teachers’ assessment training needs were prioritised differently depending on their varied educational contexts.

Due to the context-specific nature, language teachers’ training needs vary individually, especially when individual teachers bring their unique assessment knowledge, skills, and perceptions into the teaching and assessment process (Scarino, [2013](#)). For instance, language teachers preferred training for assessment knowledge, skills, and practices to assessment-relevant theories (Firoozi et al., [2019](#); Lam, [2019](#)), but they also had greater interests in assessment concepts and content and psychometrical analysis of language assessment (Zhang et al., [2024](#)), as understanding the theoretical underpinnings of assessment could boost their confidence in critiquing the effectiveness of assessment (Lan & Fan, [2019](#)). Although these studies have provided valuable insights into teachers’ LAL status quo and needs, they rely much on self-reported questionnaires or interviews, overlooking teachers’ actual LAL needs in their assessment practices.

2.4.2 Teachers’ LAL Development

A clear lack of language teachers’ LAL in assessment activities and their concerns with the suitability of assessment training has pushed research on LAL development to the forefront. Assessment training or courses are foremost believed to empower teachers with language

assessment knowledge, skills, practices, and theories in a professional way (Malone, [2017](#)). Thus, enhancing teachers' LAL through assessment training or courses has been applauded in LAL scholarship. Take Baker and Riches's ([2018](#)) study as an example, they found that Haitian language teachers' LAL had been facilitated through short-term assessment training workshops. This facilitative impact was salient in assessment skills, a better understanding of the relationship between assessment and teaching/learning, and more confidence in teacher learning. Similar findings were reported in the study by Lee ([2019](#)), in which an assessment training project was introduced to promote pre-service ESOL teachers' LAL. In terms of assessment courses, Levi and Inbar-Lourie ([2020](#)) discovered that teachers had gained general and language-specific assessment literacies by attending a general course on classroom-based formative assessment.

Nonetheless, problems of assessment training/courses have emerged when they are used to enhance teachers' LAL. For example, the time allocated for conceptual knowledge and hands-on practices is not evenly distributed (Jin, [2010](#)). Another problem is the imbalance in the content of classroom-based and large-scale assessments (Lam, [2015](#)). The biggest challenge is that LAL teaching and learning methods used in assessment training/courses have been unravelled (Fulcher, [2020](#)). All these pedagogical problems may lead to questions about the effectiveness of training or courses adopted for LAL enhancement.

As a result, a number of alternative methods other than assessment training programmes/courses have been proposed to facilitate LAL development for language teachers at different career stages. For pre-service teachers, learning journals could monitor prospective teachers' LAL development progress, as writing journals helped student teachers re-examine their understanding of assessment theories and practices (Bolivar, [2020](#)). Individualised tutoring and peer debriefing could also be effective for providing instructional feedback and techniques to support teacher candidates' LAL development (Odo, [2016](#)).

For in-service teachers, language teachers' involvement in test design and development has enhanced their LAL. Cases in point are the studies by Koh et al. ([2018](#)) (the development of test design), Kremmel et al. ([2018](#)) (writing items for high-stakes language exams), Walters ([2010](#)) (developing language test specifications), Ho and Yan ([2021](#))

(writing essay prompts through a community of practice), and Janssen ([2022](#)) (joining a test retrofit). By involvement in test design and development, teachers can familiarise themselves with the assessment process in which they acquire theoretical concepts and accumulate test design skills, ultimately increasing their LAL. In the process of test development, collaboration between language teachers and language assessment specialists (Baker & Riches, [2018](#)) or researchers (Harding & Brunfaut, [2020](#)) has been promoted to ameliorate teachers' LAL. Collaboration requires interaction, which can generate a shared understanding of LAL growth between language teachers and other stakeholders (Gan & Lam, [2022](#)).

In addition, as a self-directed method, reflection on assessment practices has been proposed for in-service teachers to construct and sustain their LAL on the job (Babaii & Asadnia, [2019](#); Tian et al., [2021](#)). Through reflection on their assessment practices, teachers strengthen their autonomy and feel empowered with agency so as to consciously bridge the gaps between theories and practices and look for solutions to assessment problems (Gan & Lam, [2024b](#)).

To date, studies on teachers' LAL development have made significant contributions to strategies or methods used for LAL development. Both assessment training/courses and alternative methods are touted as a lever to uplift teachers' LAL. Yet there is an obvious deficiency among these studies: language teachers are not viewed as an active agent to construct their LAL knowledge base and acquire a LAL know-how. Instead, they receive assessment knowledge and skills imparted from teacher educators, assessment trainers, or researchers. In other words, the existing studies on teachers' LAL enhancement were not conducted in their own instruction settings but in those for research purposes (Gan & Lam, [2024a](#)). As such, these studies have neglected the role individual language teachers play in their LAL development and paid no heed to "how LAL develops within individuals" in their assessment practices (Taylor, [2013](#), p. 407).

A pinch of studies have explored the patterns or features of individual teachers' LAL development in their assessment practices by giving in-depth descriptions of teachers' natural milieu (i.e., instruction settings) (Giraldo, [2020](#)). Xu and Liu ([2009](#)) found that university English teachers constructed their LAL through self-reflection and

collaboration with students and colleagues. They underscored the contributions to LAL development of teachers' agency, prior experiences, assessment contexts, and power relationship. Similarly, Yastıbaş and Takkaç ([2018](#)) discovered that the structure of student and coursebook-driven LAL development enabled EFL teachers to validate their assessment, which would exert positive washback on students' learning. However, they failed to illustrate the dynamic nature of LAL development. This gap was filled by the study of Xu ([2017a](#)). He observed that novice EFL teachers experienced three stages in developing their LAL. The first stage was characterised by the acquisition of practical assessment skills. In the second stage, teachers were aware of connecting assessment with teaching objectives, but they confined this connection to using assessment results to improve future teaching. The third stage featured the inclination to conduct improvised assessment practices. His study has provided in-depth descriptions of teachers' natural assessment milieu, but it did not further examine how the developmental stages were formed.

LAL is regarded as a process from the practical perspective (Su et al., [2024](#)), which can be illustrated in the study by Yan and Fan ([2021](#)). Their study portrayed a LAL developmental process by proposing an apprenticeship-based and experience-mediated model of LAL development. They posited that stakeholders' LAL development was bottom-up and cyclical. In this process, stakeholders acquire their initial LAL from prior assessment experiences and enhance it by creating and utilising opportunities or resources and testing assessment practices to meet the requirements and challenges posed by assessment contexts. They also reflect on and evaluate their assessment knowledge, skills, and practices, from which, they accumulate assessment experiences. This experience accumulation, in turn, results in stakeholders' LAL development. Yan and Fan ([2021](#)) underlined the influence of assessment context and experience on LAL development, highlighting stakeholders' agentive actions including reflection, evaluation, familiarisation, adaptation, and application. Their model gives innovative theoretical clarifications of stakeholders' LAL development process and practical implications for LAL development. Nevertheless, this model should be attested with more empirical evidence.

In sum, the qualitative studies in the LAL literature have probed into language teachers' classrooms with a focus on their assessment practices, providing a thick description of teachers' LAL development in a bottom-up fashion. Nevertheless, these studies, except Xu ([2017a](#)) study, are mainly cross-sectional, lacking long-term research engagement. The paucity of longitudinal research into LAL development may result in insufficient in-depth descriptions and incomplete observations on LAL developmental patterns, which would undermine research trustworthiness (Gan & Lam, [2022](#)).

This gap can be addressed in our study by adopting a longitudinal qualitative approach to look into the LAL developmental process of four university English instructors in their teaching contexts. With the longitudinal approach, our study aims to provide a more comprehensive picture of LAL enhancement.

2.4.3 Teacher Assessment Identity in LAL Development

To better understand the complexity of teachers' LAL development, it is essential to explore this construct with other concepts, including teachers' professional identity in language assessment. The professional identity related to assessment was introduced by Adie ([2013](#)) as the term "teacher assessment identity" in her study of observing teachers' assessment process in an online course. It is used to denote teachers' perceptions of themselves as an assessor and their concerns about how they are judged by others when they are making professional judgements or decisions on students' performance (Adie, [2013](#)).

Assessment identity literature has underscored not only what teachers do in their assessment practices but also who they are when they conduct assessment practices (Grainger & Adie, [2014](#); Looney et al., [2018](#); McDowell & Harman, [2008](#)). For example, Looney et al. ([2018](#)) proposed that teachers' identity as assessors should be reconceptualised as a construct encompassing not only knowledge and skills but also the beliefs and feelings about assessment. Similarly, Wyatt-Smith and Adie ([2021](#)) understood teacher assessment identity as teachers' personal and professional identification of their assessment activities, including assessment knowledge, skills, practices, confidence, and beliefs in the assessment process. Implicitly, these

researchers have viewed teacher assessment identity as an entity beyond assessment literacy which focuses on teachers' assessment knowledge, skills, conceptions and context-dependent assessment practices.

Based on these conceptual discussions, researchers began to explore the underlying components of teacher assessment identity. Both Estaji and Ghiasvand ([2021](#)) and Moqaddam et al. ([2021](#)) investigated what should be included in teacher assessment identity by developing and validating a Teacher Assessment Identity questionnaire with EFL teachers in Iranian contexts. Their studies revealed that teacher assessment identity was multidimensional and closely relevant to and beyond teachers' assessment literacy, with different terminologies used to indicate its context-specific nature.

With its multiple components (e.g., assessment responsibilities, dispositions, knowledge domains, or social context), teachers' assessment identity can also be categorised into different types. Coombs et al. ([2020](#)) identified three subgroups of pre-service teachers as assessors: eager, contemporary, and hesitant assessors. Differently, Leonardsen et al. ([2022](#)) portrayed five faces of being an assessor enacted in assessment practices: the quality controller, the educator, the fosterer, the motivator, and the negotiator. Diverse types of assessment identities require teachers' varying degrees of assessment literacy/LAL when they enact these identities in assessment practices (Moqaddam et al., [2021](#)).

The construction of these assessment identities is progressive and dynamic (Adie et al., [2020](#)), possibly shifting over time owing to a host of internal and external factors. The dynamics are evinced in the studies by Cowie et al. ([2014](#)) and Dao ([2021](#)). The former discovered student teachers (re)shaped their assessment identity through the negotiation of the differences or even conflicts caused by personal and contextual experiences. The latter explored the process of an experienced EFL teacher's construction of his assessment identity temporally, socially, and spatially, highlighting the discontinuity of teacher assessment identity and the active interaction of personal and professional factors in shaping the identity. The development of teachers' assessor identity is not a straightforward process (Sheehan & Munro, [2019](#)), but is "responsive to events and circumstances" (Looney

et al., [2018](#), p. 446). These events or circumstances may include the policy context, the provision of targeted resources and teachers' collaborations that contribute to grading decisions (Adie et al., [2024](#)).

(Re)conceptualisation of teacher assessment identity also indicates that teacher assessment identity is closely linked with teachers' LAL. Teachers' identity as assessors can provoke teachers' thinking, stimulate the interaction between teachers and students, and help them with making professional decisions (Adie, [2013](#); Looney et al., [2018](#)). Also, the assessment discourse teachers create may be "potentially productive and facilitative as well as inhibiting to effective classroom assessment" (Rea-Dickins, [2004](#), p. 255). In view of these advantages, Scarino ([2013](#)) considered teachers' self-awareness of the identity as an assessor to be an indispensable component of LAL, whereas Xu and Brown ([2016](#)) took teachers' assessment identity (re)construction as the pinnacle of teacher assessment literacy in practice. These scholars argued that in order to achieve assessment literacy enhancement, teachers should reframe their traditional view of teachers-as-instructors into the view of teachers-as-assessors. This shift in teacher assessment identity can be realised through contextualised reflective learning (DeLuca & Braund, [2019](#)), a learning method grounded on teachers' understanding that assessment literacy/LAL is an evolving entity associated with contexts and experiences.

Yet few empirical studies have been conducted to explore how teachers' construction of assessment identity boosts their LAL development. Only Xu ([2019a](#)) adopted a case study to touch upon this research topic. Her study discovered that teachers' growing self-awareness of the identity as an assessor prompted them to conduct action research, which provided an impetus for teachers' informal learning about assessment and this informal learning led to the participant's LAL development. Apparently, constructing the identity as an assessor became a catalyst for teachers' assessment literacy development.

The aforesaid studies have generated insights to advance our nuanced understanding of the connection between teacher assessment identity and LAL development. The connection is complex, as these two constructs which seem to be different entities, are in fact closely related and even interchangeable without clear definitions (Coombs & DeLuca,

[2022](#); Koh et al., [2018](#)). As such, there is a need for researchers to conceptualise these two constructs through empirical studies.

Also, it should be noted that in the existing literature, teachers seem to be endowed with the identity as an assessor the moment they undertake the assessment responsibility or accountability. Few studies have been conducted to delve into how language teachers perceive this identity, and more importantly, how they fabricate their assessment identity over time, a process engrained in teachers' LAL development. The current study, situating the understanding of teacher assessment identity in Adie's ([2013](#)) conceptualization, attempts to examine the construction of university English teachers' assessment identity over time along with their LAL development.

2.4.4 Factors that Influence Teachers' LAL Development

Assuming that teachers' conceptions of assessment and practices are shaped by various social, cultural, and educational beliefs, researchers perceive LAL as a dynamic, situated, and process-orientated practice in which stakeholders' personal experiences, knowledge, or beliefs interplay with social, historical, cultural, and professional variables (Inbar-Lourie, [2017](#); Yan & Fan, [2021](#)). LAL requires the involvement of teachers' knowledge, skills, principles, and self-awareness through which teachers negotiate their assessment practices and assessment identities with learners or other teachers within a learning community (Harding & Brunfaut, [2020](#); Koh et al., [2019](#); Willis et al., [2013](#)). As teachers' knowledge, skills, principles, and self-awareness are value-laden (Taylor, [2013](#)), LAL itself is definitely inseparable from the social, cultural, historical, and educational considerations. Therefore, teachers' LAL development is situated in specific educational settings where assessment culture may or may not be compatible with institutional or social values (Giraldo, [2020](#); Inbar-Lourie, [2017](#)).

Crusan et al. ([2016](#)) discovered two factors that involved in teachers' LAL development. One was related to assessment context and the other to assessment experience. Yan et al. ([2018](#)) categorised these two factors as contextual and experiential factors respectively. Contextual factors refer to those relevant to education, institutions, history, culture, or politics within a specific context and experiential

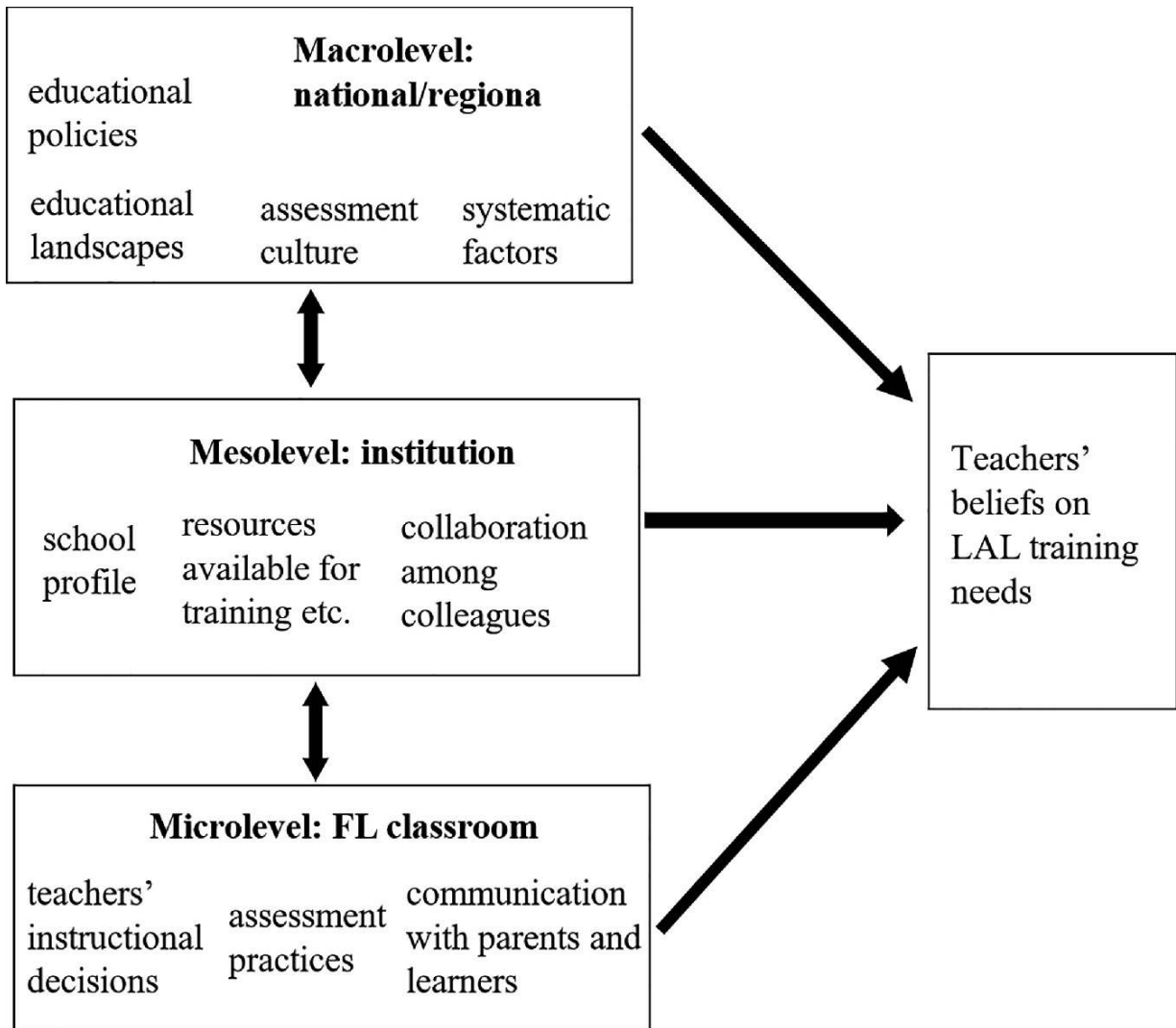
factors are teachers' prior experiences with language teaching, learning, and assessment.

Xu ([2019a](#)) underlined the macro-, meso-, and micro-contextual factors in teachers' language assessment practices. The macro contexts subsume broader influences caused by historical, social, cultural, and political factors, including the cultural norms about assessment and education and the educational policies at various levels. These macro contextual variables, Xu ([2019b](#)) acknowledged, impacted teachers' LAL indirectly. The meso-contextual factors, instead, had a direct impact on teachers' LAL. They refer to those factors at the institutional level such as school management or interpersonal relationship in workplaces. Comparatively, owing to their direct connections to classroom practices, micro-level contextual factors related to local classroom contexts have been given greater attention in LAL scholarship, including class size and assignments. However, as Xu ([2019b](#)) noted, studies on the three levels of contextual factors did not generate data from direct observations of teachers' assessment practices.

Xu's triple-level contextual factors were identical to those outlined by Yan et al. ([2018](#)). They defined three categories of contextual factors as "larger educational, social, cultural, political and historical factors that collectively form the assessment culture in a particular context" (2018, p. 159). The three categories include educational landscape and policies which refer to the assessment culture such as an exam-driven culture and local or national assessment policies, institutional mandates including requirements of using assessment to improve teaching and learning, and local contexts pertinent to local assessment training resources or constraints (e.g., workload, administrative load, or collaborative working environment) teachers are confronted with.

However, these categories were criticised for being ambiguous and overlapping (Vogt et al., [2020](#)). Vogt et al. ([2020](#)) suggested a context-categorised theoretical framework from different contextual levels (see Fig. [2.3](#)). The macro-level refers to a national or regional level, encompassing national educational policies and assessment culture shaped by historical, political, cultural, and social factors. The meso-level contextual factor is mainly characterised by the institutions, including the school culture, local resources available for training, and

communication with colleagues or principals. The micro-level is associated with the classroom context, including teachers' instructional decisions informed by assessment procedures, their prioritised assessment practices, and interaction with learners or parents. The framework also shows that these three levels of contextual factors are independent but interactive to exert the impact on teachers' LAL training needs.



Note. Adapted from Vogt, Tsagari, and Spanoudis (2020, p. 6).

Fig. 2.3 Contextual Levels Impacting on Teacher Beliefs of LAL Training Needs. *Note* Adapted from Vogt, Tsagari, and Spanoudis (2020, p. 6)

Contextual factors exerted an impact on LAL development through the mediation of experiential factors (Yan et al., [2018](#)). As mentioned previously, experiential factors signify teachers' prior experiences of language teaching, learning, and assessment (Yan et al., [2018](#)). For instance, teachers' linguistic background and their differing length of teaching experience made a difference in their self-reported assessment knowledge and practices (Crusan et al., [2016](#)). Specifically, novice teachers were more confident with their assessment knowledge in classroom than their experienced counterparts (Xie & Tan, [2019](#)).

Yan et al. ([2018](#)) also found that, while those contextual factors might gear teachers' attention to integrating teaching with testing, the experiences of assessment activities helped them operationalise this alignment. Meanwhile, the experiences in assessment practices, such as item analysis and communication with stakeholder groups, contributed to teachers' intuition about assessment concepts and principles. When contextual factors provide motivations and opportunities for teachers' assessment practices, those experiential factors reinforce their assessment knowledge and skills. These findings confirmed that teachers' LAL development "is largely contextualised and experiential" (Yan & Fan, [2021](#), p. 25).

Personal factors also matter in teachers' LAL development (Sultana, [2019](#); Xu, [2017b](#)). These factors have never been delineated in LAL literature, but they are usually linked with those factors that are brought about by individual themselves, including conceptions of assessment (Cowie et al., [2014](#); Edwards, [2017b](#)) which refers to teachers' personal beliefs, attitudes, or dispositions towards purposes and functions of assessment (Coombs et al., [2020](#); Edwards, [2017a](#); Fulmer et al., [2019](#)), self-awareness of identity as assessors (Adie, [2013](#); Scarino, [2013](#); Xu, [2017b](#)), and personal dispositions such as teachers' emotions or feelings towards assessment and teacher self-efficacy (Babaii et al., [2021](#)) and agency (Mansouri et al., [2021](#); Xu & Liu, [2009](#)).

Xu ([2017b](#)) revealed that conceptions of assessment shaped teachers' LAL from both cognitive and emotional perspectives. Cognitively, teachers' conceptions of assessment filter out those assessment theories and strategies that contradicted their conceptions and retained those that did not. Emotionally, these conceptions might

push teachers to avoid or keep using certain assessment activities owing to their positive or negative emotions or feelings with assessment as learners or teachers. Similar findings were found in Edwards' study (Edwards, [2017a](#)) where teachers' emotional engagement with students worked as either an amplifier or filter in teachers' assessment decisions, such as reducing the use of tests to mitigate students' stress.

In the current study, we differentiated teachers' prior experiences of language teaching, learning, and assessment (i.e., experiential factors) from personal factors whose focus is on teachers' beliefs, conceptions, emotions, and other dispositions. These two factors, together with contextual factors, have been addressed separately in LAL scholarship. Yet they are not mutually exclusive. Instead, they may interact with each other to jointly shape teachers' LAL and effect their assessment identity. Although studies have showcased that contextual factors are mediated through experiential factors to influence assessment training needs (e.g., Yan et al., [2018](#)), there is still limited knowledge about to what extent, the three factors are involved to affect teachers' LAL construction in a specific educational context.

2.5 Theoretical Framework of the Study and Research Questions

In view of research gaps in existing LAL studies, the current study adopted Taylor's LAL profile for classroom teachers (2013) as an overarching framework to explore the LAL development of university English instructors in the Chinese context. Taylor's LAL profile pinpoints not only LAL components but also LAL levels for each component language teachers require. Both LAL components and levels are the two key issues that LAL researchers have been attempting to address (Baker & Riches, [2018](#); Harding & Kremmel, [2016](#)). Moreover, Taylor's profile highlights that LAL is a socially situated practice. This emphasis together with its differentiation of LAL needs for language teachers makes Taylor's profile appropriate to address the two main research questions in our study.

However, this profile does not showcase a LAL development process in which language teachers construct their LAL and their assessment

identity by interacting with various factors pertinent to culture, history, education, politics, and individuals (Scarino, [2013](#)). To fill this void, we proposed an LAL framework (see Fig. [2.4](#)) to guide the present study by embedding contextual, experiential, and personal factors into Taylor's LAL profile for classroom teachers.

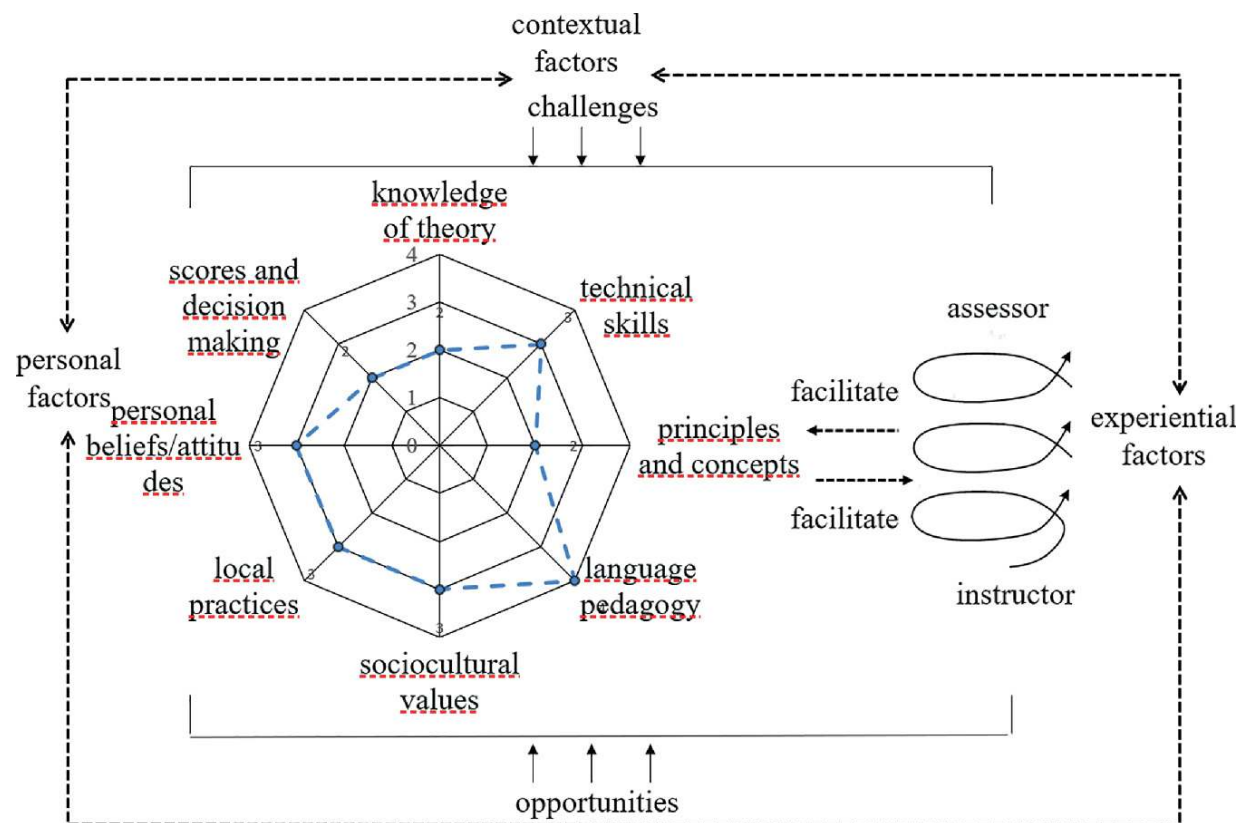


Fig. 2.4 An Adapted LAL Framework for Language Teachers

The framework posits that language teachers should develop their LAL components to an assumed level (0 to 4 stands for the continuum level described in Table [2.1](#)). These components may cover eight dimensions: knowledge of theory, technical skills, principles and concepts, language pedagogy, sociocultural values, local practices, personal beliefs/attitudes, and scores and decision-making. The dotted red underlines of each dimension, however, imply that not all of these eight dimensions are needed for language teachers in specific educational contexts. The dotted blue lines in the spiderweb show that the assumed levels may change in alignment with the LAL components needed in different educational settings. The needs for LAL

components and assumed levels may also fluctuate over time owing to the influence of contextual, experiential, and personal factors in given situated contexts. These factors may interact with each other to bring about possible opportunities and unexpected challenges for language teachers when they construct their LAL expertise and their identity as assessors. The identity construction may or may not facilitate teachers' LAL development, and vice versa. The construction process may not be linear, but spiral or even recurring.

Table 2.1 Operationalisation of the Eight Components in Taylor's LAL Profile

Dimension	Broad definition
Knowledge of theory	• Knowledge of languages, language learning, and assessment in a given context (Bøhn & Tsagari, 2021) • Knowledge of how languages are used in society (Kremmel & Harding, 2020)
Technical skills	• Developing and administrating assessments such as selecting appropriate tasks for an LTA purpose • Selecting appropriate ready-made assessments (Giraldo, 2018) • Giving appropriate feedback to students' performances (Xu, 2017b; Xu, 2019b) • Writing items (Davies, 2008) • Using self/peer assessment and other alternative assessments (Vogt & Tsagari, 2014) • Using techniques (e.g., statistics, questionnaires, interviews, analysis of language) to get information about the quality of LTA (Bøhn & Tsagari, 2021; Kremmel & Harding, 2020)
Principles & concepts	• Understanding the concept of reliability, validity, and ethics, and the way to interpret the test results about students' language ability (Davies, 2008; Fulcher, 2012; Giraldo, 2018)
Language pedagogy	• Knowing the relationship between assessment, learning, and teaching (Kremmel et al., 2018) • Knowing the use of assessment to promote learning and the teaching of disciplinary knowledge such as language or culture (Bøhn & Tsagari, 2021) • Knowing washback of assessment on teaching and learning (Kremmel & Harding, 2020; Xu & Brown, 2016)

Dimension	Broad definition
Sociocultural values	• Determining if a language assessment aligns with a local educational system (Giraldo & Murcia, 2018) • Determining if the results from a language assessment are relevant to the local context (Giraldo & Murcia, 2018) • Determining what impact the social and cultural values have on the use and development of tests and assessments in a given context (Bøhn & Tsagari, 2021) • Determining how assessments can be used to enforce social policies (e.g., certification, immigration, citizenship) (Jeong, 2013; Kremmel & Harding, 2020; Scarino, 2013)
Local practices	• Understanding local educational policies or locally developed assessment criteria (Bøhn & Tsagari, 2021) • Understanding school, classroom, or institutional dimensions of assessment acts (Scarino, 2013) • Understanding curriculum/course regulations or requirements (Giraldo, 2018)
Personal beliefs/attitudes	• Understanding how one's beliefs/attitudes might influence one's assessment practices (Scarino, 2013) • Understanding how one's beliefs/attitudes might conflict with those of other groups involved in assessment (Xu, 2017b; Xu & Brown, 2016)
Grading & decision-making	• Using rating scales to score speaking or writing performances; grading open-ended questions (e.g., short answer questions) (Kremmel & Harding, 2020). • Grading students at the end of the course (Xu, 2017b).

Additionally, contextual, experiential, and personal factors may signify different foci aligned with the purposes of the current study. Contextual factors are related to educational policies at national, regional, or university levels, university or department structures or administrative systems, and local classroom or instructional contexts. Experiential factors underline teachers' experiences with language teaching, learning and assessment, whereas personal factors stress teachers' beliefs or conceptions of assessment, self-awareness of being an assessor, and their emotional responses towards assessment. These factors, with various foci, are not mutually exclusive. Instead, they may overlap and cannot be differentiated distinctly.

The components in Taylor's LAL profile were operationalised to cater to the current study. Taylor ([2013](#)) did not specify the eight components in the profile and articulated that they were hypothesised

and referred to a conjecture by researchers in accordance with local educational contexts. Therefore, in the current framework, the eight LAL components are operationalised by referring to published LAL literature and presented in Table [2.1](#) below.

To better serve the purpose of the current study, we also operationalised the LAL continuum proposed by Pill and Harding ([2013](#)). In light of the unclear descriptions, we redefined the five levels of LAL by referring to the study by Kaiser and Willander ([2005](#)). The descriptions of the five levels are all anchored in the LAL research literature on language teachers (e.g., Baker & Riches, [2018](#); Lam, [2015](#); Yan et al., [2018](#)) and assessment literacy rubrics (e.g., Edwards, [2017b](#)). Table [2.2](#) summarises the broad descriptions of the five levels.

Table 2.2 Operationalisation of the Five Levels in Taylor’s LAL Profile

Level	Description
Illiteracy	Ignorance of the dimensions of language testing and assessment
Nominal literacy	A limited understanding of the dimensions of language testing and assessment, conducting basic assessment tasks guided by others, using naïve theoretical explanations and misconceptions
Functional literacy	A general understanding of the dimensions of language testing and assessment but yet to demonstrate some nuanced details in these dimensions, using the understanding to solve simple problems
Procedural and conceptual literacy	A clear understanding of the complexities of the dimensions of language testing and assessment within a given context, applying the understanding into practices, approaching problems by analytic reasoning
Multidimensional/expert literacy	A thorough understanding of the details of dimensions in language testing and assessment within a given context, using the dimensions in practice effectively, and being able to expand the understanding to wider assessment communities

An ever-increasing need to understand language teachers’ LAL development calls for more research undertakings to explore how teachers’ LAL is contextually situated in local realities. These research activities are likely to garner a comprehensive picture of the intricacy of LAL (Inbar-Lourie, [2017](#); Scarino, [2013](#)). Guided by the research gaps and the conceptual framework, our study addressed two main research questions:

1. How do university English instructors develop their LAL over time?
2. How do contextual, experiential and personal factors influence university English instructors' LAL development?

The first research question was developed further into three sub-questions:

- (1a) What are LAL needs for university English instructors?
- (1b) What are LAL developmental trajectories of university English instructors?
- (1c) How do university English instructors construct their assessment identity through LAL development?

The second main research question was also divided into two sub-questions:

- (2a) What are the opportunities and challenges mediated by contextual, experiential, and personal factors for university English instructors' LAL development?
- (2b) How do university English instructors mobilise the opportunities and challenges for their LAL development?

2.6 Summary

Although there is not yet a consensus on the LAL definition, this construct has sparked a considerable number of conceptual discussions and empirical studies. Conceptually, LAL has been sketched and framed from different perspectives in diverse educational contexts and for various research purposes. The conceptual discussions mainly centre around teachers' LAL needs which entail two key elements, namely, what components should be included in LAL and what LAL proficiency/level teachers should achieve. More research has been called for to identify these needs in a local educational context.

Empirically, three major research gaps have been identified. First, a majority of LAL studies are orientated to enhancing teachers' LAL through external interventions, and few have explored teachers' LAL developmental process in natural teaching contexts. Second, how teachers construct their identity as assessors in the process of LAL development remains under-researched. Third, as LAL development is mediated by various factors, what factors and how they influence teachers' LAL development in particular educational settings are underexplored. Adapting Taylor's LAL profile for classroom teachers (Taylor, 2013), we proposed a conceptual framework that synthesises the research gaps to guide the current study. These gaps will be addressed by adopting an ethnographic case approach to be introduced in the next chapter.

References

- Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, 20(1), 91–106. <https://doi.org/10.1080/0969594X.2011.650150> [Crossref]
- Adie, L., Stobart, G., & Cumming, J. (2020). The construction of the teacher as expert assessor. *Asia-Pacific Journal of Teacher Education*, 48(4), 436–453. <https://doi.org/10.1080/1359866X.2019.1633623> [Crossref]
- Adie, L., Wyatt-Smith, C., Finch, M., & DeLuca, C. (2024). Evidence of teacher assessment work and its relationship to their assessment identity. *Teaching and Teacher Education*, 141. <https://doi.org/10.1016/j.tate.2024.104518>
- 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> [Crossref]
- Babaii, E., Molana, K., & Nazari, M. (2021). Contributions of assessment-related critical incidents to language teacher identity development. *Innovation in Language Learning and Teaching*, 15(5), 442–457. <https://doi.org/10.1080/17501229.2020.1824234> [Crossref]
- Baker, B. (2024). Introduction. In B. T. Baker & Lynda (Eds.), *Language assessment literacy and competence volume 2: Case studies from around the world* (pp. 1–7). Cambridge University Press & Assessment.
- Baker, B. A., & Riches, C. (2018). 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>
[Crossref]

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>
[Crossref]

Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research*, 12(2), 222–233. <https://doi.org/10.17507/jltr.1202.02>
[Crossref]

Bolivar, E. M. R. (2020). Monitoring preservice teachers' language assessment literacy development through journal writing. *Malaysian Journal of Elt Research*, 17(1), 38–52.

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.

Coombs, A., & DeLuca, C. (2022). Mapping the constellation of assessment discourses: A scoping review study on assessment competence, literacy, capability, and identity. *Educational Assessment, Evaluation and Accountability*, 34(3), 279–301. <https://doi.org/10.1007/s11092-022-09389-9>
[Crossref]

Coombs, A., DeLuca, C., & MacGregor, S. (2020). A person-centered analysis of teacher candidates' approaches to assessment. *Teaching and Teacher Education*, 87, 1–13. <https://doi.org/10.1016/j.tate.2019.102952>
[Crossref]

Cowie, B., Cooper, B., & Ussher, B. (2014). Developing an identity as a teacher-assessor: Three student teacher case studies. *Assessment Matters*, 7, 64–89. <https://doi.org/10.18296/am.0128>
[Crossref]

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>
[Crossref]

Dao, T. B. N. (2021). A narrative inquiry of a Vietnamese university EFL teacher's assessment identity. *International Journal of TESOL & Education*, 1(3), 260–277. <http://i-jte.org/index.php/journal/article/view/109>

Davies, A. (2008). Textbooks in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
[Crossref]

DeLuca, C., & Braund, H. (2019). Preparing assessment literate teachers. *Oxford Research Encyclopedia of Education*, 1–21. <https://doi.org/10.1093/acrefore/9780190264093.013.487>

Edwards, F. (2017a). *The development of summative assessment literacy: An exploration of the experiences of beginner secondary science teachers in New Zealand* [Doctoral dissertation, The

University of Waikato]. The University of Waikato Research Commons. <https://researchcommons.waikato.ac.nz/bitstream/handle/10289/11567/thesis.pdf?sequence=4&isAllowed=y>

Edwards, F. (2017b). 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>
[Crossref]

Elder, C. (2017). *Towards an ILTA policy on language assessment literacy: How to define the construct and for whom? The 39th language testing research colloquium*. Universidad de los Andes.

Estaji, M., & Ghiasvand, F. (2021). Assessment perceptions and practices in academic domain: The design and validation of an assessment identity questionnaire (TAIQ) for EFL teachers. *International Journal of Language Testing*, 11(1), 103–131. <https://files.eric.ed.gov/fulltext/EJ1291317.pdf>

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, 2. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061970742&doi=10.1186%2fs40468-019-0078-7&partnerID=40&md5=c82c741237f8e8cbcce44c74ef6c6985>
[Crossref]

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
[Crossref]

Fulcher, G. (2020). Operationalising language assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: From theory to practice* (pp. 8–28). Cambridge Scholars Publishing.

Fulmer, G. W., Tan, K. H. K., & Lee, I. C. H. (2019). Relationships among Singaporean secondary teachers' conceptions of assessment and school and policy contextual factors. *Assessment in Education: Principles, Policy & Practice*, 26(2), 166–183. <https://doi.org/10.1080/0969594X.2017.1336427>
[Crossref]

Gan, L., & Lam, R. (2020). Understanding university English instructors' assessment training needs in the Chinese context. *Language Testing in Asia*, 10(11), 1–18. <https://doi.org/10.1186/s40468-020-00109-y>
[Crossref]

Gan, L., & Lam, R. (2022). A review on language assessment literacy: Trends, foci and contributions. *Language Assessment Quarterly*, 19(5), 503–525. <https://doi.org/10.1080/15434303.2022.2128802>
[Crossref]

Gan, L., & Lam, R. (2024a). *Language assessment literacy development of a novice university English teacher in the Chinese context*. RELC Journal, online first publication. <https://doi.org/10.1177/00336882241231304>.

Gan, L., & Lam, R. (2024b). Unveiling classroom assessment literacy: Does teachers' self-directed development play out? *Education Sciences*, 14(9), 961. <https://doi.org/10.3390/>

[educsci14090961](#)

[Crossref]

Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195. <https://doi.org/10.15446/profile.v20n1.62089>

[Crossref]

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>

[Crossref]

Giraldo, F., & Murcia, D. (2018). Language assessment literacy for pre-service teachers: Course expectations from different stakeholders. *Gist-Education and Learning Research Journal*, 16, 56–75. <https://doi.org/10.26817/16925777.425>

[Crossref]

Grainger, P. R., & Adie, L. (2014). How do preservice 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>

[Crossref]

[Crossref]

Gu, Y. Q. (2021). An argument-based framework for validating formative assessment in the classroom. *Frontiers in Education*, 6, 60599. <https://doi.org/10.3389/feduc.2021.605999>

[Crossref]

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 Nature. https://doi.org/10.1007/978-3-030-35081-9_4

[Crossref]

Harding, L. & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. T. J. Banerjee (Ed.), *Handbook of second language assessment* (pp. 413–427). DE Gruyter Mouton.

[Crossref]

Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs survey report: Part one-general findings*. <http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf>

Ho, E. C., & Yan, X. (2021). Using community of practice to characterize collaborative essay prompt writing and its role in developing language assessment literacy for pre-service language teachers. *System*, 101. <https://doi.org/10.1016/j.system.2021.102569>

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>

[Crossref]

[Crossref]

Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing*, 30(3), 301–307. <https://doi.org/10.1177/0265532213480126>

[Crossref]

Inbar-Lourie, O. (2017). Language assessment literacy. In E. S. et al. (Eds.), *Language testing and assessment, encyclopaedia of language and education* (pp. 257–270). Springer International Publishing AG.

[Crossref]

Janssen, G. (2022). Local placement test retrofit and building language assessment literacy with teacher stakeholders: A case study from Colombia. *Language Testing*, 39(3), 377–400. <https://doi.org/10.1177/02655322221076153>

[Crossref]

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>

[Crossref]

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>

[Crossref]

Kaiser, G., & Willander, T. (2005). Development of mathematical literacy: Results of an empirical study. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 24(2–3), 48–60.

[Crossref]

Koh, K., Burke, L., Luke, A., Gong, W. G., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 22(3), 264–288. <https://doi.org/10.1177/1362168816684366>

[Crossref]

Koh, K., DePass, C., & Steel, S. (2019). *Developing teachers assessment literacy: A tapestry of ideas and inquiries*. Brill Sense.

[Crossref]

Kremmel, B., Eberharter, K., & Harding, L. (2017). *Putting the 'language' into language assessment literacy. The 39th language testing research colloquium*. Universidad de los Andes.

Kremmel, B., Eberharter, K., Holzknecht, F., & Konrad, E. (2018). Fostering language assessment literacy through teacher involvement in high—Stakes test development. In D. Xerri & P. V. Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 173–194). Springer International Publishing AG. <https://doi.org/10.1007/978-3-319-77177-9>

[Crossref]

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>

[Crossref]

Kunnan, A. J., Mathew, R., & Zhang, D. (2021). Language assessment literacy among teachers in Goa: Analyses of a survey questionnaire and teacher-made tests. *The Journal of AsiaTEFL*, 18(3), 904–

922.

[[Crossref](#)]

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>
[[Crossref](#)]

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>
[[Crossref](#)]

Lan, C. S., & Fan, S. Y. (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>
[[Crossref](#)]

Lee, J. (2019). A training project to develop teachers' assessment literacy. In E. White & T. Delaney (Eds.), *Handbook of research on assessment literacy and teacher-made testing in the language classroom* (pp. 58–80). IGI Global.
[[Crossref](#)]

Leonardsen, J. K., Utvær, B. K. S., & Fjørtoft, H. (2022). *The five faces of an assessor: Conceptualizing the enactment of teacher assessment identity in vocational education and training*. Educational Assessment, Advanced online publication. <https://doi.org/10.1080/10627197.2022.2106967>

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>
[[Crossref](#)]

Looney, A., Cumming, J., Keij, F. V. D., & 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>
[[Crossref](#)]

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344.
[[Crossref](#)]

Malone, M. E. (2017). Training in language assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment, encyclopaedia of language and education* (pp. 225–239). Springer International Publishing AG.
[[Crossref](#)]

Mansouri, B., Molana, K., & Nazari, M. (2021). The interconnection between second language teachers' language assessment literacy and professional agency: The mediating role of institutional policies. *System*, 103. <https://doi.org/10.1016/j.system.2021.102674>

McDowell, L., & Harman, K. (2008). *'The designer', 'the teacher' and 'the assessor': Changing academic identities the SRHE annual conference: Valuing higher education*. Liverpool.

Moqaddam, Q. J., Khodabakhshzadeh, H., Motallebzadeh, K., & Khajavy, G. H. (2021). Measuring EFL teachers' assessment identity: Development and validation of a questionnaire. *Journal of Language and Translation*, 11(1), 29–46. <https://journals.iau.ir/article/679444-6270eb6f508688975ac0f2bec03a2081.pdf>

Odo, D. M. (2016). An investigation of the development of pre-service teacher assessment literacy through individualized tutoring and peer debriefing. *Journal of Inquiry & Action in Education*, 7(2), 31–61. <https://files.eric.ed.gov/fulltext/EJ1133529.pdf>

O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363–380. <https://doi.org/10.1177/0265532213480336> [Crossref]

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> [Crossref]

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249–258. <https://doi.org/10.1191/0265532204lt283ed> [Crossref]

Sayyadi, A. (2022). In-service university-level EFL instructors' language assessment literacy and training needs. *Profile: Issues in Teachers' Professional Development*, 24(1), 77–95. <https://doi.org/10.15446/profile.v24n1.93676> [Crossref]

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> [Crossref]

Sheehan, S., & Munro, S. (2019). *Classroom assessment: The development of teachers' cognition*. <https://www.teachingenglish.org.uk/sites/teacheng/files/J175%20ELT%20Research%20Papers%20-%20teacher%20cognition%20and%20assessment%20A4%20FINAL%20v2%20Web.pdf>

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534–539.

Su, X. L., Yin, H. B., & Lee, I. (2024). Let a hundred flowers bloom: Towards a coexistence of paradigms in language assessment literacy. *Linguistics and Education*. <https://doi.org/10.1016/j.linged.2024.101346>

Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9, Article 1. <https://doi.org/10.1186/s40468-019-0077-8>

Sun, H., & Zhang, J. (2022). Assessment literacy of college EFL teachers in China: Status quo and mediating factors. *Studies in Educational Evaluation*, 74. <https://doi.org/10.1016/j.stueduc.2022.101157>

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36.

<https://doi.org/10.1017/s0267190509090035>
[Crossref]

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>
[Crossref]

Tian, W., Louw, S., & Khan, M. K. (2021). Covid-19 as a critical incident: Reflection on language assessment literacy and the need for radical changes. *System*, 103. <https://doi.org/10.1016/j.system.2021.102682>

Tsagari, D. (2021). Language assessment literacy: Concepts, challenges, and prospects. In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for improved student learning* (pp. 13–32). Routledge.

Vogt, K., Bøhn, H., & Tsagari, D. (2024). Language assessment literacy. *Language Teaching*, 57(3), 325–340. <https://doi.org/10.1017/S0261444824000090>
[Crossref]

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>
[Crossref]

Vogt, K., Tsagari, D., & Spanoudis, G. (2020). What do teachers think they want? A comparative study of in-service language teachers' beliefs on LAL training needs. *Language Assessment Quarterly*, 17(4), 410–433. <https://doi.org/10.1080/15434303.2020.1781128>
[Crossref]

Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly*, 7(4), 317–342. <https://doi.org/10.1080/15434303.2010.516042>

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *Australian Educational Researcher*, 40(2), 241–256. <https://doi.org/10.1007/s13384-013-0089-9>
[Crossref]

Wyatt-Smith, C., & Adie, L. (2021). The development of students' evaluative expertise: Enabling conditions for integrating criteria into pedagogic practice. *Journal of Curriculum Studies*, 53(4), 399–419. <https://doi.org/10.1080/00220272.2019.1624831>
[Crossref]

Xie, Q., & Tan, S. Y. (2019). Preparing primary English teachers in Hong Kong: Focusing on language assessment literacy. *Journal of AsiaTEFL*, 16(2), 653–673. <https://doi.org/10.18823/asiatefl.2019.16.2.14.653>
[Crossref]

Xu, Y. T. (2017). Assessment literacy of university English instructors in China: A mixed-methods study [Doctoral dissertation, The University of Hong Kong]. The HKU Scholars Hub.

- Xu, H. (2017a). Exploring novice EFL teachers' classroom assessment literacy development: A three-year longitudinal study. *The Asia-Pacific Education Researcher*, 26(3), 219–226. <https://doi.org/10.1007/s40299-017-0342-5>
[Crossref]
- Xu, Y. T. (2017b). *Assessment literacy of university English instructors in China: A mixed-methods study* [Doctoral dissertation, The University of Hong Kong]. The HKU Scholars Hub.
- Xu, Y. T. (2019a). English language teacher assessment literacy in practice. In X. S. Gao (Ed.), *Second handbook of English language teaching* (pp. 517–539). Springer Nature. https://doi.org/10.1007/978-3-030-02899-2_30
[Crossref]
- Xu, Y. T. (2019b). How teacher conception of assessment mediate assessment literacy: A case study of university English teacher in China. In S. Papageorgiou & K. M. Bailey (Eds.), *Global perspectives on language assessment* (pp. 197–211). Routledge.
[Crossref]
- Xu, Y. T., & Brown, G. T. L. (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>
[Crossref]
- Xu, Y. T., & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133–158. https://www.altanz.org/uploads/5/9/0/8/5908292/9.si7xubrown_final_formatted_proofed.pdf
- Xu, Y. T., & Liu, Y. C. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 493–513. <https://doi.org/10.1002/j.1545-7249.2009.tb00246.x>
[Crossref]
- Yan, X., & Fan, J. (2021). “Am I qualified to be a language tester?”: Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246. <https://doi.org/10.1177/0265532220929924>
[Crossref]
- 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>
[Crossref]
- Yastibaş, A. E., & Takkaç, M. (2018). Understanding language assessment literacy: Developing language assessments. *Journal of Language and Linguistic Studies*, 14(1), 178–193. <https://files.eric.ed.gov/fulltext/EJ1175787.pdf>
- Zhang, L., Zheng, Y., & Yang, D. (2024). Exploring Chinese university English teachers' language assessment literacy: A mixed-method study. *Asia Pacific Journal of Education*, 1–19. <https://doi.org/10.1080/02188791.2024.2354684>

3. Research Methodology

Ling Gan¹✉ and Ricky Lam²✉

- (1) School of Languages and Communication, Beijing Technology and Business University, Beijing, China
- (2) Department of Education and Psychology, Hong Kong Baptist University, Hong Kong, China

✉ Ling Gan

Email: linggan@btbu.edu.cn

✉ Ricky Lam (Corresponding author)

Email: rickylam@hkbu.edu.hk

3.1 Overview

The purpose of this study is to present the process of university English instructors' LAL development in their teaching contexts. This chapter first delineates the research context where English language assessment at the tertiary level in China and assessment requirements on university English instructors are introduced. It then justifies the selection of an ethnographic case study, research sites and participants. It describes the process of data collection with multiple instruments including narrative frames, interviews, field observations, post-observation conversations and documents. This chapter also presents the step-by-step data analysis within datasets and across cases in detail. Then trustworthiness, ethical considerations and reflections on the role of researchers are discussed in this chapter.

3.2 Research Context

In China, College English is a mandatory course for non-English majors at the tertiary level and has undergone significant reform over the past few decades. The primary aim of the College English curriculum reform is to enhance the quality of instruction amid a rapidly expanding university enrolment. Assessment, a critical component for ensuring quality, has also been restructured in tandem. Prior to the College English Curriculum Reform initiated by the Ministry of Education in 2007, College English instruction focused on preparing students for the standardised College English Test (CET), which has two levels: Bands 4 and 6 (henceforth CET4/6), implemented since 1987. However, the detrimental effects of high-stakes testing prompted the Ministry to overhaul both the teaching and assessment frameworks. Since 2007, College English Curriculum Reform has integrated formative assessment alongside summative assessment in College English instruction. This shift aims to diminish the emphasis on standardised testing and to re-establish the roles of teachers and students in the processes of teaching, learning, and assessment.

Since the introduction of the College English Curriculum Reform, Chinese universities have established localised and contextualised requirements for assessment in College English instruction to promote formative assessment among English instructors. The latest Guidelines on College English Teaching published by the Ministry of Education in 2020 further emphasise the need for university instructors to assess students' language performance, provide constructive feedback, and actively involve students in the assessment process using a variety of methods. Through effective assessment practices, College English seeks to inspire student learning and enhance the overall curriculum.

However, this assessment reform has raised concerns among scholars regarding the assessment expertise of university English instructors. Research indicates that while assessment practices in College English classes are nominally formative, they often function summatively in practice (Liu & Xu, [2017](#)). Additionally, there is limited understanding of whether university English instructors possess the necessary qualifications to effectively integrate formative assessment and summative assessment in their classrooms (Wang, [2017](#)).

In response to these challenges, the Guidelines on College English Teaching underscores the importance of developing LAL among university English instructors. It advocates for targeted training programmes that focus on enhancing instructors' knowledge and skills in language assessment. Such training would enable instructors to effectively implement diverse assessment methods for FA, employ various analytical techniques to report assessment results, and, crucially, manage the interplay between teaching and assessment more effectively.

Admittedly, university English instructors in China have a low LAL level and a dire need for assessment training (Sun & Zhang, [2022](#)). They have fewer assessment training opportunities than their primary or secondary level counterparts who are provided with systematic assessment training programmes (e.g., courses, teaching practicum, certificates) (Yan & Fan, [2021](#)). Although they receive short-term orientation training before starting their university teaching career, the training is not tailored for English instructors but for all new instructors. This two-or-three-week orientation session often covers higher education laws and regulations and educational psychology, while assessment has been minimally mentioned. English instructors without prior assessment learning or training experience cannot rely on the orientation training to obtain assessment knowledge, skills, or practices for College English teaching. Instead, they need to learn to become assessment-literate on their job.

Most training opportunities for in-service English instructors at Chinese universities are primarily offered by several reputable publishers (Yan & Fan, [2021](#)). These publishers often invite experts in language testing and assessment to conduct lectures and workshops. Typically, these training programmes last from three to seven days and are scheduled during winter or summer breaks. Due to the high costs associated with these programmes, candidates are selectively chosen by their departments based on various criteria. More often than not, instructors teaching non-English majors tend to receive more assessment training than those teaching English majors, reflecting the market demand for language education (Yan & Fan, [2021](#)).

University English instructors in China can be categorised into two groups. One group operates within the College English department,

focusing on language proficiency courses for non-English majors, while the other group is situated in the English department, where they teach a range of content courses, including language proficiency courses for English majors (Yan & Fan, [2021](#)). Given their differing student populations, these two groups are expected to fulfil varying job responsibilities and performance expectations. Generally, College English instructors bear a heavier teaching workload and engage less in academic research, with their institutional efforts primarily linked to teaching and learning (Zhang & Li, [2020](#)). In contrast, English department instructors typically have fewer teaching responsibilities and devote more time to academic research.

In this study, university English instructors focus on those in College English departments, whose LAL development is crucial for enhancing English learning among a significant population of adult learners in China.

3.3 An Ethnographic Case Study

The current study took place within an interpretative research paradigm, as interpretive research can gain insiders' perspectives and focus on understanding the sociocultural processes in natural settings (Denzin & Lincoln, [2017](#)). This paradigm can generate insights into the process of university English instructors' LAL development in their assessment practices in the Chinese context. As mentioned in Chap. 2, LAL is a context-situated practice and LAL researchers should take a close look at language teachers' LAL reality in their teaching settings (Giraldo, [2020](#); Hill, [2017](#)). Given this precondition for LAL research, qualitative approaches, strongly associated with the interpretive research paradigm, are considered appropriate to explore university English instructors' LAL development, an important aspect of teachers' professional development in language assessment (Harding & Kremmel, [2016](#)). This paradigm is aligned with the post-positivistic view that LAL development is a dynamic situated process (Inbar-Lourie, [2017](#); Koh et al., [2019](#)). It also echoes a long call for an interpretative qualitative approach to closely examine teachers' LAL development from their natural teaching contexts (Giraldo, [2020](#); Mosquera & Macias, [2015](#)).

Ethnography is a qualitative approach used to describe cultural patterns and structural regularities of a social group under investigation by exploring the shared feelings, perceptions, and experiences of a group of people within their culture (Creswell, [2007](#); Fusch et al., [2017](#); Green & Bloome, [1997](#)). Two key concepts in an ethnography are contexts and patterns within the frame of culture (Angrosino, [2008](#); Merriam & Tisdell, [2016](#)). Ethnography emphasises collecting data from a real-world context rather than from experimental conditions. To understand the culture of a group from the real-world context, ethnographers need to enmesh themselves in the day-to-day life of the entire group through extended observations and in-depth interviews. This extended engagement helps to generate data with thick descriptions, a crucial aspect rooted at the heart of an ethnography (Merriam & Tisdell, [2016](#)).

The case study approach evolved out of an ethnographic research design (Fusch et al., [2017](#)). As an interpretative research method, a case study is defined as “an empirical method that investigates a contemporary phenomenon (the “case”) in-depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident” (Yin, [2018](#), p. 15). The case study underscores the context where a bounded system is investigated in-depth. The focus of the investigation is on the case situated in the context (Walters, [2007](#)). Moreover, a case study can be used to document changes in a complex phenomenon over time (van Lier, [2005](#)). Both the case study and the ethnographic approach aim to identify and understand the unique characteristics of a phenomenon from the real-world context through thorough examination. They have been interrelated to each other in this study to address the research questions.

In the current study, an ethnographic case approach, interpretatively qualitative in nature (Patton, [2015](#)), was used to gain an in-depth and holistic understanding of the process of university English instructors’ LAL development. This methodology is featured with a nuanced portrayal of multiple cases based on in-depth data collection over an extended timeframe. As a “case study” is a choice of an object to be examined, which can be an individual, an event, or an institute or a cultural group (Stake, [1995](#); Walters, [2007](#)), the cases in

the current study are four English instructors from different universities in China.

A specific approach to ethnography was adopted in our study. According to Green and Bloome ([1997](#)), there are three approaches to ethnography: (1) doing ethnography, (2) adopting an ethnographic perspective, and (3) using ethnographic tools. The third approach was employed. That is, we used ethnographic methods and techniques that are usually associated with fieldwork but may or may not be guided by cultural theories or questions about the social life of a group of people. Specifically, typical fieldwork methods of field observations, interviews, and documents were used for data collection.

Instead of being a full-fledged ethnography, the present ethnographic case study aims to document the LAL development process of four university English instructors over three semesters. It attempts to keep a thick description of individual cases and identify significant patterns of individual instructors' LAL development in the Chinese context. Although the participants in the study share features in their LAL development, the present study does not intend to generalise this development. Instead, it is to explore and understand the particularity and complexity of cases within specific contexts (Stake, [2010](#)) by means of multiple data sources over an extended period.

3.4 Selection of Participants

Purposeful sampling was adopted in this study to select appropriate participants (Duff, [2008](#); Merriam & Tisdell, [2016](#)). This sampling strategy is used for the selection of information-rich cases, from which "one can learn a great deal about issues of central importance to the purpose of the inquiry" (Patton, [2015](#), p. 53). That is, the selected cases should most likely address research questions (Yin, [2018](#)), by making the selection on "conceptual grounds, not on representative grounds" (Miles et al., [2020](#), p. 51). Following this selection criterion, we chose those university English instructors who were information-rich conceptually but were contrasting with each other potentially. Participants were then selected initially based on the following two criteria:

- They are university instructors teaching English courses for non-English major undergraduates.
- They are willing to participate in the current study.

The first criterion could ensure the selected participants had been experiencing or were being involved in the cultural context of teaching College English as a foreign language, while the second one could guarantee the availability of steady information from the participants. It is assumed that the shared culture context can help the researcher identify the common patterns of English instructors' LAL development. It is also posited that the ongoing commitment to this study makes it possible to obtain more valid and trustworthy data when participants take this piece of research as an opportunity to reflect on their own professional knowledge, skills, and practices.

Informed by these two criteria, we sent invitation letters to English instructors at universities in Wuhan, the capital of Hubei Province, China, after obtaining the Research Ethics Committee approval from Hong Kong Baptist University. Our selection of Wuhan was motivated by two key factors. First, the first author had established academic connections in the city from previous work, enabling access to the research sites through relationships with deans and heads of foreign language departments. Second, as the third largest city in China, Wuhan is home to a diverse array of universities, offering opportunities to select participants with rich, varied educational backgrounds. This diversity is essential, as different assessment practices across institutions may influence the development processes of LAL among participants.

From the initial outreach, we identified seven university English instructors willing to participate in the study. We then applied a maximum variation strategy (Duff, [2008](#); Miles et al., [2020](#)) to refine the selection of participants. This approach focuses on identifying cases with diverse characteristics, allowing for the discovery of common patterns amidst the variety (Patton, [2015](#)). The diversity among participants also enhances reader generalisability (Merriam & Tisdell, [2016](#)). As outlined in the literature review, factors such as teachers' prior assessment experiences (Crusan et al., [2016](#); Yan et al., [2018](#)), conceptions of assessment (Y. Xu, [2017](#)), research interests (Sultana, [2019](#)), gender, and teaching experience (Jiang, [2020](#)) are crucial for

understanding their LAL development. The criteria for further selection are presented as follows:

- different genders are considered.
- participants have varied educational backgrounds.
- participants have various lengths of teaching experience.
- participants have different assessment training/learning experiences.
- participants have diverse research interests or areas.
- participants come from different types of universities.

In ethnographic case studies, engaging three to four participants is generally considered ideal, as this number facilitates in-depth analysis (He, [2013](#); Mu, [2020](#)). A smaller, diverse group provides richer insights while effectively managing the time and effort required for data collection and analysis. From an initial pool of seven candidates, we selected four focal cases: Emma, Rachel, Ryan, and Tim (pseudonyms; see Table [3.1](#)). Emma was chosen for her strong interest in language assessment, demonstrated during her participation in a prior research project on English instructors' assessment training (Gan & Lam, [2020](#)). Her passion equipped her to offer valuable insights into her LAL development. Additionally, her experience in College English teaching and her recognition as the first runner-up in a provincial teaching contest¹ contributed to her unique perspectives on language assessment.

Table 3.1 Biographic Information of the Participants

Instructors	Emma	Rachel	Ryan	Tim
Gender	Female	Female	Male	Male
Working university	AU, key university ² /agricultural university	WU, non-key university/comprehensive university	NU, key university/normal university	LU, key university/economics & law university
Location of university	Wuhan	Wuhan	Wuhan	Wuhan
Years of teaching experience	12 years	Two years	Eight years	Four years

Instructors	Emma	Rachel	Ryan	Tim
Educational background	Master's degree in second language acquisition	Master's degree in interpretation	Master's degree in educational research, pursuing a doctor of education degree	Master's degree in corpus linguistics
Research areas	English language teaching	Translation	Teacher development	Corpus linguistics

Rachel, in contrast, was selected through a recommendation from a friend who heads the College English Section at her university. Eager to foster professional growth among novice instructors, my friend encouraged Rachel's participation in various research projects. As a new instructor, Rachel was a quick learner, hoping to enhance her skills in language assessment through this study. Similarly, Ryan and Tim were chosen based on recommendations by colleagues who received our project invitation. Ryan, a competitive instructor pursuing a Doctor of Education degree at a Sino-foreign university, engaged in College English teaching contests at local and national levels, experiences that likely influenced his LAL construction. Tim, who was my colleague at a previous university, was a popular novice instructor known for his exceptional teaching. His enthusiasm for research methods in applied linguistics promised a diverse perspective on LAL development.

In this study, both novice and experienced instructors were included to compare their LAL development. Novice teachers have less than five years of experience (Kim & Roth, [2011](#)), while experienced teachers have six or more. Rachel, at the beginning of her teaching journey, represents a novice, while Tim, nearing the end of his novice phase, offers a slightly different perspective. Meanwhile, Emma and Ryan are experienced instructors with 12 and 8 years of service, respectively. The varied levels of experience allow for a nuanced observation of LAL development trajectories among novice and experienced instructors.

The participants' universities also play a vital role in the study. Key universities typically attract more funding from government sources and recruit higher-quality educators. These institutions often impose higher standards on their instructors, particularly regarding assessment practices, and have the capability to admit students with superior English proficiency, which can enhance teaching outcomes. In

contrast, non-key universities tend to employ fewer high-quality teachers and enrol students with lower English proficiency. The differences between these types of universities provide a context for exploring the challenges and opportunities that participants experience while developing their LAL and assessment identities in their everyday practices.

3.5 Data Collection

As outlined in Sect. [3.3](#), this study spanned one and a half academic years, conducting longitudinal investigations in three stages from October 2020 to January 2022. Each stage was scheduled according to the teaching timelines of the four instructors. Data collection included narrative frames, interviews with instructors and students, field observations, post-observation discussions, and relevant documents. This phase concluded in January 2022, as access to the observed settings became increasingly challenging due to stricter COVID-19 policies. Despite the official end of data collection, we maintained regular contacts with the participants and continued to receive updates on their assessment practices.

3.5.1 Narrative Frames

Narrative frames are tools designed to capture participants' professional experiences in a structured story format (Barkhuizen & Wette, [2008](#)). A frame works as a scaffolding for participants to share their language assessment stories using the structures designed by the researcher. This structured approach allows participants to express their experiences coherently in their preferred language. In this study, all participants used standard written Chinese enabling them to communicate their thoughts and experiences thoroughly.

The study involved four narrative frames. The first frame (see Appendix 1) was distributed at the study's outset to gather demographic information, such as educational backgrounds, professional experiences, and research interests. It also collected autobiographical accounts related to language learning and teaching experiences. Participants described their experiences as students undergoing assessments, assessment practices in their classes, and

their views on the interplay between language teaching, learning, and assessment. They also discussed their roles as assessors and the assessment policies they encountered.

The remaining three narrative frames (see Appendix 2) were completed at the end of each semester. These frames focused on specific events related to the development of participants' LAL, highlighting challenges, opportunities, and strategies in assessment practice. Participants reflected on their assessment methods each semester and further explored their identity as assessors. Each subsequent frame built on information from the previous ones, and initial instructions were provided to ensure a smooth narrative flow.

3.5.2 Interviews

Narrative frames, though informative, may constrain the participants to develop their own stories due to the prescribed prompts in the frames (Barkhuizen & Wette, [2008](#)). To fill this void, semi-structured interviews were implemented to gather detailed insights into participants' LAL development. Interviews aim to explore individuals' lived experiences and the meanings they construct from their stories (Seidman, [2019](#)), which allow researchers to probe into participants' feelings and perceptions, thus enhancing mutual understanding (Miller & Glassner, [2004](#)).

Over three semesters, six semi-structured interviews were conducted with each instructor, consisting of two interviews per semester. Each interview, held either mid-semester or at the end, focused on exploring university English instructors' beliefs and perceptions regarding language teaching, learning, assessment, their assessor identity, self-reported LAL, and the rationale behind their assessment practices. The interviews also provided insights into critical incidents, challenges, opportunities encountered, and factors influencing their LAL development.

The first interview in October 2020 set a baseline, focusing on participants' previous assessment experiences and perceptions of being an assessor, forming a foundation to observe their LAL evolution. The second and fourth interviews, at the semester's end, examined participants' motivations for observed assessment practices, challenges faced, opportunities for LAL development, and semester-specific

assessor perceptions. The third and fifth interviews, conducted mid-semester, centred on assessment schedules, plans for LAL development, and assessor identity.

The sixth, also the final interview, encouraged reflection on what participants gained or missed in LAL development, challenges faced, opportunities embraced, and their evolving assessment identity or any salient assessment episodes they have not mentioned previously. Each interview was interconnected, with questions tailored to individual responses from preceding interviews, aside from general guiding queries (see Appendix 3). These six interviews were connected, as subsequent interview questions were based on participants' responses to previous questions. Some particular quotes and uncovered questions in previous interviews were addressed again in subsequent interviews. Consequently, interview questions varied from individual to individual, apart from common guiding questions in the protocol (see Appendix 3).

Additionally, a focus group interview with students was conducted after each round of field observations at the end of each semester, aiming to understand instructors' LAL development from students' perspectives. Focus group interviews, ideal for collecting collective insights, encouraged students to share and exchange views on their instructors' assessment practices, providing alternative perspectives (Chadwick et al., [2008](#); Gill & Baillie, [2018](#)). This method also saved time and generated richer data than individual interviews.

Each focus group consisted of six or seven randomly selected students based on a list provided by instructors who sent invitations. During those interviews, as a moderator, the first author facilitated open discussions and adhered to the research agendas, while maintaining an observer role to minimise her influence on discussions.

Both instructor and student interviews were conducted in standard spoken Chinese for clarity and ease of expression. Individual interviews lasted from 50 to 100 minutes, while group interviews were at least 80 minutes, with audio and video recordings respectively, transcribed immediately afterwards.

3.5.3 Field Observations and Post-Observation Conversations

To triangulate data from narrative frames and interviews, four rounds of field observations were conducted to examine how instructors developed and enhanced their LAL in real work contexts. Field observation is crucial in ethnographic studies, as it requires researchers to immerse themselves in the environment and use their senses to understand cultural practices and interactions (Angrosino, [2008](#); Holloway et al., [2010](#)). This method helps gain in-depth insights into how practices are formed through sociocultural interactions.

Observations occurred over three semesters, including classes, student tutorials and English-speaking tests, guided by a structured observation protocol (see Appendix 4). These observations captured the instructors' in-class assessment practices and how they built their LAL knowledge and assessment identity. By observing tutorials, we noted how instructors disseminated LAL and shaped their identities through various interactions. Prolonged engagement in these environments provided a nuanced understanding of LAL development in authentic teaching settings.

Throughout the observations, we took the role of an observer-as-participant, maintaining a researcher-participant relationship but retaining limited interactions. This approach helped evade threats posed by a "complete participant" and avoid information loss caused by a "complete observer" (Creswell & Poth, [2018](#), p. 167). It also allowed for building rapport, enabling to capture comprehensive data (Patton, [2015](#)).

Classroom observations were videotaped with a camera placed at the back of the classroom, where we took field notes. These notes, along with transcribed observations, enriched the interpretation of observed data. Audio recordings were used for tutorials and speaking tests to avoid discomfort from our closer presence.

Following observations, we engaged in brief, informal conversations with participants during breaks or en route to their next classes. Such post-observation discussions fostered professional reflection by emphasising collaborative and oral exchanges (Kim & Silver, [2016](#); Mann & Walsh, [2013](#)). Each conversation, lasting about 10 minutes, aimed to clarify or reflect on observed practices. With participants' consent, these conversations were audio recorded in standard spoken Chinese to promote clear and easy communication.

3.5.4 Documents

Documents constituted another important source of data in this qualitative research, offering several benefits. They are easily accessible and physically stable, as their content remains unaffected by researchers' presence (Merriam & Tisdell, [2016](#)). Additionally, documents can provide information unavailable from other sources, helping to confirm or challenge existing data (Hammersley & Atkinson, [2007](#)). They can also offer specific details from which researchers make inferences (Yin, [2018](#)).

In this study, documents were categorised into three types. The first category included national teaching guides, course syllabi, and curriculum requirements, which provided insights into assessment policies and teacher development. The second category comprised assessment-related documents such as assessment rubrics, criteria, grading sheets, plans, student assignments, and presentation slides. The third category involved communications about assessment between teachers, students, or colleagues, via emails or instant messaging apps like WeChat and QQ (popular social media apps used in China). These documents were collected at various junctures during observations at the research sites.

3.5.5 Summary of Data Collection

To recap, data were collected over a period of one and a half academic years and sourced from various methods. The collected data are summarised in Table [3.2](#).

Table 3.2 A Summary of Collected Data

	Narrative	Interview	Field observation & post-observation conversation	Document
Emma	Four narrative accounts	Individual interviews (six audios) students' focus group interviews (three videos)	Classroom lessons (11 videos) speaking tests (two videos) conversations (11 audios)	Course syllabus, lesson plans, assessment requirement documents, speaking assessment criteria sheets, students' peer-assessment sheets, grading sheets, students' portfolios, and QQ messages

	Narrative	Interview	Field observation & post-observation conversation	Document
Rachel	Four narrative accounts	Individual interviews (six audios) students' focus group interviews (three videos)	Classroom lessons (12 videos) speaking tests (two videos) conversations (12 audios)	Course syllabus, lesson plans, teaching guides, grading sheets, speaking grading criteria, students' quizzes, and dictations
Ryan	Four narrative accounts	Individual interviews (six audios) students' focus group interviews (three videos)	Classroom lessons (13 videos) tutorials with students (two videos) conversations (12 audios)	Course syllabus, lesson plans, assessment requirement documents, students' writing assignments, quizzes, grading sheets, emails, questionnaires, and reflective notes
Tim	Four narrative accounts	Individual interviews (six audios) students' focus group interviews (three videos)	Classroom lessons (12 videos) speaking tests (two videos) conversations (10 audios)	Course syllabus, lesson plans, students' writing assignments, and grading sheets

3.6 Data Analysis

Data analysis is a process of making meanings and representation of data to unfold participants' stories (Merriam & Tisdell, [2016](#)). In this study, analysis began concurrently with data collection and continued afterwards, focusing on understanding the data through repeated readings of transcripts from narrative frames, interviews, field notes, conversation scripts, and documentary artefacts. Thematic analysis was employed to organise, reduce, categorise and develop codes and themes within data, offering flexibility and theoretical freedom unlike other approaches (e.g., grounded theory) that aim to develop a theory (Braun & Clarke, [2006](#)).

This study used both inductive and deductive methods for analysis. Inductive analysis means deriving themes directly from raw data without relying on pre-existing theories or researchers' conceptions

(Thomas, [2006](#)). However, data analysis is usually driven by researchers' analytic perceptions or theoretical frameworks, i.e., practices of deductive analysis (Braun & Clarke, [2006](#)). This deductive method complemented the inductive analysis within datasets and across cases.

3.6.1 Analysis within Data Sets

Within-datasets analysis began with interview and conversation audios as well as observation videos. These data were uploaded into the website *iflyrec.com* for transcription. The interviews and conversations were transcribed verbatim, and those transcripts used for analysis were translated into English and were cross-checked by a colleague and the four instructors to ensure consistency with the original Chinese versions (Creswell & Poth, [2018](#)).

Observation data were analysed somewhat differently. First, we reviewed every ten-minute segment of observation recordings for assessment-related activities and transcribed these verbatim. Since classroom instruction was in English, further translation was unnecessary. Once verified by participants, transcripts of observations were uploaded to MAXQDA20.0.3 for coding.

Narrative frames were uploaded to MAXQDA20.0.3, and through repeated reading, stories for each case were developed and translated into English. These translations were cross-checked with a colleague and the participants. Additionally, course syllabi, assessment materials, student assignments, and grading sheets were scanned and uploaded for analysis.

During the transcription and transformation, data were read repeatedly to extract and track ideas related to LAL aspects for addressing research questions in coding software. All data underwent primary coding in MAXQDA20.0.3 using line-by-line, descriptive, and open coding methods (Tracy, [2020](#)).

Narrative frames and interviews were coded line-by-line to explore instructors' perceptions of their assessment experiences. This approach enabled close examination of the data, selection of specific quotes, and obtaining further comments from instructors in subsequent sessions (Charmaz, [2014](#)). Iterative analysis refined codes, documenting

instructors' LAL development over time and avoiding definition drift (Gibbs, [2018](#)).

Field observations and post-observation conversations were coded in a different approach. Descriptive codes were assigned to the transcripts to look for what we called “assessment scenes” (Xu, [2017](#)) (see Table [3.3](#)), from which university English instructors' LAL were constructed and enacted. These scenes mainly refer to classroom assessment practices, including formal (e.g., written tests, presentations, grading) and informal assessment (e.g., peer assessment/feedback, debriefing of assessment criteria). Field notes supplemented this process, aiding the identification and coding of assessment practices.

Table 3.3 Assessment Scenes from Classroom Observations

Assessment scenes	
Formal assessment	Presentations
	Quizzes
	Dictations
	Mid-term/final tests
	Teacher questioning
	Teacher feedback
	Grading
Informal assessment	Self- and peer assessment
	Peer feedback
	Group assessment
	Criteria introduction

Documentary texts were processed through open coding to generate “assessment realities”, such as requirements, activities, and communications. There were three procedures for identifying “assessment realities” from the documents. This involved analysing general documents, reviewing assessment-related materials, and examining communication messages.

In the secondary coding cycle, codes from the primary cycle were synthesised into interpretative themes (Saldaña, [2016](#)), with axial coding grouping related codes under broader categories (Tracy, [2020](#)). This process involved pattern identification and interpretation (Miles et al., [2020](#)).

Three sources were used for thematising in this study. Taylor's eight LAL dimensions (2013) served as a "start list" (Miles et al., [2020](#)). For example, the codes "English teaching and learning" and "language use" and the codes "using assessment criteria" and "diagnostic assessment" were categorised as "knowledge of language pedagogy and theory" and "knowledge of language assessment theory" respectively. These two subcategories were under the theme of "knowledge of language assessment". Research questions also structured theme development. Documents were used as the complementary source to generate themes during the analysis. For instance, themes like "preparing students for standardised tests" were developed from document analysis and used across various data sources. Constantly refining and regrouping codes was an ongoing practice throughout the analysis.

3.6.2 Cross-Cases Analysis

Cross-case analysis was conducted by comparing the data of the cases to "look for underlying similarities and constant associations, compare cases with different outcomes, and begin to form more general explanations" (Miles et al., [2020](#), p. 96). The within-case and cross-case analysis was carried out in a cycling fashion. Initially, data from each case were synthesised and interpreted. For example, in Emma's case, themes and subthemes were identified and refined, leading to tentative interpretations of her LAL development and assessment identity, as well as preliminary assumptions about influencing factors.

After analysing Emma's data, we proceeded to analyse the other cases: Rachel, Ryan, and Tim. Whenever a discrepant "negative case" (Saldaña, [2016](#); Tracy, [2020](#)) arose, we revised original assumptions and reformulated explanations to establish a common pattern and general explanation, which contained different situations and experiences of the four focal cases.

Throughout this process, we frequently revisited the data within each case, repeatedly reviewing the original data to explore meaning

connections across categories. Writing memos and reflections facilitated the creation of “within-case displays” (Miles et al., [2020](#)), helping to establish detailed profiles and refine our understanding of each case through a constant comparative method (Charmaz, [2014](#)).

3.7 Trustworthiness, Ethical Issues and Roles of the Researcher

Trustworthiness in qualitative studies ensures the truthfulness of study findings within specific contexts (Hays & Singh, [2012](#)). The current study employed three major strategies to ensure trustworthy data collection, analysis, and interpretations.

First, credibility was enhanced through data and viewpoint triangulation. Data were collected from multiple sources, including narrative frames, interviews, field observations, post-observation conversations, and documents. Viewpoints from the four English instructors and their students were compared across three semesters. In addition, the extended fieldwork allowed triangulation of the same participants’ perceptions at different stages of the study.

Second, member checking was employed for data collection and analysis. Tools like narrative frames, interview questions, and observation protocols were discussed with research peers to reduce bias and increase clarity. During data analysis, participants verified interview and observation transcripts and a colleague with a background in qualitative research checked coding and themes to ensure consistency. In addition, draft findings were reviewed by instructors to identify any biases.

Third, detailed context descriptions helped readers assess the transferability of findings. A well-organised case database supported easy data retrieval and audit trails (Yin, [2018](#)). We kept case databases well organised with case name/time/event for easy retrieval (e.g., Data collection/Stage3(202110–202201)/ Rachel/ interview/ Rachel-interview 5–20211104).

Ethical standards, crucial in qualitative research, were also upheld (Creswell & Poth, [2018](#); Hays & Singh, [2012](#)). We obtained ethical approval from the Research Ethics Committee of Hong Kong Baptist University. We informed participants of the study’s purpose,

confidentiality, and voluntary nature, with all signing their consent forms. During the first observation and focus group interview, we explained to the students the study's purpose and assured them of confidentiality. Initial classroom observations incurred students' curiosity, but over time, students became indifferent to the researchers' presence. For instance, even with the presence, students in Rachel's and Tim's classes did not conceal their boredom and non-cooperative attitudes.

Researchers' influence on the cases or the setting under study is a major concern in qualitative enquiry (Maxwell, [2013](#)). As a former university English instructor, we tried to minimise preconceptions affecting the data sets by maintaining openness and actively engaging with participants' perspectives (Patton, [2015](#)). Perceived academic authority as a researcher was managed by means of fostering reflective thinking rather than giving direct advice. This encouraged critical thinking, as demonstrated by Emma, who was guided to reflect on her assessment practices. To balance power dynamics, we invited instructors to comment on the data interpretations. This reciprocal interaction showed our respect towards instructors. Instructors such as Rachel and Emma also acknowledged a deeper interest in assessment inspired by our field observations. During observations, we provided resource recommendations upon request. This built trust and openness, especially with Rachel, who became more communicative and informative in later interviews.

3.8 Summary

In this chapter, we provided an overview of the methodological issues for this study. We presented the research context first and then delineated the research design of an ethnographic case study to document instructors' LAL development. We also specified the procedures of data collection and analysis, which lay a solid foundation for generating useful information from multiple sources in a trustworthy and ethical manner. In addition, three trustworthiness strategies were adopted to ensure the quality of the study, followed by the discussion of ethical considerations and reflections on the researchers' dual roles in the present study. The following four chapters

present the stories of the four university English instructors in their LAL development respectively.

References

- Angrosino, M. (2008). *Doing ethnographic and observational research*. Sage.
- Barkhuizen, G., & Wette, R. (2008). Narrative frames for investigating the experiences of language teachers. *System*, 36(3), 372–387. [\[Crossref\]](#)
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa> [\[Crossref\]](#)
- Chadwick, B., Gill, P., Stewart, K., & Treasure, E. (2008). Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal*, 204(6), 291–295. <https://doi.org/10.1038/bdj.2008.192> [\[Crossref\]](#)
- Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.).
- Creswell, J. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.).
- Creswell, J., & Poth, C. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.).
- 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> [\[Crossref\]](#)
- Denzin, N. K., & Lincoln, Y. S. (2017). *The SAGE handbook of qualitative research* (5th ed.).
- Duff, P. A. (2008). *Case study research in applied linguistics*. Lawrence Erlbaum Associates.
- Fusch, P. I., Fusch, G. E., & Ness, L. R. (2017). How to conduct a mini-ethnographic case study: A guide for novice researchers. *The Qualitative Report*, 22(3), 923–941. <https://doi.org/10.46743/2160-3715/2017.2580> [\[Crossref\]](#)
- Gan, L., & Lam, R. (2020). Understanding university English instructors' assessment training needs in the Chinese context. *Language Testing in Asia*, 10(11), 1–18. <https://doi.org/10.1186/s40468-020-00109-y> [\[Crossref\]](#)
- Gibbs, G. R. (2018). *Analyzing qualitative data* (2nd ed.). [\[Crossref\]](#)

Gill, P., & Baillie, J. (2018). Interviews and focus groups in qualitative research: An update for the digital age. *British Dental Journal*, 225(7), 668–672. <https://doi.org/10.1038/sj.bdj.2018.815> [Crossref]

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> [Crossref]

Green, J., & Bloome, D. (1997). Ethnography and ethnographers of and in education: A situated perspective. In J. Flood, S. B. Heath, & D. Lapp (Eds.), *Handbook of research on teaching literacy through the communicative and visual arts* (pp. 181–202). Macmillan Publishers.

Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice* (3rd ed.). Routledge.

Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 413–427). DeGruyter Mouton. [Crossref]

Hays, D. G., & Singh, A. A. (2012). Qualitative inquiry in clinical and educational settings..

He, P. C. (2013). *Learning to teach in school-university partnership: Tensions, agency and identity* [Doctoral dissertation, The University of Hong Kong]. The HKU Scholars Hub.

Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment*, 6(1), 1–17.

Holloway, I., Brown, L., & Shipway, R. (2010). Meaning not measurement: Using ethnography to bring a deeper understanding to the participant experience of festivals and events. *International Journal of Event and Festival Management*, 1(1), 74–85. <https://doi.org/10.1108/17852951011029315> [Crossref]

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 257–270). Springer International Publishing AG. [Crossref]

Jiang, J. (2020). Gaoxiao yingyu jiaoshi ceping suyang xianzhuang jiqi yingxiang yinsu yanjiu [Study of college English teachers' assessment literacy and its influencing factors]. *Foreign Language World*, 6, 18–26.

Kim, K., & Roth, G. (2011). Novice teachers and their acquisition of work-related information. *Current Issues in Education*, 14(1), 1–28. <https://cie.asu.edu/ojs/index.php/cieatasu/article/view/268/152>

Kim, Y., & Silver, R. E. (2016). Provoking reflective thinking in post observation conversations. *Journal of Teacher Education*, 67(3), 203–219. <https://doi.org/10.1177/0022487116637120> [Crossref]

Koh, K., DePass, C., & Steel, S. (2019). *Developing teachers assessment literacy: A tapestry of ideas and inquiries*. Brill Sense.

[[Crossref](#)]

Liu, J., & Xu, Y. (2017). Assessment for learning in English language classrooms in China: Contexts, problems and solutions. In H. Reinders, D. Nunan, & B. Zou (Eds.), *Innovation in language learning and teaching: The case of China* (pp. 17–38). Palgrave Macmillan.

[[Crossref](#)]

Mann, S., & Walsh, S. (2013). RP or 'RIP': A critical perspective on reflective practice. *Applied Linguistics Review*, 4(2), 291–315. <https://doi.org/10.1515/applirev-2013-0013>

[[Crossref](#)]

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). SAGE.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Miles, M., Michael, M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.).

Miller, J., & Glassner, B. (2004). The inside and the outside: Finding realities in interviews. In D. Silverman (Ed.), *Qualitative research* (2nd ed., pp. 125–139). Sage.

Mosquera, L. H., & Macias, D. F. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302–312. <https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09>

[[Crossref](#)]

Mu, C. J. (2020). *Understanding Chinese multilingual scholars experiences of writing and publishing in English: A social-cognitive perspective*. Palgrave Macmillan.

[[Crossref](#)]

Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.).

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). SAGE.

Seidman, I. (2019). *Interviewing as qualitative research: A guide for researchers in education and social sciences* (5th ed.). Teachers College Press.

Stake, R. (2010). *Qualitative research: Studying how things work*. The Guilford Press.

Stake, R. E. (1995). *The art of case study research*. SAGE.

Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9, Article 1. <https://doi.org/10.1186/s40468-019-0077-8>

[[Crossref](#)]

Sun, H., & Zhang, J. (2022). Assessment literacy of college EFL teachers in China: Status quo and mediating factors. *Studies in Educational Evaluation*, 74. <https://doi.org/10.1016/j.stueduc.2022.101157>

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *The*

American journal of evaluation, 27(2), 237–246. <https://doi.org/10.1177/1098214005283748> [Crossref]

Tracy, S. (2020). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact* (2nd ed.). John Wiley & Sons.

van Lier, L. (2005). Case study. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 195–208). Erlbaum.

Walters, S. (2007). Case study or Ethnography? Defining terms, making choices and defending the worth of a case. In W. Geoffrey (Ed.), *Methodological developments in ethnography* (Vol. 12, pp. 89–108). Emerald Group Publishing Limited. doi:[https://doi.org/10.1016/S1529-210X\(06\)12006-9](https://doi.org/10.1016/S1529-210X(06)12006-9).

Wang, X. Y. (2017). A Chinese EFL teacher's classroom assessment practices. *Language Assessment Quarterly*, 14(4), 312–327. <https://doi.org/10.1080/15434303.2017.1393819> [Crossref]

Xu, Y. T. (2017). *Assessment literacy of university English instructors in China: A mixed-methods study* [Doctoral dissertation, The University of Hong Kong]. The HKU Scholars Hub.

Yan, X., & Fan, J. (2021). Am I qualified to be a language tester?: Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246, Article 0265532220929924. doi:<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> [Crossref]

Yin, R. Y. (2018). *Case study research and applications: Design and methods* (6th ed.).

Zhang, W. X., & Li, S. J. (2020). Xin shidai daxue yingyu jiaoxue guanli he jiaoshi fazhan de wenti yu duice: *Daxue yingyu jiaoxue zhinan xiangguan yaodian jiedu* [College English teaching administration and teacher development: Analysis of the *College English Teaching Guidelines*]. *Foreign Language World*, (5), 17–23.

Footnotes

¹ To promote university English teachers' professional development, English teaching contests have been held at institutional, municipal, and national levels since 2010. Teachers are required to demonstrate one part of their teaching within limited time before experts who assess participants' teaching performance based on specific criteria. Winning an award in municipal contests or above is one of the key criteria for promotion.

² Universities in China are categorised as key universities and non-key universities. The key universities refer to 100 universities under Project 985 and Project 211 while non-key universities are those common public and private universities. Now, the key universities are 42 universities

that are designated to develop into the world first-class levels and 95 universities including the 42 universities above that are qualified to develop first-class courses under the Double First-Class Project, Shuangyiliu, initiated in 2015.

4. The Case of Emma

Ling Gan¹✉ and Ricky Lam²✉

(1) School of Languages and Communication, Beijing Technology and Business University, Beijing, China

(2) Department of Education and Psychology, Hong Kong Baptist University, Hong Kong, China

✉ **Ling Gan**

Email: linggan@btbu.edu.cn

✉ **Ricky Lam (Corresponding author)**

Email: rickylam@hkbu.edu.hk

4.1 Overview

This chapter reports how Emma, an experienced university English instructor, develops her LAL and constructs her assessment identity throughout three semesters in her daily assessment practices. Emma's previous assessment education empowered her with relatively advanced assessment knowledge. She proficiently used both traditional and alternative assessment methods and made agentic efforts to improve their effectiveness on student learning. She was confident in her unbiased grading of students, although administrative requirements and non-achievement factors influenced her grading decisions. Not acknowledging her accountability, she prepared students for standardised proficiency tests. Emma was also keen on assessment research which facilitated her nuanced understanding of peer assessment. She did not identify herself as an assessor at the outset, despite implementing so many assessment activities. Yet along with her increasing assessment awareness, she recognised her assessor identity

but diminished it deliberately owing to students' negative feedback on her teaching. Throughout the three semesters, Emma had to face heavy workloads, students' negative feedback and institutional mandates during her LAL development. Against all the odds, she took the initiative in enhancing her LAL and constructing her assessment identity.

4.2 Background Information

In her late 30s, Emma has been working as an English instructor to teach College English for 12 years in AU, a key university with a focus on the agriculture fields, after she earned a master's degree in Linguistics and Applied Linguistics. Emma was promoted to be an associate professor when she was awarded with the first runner-up of a national English teaching contest. Now, she has been struggling with another promotion in her career ladder by academic publications.

In Emma's university, College English teaching is level- and phase-based. Students are placed into A and B levels respectively based on their scores of an in-house placement test, with higher English proficiency of level-A students than level-B ones. In terms of the curriculum, College English teaching is divided into three phases. In the first academic year, students learn fundamental English skills in the classes of general English, except the listening skills acquired by students out of class. In the second academic year, students begin to learn courses for specific purposes including English for General Academic Purposes in the first semester and follow-up courses in the second semester. Level A students start ESP and follow-up courses one semester earlier than level B students who learn these courses until the fourth semester. Emma's class we observed over three semesters consisted of level A students who completed General English in the first two semesters and General Academic Writing (GAW) in the third semester, with two lectures (45 min for each) of English learning each week.

4.3 Prior Assessment Experiences

Like most students in China, Emma was assessed through “tons of paper-and-pencil tests in primary and secondary schools” (Emma-IT1, pos.37), as these were widely regarded as the only valid assessment of achievement. The tests, as she claimed,

... cannot assess students’ learning performance comprehensively. They are simply product-oriented, paying little attention to assessing learning process. What’s more, I don’t think pen-and-pencil tests contribute to students’ learning autonomy. That’s why I have thrown away all these traditional test papers. They are useless for me. (Emma-IT1, pos.37)

As a student, Emma expressed aversion towards paper-and-pencil tests, claiming they failed to document her learning progress and overlooked her continuous efforts and positive attitudes.

Once Emma entered university, her language skills were evaluated using diverse methods that offered a more “comprehensive and authentic assessment” than traditional testing (Emma-IT1, pos.37). She was particularly impressed by portfolio assessment used by her writing teacher from an English-speaking country. These portfolios, which she has treasured for 20 years, documented her learning process, academic growth, and achievements, teaching her to self-regulate her learning—an invaluable skill for her future career. Such benefits were not available through traditional paper-and-pencil tests.

As a postgraduate student, Emma took an elective course focusing on fundamental testing concepts including validity, reliability, and purposes of testing, but it excluded hands-on assessment practices. As a result, she nearly forgot the course content when she started working, leaving her disheartened by its unbalanced focus.

Emma furthered her understanding of language assessment through in-service training programmes at AU. She participated in three training sessions focused on grading writing and translation for standardised tests like CET4/6 and the National English Entrance Examination for Postgraduate Studies, and also received training for grading speaking in the National English University Entrance Examination. Although these sessions, ranging from two hours to half a day, aimed to enhance grading efficiency rather than classroom-based assessment skills, they were still valuable opportunities.

... Understanding the criteria can help me assess students' language performance from different dimensions. Based on the criteria, I can attend to the assessment of students' learning outputs accurately, especially the speaking outputs. (Emma-IT1, pos.31)

Inspired by the training content, Emma became increasingly aware of the significant role of criteria in assessment and familiarised herself with more assessment criteria, as she said, "through the training, I know there are criteria-referenced and norm-referenced tests ..." (Emma-IT1, pos.44).

Emma practices were enriched not only by formal training but also by her self-study of language theories and pedagogy, and most importantly, her teaching experience. For example, teaching an IELTS speaking course to AU students in a study-abroad programme pushed her to learn IELTS assessment criteria and innovate effective methods for assessing speaking skills. She was also offered an opportunity to attend a training session with IELTS trainers in the UK, which led her to frequently use IELTS criteria for assessing College English students' speaking abilities.

Overall, at the outset of the study, Emma had a relatively high LAL profile. Her knowledge of language testing and assessment and strong grading skills were a result of professional training and ongoing self-study. Influenced by her own assessment experiences, she had a negative view of traditional tests and favoured alternative assessment methods that tracked students' learning progress.

4.4 An Active Assessment Practitioner

4.4.1 Understanding Assessment Requirements

Emma held a positive attitude towards assessment in English teaching and learning. She took assessment as an essential ongoing part of English teaching practices and as the most direct tool to examine teaching and learning, for assessment identifies learning deficiency and direct learning.

Students are always score-driven. Where there is an assessment, there is a force that drives them to learn. Whatever force

assessment may engender intrinsically or extrinsically, it works for teaching and learning. (Emma-IT2, pos.127)

This quote showed that Emma clearly understood Chinese students' perceptions of assessment and the relationship between teaching, learning and assessment. Thus, she considered language assessment indispensable for College English teaching and learning.

Emma's proposal of "comprehensively assessing students" (Emma-IT1, pos.58) found its way in her self-reported assessment practices. Regardless of a negative attitude (see Sect. 4.3), she did not abandon traditional tests, as using them is "a commonplace in China" (Emma-IT2, pos.105), and these tests could assess students to a certain extent.

I think traditional tests... can motivate students to learn. I also highlight the results of the summative exams. If a student performs well in formative assessments but poorly in summative exams, it may suggest a lack of dedication to learning. Exam scores can somewhat reflect a student's current learning status. (Emma-IT2, pos.105)

For Emma, summative exams could also be used formatively. She valued the information summative tests provided, as this information could help her "adjust the teaching" (Emma-IT2, pos.106).

Aware of assessment's impact on teaching and learning, Emma adhered to the department's proposed assessment tasks to enhance their effectiveness (see Table 4.1). Since these tasks were developed through collaboration with colleagues, Emma believed they were suitable for level A students in College English classes.

Table 4.1 Assessment Requirements for College English by Emma's Department

Assessment tasks	Weight (%)	Descriptions of the tasks
Online autonomous learning	30	Tasks on Unipus, a learning app, and at least three listening assignments on this app should be completed out of classes each semester
Assignments	30	Assignments include students' participation in class activities such as oral presentations or group work and written assignments

Assessment tasks	Weight (%)	Descriptions of the tasks
Oral testing	10	An oral test takes place at the end of each semester. The testing formats and methods are decided by the instructor
Final exam	30	It is a campus-wide achievement test with listening, reading, translation, and writing modules

The department prioritises formative assessment with a 60% weighting over summative assessment at 40%, reflecting a shift towards greater emphasis on formative assessment (China Ministry of Education, [2020](#)). Additionally, teachers have flexibility in defining “Assignments” components (see Table [4.1](#)). Emma customised these components (see Table [4.2](#)) and prioritised them over online autonomous learning due to network issues. She emphasised student involvement and interactions, focusing on productive language skills like speaking and writing. Her assessment plan aligned with both departmental requirements and her views on language teaching and learning assessments.

Table 4.2 Components of “Assignments”

Assessment task	Components	Weight (%)
Assignments	1. Students’ participations: (1) responses to questions by sending comments via “Rain Classroom”, a teaching platform with MOOC (2) oral presentations (3) group work (4) involvement in self/peer-assessment	50
	2. Written assignments (1) writing on <i>Pigai.org</i> , an automatic grading writing website (2) notes as portfolios	50

4.4.2 Integrating Peer Assessment into Speaking

From the beginning of this study, Emma emphasised assessing students’ productive skills, particularly speaking, as noted in Sect. [4.3](#).

... I focus on assessing students' speaking ability because effective English expression indicates strong writing and listening skills. Speaking reflects comprehensive language proficiency and should be both the starting and ultimate goal of English learning. (Emma-IT1, pos.19)

These comments showed that her focus on speaking reflected her belief that English learning should prioritise productive skills, which in turn shaped her assessment approach. To evaluate these skills, she used various methods such as oral group reports, role plays, and individual presentations, often incorporating peer assessment.

Emma's students reported that speaking activities were frequent in her classes, with peer assessment used to evaluate their peers' speaking skills.

GLJ: Our English instructor often asks us to have a group discussion and then evaluate each member's speaking performance. The two major criteria for evaluation are clarity and grammar correctness. I think the instructor intends to encourage us to speak and think in English, as we seldom have the opportunity to use English in our daily lives. (Emma-student-FGI-1, pos.7)

LY:... I can feel that our English instructor emphasises assessing our speaking skills. She assigns speeches, and if we can't present them in class, we are required to record them as video or audio and submit them later. In the following class, we collectively evaluate our peers' performances using IELTS speaking criteria. (Emma-student-FGI-2, pos.22)

Students' comments indicated that Emma effectively organised these speaking assessments. By incorporating peer assessment and providing clear criteria, she aimed to enhance students' understanding of the assessment process.

Emma integrated peer assessment into speaking activities to explore its potential to enhance student autonomy in learning. However, this strategy did not align with her classroom practices. In one observed class, peer assessment was included, but it was not

effectively implemented. This mismatch could be illustrated in the following classroom episode:

The topic for today's class was "love your neighbour". At first, Emma organised a group discussion about students' likes and dislikes in their neighbourhood. After the discussion, students were asked to describe their neighbourhood while other students listened and drew pictures based on the descriptions. Then, students would vote for one picture that was closest to the descriptions. The student who gave the description was asked to justify why he or she believed the picture portrayed his/her description. When this activity was completed, the speaking came to the role-play part.

Emma: Now, we're going to move into another activity, that is a role play. You are still required to work in pairs and groups, okay? In the very first step, you are supposed to tell a story. Firstly, you get prepared to tell a story and choose one of the situations below (shown on the screen) and talk about when you have some funny or amusing experiences, for example, you get stuck in a lift or miss a flight or train. Yes, or some embarrassing moment... And in the second step, you're going to take turns to play the roles of a storyteller and a listener. Those listeners should write down several keywords (Emma wrote down "clarity, details, errors, and funny/not funny" on the blackboard. See Table [4.3](#)) on a piece of paper and evaluate the partner's story by giving a tick.

Students began to tell their stories in pairs within a group. There were seven groups with six students in each. Meanwhile, Emma walked around each group and sat among students to listen and talk. About 10 minutes later, one student from each pair was asked to share his/her story.

Emma: (After one student from group one finishes her story) Thank you. I'm impressed by the consistency in keeping the past tense. You don't need to tell a complicated story. Even though the story is very simple, if you can keep it in the past tense, that's a great job. Okay, now it comes to group two. Who is the group leader? Who would you want to share the story?

Student A: (Stand up and point to student B): ZQ, please.

Emma: Okay.

ZQ stood up slowly, beaming, and began to tell her story. After she finished the story, Emma picked another student from other groups until all groups completed their stories. However, no group members were asked to give oral or written comments on the stories with the given criteria during the whole period. The class ended up with seven students' storytelling. (Emma-F03)

This observation excerpt illustrates that Emma provided assessment criteria but neither collected the assessment sheets nor asked students to give oral feedback on their peers' performance. Without peer feedback, the assessment did not enhance learning autonomy (Miao et al., [2006](#)). This gap was also noted in student interviews, with students highlighting the lack of feedback.

We did not know how our speaking was graded by our peers. The scores were only shown to the instructor but not us. Students who were peer-assessed didn't know their scores at all. What's more, we didn't see or hear any feedback on our speaking from our peers. (Emma-student-FGI-3, pos.114)

Table 4.3 Criteria for Peer Assessment on Speaking

Criteria	Student 1 (name)	Student 2 (name)
Clarity		
Details		
Errors		
Funny/not funny		

At the end of the first semester, Emma realised that peer assessment was not as effective as she had expected. Most students simply gave a holistic score without identifying problems and giving peer feedback on their peers' speaking performance (see Fig. [4.1](#)). This led Emma to question the effectiveness of peer assessment adopted in her class. Although she considered providing students with a training session on

In the second semester, Emma made three attempts to address the effectiveness of peer assessment. First, she employed a new app recommended by the department to store and categorise students' assignments and record their peer assessment participation (see Fig. 4.2). This learning-recording app “reduced the workload, increased the efficiency of peer grading, and raised students’ awareness of peer assessment” (Emma-IT3, pos.7).

姓名	总分	教师分	学生分	互评数
[blurred]	7	6	9	3/3
[blurred]	7	7	7	0/3
[blurred]	7	8	6	0/2
[blurred]	7	7	7	0/3
[blurred]	7	7	7	1/3
[blurred]	7	7	7	0/3
[blurred]	7	7	7	0/3
[blurred]	6	7	6	0/3
[blurred]	6	7	6	0/3

Student name	Speaking score	Score given by teachers	Score given by peers	The number of peer assessment
[blurred]	7	6	9	3/3
[blurred]	7	7	7	0/3
[blurred]	7	8	6	0/2
[blurred]	7	7	7	0/3
[blurred]	7	7	7	1/3
[blurred]	7	7	7	0/3
[blurred]	7	7	7	0/3
[blurred]	6	7	6	0/3
[blurred]	6	7	6	0/3

Fig. 4.2 A Record of Peer Assessment with a New App

Second, she added a new component to the peer assessment criteria. At the beginning of the second semester, one student approached Emma to discuss value learning¹ in his National Undergraduate Training Programme for Innovation and Entrepreneurship. The discussion stimulated Emma’s awareness that this topic was consistent with the current call for value learning through the curriculum in College English teaching, as she recounted,

Value elements can make College English teaching look trendy. The speaking assessment criteria I used for students stem from IELTS, but IELTS is more catered to the western examiners. Yet, College English teaching occurs in the Chinese context where

students are taught to speak about what happens in China in English. I think the value learning through curriculum can combine English teaching with Chinese culture in an appropriate way. (Emma-IT3, pos.39)

This change reflected Emma's agency to promote her assessment practices consciously and selectively by drawing on her sharp perception of the curriculum and accumulated assessment experience. The growing assessment experience, as Emma responded, was relevant to her own "reflections on the assessment practices and processes in the class" (Emma-IT3, pos.72).

Third, Emma conducted a 45-min training session on speaking assessment criteria, explaining them with examples. This training aimed to enhance students' ability to assess peers and improve their speaking. The brief worked effectively, as most students could assess their peers with improved accuracy and at least "70% of students give the same scores as I do" (Emma-IT4, pos.47). This outcome is evident in Fig. [4.2](#), where teacher and student scores are closely aligned in the second and third columns.

Despite the improved reliability between peer and teacher assessment, Emma was disheartened by an incident, as she narrated,

On my way to another class, I found a sheet on the ground. It was exactly the sheet of the speaking assessment criteria I just showed students in the class! It is an important material and took me a long time to prepare. Look, students just threw it away like rubbish. You can't imagine how I felt (shaking her head). (Emma-IT4, pos.53)

Emma was hurt by the student's perfunctory attitude towards the assessment criteria. She also felt frustrated, believing that her efforts to enhance students' understanding of assessment criteria seemed to be disregarded.

Emma's frustration deteriorated with two more incidents. First, her research on peer assessment's impact on learning autonomy revealed unexpected results: it did not significantly affect autonomy. She noted that peer assessment benefited students with strong autonomy but not

those needing more teacher guidance. This observation led Emma to conclude that “even good assessment could not solve all learning problems” (Emma-IT4, pos.17). Second, at the semester’s end, student evaluations did not mention her assessment practices, highlighting teaching design instead. In addition, an informal survey revealed that students did not know how to improve their language learning through peer assessment, indicating their scepticism about alternative assessment methods (Chen et al., [2014](#)). Students’ responses and scepticism led Emma to further question the effectiveness of peer assessment on students’ learning. Despite these challenges, she continued to use this method, as “it still works for some students” (Emma-IT2, pos.201).

In the third semester, Emma reduced peer assessment in her General Academic Writing course, instead using multiple-choice questions (MCQs) to assess students’ understanding of academic writing concepts (see Table [4.4](#)). This shift was driven by the course coordinator, a veteran instructor who was very confident with and proud of this approach and pushed every instructor of this course to follow suit. Emma felt uncomfortable with this approach, as it was “very teacher-centred and lacks teacher-student interaction” (Emma-POC5, pos.2). Nevertheless, she complied out of respect for the coordinator.

Table 4.4 The Frequency of Using MCQs in the Observed Class Sessions

Field observation (the third semester)	Number of MCQs
F09	10
F010	2
F011	10
F012	1

Students held different attitudes towards the MCQs. Most interviewed students applauded them, as these “mini tests” could prevent them from distraction. However, some students frowned on these MCQ tests, as there were “fewer interactions between the instructor and us [students]” and these tests made them “bored and uncomfortable” and “restless” (Emma-student-FGI-3, pos.4, 5 & 9).

At the semester's end, Emma managed to design a peer assessment task on students' oral presentation of academic writing. While students were making presentations, their peers assessed their performance by scoring on an electronic sheet in a Word file based on given criteria. The peer scores were directly delivered to Emma via an app. The average of their peers' and Emma's scores was the students' final result of their oral presentation. However, again, peer scores and feedback were not shown to students after peer assessment.

4.4.3 Conducting Portfolio Assessment

Apart from peer assessment, Emma utilised portfolios as an alternative assessment method. Introduced by her writing teacher (see Sect. [4.3](#)), portfolio assessment had made a lasting impression on Emma as a student.

My foreign teacher used portfolio assessment in the writing course, with a clear objective to record students' progress in different writing genres. Now, as College English is a comprehensive course, I do not record students' learning process of one language skill but all the skills. I ask every student to have a notebook where they note down what they have learned, including vocabulary, mind mapping, and their reflections on learning. Their notebooks help me assess their learning attitudes and progress. (Emma-IT1, pos.41)

This excerpt had at least two indications for her understanding of portfolio assessment and assessment in general. First, Emma used portfolios to record her students' learning progress, through which she could make assessment decisions on students' learning process instead of learning products only. Second, she adapted portfolios to meet teaching needs, choosing a notebook over a leaflet to effectively display students' individualised learning (see Fig. [4.3](#)).

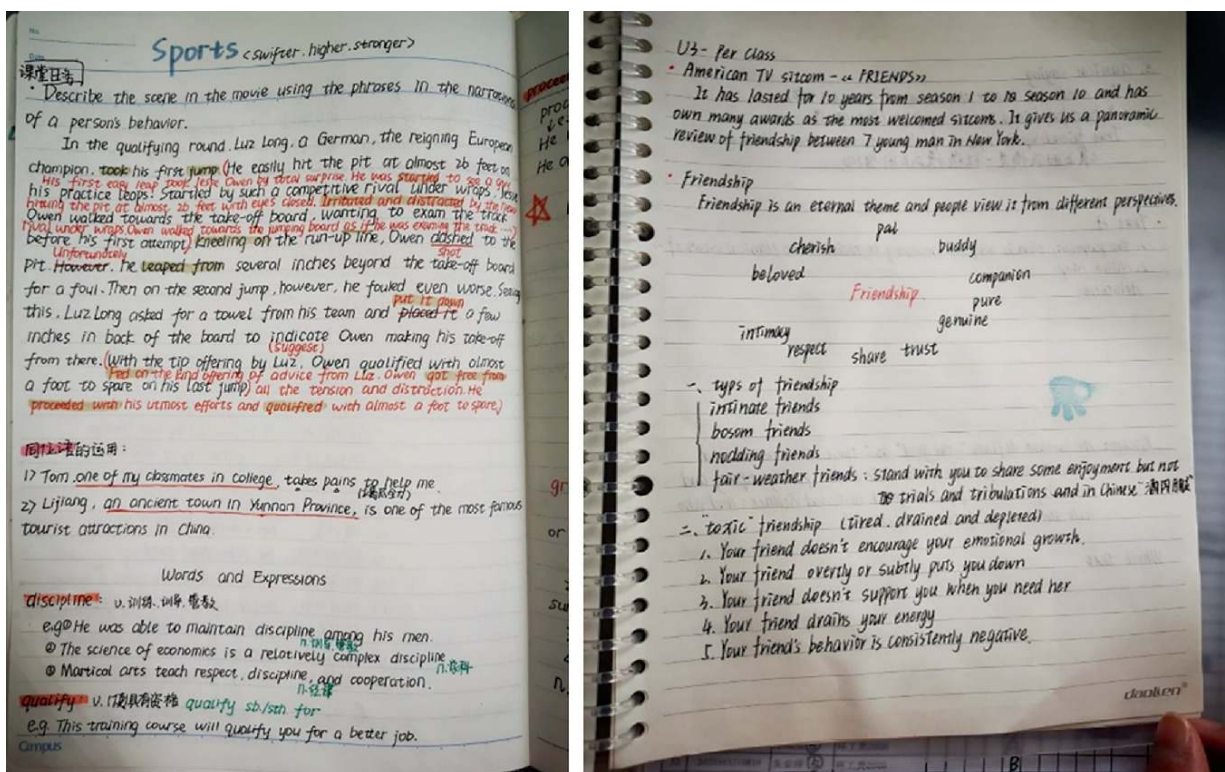


Fig. 4.3 A Snapshot of Students' Notebook Portfolio

However, this approach faced setbacks. Emma did not explicitly communicate the purpose of portfolios to students (Emma-student-FGI 2 & 3). Without a clearly stated objective, students would not reflect on their learning progress through portfolios. Additionally, she provided no specific feedback, merely marking the notebooks with a tick due to time constraints. Consequently, the portfolios did not effectively serve learning, grading, or reporting functions (Lam, 2020), just as some students complained, "it is pointless to hand in the notebooks which are returned without any specific feedback" (Emma-student-FGI-2, pos.166).

In the third semester, as physical notebooks were too heavy to carry, Emma switched them into e-portfolios for convenience. She asked students to submit self-edited writing drafts to an online platform, creating digital portfolios. These drafts highlighted revisions focused on elements like linguistic features and academic structure, with submissions required throughout the semester. In addition, the drafted versions should be submitted while the course was in progress rather than at the end of the semester. In this process, Emma showcased

exceptional portfolios to set benchmarks, believing this practice could “prompt students to learn better by telling them what the good examples should be” (Emma-IT5, pos.97).

However, Emma cast doubt on “good examples” in portfolios, recognising that they were not good indicators of students’ language proficiency, as she commented,

Actually, I feel a bit...eh every time I present good examples of portfolios, how to say, a bit diffident... Some boys don’t enjoy taking notes, but they are very proficient in language skills... Those who take very detailed and clean notes look least active. You know, they do not necessarily have better language skills than those who can’t do well in taking notes. So, I am thinking if there is a point of telling students what good examples are... (Emma-IT5, pos.97)

It could be shown that the ongoing reflection led Emma to evaluate her portfolio practices critically, providing insights into the effectiveness of various portfolio formats for enhancing student learning.

4.4.4 Grading

While adeptly using alternative assessment methods, Emma was confident in her ability to grade students’ performance effectively, drawing on her extensive experience, including prior training in rating and teaching IELTS speaking. She claimed that “at the sight of a [writing/speaking] sample ..., I can judge immediately which level it should be placed into. My judgement is always right” (Emma-IT5, pos.147).

Her confidence grew after attending a two-phase training programme by the British Council in the third semester. The first phase was an online training focused on the assessment criteria of IELTS components aligned with the Common European Framework of Reference for Languages (CEFR). The second phase involved testing and assessment workshops by Chinese scholars. Although the programme was a chance to enhance her LAL, Emma felt it offered no new insights or changes to her assessment practices, because she had “paid much attention to language assessment before this [assessment training]”

and she “assess[ed] in a way probably based on my [her] previous understanding of language assessment” but not because of the training (Emma-IT5, pos.167).

Insisting her grading was criteria-based, Emma expressed her concern for fairness in grading. If assessment tools or methods were not used properly, students could not be assessed fairly. As such, an assessment was unfair to those who “work very hard, but perform poorly” (Emma-IT2, pos.183).

She was also concerned when her grading failed to differentiate students’ English proficiency levels, especially during the final achievement exam. While closed-ended items like MCQs in listening and reading were automatically scored, open-ended items like translation and writing, meant to distinguish proficiency levels, did not achieve this purpose.

We, instructors, are required not to grade too strictly, or a lot of students will fail the final exam, which is not allowed by the Academic Registry. The Registry office has a policy that instructors cannot fail one-third of students in a class. In addition, the department requires that the average score in writing should come to 11 out of 15. I think I am a strict marker, but no matter how strict I am, the grading cannot separate students, as most students can get 11 or 13 in the writing section. No difference in students’ final scores. (Emma-IT2, pos.184)

Emma’s grading was influenced by administrative requirements from both the university and her department. Despite aiming for objectivity, she had to adjust scores higher to meet these standards so that “[students’] score sheets all look good” (Emma-IT2, pos.186). This conflict between fair grading and institutional mandates led her to question the reliability of the final exam.

In discussions with colleagues during collective grading, Emma found the conversations “simple and superficial” (Emma-IT2, pos.222), focusing mainly on test item difficulty and student performance. Thus, these discussions did not resolve her concerns or deepen her understanding of the grading issues.

Although Emma claimed her grading was unbiased, non-achievement factors influenced her decisions. During a 45-min oral test in the second semester, observed firsthand, students were required to give a three-minute speech on one of topics covered in the semester. They prepared before the testing and selected one topic by drawing lots during the testing. Students' performances were scored on a paper, which is shown in Fig. 4.4.

2020 级 A 班四级期末口语考试		
日期: 2021 年 6 月 21 日	地点: 教学楼 211	监考老师: [redacted]
	小演讲 (100 分)	
210102	75	
210115	83	
210117	90	
210119	70	
210120	78	
210121	82	
210123	90	
210127	75	
210129	78	
210130	88 90	
210132	88	
210202	75	
210205	78	
210212	75	
210215	73	
210216	70	
210217	70	
210219	75	
210220	85	
210223	85	
210224	70	
210225	70	
210226	87	
210227	72	
210228	76	
210230	85 88	
210231	85	
210303	85	
210306	85	
210309	70	
210318	78	
210319	75	
210320	85	
210323	85	
210324	75	
210325	75	
210327	70	

Speaking test (Band 4) for year-one students		
Date: 21 June 2021	Location: s11	
	Invigilator: [redacted]	
	Mini speech (100 full mark)	
	210102	95
	210115	83
	210117	90
	210119	90
	210120	98
	210121	82
	210123	90
	210127	95
	210129	98
	210130	88 90
	210132	88
	210202	95
	210205	98
	210212	95
	210215	93
	210216	90
	210217	95
	210219	95
	210220	85
	210223	95
	210224	90
	210225	90
	210226	89
	210227	92
	210228	96
	210230	85 88
	210231	85
	210303	85
	210306	85
	210309	90
	210318	98
	210319	95
	210320	85
	210323	85
	210324	95
	210325	95
	210327	90

Fig. 4.4 A Grading Sheet of an Oral Test

By adopting a holistic grading, Emma assigned scores “in a decisive way” (Emma-IT4, pos.41) except for two students. Their scores had been reconsidered and raised. She explained that the two students were positive towards English learning and fully engaged throughout the semester. Considering their learning attitudes and efforts, she gave extra scores. This reflects a common practice among EFL teachers, considering factors like effort and motivation in grading (Arrafii, 2020; Sun & Cheng, 2014).

The same approach was used in the third semester for oral presentations on academic writing, where she again increased scores

for two students whom she felt deserved it.

... I am very clear why I raised the two students' scores. They have positive learning attitudes and engage themselves in every class attentively. Instead, you can see I did not raise the score of another student right below the two students. He got 89 but I did not raise it to 90 since the GPA from 89 is different from that at 90 or above. He is not on the list of being considered to get a score up... (Emma-IT6, pos.52)

Thus, Emma found herself in a grading paradox: aiming for fairness and objectivity in summative assessment but needing to adjust scores to meet institutional requirements and account for non-performance factors. However, Emma remained confident in her grading, partly due to her familiarity with her students and understanding of the speaking assessment criteria, and partly because she had a heightened awareness of using assessments effectively.

4.4.5 Preparing Students for CET4

In most Chinese universities, preparing students for CET4/6 exams is often part of College English teaching (Li, [2021](#)). Although CET4/6 certificates are no longer mandatory for graduation, they are relevant to employment and further study opportunities, such as postgraduate programme recommendations. Despite its current low-stakes nature, CET4/6 remains a major concern for students. However, Emma's course syllabus did not include CET 4/6 preparation. She believed that preparing for these exams was "not a responsibility" of university English instructors but rather "students' own job" (Emma-IT4, pos.175).

After observing her students struggling with an online listening test, Emma realised that they were not adequately prepared for the upcoming CET4. Colleagues' rhetorical questions, like "You are not going to prepare students for CET4/6?" also prompted her to reconsider, believing that "... if I [she] coach[es] CET4 a bit in class, that may arouse students' attention to the test and motivate them to learn English" (Emma-POC4, pos.28).

Emma understood the significance of CET4 as more than a certificate. “It is not a matter of certificate but a matter of qualification. If students fail to arrive at the cut-off score of CET 4, they will be disqualified from a lot of programmes” (Emma-IT4, pos.25). This awareness, coupled with the intent to motivate students, led Emma to prepare them for the CET 4 in the second semester. This decision aligned with her understanding of the positive washback effect of assessment mentioned in Sect. [4.3](#).

To ready her students for CET 4, Emma dedicated several classes to coaching, focusing particularly on listening. The following observation excerpt notes her test-taking strategies for CET 4 listening preparation.

Emma:... Today, I am going to give a lecture on preparing for CET 4 listening. I expect that after this lecture you can know how to choose answers before listening by making predictions through reading the items. We can practice while I talk about the strategies. First of all, I will summarise the problems you have in listening. Second,...

Emma: The first problem in listening is related to your pronunciation... Generally, if you can pronounce a sound correctly, you recognise the sound quickly. If you can't, you will be late for catching the sounds and miss the message. The second problem is concerned with the understanding of tone and intonation...

After Emma identified four problems in CET 4 listening, she provided targeted strategies to solve these problems. She then began to play the listening audio files.

Emma: Next, we are going to hear a news report (Three item stems with options were presented on the computer screen). While you are listening, please pay attention to the speech features I mentioned just now. In this report, you will hear some linking sounds. Let's see if these sounds would impact your listening comprehension. In addition, you need to change the habit of doing multiple choices. I expect that you CAN [emphasis added] try to choose the correct answer by reading the four multiple choices WITHOUT [emphasis added] listening to the questions related to the news. Now, let's have a try.

Emma played the news report audio once. When it was finished, Emma asked students to show their answers immediately. (Emma-F05)

This excerpt from a classroom observation captures part of Emma's "problems-strategies-practices" session aimed at preparing students for CET 4. First, Emma emphasised the link between speaking and listening skills, aligning with her views in Sect. [4.3](#). Second, she sought to boost students' listening scores using the "read-to-tick" strategy. However, students had mixed opinions on its effectiveness.

WQG: I remain neutral toward this strategy. I think it works for some listening passages but not for all. (Emma-student-FGI-1, pos.41)

ZXY:... This strategy can get us ready enough for CET 4, but I don't think it can qualify us for CET 6, which is much more difficult than CET 4. (Emma-student-FGI-2, pos.96)

Emma's focus on CET 4 listening stems from two primary reasons. First, students often have lower listening proficiency, particularly those from areas where listening is not part of *Gaokao* (an annual national undergraduate admission exam of China) and thus not intensively practiced. This prompted Emma to "provide more guidance on test-taking strategies for this sub-skill" (Emma-IT4, pos.27). Second, Emma believes there is greater potential for improvement in listening, as it has been underemphasised in College English teaching (He & Chen, [2017](#)). Therefore, students need more strategies and practice in this area.

However, in the third semester, Emma was observed not to prepare students for CET6. It was partially because "time is really tight to cover all required contents in General Academic Writing with no extra time for CET6 preparation" (Emma-IT6, pos.81). More importantly, Emma viewed test preparation as "students' own job", believing she was not accountable for preparing students for CET4/6, and thus saw low necessity for this focus.

4.5 An Earnest Assessment Researcher

Emma's interest in peer assessment and its integration into speaking activities motivated her to engage more deeply in research on this method. She hypothesised that assessment could enhance students' learning autonomy, a skill often underdeveloped in Chinese primary and secondary education.

To explore this, Emma conducted a quasi-experiment study² to examine whether peer-assessment could impact students' learning autonomy. This research topic emerged when she pivoted to online teaching due to the COVID-19 pandemic in early 2020. She gathered data using questionnaires and student interviews.

Despite collecting data, Emma had not yet transformed it into a manuscript by the start of my study. She was occupied with administrative tasks, such as preparing a colleague for a teaching contest as part of an expert panel, which "pushed [her] to revisit [her] assessment knowledge, skills, and practices from a renewed perspective" (Emma-IT4, pos. 53). Additionally, as an associate professor, she was frequently involved in various non-teaching activities.

As an associate professor, you have to serve the department or the university, even without payment. Otherwise, you are not considered and recommended by the department or the university if you want to be promoted in the academy or administration. (Emma-IT2, pos.139)

Emma ascribed the failure of the manuscript writing to her under-designed study, which she felt stemmed from a lack of theoretical foundation in language assessment and limited research methodology knowledge. She acknowledged the need to improve through sustained literature reading but admitted to struggling with finding relevant materials.

Participation in our study marked a turning point. As Emma recounted, she "caught some keywords [from interviews]. Using them, I [she] searched more assessment-related papers through the databases in the university's library" (Emma-NF2, pos.20). This experience catalysed Emma's LAL development, as evidenced in her eagerness to

access more assessment knowledge and skills and her commitment to conducting well-designed research.

By the second semester, Emma initiated a “well-designed” study on peer assessment in her IELTS class. The promising data motivated her to begin writing, viewing the process as a reflection on assessment practices enhancing her understanding of peer assessment in English teaching. Before drafting the manuscript, she submitted an abstract to an international conference, hoping its acceptance would further motivate her to complete the manuscript.

It [writing the manuscript] prompts me to think about the study thoroughly... I think it is a part of my reflection and an indicator of how far I can reflect. It is also a part of my teaching practice. Anyway, I would like to move on, no matter what result there will be. The research has tickled my curiosity [about peer-assessment]. (Emma-IT5, pos.145)

In the middle of the third semester, Emma was thrilled to learn that her abstract was accepted for the targeted conference. Motivated by this success, she committed herself to writing the manuscript. However, the writing process was not continuous; interruptions came from teaching, administrative duties, and family matters. Despite these challenges, Emma persisted in writing whenever possible. The conference acceptance fuelled her motivation to advance in academic writing focused on language assessment.

... I have the motivation to work on the manuscript when the abstract was accepted. The research prompts me to think more about two sides of assessment on learning and the research won't be an end when the manuscript is finalised. Instead, I will consider how I can continue and improve the research in this agenda next semester... (Emma-IT6, pos.100)

Emma also realised that while “doing research does not necessarily lead to the changes of assessment practices” (Emma-IT6, pos.102), it promoted reflection and exploration, enriching her LAL. This awareness aroused her determination to progress in assessment research, which in turn, contributed to her classroom teaching.

Seemingly, Emma gained growing interests in integrating assessment research with her practices.

Additionally, as she accessed more literature during the manuscript writing, Emma deepened her understanding of peer assessment. She recognised its significant impact on students, noting that “students have a self-growth from their peers, as it [peer assessment] has provoked them emotionally and intellectually. The subsequent impact is huge” (Emma-IT5, pos.136).

4.6 Perceived Assessment Identity

4.6.1 “I Am not an Assessor but an Explorer”

Along with her LAL development in assessment practices, it was surprising that Emma did not identify herself as an assessor at the study’s outset, although she assessed students on a daily basis.

I don’t have a strong identity as an assessor. The conventional assessment is score-oriented. If I do not grade students, it seems that I am not conducting assessment. However, I don’t grade students all the time. Maybe my understanding of being an assessor is limited, but I believe that assessment is always closely related to scores or grades. (Emma-IT1, pos.106)

Two inferences about her assessment identity emerge: First, the label of being an assessor is always connected with a score or a grade. For Emma, being an assessor means grading or marking students’ language performance. Secondly, not all assessment activities in her class involve scores or grades. Consequently, she viewed informal, instruction-embedded assessments as part of teaching rather than assessment (Rea-Dickins, [2001](#)).

In the meantime, Emma intentionally downplayed her assessor identity. She recognised that students, accustomed to score-focused learning from *Gaokao* preparation, often equated learning with grades. This mindset hindered deeper learning necessary for university adaptation. To counteract this, Emma purposively deemphasised scores by not explicitly informing students that their language performance was being assessed. For instance, she required students to use

portfolios to record their learning without giving any scores or grades. She expected students to shift their focus away from grades.

Instead of identifying as an assessor, Emma preferred seeing herself as an “explorer” seeking “the shining points of students” (Emma-IT1, pos.100). This updated identity stemmed from her evolving understanding of language assessment as a process intertwined with teaching and learning. Emma emphasised providing emotional support alongside feedback, which she believed was crucial.

When giving feedback or comments, we must be cautious, because it is a good time to explore students’ shining points in assessment activities. The shining points can fulfil students’ self-esteem and give them confidence in assessment engagement. They are also encouraged to advance their learning by making good use of the feedback or comments. (Emma-IT1, pos.100)

Emma insisted that her feedback not only guided students on improving language performance but also motivated them by acknowledging their efforts, fostering learner autonomy. This perspective led Emma to view herself as a feedback giver, who “diagnose[s] problems and offers help in students’ autonomous learning” (Emma-NF1, pos.44).

4.6.2 “I Am an Assessment Facilitator”

At the end of the first semester, Emma began to highlight her identity as an assessment facilitator, focusing on promoting students’ autonomous learning, as she claimed,

Assessment itself is an approach to develop students’ learning autonomy by involving them in self-/peer assessment. My job, as an instructor, is to facilitate the use of assessment to achieve this goal so that students can view the assessment results and washback in a proper way. (Emma-NF2, pos.34)

Emma’s identity as an assessment facilitator became more pronounced in the second semester as her awareness of assessment grew. She implemented more assessment tasks, used an app to document peer assessments, and increased writing tasks on *Pigai.net*, an automated

writing assessment platform. These actions helped her as an assessor, guiding students in reflecting on their learning process and progress, as she described,

As an instructor, I have to evaluate the assessment from the perspective of an assessor. I need to guide students to participate in assessment tasks such as self or peer assessment. When they are doing peer assessment, I need to give them linguistic support. After the assessment, I evaluate if their self or peer assessment tasks are appropriate or not. (Emma-IT3, pos.76)

Emma played various roles in assessment, acting as a “supporter or guide” to provide scaffolding on content, methods, and criteria, and as a “judge” evaluating students’ performance. Along with the shift of roles, Emma immersed herself in assessment activities as an assessor. Her evolving understanding that “assessment is not always manifested with scores or grades” (Emma-IT4, pos.111) and her participation in our research reinforced this identity. She noted that “your interviews remind me that I need to be aware of the strong washback assessment exerts on the teaching and learning” (Emma-IT3, pos.68), and that she was “clear about what to assess and how to assess” (Emma-IT3, pos.68).

Over the three semesters, Emma consistently identified as a facilitator of language assessment. Despite this identity weakening by the study’s end (see Sect. [4.6.3](#)), she aimed to empower students to become assessment facilitators by involving them in designing tasks and criteria, ensuring that “students can be placed to a greater extent in the core of autonomous learning in which both teachers and students are key agents.” (Emma-IT6, pos.117).

4.6.3 “Diminish My Identity as an Assessor”

At the outset, Emma did not identify as an assessor. However, over three semesters, she gradually developed her identity as an active assessment practitioner. She implemented formative assessment methods, utilised peer assessment for speaking, and designed assessment criteria. These activities slowly shifted her perception of

herself as an assessor and made her aware of the accompanying responsibilities.

Maybe I am also an assessor... Being an assessor, I need to evaluate if students can appropriately assess themselves and their peers. I may need to observe and assess students' changes in language, language knowledge, and competence as well as the competence of self and peer assessment... after all, students need my help and judgement as an assessor. (Emma-IT5, pos.216)

By the third semester, Emma's confidence in this identity wavered. She was required to increase the use of MCQs to test students' understanding of academic writing, as the General Academic Writing course coordinator favoured this approach. Emma disliked this style, feeling it placed excessive pressure on students and turned her into a "student controller" (Emma-NF4, pos.27). Unsurprisingly, she received negative feedback in student evaluations, which left her feeling depressed rather than agitated.

... Saying "I'm not an assessor" is my instinct response. There used to be a rapport between me and students. I didn't assign students many assessment tasks, and they were very happy. Over these three semesters, when I increased the frequency of assessment and utilised a variety of assessment methods, I found that students were put under great pressure. They became extremely anxious when they got low scores. The consequence was that they gave negative responses in the teacher evaluation. I think this is bad for the teacher-and-student relationship. I am quite clear that those teachers who rank first or second in the department in the teacher evaluation won't assign many assessment tasks to students. (Emma-IT6, pos.119)

Previously, Emma had avoided identifying herself as an assessor, fostering positive rapport with students by not assigning many assessment tasks. However, the increased assessment led to negative

student feedback, discouraging her and undermining her relationship with them.

As a result, Emma determined to diminish her identity as an assessor in her future teaching. She planned to limit herself to departmental assessment tasks rather than develop additional activities, stating “I can’t make students realise that I am an assessor. This damages the teacher-and-student relationship” (Emma-IT6, pos.127).

4.7 Emma’s Case in Summary

4.7.1 LAL Profile

Emma’s case over three semesters illustrates her LAL profile. She acquired conceptual knowledge about language testing and assessment during her MA programme and was well-versed in fundamental principles and concepts (principles & concepts). Emma understood the washback effects of language testing and aimed to enhance peer assessment to support students’ autonomous learning (language pedagogy). She clearly articulated and adhered to departmental assessment requirements (local practices) and supported the emphasis on assessing productive skills like speaking, which are often weak in Chinese students (knowledge of theory). Accordingly, she embedded peer assessment with speaking teaching, as she firmly believed that peer assessment lent itself to students’ autonomous learning (personal beliefs/attitudes).

Despite opposing paper-and-pencil tests due to her own student experiences, Emma used traditional tests and prepares students for standardised exams, recognising the exam-driven culture in China (sociocultural values). Meanwhile, she prepared effective test-taking strategies (technical skills) and utilised portfolios to document students’ learning process and progress (language pedagogy), although they were not as successful as anticipated.

Confident in grading, Emma drew on her rating and IELTS training to apply assessment criteria effectively (grading & decision-making). She strived for objective grading but had to balance the Academic Registry’s grading policy with her own consideration of students’ non-performance factors (personal beliefs/attitudes). Emma’s LAL profile,

detailed in Fig. 4.5, reflects her understanding of LAL components and her practical application within her teaching context throughout the three semesters. It should be noted that Emma's LAL profile is not used to compare with Rachel's, Ryan's and Tim's shown in the subsequent chapters. Instead, all these profiles simply depict what components the four instructors' LAL include and what LAL level they have reached at these components by referring to Taylor's LAL profile and the five LAL proficiency levels proposed by Pill and Harding (2013) (see Tables 2.1 and 2.2).

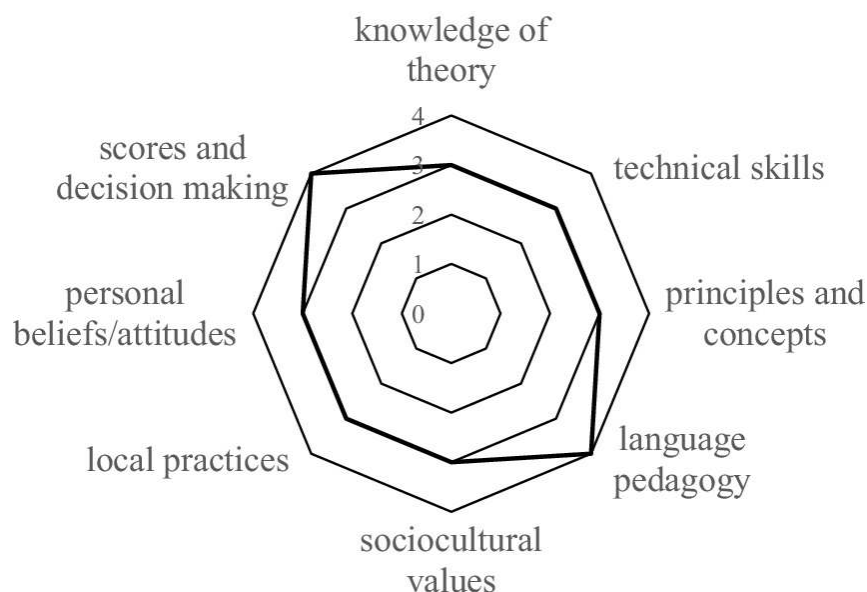


Fig. 4.5 Emma's LAL Profile

4.7.2 LAL Developmental Trajectory

Emma's LAL development followed an individualised trajectory. She acquired her initial knowledge base, and skill sets from the assessment course in her MA programme and her experiences of being assessed as a student. She was skilled in pinpointing basic assessment concepts and utilising traditional and alternative assessment methods in the class right from the beginning of the current study. Throughout the three semesters, Emma increased the assessment frequency and gained growth in assessment awareness along with her accumulation of teaching experience. She also achieved a nuanced understanding of the synergy of assessment with language teaching and learning when she worked as an expert to prepare her colleague for a teaching contest. To

solve problems in the effectiveness of peer assessment, she turned to webinars and made agentive efforts, which reinforced her sense of using assessment to promote students' learning.

Emma's LAL was accelerated and advanced over the three semesters. She became more confident with her grading ability in students' speaking as she gained more access to assessment criteria in an IELTS training programme in the third semester, yet this training session did not significantly change her grading practice. Notably, Emma developed and maintained her interest in assessment research over the three semesters when her abstract with the theme of peer assessment was accepted by an international conference. Moreover, it seemed that Emma became more concerned with students' emotional responses towards assessment when she reflected that the increasing assessment had brought great anxiety and pressure for students. Based on the discussion, Emma's LAL developmental trajectory could be presented as follows in Figure. [4.6](#).

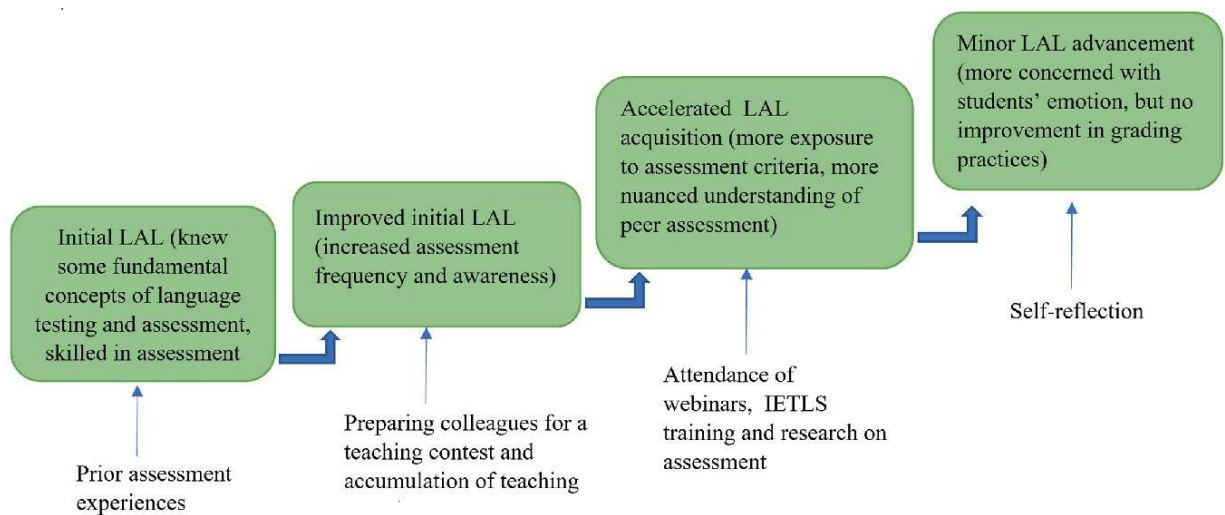


Fig. 4.6 A Trajectory of Emma's LAL Development

4.7.3 Assessment Identity

At the beginning of the study, Emma did not see herself as an assessor. She even avoided constant grading to prevent emotional stress in students. Instead, she viewed herself as an explorer, focusing on discovering students' "shining points" through assessment activities, which allowed her to offer emotional support to anxious students.

Instead, Emma became an active assessment practitioner, integrating formative and summative assessment in alignment with departmental requirements. She graded speaking tests using self-designed criteria, utilised portfolios to track learning, and improved peer assessment effectiveness by trying various methods.

Despite this realisation, Emma maintained a low profile as an assessor in class, concerned that increased assessment frequency was provoking student complaints and negatively affecting the teacher-student relationship. Consequently, she decided to reduce assessment activities to minimise anxiety and negative feedback in student evaluations.

Emma also pursued becoming an assessment researcher, particularly interested in how peer assessment could foster learning autonomy. She conducted research, submitted an abstract to an international conference, and was inspired by its acceptance to craft a manuscript. This process deepened her understanding of peer assessment as an alternative assessment method.

4.7.4 Challenges and Opportunities for LAL Development

During her LAL development, Emma faced three major challenges. Firstly, her heavy teaching and non-teaching workload restricted her LAL input. Due to tight schedules, administrative duties, and family pressure, she struggled to consistently read relevant literature, limiting her theoretical understanding of language assessment. Additionally, Emma found it difficult to locate appropriate literature, a common issue among university English instructors in China who often lack the knowledge to identify necessary papers (Borg & Liu, [2013](#)).

The second challenge came from her students. Emma's increasing assessment tasks generated negative feelings among them, resulting in unfriendly comments in evaluations, which undermined her confidence in language assessment. Consequently, she chose to reduce her explicit assessor role.

Thirdly, institutional mandates impacted her assessment practices. The university's grading policy compelled Emma to grade student writing loosely, hindering her progress in fair grading practices.

Despite these challenges, Emma had opportunities to advance her LAL over the three semesters. Her experience teaching IELTS speaking

led to an opportunity to attend an online IELTS training session, which deepened her understanding of assessment criteria. She also participated in a peer assessment webinar recommended by colleagues, although it did not directly solve her assessment issues.

References

- Arrafii, A. M. (2020). Grades and grade inflation: Exploring teachers' grading practices in Indonesian EFL secondary school classrooms. *Pedagogy, Culture & Society*, 28(3), 477–499. <https://doi.org/10.1080/14681366.2019.1663246> [Crossref]
- Borg, S., & Liu, Y. I. (2013). Chinese college English teachers' research engagement. *TESOL Quarterly*, 47(2), 270–299. <http://www.jstor.org/stable/43267792>
- Chen, Q., May, L., Klenowski, V., & Kettle, M. (2014). The enactment of formative assessment in English language classrooms in two Chinese universities: Teacher and student responses. *Assessment in Education: Principles, Policy & Practice*, 21(3), 271–285. <https://doi.org/10.1080/0969594X.2013.790308> [Crossref]
- CMoE. (2007). *College English curriculum requirements*. Foreign Language Teaching and Research Press.
- CMoE. (2020). *Guidelines on college English teaching*. Higher Education Press.
- He, L., & Chen, D. (2017). Developing common listening ability scales for Chinese learners of English. *Language Testing in Asia*, 7, Article 1. <https://doi.org/10.1186/s40468-017-0033-4>
- Lam, R. (2020). Writing portfolio assessment in practice: Individual, institutional, and systemic issues. *Pedagogies: An International Journal*, 15(3), 169–182. <https://doi.org/10.1080/1554480X.2019.1696197>
- Li, J. (2021). Perceived effects of CET4 test preparation, language ability, and test performance: An exploratory study of Chinese EFL learners. *Language Education & Assessment*, 4(2), 38–58. [Crossref]
- Miao, Y., Richard, B., & Zhen, Y. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, 15(3), 179–200. <https://doi.org/10.1016/j.jslw.2006.09.004>
- 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>
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429–462. [Crossref]

Sun, Y., & Cheng, L. (2014). Teachers' grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326–343. <https://doi.org/10.1080/0969594X.2013.768207>
[Crossref]

Footnotes

¹ Value learning through curriculum has been advocated by China Ministry of Education in recent years to raise university students' moral consciousness and their ideological quality. It is highly proposed in College English to cultivate students' ability to tell good "Chinese stories" in English.

² This study was conducted before Emma joined our research project.

5. The Case of Rachel

Ling Gan¹✉ and Ricky Lam²✉

- (1) School of Languages and Communication, Beijing Technology and Business University, Beijing, China
- (2) Department of Education and Psychology, Hong Kong Baptist University, Hong Kong, China

✉ **Ling Gan**

Email: linggan@btbu.edu.cn

✉ **Ricky Lam (Corresponding author)**

Email: rickylam@hkbu.edu.hk

5.1 Overview

This chapter reports Rachel's LAL development, her construction of assessment identity and the difficulties and challenges she encountered over three semesters in her daily assessment practices. As a novice university English instructor with limited prior assessment education, Rachel initially modelled her assessment techniques on those of her former teachers. Recognising her weaknesses in assessment, she proactively attended lectures and conferences on assessment and read relevant books. These efforts enabled her to acquire more assessment knowledge and apply it to practice effectively. Rachel first implemented peer assessment in writing and later in speaking. Through iterative trials, she significantly improved her assessment practices.

Participation in a teaching contest further enhanced her understanding of language assessment. With increased awareness and advanced practices, Rachel gradually identified herself as an assessor, despite initially not recognising this identity and considering herself a

stumbling novice assessor. Through her LAL development, Rachel encountered challenges such as a lack of institutional support and students' low language proficiency. Nevertheless, she took her agency to address these difficulties.

5.2 Background Information

Rachel had been working at WU for two years when she joined our study. She pursued her master's degree in translation, as her enthusiasm for this research area has continued since her undergraduate study. After earning her master's degree, she had two job options, namely, either working in a translation company or teaching English at a non-key university. Rachel finally chose to become a university English instructor at WU, a non-key university, considering being a translator is not a good full-time job.

In terms of College English, Rachel's department has created its own teaching and learning programme. Taking students' low English proficiency into account, the programme for non-English majors intends to cultivate students' integrated skills of listening, speaking, reading, writing, and translating, and develop their awareness of both Chinese and Western cultures. The programme also experienced a structural reform in the 2020–2021 academic year. A placement test was developed to place students into ordinary and excellent levels respectively. Students in the ordinary classes learn General College English with integrated skills in the first academic year. From the second year on, they have two options. One is to continue the General College English, and the other is to select the follow-up courses including academic English, advanced English speaking or writing, etc. Because of the strict entry rules for follow-up courses, most students choose to continue General College English. As a novice instructor, Rachel was assigned to teach General College English over the observed three semesters. She taught year two students in the previous two semesters and year one students in the third semester. Different from Emma, Rachel had four lectures (45 min for each) every week.

5.3 Prior Assessment Experiences

Like Emma, Rachel had more experiences of being assessed at primary, secondary and tertiary education. In her early schooling, Rachel was largely assessed through pen-and-pencil tests. Upon entering university, she discovered that her language performance was evaluated by using diverse methods, including Q&A sessions, presentations, speeches, and term papers. Reflecting on her experiences as an assessee, Rachel expressed mixed feelings: while assessment tasks were often strenuous and unpleasant, she recognised their role in motivating students to learn harder. She acknowledged that assessment could provoke negative emotions despite its educational benefits.

... To be frank, as a student, I feel stressed when being assessed. Assessment always reminds me that I should do my best to learn. Of course, assessment also helps me identify my learning weaknesses through the teachers' feedback. Anyway, assessment benefits my learning, but I cannot deny that I turn against all the assessments more or less as a student... (Rachel-IT1, pos.43)

With her "superficial" understanding of assessment, Rachel initially approached assessment tasks by emulating her former teachers.

I am not backgrounded in language education. My assessment of students is probably influenced by past assessment experiences as a student. How I ask questions and how I organise assessment tasks are all learned from my teachers... (Rachel-IT1, pos.16)

Unlike some new instructors, she had limited assessment training. Although she underwent orientation training before teaching at WU, the programme was not tailored for foreign language teachers and was primarily "pedagogy-orientated" (Rachel-IT1, pos.20). Additionally, while observing and assessing other teachers, she noted the evaluation scale did not address assessment practices, and the observed teachers rarely engaged in assessment activities with students. Consequently, Rachel missed opportunities to learn about effective assessment practices during these observations.

Comparatively, Rachel did not have much assessment learning or training experience. Similar to all new instructors, Rachel joined an orientation training before her teaching at WU. As the programme aimed at instructors from different disciplines, the training content was not tailored to foreign language teachers' needs, but "pedagogy-orientated" (Rachel-IT1, pos.20). Apart from taking the training course, Rachel was required to observe and assess other teachers' classrooms with a given evaluation scale. Yet, the scale did not include teachers' assessment practices. Moreover, all the observed teachers did not use assessment activities to interact with students. In this case, it was less likely for her to learn about assessment knowledge and practices from the observation session.

Rachel once attended a one-hour lecture regarding language testing and assessment. As Rachel's department often invited experts in different fields to give lectures, instructors could freely attend lectures they were interested in. Rachel accidentally attended this lecture before she realised its central theme was language assessment. The lecture covered formative, summative, and diagnostic assessments on English teaching and introduced assessment methods applied to these three assessment approaches. Although this lecture was not specific enough to help Rachel use these methods, it increased Rachel's understanding of basic assessment concepts and sparked her implicit knowledge of language testing and assessment, prompting her to connect this knowledge to her practices:

Before the online lecture, I knew little about language assessment, not even about the concepts of language assessment. The lecture is interesting!... I learned the concepts like formative, summative, or diagnostic assessment and understood their implications. I also realise that I have used some assessment methods unconsciously in my teaching before, but I have never associated them with the assessment approaches... I also know that methods which look like teaching methods can be used to assess students. (Rachel-IT1, pos.24)

Overall, Rachel's lack of formal training in language assessment resulted in limited initial LAL development. She had moderate

awareness of the importance of language assessment in teaching and learning, and the basic knowledge of assessment concepts. She also implemented assessment tasks by modelling her former teachers' practices.

5.4 An Unswerving Assessment Practitioner

5.4.1 Understanding Assessment Requirements

Despite limited assessment knowledge, Rachel intuitively considered language assessment prominent in students' learning at least in three ways. Firstly, it examines students' learning process and outcomes, assessing proficiency and promoting learning. Secondly, it serves as a reminder for students, spurring them to learn the required content. Thirdly, it works as "a sucker" (Rachel-NF1, pos.30), drawing students' attention back to what is being learned.

Rachel saw two purposes for language assessment from a teacher's perspective. First, assessment helps diagnose students' learning progress and adjust teaching. "If most students do well in this or that part of tests, I assume that they have mastered this knowledge and my teaching works" (Rachel-IT1, pos.13). Second, it "checks what's missing and what's needed in learning" (Rachel-IT1, pos.46). To maximise the effect of assessment on teaching and learning, Rachel strictly followed assessment requirements by her department to implement assessment tasks for fear of "making mistakes" as a novice instructor (Rachel-IT1, pos.57).

Table [5.1](#) summarises assessment tasks, their weights, descriptions and detailed guidelines. Four inferences include. First, tests specified for formative purposes are CET4-orientated. Preparing students for CET4 was a major task of College English teaching at Rachel's university, with a module integrated into daily teaching to capture student attention and effort. Second, the requirements offer detailed guidance and examples, enabling even novice teachers like Rachel to conduct assessments effectively. Third, they motivate students by awarding bonus scores for in-class performance. Fourth, with 70% of the final score based on exams, summative assessment plays an influential role in determining final grades at WU.

Table 5.1 Assessment Requirements for College English by Rachel's Department

Assessment task		Weight (%)	Descriptions of the tasks	Guidelines for the assessment	Remarks
Formative assessment (With full mark being 100, accounting for 30% of the total assessment at the end of the semester)	Class attendance	20	Students are required to attend every class on time	Instructors can choose their own way of assessing, but they should show assessment sheets to students regularly	Classroom performance includes students' participation in in-class activities like oral presentations or group work and answering questions
	Notes taking	20	Students take notes of learning content in the textbook or a notebook	Instructors can choose their own way of assessing. It is suggested that assessment is based on different scales, such as <i>fully complete and detailed notes, complete and detailed notes, partially complete and detailed notes and no notes</i> , etc	
	Assignments	20	Assignments refer to those reading, writing, listening exercises, or with clear due dates	Instructors can choose their own way of assessing. It is suggested that assessment is based on different scales	
	Tests	40 (see attached excel sheet 4)	Tests include vocabulary, reading, speaking and dictation. Each type accounts for 10% of the total	Vocabulary (10%) (Students memorise 50 words specified by instructors from CET 4 past papers once every two weeks. 300 words are memorised within 12 weeks each semester	

Assessment task	Weight (%)	Descriptions of the tasks	Guidelines for the assessment	Remarks
			Reading (10%). Instructors decide on the number of reading passages, but the passages must come from CET 4 test papers. Instructors need to teach students how to locate topic sentences or keywords from paragraphs. (For specific guidelines, refer to attached file 2)	
			Speaking (10%). Refer to the CET 4 speaking test. It takes place between the 8th and the 10th week of the semester to prepare students for the CET 4 speaking test to be held in the 11th or 12th week	
			Dictation (10%). Dictate words from the textbook	

Rachel applauded these assessment requirements, as they solidly catered to WU students' low English proficiency and learning motivation, helping them engage with English through assessment. In addition, as the university emphasises CET4 passing rates, it was justifiable to orientate more classroom assessments to CET4. Yet, Rachel complained that the plethora of tests specified in the assessment tasks stressed students out, which would "lead students to turn against the course of College English" (Rachel-IT2, pos.59). With respect to this situation, she felt powerless to effect change as "a small potato" (Rachel-IT2, pos.60).

Meanwhile, Rachel appreciated the detailed guidelines for assessment tasks provided by the department. They set useful examples for teachers to assess students formatively. They also raised teachers' self-awareness that their assessment should be evidence-

based, which would guarantee the reliability and validity of teachers' grading, as Rachel commented,

... Without these guidelines, I may grade students randomly and simply give a holistic score for students' daily performance and the final exam...Therefore, the department's assessment requirements have helped me a lot. They have explicitly instructed me what to assess and how to assess... (Rachel-IT5, pos.30)

5.4.2 Utilising Traditional Vocabulary Tests

Although the department guidelines paved the way for Rachel to assess students without “making mistakes”, she did not innovate her assessment activities, instead adhering to prescribed assessment tasks. She relied on traditional tests such as vocabulary, dictations, and speaking and listening/reading tests. With the flexibility to choose her vocabulary assessment method, Rachel preferred a more conventional translation method. She selected 50 words from past CET4 test papers, requiring students to provide their Chinese meanings (see Fig. [5.1](#)).

2020-2021-1 Vocabulary Test 1 CET4 Test Paper (12/2017) (No.3)			
Student Name:		Student No.:	Class:
Teacher Name:			
Time:			
Number	Words	Part of speech	Chinese meaning
1	construction	n.	
2	generate	v.	
3	vehicle	n.	
4	efficiency	n.	
5	element	n.	
6	material	n.	
7	inadequate	a.	
8	hazard	n./v.	
9	identify	v.	
10	evidence	n./v.	
11	facility	n.	
12	track	n./v.	

Fig. 5.1 A Snapshot of a Vocabulary Test Used by Rachel

Rachel employed two methods to examine the outcomes of vocabulary tests. Initially, students chose appropriate words to complete sentences shown via PowerPoint. They had 8 to 10 min for this task. Alternatively, they wrote 20 English words based on given Chinese meanings and parts of speech within 10 min. Due to potential

technical issues with PowerPoint, Rachel favoured the second method. These vocabulary tests were reported to work for students' learning, as they prompted students to memorise new words biweekly. In addition, Rachel noted improvement in students' vocabulary, as "the words they use to answer my[her] questions seem to be more versatile" (Rachel-IT3, pos.19).

Rachel also identified several problems from the vocabulary test results. The tests did not help students learn to use words in context, and she was unsure whether the difficulty level suited their varied proficiency. Additionally, students grew tired of these tests. These problems posed great challenges to Rachel, because she was completely in the dark of solving them. She felt helpless, as novice teachers like her lacked sufficient tacit knowledge about students' language ability and affection, and appropriate mastery of learning objectives and curriculum targets (Xie & Tan, [2019](#)).

Being anxious about this worrying situation, Rachel turned to the first author for help during our first interview, requesting assessment book recommendations. As a researcher, we were aware that our response might intervene with Rachel to some extent. However, not disappointing her, especially as a novice teacher asking for help, we provided a list of books related to language testing and assessment.

Despite the recommendations, Rachel continued using the same vocabulary tests in the second semester, puzzled by the same issues. In the third semester, she reduced vocabulary tests due to time constraints from preparing for a teaching competition, which led to student complaints.

LXT: ... She [Rachel] mentioned there would be several vocabulary tests this semester in the first class, but she actually just gave us one test... I think she should use more tests to practise our vocabulary. If we are not tested, we will forget newly learned words quickly...

TDCQ: ... I also think that vocabulary tests should be used frequently. The tests can motivate us to memorise more words and improve our vocabulary. Eh, they are quite helpful.

XU: I think our instructor gave us very few assignments, given our current learning tasks. The vocabulary tests are the least in particular. Without the tests, we are not motivated to learn and memorise English words.

(Rachel-student-FGI-3, pos.86,88, &90)

However, Rachel resolved the vocabulary difficulty issue by using a “Vocabulary Talents” app introduced and recommended by her department. This app, from a reputable language publisher, helps university students with vocabulary learning and allows difficulty level adjustments to accommodate different proficiency levels.

5.4.3 Giving Oral Feedback

Rachel faced challenges with traditional tests and felt insecure about providing feedback, especially orally, as she “can’t find an entry to give feedback to students’ performance” (Rachel-IT2, pos.49). Besides, as a novice teacher, Rachel prioritised her teaching tasks as scheduled. She would rather spend more time lecturing than providing feedback.

A common informal assessment task in Rachel's classes was Q&A, which highlighted her feedback difficulties. This could be shown in the following observation episode in the first semester.

It was a listening and speaking class. Rachel played an audio file of a news report about the impact of technology development on people’s lives. After the first play, Rachel raised questions.

Rachel: Could anyone tell me what’s the main idea of this news report?

(Students kept silent).

Rachel (Call a boy’s name): Okay. XX, could you please tell me what’s the main idea of this news report?

Student A: (Stand up, lowering his head and saying nothing):...

Rachel: What is the news report about? Did you hear any information?

Student A (In a low voice): Technology

Rachel: Yes! (Raise her tone). The development of future requires the technological development. Did you get any other information from the news? (Then, she repeated the question in Chinese).

Student A: (Keep silent):...

Rachel (Keep smiling): How about these two questions? What did you choose?

Student A (In a very low voice)

Rachel: Okay. B. The second one?

Student A: (In a very low voice): A

Rachel: A? Okay. A. Thank you for your answer. (Rachel turned to another student).
(Rachel-F03)

This episode showed how Rachel gave oral feedback to students' answers. First, she repeated students' answers without scaffolding them to provide extra input. Probably, Rachel did not have sufficient word bank to give comments, so that she described her responses as "monotonous". Second, she struggled to guide students with hints or clues, often moving to the next question without waiting for students' input.

In the second semester, Rachel reported a slight improvement in giving oral feedback, although she still struggled to find time due to lengthy lectures. However, observations showed that she remained ineffective in guiding students to answers her questions, only waiting for students' responses without providing any assistance.

By the third semester, Rachel showed improvement by seeking advice from experienced colleagues who prepared her for a teaching contest. Her colleagues suggested that she should directly point out students' deficiencies and propose improvement strategies. Rachel

accepted this suggestion with minor adjustments, aiming to raise students' self-awareness of learning deficiencies through feedback. Rachel's development in giving feedback could be demonstrated in the following observation episode.

Six groups made a presentation on special food in their hometowns. When one group was presenting, the other group members needed to score the group's performance with an assessment sheet. When all the presentations and peer assessment were completed, Rachel gave her feedback.

Rachel: Thank you for your work. I suppose this is the first time you use PowerPoint slides to make presentations, right? Next semester, we will have more similar presentations...

Based on your work, I want to give you some suggestions. Your presentations show that you're not experienced in this area, as a lot of problems have emerged in the presentation of each group... First, in terms of PowerPoint slides, there are too many words on each slide... You don't need to display ALL YOUR WORDS [emphasis added] but just some key phrases and words... Additionally, did you notice that the background colours are hard to make the words seen? I suggest... The second issue is concerned with your presentation. Some groups did well in coherence. For instance, I heard one group... I hope other groups can learn from this group. The last problem is very serious. That is, the presenters read the scripts instead of talking with eye contact... you can have a glimpse of your script but don't read it.... (Rachel-F010)

The episode revealed that Rachel had found the entry to giving feedback to her students' presentation tasks. Her improved skill was probably ascribed to the assessment criteria used to assess presentations, as these rubrics provided a template to underpin her feedback.

However, as Rachel did not elaborate and personalise feedback for individual groups, the effectiveness of feedback was challenged by students.

LXT: ... Feedback was not given to our group at all. I am quite concerned about the result of the assessment because I have devoted a lot of time to making the PowerPoint slides and practicing my speaking. I am a bit depressed with no feedback on whether or not my group had done well. I think I am not motivated to do another group work. If our instructor can give specific feedback to each group's presentation, we can identify the problems easily and be motivated to do better.

XU: Yes. I also expect that our English instructor can identify and analyse the problems we have in the presentation and then offer some targeted suggestions to help us improve.

(Rachel-student-FGI-3, pos.68&70)

These responses revealed that Rachel failed to adopt effective feedback-giving strategies. This issue arose from limited class time and her lack of training. As such, Rachel reported that she “made limited progress” and was still “less proficient in giving feedback” (Rachel-IT5, pos.66).

5.4.4 Grading

While less confident in giving feedback, Rachel seemed to be more comfortable and proficient in grading students with the help of grading templates provided by the department. Figure [5.2](#) shows a snapshot of an overview of Rachel's grading sheet on students' performance.

2020-2021-1 Formative Assessment Records (For Instructors who teach from 2019)											
Class :		Instructor :									
Number	Student ID	Name	Attendance (20%)	Notes taking (20%)	Assignment (20%)	Tests(40%)				In-class participation	the final score for formative performance
						Vocabulary (10%)	Reading (10%)	Speaking(10%)	Dictation (10%)		
1	1902		20	18	19	9	7	7	8	10	98
2	1902		16	17	17	7	7	6	6	6	82
3	1907		20	15	18	8	6	0	5	0	72
4	1907		19	18	16	8	8	7	6	4	86
5	1907		20	20	20	10	10	6	10	0	96
6	1907		20	20	20	10	10	8	8	4	100
7	1907		20	16	19	9	6	6	5	3	84
8	1907		20	16	17	7	6	7	9	4	86
9	1907		20	16	18	8	6	0	10	0	78
10	1907		20	20	20	10	10	7	10	3	100
11	1907		20	18	18	8	8	7	8	11	98
12	1907		19	16	20	10	6	8	8	1	88
13	1907		19	18	17	8	8	7	8	13	98
14	1907		19	19	18	8	9	7	9	0	89
15	1907		19	20	17	7	10	8	5	8	94
16	1907		20	20	19	9	10	7	9	0	94
17	1907		20	20	18	8	10	7	7	3	93
18	1907		20	20	18	8	10	8	10	3	97
19	1907		20	19	20	10	9	7	10	3	98

Fig. 5.2 A Screenshot of an Overview of Rachel's Grading. *Note.* From the fourth column on (left to right), the eight columns represent the scores of class attendance, notes taking, assignments, vocabulary tests, reading tests, speaking tests, dictations, and in-class performance. The last column is the final score for students' formative performance

Rachel kept the scores on each single component of assessment tasks based on students' performance. With respect to her grading, Rachel commented,

My grading of students' assessment tasks incorporates both objective and subjective elements. For example, class attendance and tests on vocabulary, reading, and dictation are graded objectively; students earn full marks for perfect attendance, and tests are scored using answer keys. However, I do include subjective considerations in my grading process. For instance, my assessment of students' note-taking and in-class performance relies on personal judgment, as there are no specific criteria for what constitutes good or poor notes. Additionally, evaluating student involvement in Q&A sessions, group activities, and language output involves my interpretations, leading to potential fluctuations in grading. I often adjust my grading based on the understanding that some students may risk failing the course if they perform poorly on the final achievement test. In such cases, I am willing to assign

more formative assessment scores to help them pass. (Rachel-IT2, pos.79)

Rachel conducted meticulous grading of students' assessment tasks and documented their performance in Excel sheets, which could guarantee fair evaluations. However, she also adopted a lenient grading approach, often considering non-performance factors, common among Chinese EFL teachers (Sun & Cheng, [2014](#)).

Similar to Emma, Rachel felt pressured to adhere to the unwritten expectations of the Academic Registry, which led her to loosen her grading standards so that most students in her class could pass. She commented,

... There are always uncontrollable factors in students' formative performance. I need to adjust scores in the Excel sheet, because the university requires a pass rate and a "good-looking" grade sheet. It is impossible to adjust the scores in the summative final exam, so I have to turn to students' formative tasks... (Rachel-IT6, pos.14)

However, most students in the focus group interviews expressed their negative comments on Rachel's grading, as could be seen as follows.

JX: I disagree with LXT (who is satisfied with her score given by Rachel and believes that Rachel's grading is objective). Although I recognise my English needs improvement, I don't understand why I failed the final exam, especially after completing all the formative assessment tasks with great effort. I am unclear about the grading criteria and have no idea of how our English instructor arrived at our scores.

TDCQ: I think the scores should be specified, including all scores in the formative assessment components and the scores in the final exam. It would be better if we know how we are graded in a formative and summative way.

XT: You see, we are given only one score which does not suggest anything for our learning.

JX: Agreed. The score does not accurately reflect my performance on the final achievement test or in formative assessments. I am aware that I did poorly on the listening section due to a headset issue, but I'm still unclear about the reasons for my overall failure. I would appreciate it if the instructor could provide detailed scores and share them with us.

XJK: Although I passed, I am uncertain about my weaknesses. I'm unsure if they lie in the formative assessments or the final exam, as I do not clearly understand the grading criteria. Consequently, I cannot accurately assess whether the scores reflect my true language performance.

(Rachel-student-FIG-3, pos.184–188)

This excerpt highlights two issues in Rachel's grading practices. First, despite her claim of using criteria-referenced grading, students were not provided with the assessment criteria. Second, while she carefully graded and recorded students' language performance for each assessment task, she did not share these specific scores with the students. Therefore, students lacked insight into their learning progress, and the non-transparent grading practices did not enhance their educational experience.

While Rachel adjusted students' scores on formative tasks subjectively due to administrative mandates, her grading for the speaking test was objective and criteria based. The following sheet illustrates how Rachel assessed students' speaking during the test.

As shown in Fig. [5.3](#), Rachel rated each student by assigning precise decimal scores for every criterion, demonstrating her adherence to a strict criteria-based grading approach rather than the holistic grading commonly used by university English teachers in China (Cheng & Wang, [2007](#)). This approach aligned with Rachel's principle that an assessor should "try his/her best to be fair and objective" (Rachel-IT4, pos.92).

Form 3: Speaking test records (20%)								
Class:		2(full mark)	3	2	3	Teacher:		
Student	Name	pronunciation & intonation	accuracy	fluency	topic relevance			
		1	2	1	2			
		2	3	2	3			
		1.5	2	0	2			
		1.6	2.1	1.8	2.5			
		1.5	2.2	2	2			
		1.6	2.6	1.8	2.6			
		1.4	2.4	1.6	2.2			
		1.6	2	1.4	2.2			
		1.4	2	1.4	2.2			
		1.2	2	1.2	2			
		1	2	1.8	2.4			
		1	2	1.8	2.4			
		1.4	2	1.6	2			
		1.4	2.2	1.4	2.4			
		1.2	2	1.2	2.4			
		1.4	2.2	1.4	2.2			
		1.2	2.4	1.4	2.6			
		1.6	2.4	1.6	2.4			
		0.8	1	0.8	2 3.8			
		0.4	1	0.4	1			
		1.8	2.4	1.2	2.4			
		1.8	2.2	1.2	2.2			
		1.2	2.2	1	1			

Fig. 5.3 A Sheet of Rachel's Grading in a Speaking Test. *Note.* The red words in the third row (from top to down) represent the four assessment criteria from right to left: pronunciation and intonation, accuracy, fluency, and relevance. The red numbers above them indicate the full score for each criterion component. The numbers in black are the scores Rachel gave to students on each component. Each student's final speaking test score is the sum of the four scores in each row

Despite this, Rachel lamented that her grading in the speaking test often felt "rushed". The criteria were so "over fine-tuned" that she struggled to provide scores for each within the limited time available. To address this, she decided to modify the criteria to allow for more

convenient grading. “I will adjust the assessment criteria and make them more convenient for my grading” (Rachel-IT4, pos.54).

By the third semester, Rachel simplified her assessment criteria for the speaking test, reducing the four components to two. This change allowed her to “judge and assess students’ speaking more quickly” (Rachel-IT6, pos.31). In addition, she shifted from analytic to holistic grading, realising that there was “no difference between an analytic grading and a holistic grading” (Rachel-IT6, pos.31). This transition indicated her improved familiarity with the assessment criteria and rubrics, allowing her to apply them more flexibly.

Regarding the grading in the speaking test, Rachel’s students agreed that her grading was careful, objective, and fair.

TX: As I know, the English instructor graded our speaking according to the fluency, accuracy, and other factors. I think her grading can objectively reflect my English-speaking proficiency to some extent...

LY: I agreed with TX. I want to add one thing. On the day of the speaking test, I noticed that our English instructor used her phone to record every student’s speaking with a timer. See, she was so careful about her grading, because I know she would play the recordings to recheck our speaking before she could give us a fair score.

(Rachel-student-FGI-1, pos.232&233).

Compared with the assessment on formative tasks, Rachel’s grading on the summative speaking test seemed to be more accepted by students, probably because it was criteria-referenced and transparent. The criteria ensured objectivity and unbiased scoring which could represent students’ authentic speaking proficiency.

5.5 An Active Assessment Explorer

5.5.1 Conducting the First Peer Assessment on Students’ Writing

Rachel accumulated her assessment knowledge and increased her assessment self-awareness not only from assessment practices but also through more assessment resources. For instance, she read an assessment handbook from the book list the first author recommended after the first interview. Rachel “found treasure” (Rachel-NF2, pos.33), as this book enriched her theoretical understanding of language assessment. This book delineated various classroom-based assessment methods and their applications and also offered tactical strategies to cope with assessment problems. Moreover, academic reading motivated her to attempt to “verify if these findings [of the studies in the handbook] can be applied in my [her] class” (Rachel-NF2, pos.20).

Apart from the handbook, Rachel actively searched academic journals to access updates on local assessment studies. She read articles published in *Language Testing and Teaching*, a Chinese professional journal in language assessment. From these articles, she understood how diagnostic assessment could be used to promote students’ English learning. She also picked up new assessment concepts including consequential validity, predictive validity, and test wiseness.

Sometimes, Rachel found it hard to internalise assessment theories covered in the articles, as she could not “understand what underpins assessment methods” and was confused about “how to analyse students’ assessment results when a certain assessment method is used” (Rachel-IT2, pos.37). This difficulty was ascribed to the fact that she was “unable to invest a large portion of time in consistent language assessment learning” (Rachel-IT4, pos.79). However, through literature reading from time to time, Rachel achieved a slow growth in understanding language assessment concepts.

In addition, a three-day assessment online conference at the beginning of the second semester enhanced Rachel’s assessment knowledge base. This conference was described to be “eye-opening”, as it exposed Rachel to various concepts and methods in language assessment. What impressed her most were the two talks. The first talk regarding critical reading assessment made her realise that critical reading assessment was beyond the basic reading skills and showed her a new direction of how to assess critical reading, although my observations discovered that she had not yet practised this assessment in the class. Through the second talk, Rachel learned a new concept

unique to the Chinese context, namely, *China's Standards of English Language Ability*, a language proficiency framework, similar to CEFR (Liu & Pan, [2019](#)). This talk also deepened her understanding of “different criteria used to assess students’ various proficiency levels” (Rachel-IT3, pos.12).

The increasing assessment knowledge and enhanced assessment awareness brought about changes in Rachel’s assessment practices in the second semester. One of them was her initial attempt to use peer assessment on students’ writing. Before the implementation, Rachel read the assessment handbook and watched demo videos to familiarise herself with the procedures. She also produced peer assessment criteria by referring to the grading rubric for the CET4 writing section. The following is an episode of peer assessment on writing in Rachel’s class.

Rachel first analysed the writing assignment that was completed two weeks ago. Then, she took about 15 minutes to instruct how to write a paragraph, including the use of topic sentences, supporting details, and transitional words. After that, she gave instructions on how to conduct peer assessment to evaluate students’ writing outputs.

Rachel: ... Now, you are going to assess your fellow students’ writing essays in groups... When you are assessing, you should note the following points (A slide of PowerPoint is shown on the screen). First of all,... you should.... Second, you should... You can make comments in the blank on the right side. Third, you need to... The last point is very important, but it is often neglected. That is, you need to check... After the assessment, put your name at the end of the essay, note down the strengths and weaknesses of the essay, and give a score based on the rating criteria. Next, I will show you the rating criteria (A slide is shown on the screen). As you can see, the rating can be divided into 6 ranges. Usually, you.... Then, you... Topic relevance, clear expressions, coherence, and linguistic errors are all delineated in this rating criteria. Now, it is your turn to assess and grade your fellow students’ essays based on all the aforementioned criteria.

Three students sitting nearby formed into groups. Then, Rachel distributed three ready-made essays to each group.

Rachel: Each group read the essays together. Each member should not just read one essay. Instead, you need to read, discuss, and assess THREE [emphasis added] essays together. If there is disagreement, you need to discuss it to agree.

Students began to discuss with each other in Chinese. Meanwhile, Rachel walked around the groups and stopped to listen and talk with students (in Chinese). Most students were actively engaged in the discussion, as they raised their heads frequently to refer to the rating criteria shown on the screen and discussed with group members in murmurs. Peer assessment lasted about 30 minutes.

Rachel: Time is up. Now, let's hear some groups how to assess their peers' writing essays. Any volunteers? (Students keep silent). Your group (Disappointed a little bit, Rachel pointed to the group in front of her). Would you recommend one member to share with us what are the good points and problems of the essays you assessed?

A girl stood up slowly and spoke in Chinese.

Girl A: First of all, this student is very good at using a variety of verbs. The verb collocation is also good. We can also find that the student uses very good sentences and phrases. For instance, she/he attempts to use eclipsing or inverted structures. This is the good point.

Rachel: Ok. Good. So, what are problems in the essays? Let's hear what linguistic errors are found in the essays. You may pay attention to problems such as the singular or plural forms of words and tense.

Rachel looked at Girl A, waiting for her reply. Yet, the girl did not. Then, Rachel nodded and signalled her to sit down. Rachel moved to another group and asked one student to share. 12 groups completed their sharing within 20 minutes.

Rachel: (After students of the last group finish their sharing): Thank you for your sharing. Now, I will collect all the essays and return to every student. Please read your peers' comments carefully and then revise your essay. Next Tuesday (four days after this class), hand in your original essay marked and commented by your peers and the revised version to me. I will score your revised essays and then you can compare the differences between my grading and your peers' grading. Thank you...

(Rachel-F06)

Three inferences can be made from this observation episode. First, Rachel provided brief peer assessment training for students including how to give peer feedback and grade peer essays by presenting assessment criteria and grading rubrics, which are considered being effective to improve the quality of peer assessment (Ashton & Davies, [2015](#); Boon, 2013). Second, she facilitated the peer assessment process by providing scaffolds through explaining rubrics and listening to and talking with students. With teachers' supports, students could generate informative and useful feedback on peers' writing (Schunn & Wu, [2019](#)). Third, she requested students to share their peer feedback, but she did not give her own comments on students' peer feedback. Overall, the first attempt to use peer assessment on students' writing was a successful initiative to step out of "monotonous" assessment tasks prescribed by the department. Rachel thereby claimed, "I can conduct peer assessment!" (Rachel-IT4, pos.7).

Nevertheless, Rachel identified two unintended problems in the first peer assessment attempt. Firstly, students would like to give scores but not write down their names as a marker in their peers' writing sheets. It was possibly because students were worried that the scores would impair their relationship with peers (Li, [2017](#)). Secondly, not all the students were involved in generating peer feedback, especially those with low English proficiency. Without receiving immediate feedback from peers, students may not benefit from the process of giving, receiving and enacting feedback in this formative assessment activity (Schunn & Wu, [2019](#)).

Despite these problems, Rachel was delighted with the expected effectiveness of peer assessment. She was surprised to see that students could identify their peers' strengths and weaknesses and that they could revise and improve their writing based on peers' comments, as "students rewrote paragraphs to make their writing more logical" (Rachel-IT4, pos.14).

This alternative assessment method was also welcomed by the students.

Canty: I think this method is good. When I assess my fellow students' writing, I can learn what is good in their writing. I can also compare my own writing with others' and find out the problems in my writing. Then, I know where my writing proficiency arrives now and how I can improve it.

YLS: Peer assessment brings competition. When I compare my own writing with others', I can identify the disparity between me and my classmates, which will bring about sort of pressure. The pressure will promote my learning interest and motivate me to learn English better. After the peer assessment, teachers' assessment and feedback on the writing can help me notice my writing proficiency. A clear understanding of my English proficiency is a good evaluation of my ability to deal with different kinds of tests.

(Rachel-Student-FGI-2, pos.59&61)

The practice of peer assessment in writing helped both students and Rachel to get a better understanding of students' writing proficiency. This advantage together with the proven improvement in students' writing quality enhanced Rachel's self-efficacy and confidence in implementing peer assessment. She would "continue to use this method in the class" (Rachel-IT4, pos.23).

In the third semester, Rachel intended to optimise the effectiveness of peer assessment. She reflected on the first attempt of using peer assessment and envisioned her further strategies in the implementation.

This semester, I am not going to have peer assessment on writing but on group presentations... What I need to improve is

that I should give students explicit assessment criteria. In the first peer assessment on writing, I simply listed some points on the screen for students' reference when they peer assess.

This semester, I plan to give them a checklist or an assessment rubric in which detailed criteria will be presented... With the checklist or rubric, I think students are much clearer to know how to conduct peer assessment. (Rachel-IT5, pos.82&83)

Rachel refined the assessment criteria of peer assessment. She assigned a group presentation in which students were required to introduce their hometowns' special dishes. While one group was making the presentation, students in other groups listened, watched, discussed, and marked the group's presentation based on a given rubric sheet. The rubric included three assessment components (see Table 5.2). After all groups completed the assessment task, the peer assessment sheets were collected by Rachel who would provide a final score for each group by averaging the other groups' marks. The criteria were sent to students when they were preparing for their presentation. In the class, before the presentations, Rachel explained the rubric in detail.

Table 5.2 Rachel's Peer Assessment Sheet for Presentations

Presentation group Leader of the grading group	Topic	
	Criteria	Score
Content	Introduce a hometown special food appropriately and in detail, including background, names and ingredients (15') (11–15; 6–10; 1–5)	
	Explain the production of a hometown special dishes explicitly, including the use of connection words like “first, next” (15')	
	Clarify the cultural significance, historical background or personal connections of a hometown special food (15')	
Organisation	Logical, cohesive, and coherent (15')	
Delivery	The presenters speak fluently, with appropriate speed and pauses (15')	
	The presenters have eye contact with the audiences, without reading PPT or scripts (15')	

Presentation group Leader of the grading group	Topic	
	Criteria	Score
	PPT is well constructed, and appropriate visuals are used to illustrate the content (10') (8–10; 5–7; 1–4)	

Compared with the first peer assessment, Rachel's peer assessment in the speaking class has improved greatly. An explicit assessment rubric for group presentations enabled students to know how to assess their peer's performance accurately. These criteria were also expected to motivate students to prepare their presentations well, as seen in Rachel's oral feedback on students' presentations in Sect. [5.4.3](#).

Nonetheless, it was not flawless in Rachel's use of this alternative method. For instance, as mentioned in Sect. [5.4.3](#), Rachel collected all the assessment sheets, but she did not return them to students. Students were not informed of the results of their presentations. Moreover, although Rachel gave feedback to students' presentations, students had low efficacy towards the learning activity, as Rachel's feedback was collective, not group-based, which could not motivate them to learn further.

5.5.2 Improving Speaking Tests

While Rachel advanced her LAL in assessment practices by trialling peer assessment, she attempted to improve speaking tests by overcoming challenges in assessing students' speaking tasks. Students' unwillingness to take tests was one of Rachel's concerns. For instance, in the speaking test of the first semester, students were required to give a two-minute public speech on a given topic. To Rachel's surprise, a few students came to her and told her that they were not going to take the test, even though Rachel informed them of all the assessment requirements and the consequences of not taking the test. "This is the first time I have known that students dare not to take tests!" (Rachel-IT3, pos.24). Students' non-cooperative attitudes towards the speaking depressed Rachel greatly.

Depressed as she was, Rachel understood those students, because making a public speech was, indeed, difficult for them, given their low English proficiency. She blamed herself for the difficulty, as before the speaking test she neither explained how to make a public speech nor provided any opportunities for students to practise in the class.

Rachel was trying to address students' reluctance to take a speaking test by discussing it with her colleagues. Their discussions seemed to be fruitful, as they enabled Rachel to know more about various assessment methods and tools and to prioritise two important elements in speaking tests: "students' English proficiency and topics relevant to textbooks" (Rachel-IT2, pos.95).

Considering these two elements, Rachel trialled a different format and task of the speaking test in the second semester. She drew on the topics in the textbook and told students to select one paragraph by drawing lots to read and then answer her questions individually. Rachel explained that the familiarity with the topics and the individual work would make students "not so nervous and stressful" (Rachel-IT4, pos.66). This trial worked, as no students were absent from the speaking test in the second semester. In this trialling process, Rachel's active reflection helped to adjust her assessment practices to be more orientated to students' learning needs.

... I would recall today's assessment practices and students' responses to the assessment. I review if the assessment content and methods are appropriately selected and used to meet students' English proficiency and their needs. I am always exploring what methods are effective and then adjusting them for the next assessment activity. (Rachel-IT3, pos.44)

Persistent reflections prompted Rachel to consider selecting appropriate speaking test methods to foster students' learning. Although these methods distressed her sometimes, agentic reflection enhanced Rachel's LTA awareness substantially.

... I don't want to use the CET 4-speaking format for the first-year students, because I don't want the speaking test to become CET4/6-orientated from the beginning. After all, the College

English course is not all for CET4/6. Instead, it should lead and motivate students to learn English... (Rachel-IT5, pos.91)

Rachel's language testing and assessment awareness enhancement was evident in this interview excerpt. First, she began to think over making good use of the speaking test to promote students' learning rather than for the test's sake only, although she was almost clueless about how-to. Second, she became more orientated to students' learning needs in choosing the test method. She tried to align the test method with students' speaking proficiency and made the test more systematic and sustainable instead of "give students whatever tests come to mind" (Rachel-IT6, pos.16).

Inspired by her reflections, Rachel innovated the speaking test in the third semester. In this semester, as the university raised the pass rate of the final exam to 70%, most of Rachel's colleagues required students to recite translation paragraphs in the speaking test so that students could memorise these versions and did well in the final exam, Rachel critiqued that reciting translation paragraphs was helpful for students' speaking. Different from her colleagues, she asked students to prepare topics learned from textbooks, as these topics could be used in the writing task in the final exam, and, thus, this preparation could achieve dual goals, both for speaking and writing. By innovating test methods, she attempted to optimise and sustain the speaking test to "authentically identify and enhance students' speaking proficiency" (Rachel-IT6, pos.22). Overall, Rachel was willing to take risks in her assessment decisions and developed a growing awareness of fostering language teaching and learning through assessment.

5.6 Participating in a Teaching Contest

As Rachel enhanced her knowledge, skills, and practices of language testing and assessment, she was presented with opportunities for professional development. At the end of the first semester, Rachel was notified that she was going to participate in an English teaching contest on behalf of her university during the third semester. In Rachel's department, it is customary for novice teachers to take turns participating in cross-institutional contests, which aim to strengthen

teaching practices. Rachel was delighted with the opportunity, as it meant more possibilities for her professional development. When preparing for this contest, Rachel repeatedly presented her class in front of colleagues who observed and critiqued her class and offered suggestions for improvement. This peer-observed lesson enabled Rachel to learn from her veteran colleagues both in teaching and assessment.

Another significant opportunity arose during the summer break following the second semester when Rachel attended the *National Training Programme for In-service EFL Teachers in Tertiary Education*, organised by a renowned language publisher in China. During this programme, she observed a teaching contest where she was “shocked to observe two or three teachers saliently demonstrate their assessment in the contest” (Rachel-NF3, pos.20). This was noteworthy since assessment is typically not showcased in contests, as “assessment is believed not to represent a teachers’ teaching competence” (Rachel-NF3, pos.20). Inspired by these peer classes, Rachel was determined to add a checklist to her presentation software to show contest judges that she assessed students’ learning, believing that “teaching, learning and assessing constitute a complete cycle” (Rachel-NF3, pos.20). Apparently, peer observations significantly advanced Rachel’s understanding of integrating assessment with classroom instruction.

On the day of the teaching contest, Rachel demonstrated her teaching talents while competing with her peers. Regarding the teaching contest, Rachel commented,

I don’t think the contest has a substantial impact on my assessment practices but on my teaching. It might probably influence my feedback on students’ language performance. Judges in the teaching contest also assessed how I gave oral feedback to students... During the contest, I used more Q&A and tried to use rich vocabulary to give comments... In addition, there were more interactions between students and me, and more feedback given to students... Another influence may be the inclusion of a checklist in my PowerPoint. I don’t know if it is helpful for the result. Anyway, I included it. (Rachel-IT6, pos.54-55)

Although she acknowledged that the contest influenced her assessment practices to some extent, she did not fully recognise the impact. It remained unknown about the long-term effects of the contest on her assessment practices in real classrooms, but it did help her understand the synergistic value between teaching and assessment.

5.7 Perceived Assessment Identity

5.7.1 “I Am not an Assessor but a Guide”

As a novice English instructor, Rachel focused more on teaching than on assessment to establish a foothold as a teacher. In view of this, at the beginning of this study, Rachel did not recognise herself as an assessor. For her, the identity as an assessor was only limited to the moment she was assessing rather than a consistent label she identified with.

While I am assessing the students, I might be an assessor. However, most of the time, I am doing the teaching job and assessment is only an auxiliary approach to my teaching. For instance, I use assessment to diagnose my teaching problems. I also utilise assessment to discover students’ learning achievement and progress, based on which I can provide feedback accordingly. Assessment is a tool used to facilitate my teaching. In this sense, I am an instructor instead of an assessor. (Rachel-IT2, pos.97)

The quotes revealed that Rachel exhibited a contradictory conception of her identity as an instructor and an assessor. On the one hand, while she acknowledged the two identities were inseparable, she also believed they were mutually exclusive. On the other hand, she felt that because assessment was linked to teaching, her identity as an instructor took precedence over that of an assessor. Only when she conducted assessments or give formative feedback to students, she felt she was wearing an assessor hat.

Rachel preferably claimed herself to be a guide in the assessment process. As a guide, she “decides what, when and how to assess based on students’ learning progress” and “leads students to take assessment step by step” (Rachel-IT1, pos.61). She believed in establishing an emotional connection with her students during assessment. While an

assessor gives an “objective and emotionless” score or grade, “an instructor needs to have more affective considerations” (Rachel-IT2, pos.101) to decide how to help students improve. This identity was also echoed in the following excerpt from one narrative frame.

I think I am a guide in language assessment. Even though we advocate that language teaching should be student-centred, instructors play a very important role in the class. Instructors need to schedule the assessment, choose assessment methods, monitor, and manage the assessment process. They also need to analyse and interpret assessment results and then give feedback to students. (Rachel-NF1, pos.56)

The quote revealed that Rachel provided students with ongoing guidance throughout assessment, including designing and conducting assessments, giving feedback, and interpreting and using assessment results. These guiding directions enabled students to work towards their desired learning results. Given all these advantages, Rachel maintained her identity as a guide in assessment until the end of the study.

5.7.2 “I am a Stumbling Novice Assessor Who Fumbles in the Assessment”

As Rachel’s understanding of language testing and assessment grew, she warily claimed “I guess I might be an assessor” (Rachel-IT3, pos.39), realising that she was assessing “students’ learning progress as well as their language proficiency” (Rachel-IT3, pos.39). The words “I guess” and “may” indicated that Rachel was tentative with her assessment identity. This uncertainty could be attributed to Rachel’s conception of language assessment. She insisted that language assessment should be subordinate to teaching and learning and “it cannot take up a large amount of time throughout the teaching” (Rachel-IT4, pos.81). Because of this mindset, she recognised her instructor identity more than an assessor.

Contradictorily, by the end of the second semester Rachel articulated her identity as an assessor more clearly in the narrative frame. She narrated, “I am responsible for what to assess for students, I

stick to objectivity and fairness while giving assessment and I also analyse the results after assessment in a scientific way” (Rachel-NF3, pos.32). Engaging in assessment practices made her more aware of an assessor’s responsibilities.

It was not surprising that Rachel had such a striking change towards her own assessment identity. The field observations had facilitated her awareness of the importance of language assessment in language teaching and learning, prompting her to give more attention to language assessment, as she explained,

I will begin to think what and how I am going to assess in each class and how or whether I can make the assessment activities more effective and efficient by consciously searching relevant materials” (Rachel-IT4, pos.108).

She began to see her role as an objective assessor, allowing her to conduct assessments that addressed students’ learning needs and communicate results effectively. Nonetheless, she still identified as “a stumbling novice assessor who fumbles in the assessment” (Rachel-IT3, pos.24).

In the third semester, Rachel still regarded herself as a stumbling assessor, even though she had notably improved her assessment practices. She was “constantly exploring how to choose appropriate assessment methods and how to use assessment to promote and guide follow-up teaching” (Rachel-IT5, pos.105). In this way, she linked her assessment identity to her role as a guide to enhance student learning.

Despite efforts to maximise assessment effectiveness by trialling assessment methods, Rachel sometimes felt lost and paralysed when she could not find solutions to assessment problems. This affective experience accompanied her as a novice assessor who “is fumbling in assessment and is making efforts to enhance assessment competence” (Rachel-IT4, pos.89). Her struggles with language assessment were caused partially by the department and the university that “pay no attention to assessment training at all for teachers, especially for novice teachers” (Rachel-IT4, pos.79).

5.8 Rachel's Case in Summary

5.8.1 LAL Profile

Rachel's LAL profile was clearly illustrated through her experiences. Initially, she had limited knowledge of language assessment due to a lack of educational background in the area. However, she recognised the strong connection between language assessment, teaching, and learning (language pedagogy). Through attending lectures and receiving departmental guidance, she improved her understanding of assessment concepts. Additionally, she became familiar with her department's assessment requirements and critically reflected on them (local practices). Although she initially lacked confidence in providing oral feedback, Rachel made progress over the three semesters, although some of her feedback remained insufficient or inappropriate (technical skills).

Rachel demonstrated proficiency in designing assessment criteria and grading students' speaking performances accordingly (grading). Believing there was no significant difference in grading methods, she shifted from analytic to holistic grading (personal beliefs/attitudes). She also utilised Excel sheets to track students' performance across assessments, while sometimes adjusting scores to meet the university's passing rate requirements (local practices).

Throughout the three semesters, Rachel expanded her grasp of assessment concepts, including critical reading and peer assessment (principles & concepts). While she had not yet applied critical reading assessment in her classes, she used peer assessment for writing and speaking tasks and sought to enhance its effectiveness (language pedagogy/technical skills). Understanding her students' English proficiency levels and low motivation (knowledge of theory), Rachel trialled various methods in speaking assessments to better address their learning needs and ensured all students had a fair chance for success (principles). However, as a novice instructor, Rachel was barely aware of the sociocultural aspects of language assessment. Thus, Rachel's LAL profile can be presented as follows (Fig. [5.4](#)).

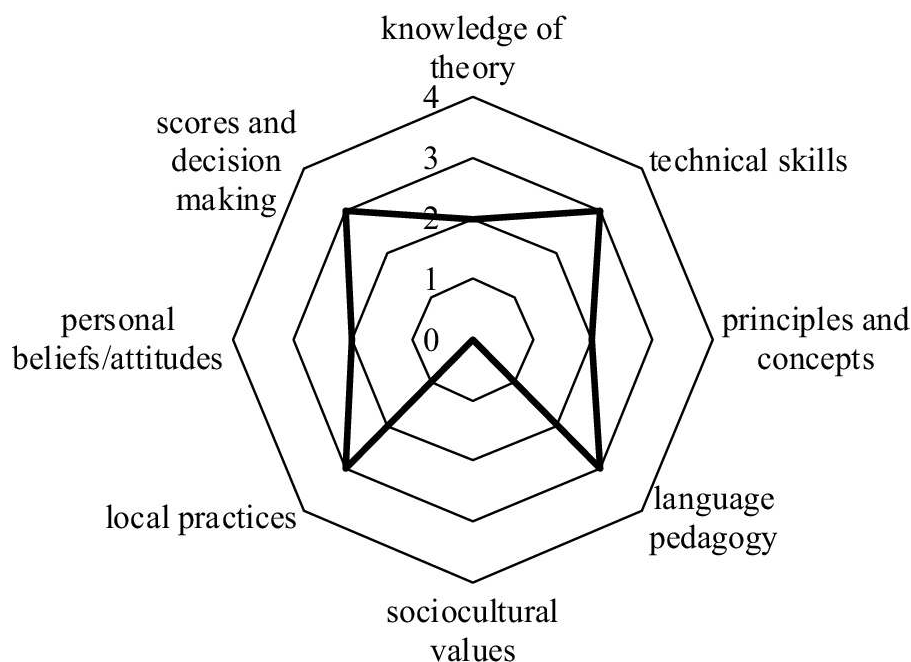


Fig. 5.4 Rachel's LAL Profile

5.8.2 LAL Developmental Trajectory

Rachel's LAL profile reveals three key areas of development. First, she advanced her conceptual knowledge of assessment. Over three semesters, Rachel deepened her understanding through readings, lectures, and conferences, which motivated her to search for additional assessment-related knowledge.

Second, she demonstrated her enhancement in crucial assessment skills. By consulting colleagues and utilising assessment criteria, Rachel provided explicit feedback on students' oral presentations. She also began incorporating peer assessment in her writing and speaking classes to engage students and promote positive learning outcomes. This experience increased her comfort with alternative assessment methods. Although her grading became more objective and was based on established criteria, she occasionally adjusted scores to meet administrative requirements.

Third, Rachel developed her LAL in the area of ethical assessment practices. She became more student-orientated when choosing methods to assess students' speaking. She took into consideration students' English proficiency and their emotional responses to assessment. This shift was credited to her constant reflections on the

assessment practices and her increasing awareness of the importance of assessment for student learning.

Rachel's LAL development followed an individual trajectory characterised by a bottom-up process. With no assessment education background before becoming an English instructor, her initial LAL was limited to the use of formal pen-and-pencil tests and simple informal assessment methods. Gradually, through faculty guidance and attendance at assessment-related events, she advanced her initial LAL when expanding her understanding of language assessment. She also exposed herself to language assessment by attending assessment lectures, seminars and conferences, accelerating her LAL acquisition and construction. While academic reading deepened her nuanced understanding of language assessment, agentic reflections and observations of peers' assessment practices significantly enhanced and innovated her assessment practices. Moreover, her consistent self-reflection and proactive problem-solving mindset further expedited her LAL development. Rachel's developmental trajectory can be illustrated as follows (Fig. 5.5).

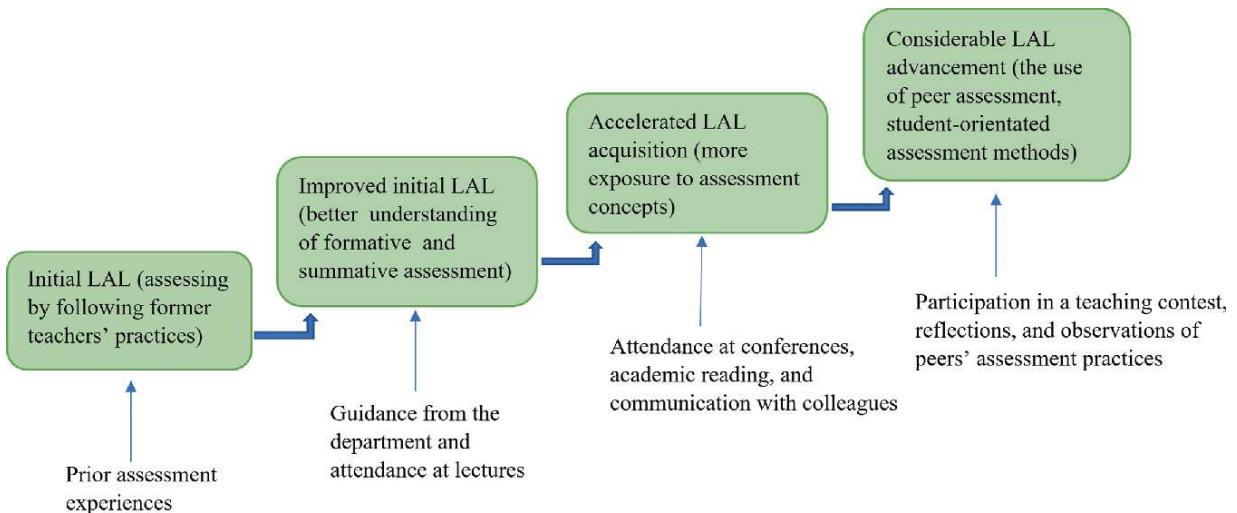


Fig. 5.5 A Trajectory of Rachel's LAL Development

5.8.3 Assessment Identity

Like Emma, Rachel did not initially see herself as an assessor. As a novice instructor placing more focus on her teaching, she was unaware of her assessment practices. Although she administered vocabulary tests, conducted speaking tests, and provided feedback, she viewed

these activities as secondary to her teaching. This perception rendered her dispatched from language assessment in her teaching, marginalising her identity as an assessor. Additionally, her limited understanding of language assessment led her to separate it from teaching and exclude the assessor's identity from her teaching role.

Throughout the study, Rachel gradually became more aware of her role as an assessor. As she gained a better understanding of language assessment through reading and attending lectures or conferences, she realised that assessment activities were integral to her teaching. Meanwhile, the field observations and interviews prompted her to rethink her responsibilities of being an assessor, leading her to focus on peer assessment as a means to engage students in their learning. As a result, Rachel's awareness of being an assessor was heightened.

Despite this growth, Rachel preferred to describe herself as a "stumbling" novice assessor. As she acquired more knowledge of language assessment, she began experimenting with alternative assessment methods in her classes. She sought support from colleagues and observed peers' classrooms to explore how to enhance the effectiveness of her assessment methods. Being aware of her weaknesses, Rachel admitted she was still in the process of exploring effective assessment practices.

Ultimately, Rachel emerged as a committed assessment practitioner. Initially not taking major steps toward developing her LAL, she unflinchingly followed institutional requirements diligently. She employed traditional paper-and-pencil vocabulary tests, graded and recorded student learning progress across all assessment components. She also redesigned speaking tests to suit her students' language proficiency levels, and scored their performances based on self-designed criteria.

5.8.4 Challenges and Opportunities for LAL Development

Over the three semesters, Rachel's LAL development did not proceed without challenges. First of all, she struggled to find time for reading assessment literature. With teaching and preparation consuming most of her schedule, Rachel could not dedicate time to academic reading, which was essential for acquiring the conceptual assessment knowledge she needed.

Second, the lack of institutional support for LAL development posed another barrier. Although Rachel's university offered pedagogical training, it did not emphasise training in language assessment. In addition, novice instructors were encouraged to participate in teaching contests, but these focused more on teaching activities than on assessment practices, as assessment skills were not viewed as beneficial in this context.

Third, students' English proficiency posed a great challenge to Rachel's LAL development. The low English proficiency made students reluctant to involve themselves in assessment practices, leading to Rachel's low efficacy towards language assessment to some extent. How to empower herself with tacit knowledge about students' language proficiency and design appropriately challenging assessment practices have set high demands on Rachel's LAL development.

Despite the absence of formal assessment training, Rachel benefited from various opportunities for LAL development. Her department provided guidance for novice instructors through examples of effective assessment implementation and encouraged her participation in a teaching contest, allowing her to observe peers' practices. This experience increased her understanding of the role of assessment in language teaching and learning.

Furthermore, Rachel proactively sought to develop her LAL through her agency. Aware of her limitations, she requested recommendations about language assessment publications and read refereed journal papers to address her assessment challenges at work. Although her colleagues offered her help, she remained reflective and sought a way out to resolve her predicaments, such as attending conferences to gain additional conceptual assessment knowledge to further her growth in language testing and assessment.

References

Ashton, S., & Davies, R. S. (2015). Using scaffolded rubrics to improve peer assessment in a MOOC writing course. *Distance Education*, 36(3), 312–334. <https://doi.org/10.1080/01587919.2015.1081733>

Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, 4(1), 85–107. <https://doi.org/10.1080/15434300701348409> [Crossref]

Li, L. (2017). The role of anonymity in peer assessment. *Assessment & Evaluation in Higher Education*, 42(4), 646–656. <https://doi.org/10.1080/02602938.2016.1174766>

Liu, J., & Pan, M. (2019). English language teaching in China: Developing language proficiency frameworks. In X. S. Gao (Ed.), *Second handbook of English language teaching* (pp. 415–432). Springer Nature Switzerland AG.
[Crossref]

Schunn, C. D., & Wu, Y. (2019). The learning science of multi-peer feedback for EFL students. *TEFLE*, 13–21. [https://www.lrdc.pitt.edu/Schunn/papers/The%20Learning%20Sci for EFL Students Schunn Wu 2019.pdf](https://www.lrdc.pitt.edu/Schunn/papers/The%20Learning%20Sci%20for%20EFL%20Students%20Schunn%20Wu%202019.pdf)

Sun, Y., & Cheng, L. (2014). Teachers' grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326–343. <https://doi.org/10.1080/0969594X.2013.768207>
[Crossref]

Xie, Q., & Tan, S. Y. (2019). Preparing primary English teachers in Hong Kong: Focusing on language assessment literacy. *Journal of AsiaTEFL*, 16(2), 653–673. <https://doi.org/10.18823/asiatefl.2019.16.2.14.653>

6. The Case of Ryan

Ling Gan¹✉ and Ricky Lam²✉

- (1) School of Languages and Communication, Beijing Technology and Business University, Beijing, China
- (2) Department of Education and Psychology, Hong Kong Baptist University, Hong Kong, China

✉ **Ling Gan**

Email: linggan@btbu.edu.cn

✉ **Ricky Lam (Corresponding author)**

Email: rickylam@hkbu.edu.hk

6.1 Overview

This chapter focuses on how Ryan enhanced his LAL, constructed his assessment identity and overcame difficulties over three semesters. As an experienced English instructor, Ryan received considerable assessment education and training which empowered him with sufficient assessment knowledge. This knowledge, in turn, enabled him to effectively use diverse assessment methods in teaching practices. For instance, he routinised quizzes as a diagnostic tool and conducted group peer assessment on writing by providing detailed criteria. Through these practices, Ryan extended his knowledge about assessment standards. He insisted on giving informative feedback which, though, was delayed due to many non-teaching commitments. Students' positive feedback also pushed him to adhere to fair grading without being influenced by non-achievement factors. Over the semesters, participation in teaching contests heightened his awareness of integrating assessment with teaching practices. Despite these

practices, Ryan did not recognise his identity as an assessor but claimed himself to be an assessment designer and facilitator.

6.2 Background Information

In his mid-30s, Ryan had been working as a university English instructor at NU, a key normal university, for eight years when joining the current study. He graduated from NU as an English major educated to be an English teacher. From the third year on as an undergraduate student, Ryan worked as a part-time lecturer in a well-known English training institute, helping students prepare for the high-stakes exams like *Zhongkao* (an examination for junior high school graduation in China) and *Gaokao*. At the same time, he found another part-time job to teach students English speaking skills in a college. He kept these two part-time jobs until he earned enough money to pursue an MA programme of education research in the UK.

Having earned the MA degree, Ryan decided to apply for a PhD programme. When preparing for the application, he was told that his alma mater, NU, was recruiting university English instructors. Considering this recruitment was a once-in-a-lifetime opportunity, Ryan quit the PhD application and went back to China to become a university English instructor.

Doctoral study was Ryan's another adventure. In the past eight years, Ryan married and had a child. In the meantime, he made an important decision to pursue his doctoral degree. In the fourth year of working in NU, Ryan applied for an EdD programme at a Sino-British university. Influenced by the enthusiasm of his MA supervisor in research, Ryan showed great interest in doing research while staying in the UK. This interest could perfectly continue in the EdD programme which provided him with a possibility to look after his family as well.

Ryan teaches College English for non-English majors at NU, which has adopted a level-based teaching in this English course. Students are placed into elementary, intermediate, and advanced levels according to their ranks by an in-house placement test. While the elementary-level students need to learn General College English for two years, the intermediate-level students learn three-semester general English and one-semester courses relevant to English for Specific or Academic

Purposes. The advanced-level students have only one-semester general English and three-semester ESP or EAP courses. In the previous two semesters of my study, Ryan was in charge of both intermediate and advanced classes, teaching general English with a focus on reading and writing. The class I observed is an advanced-level class in the first semester and an intermediate-level one in the second and the third semesters.

6.3 Prior Assessment Experiences

Ryan was assessed differently at the primary/secondary and tertiary levels. Like other participants, his early education was dominated by pen-and-pencil tests to prepare for high-stakes exams such as *Zhongkao* and *Gaokao*, with essays often marked without feedback. In contrast, assessment methods in his university were more diverse, including traditional tests and methods such as individual/group presentations, mini-theses, journal reflections, term papers, etc. Alternative assessments like self- and peer assessment were also commonly used.

Two courses left a strong impression on Ryan with their use of alternative assessments. The course *language pedagogy* required students to write weekly journals and give comments to at least one peer's journal entry. "Writing journals is a kind of self-assessment and reflection while giving comments to others' journals can be peer assessment. It is a good combination of self- and peer assessment" (Ryan-IT1, pos.40). Ryan was influenced by self-assessment in a lexicography course where students were given a checklist to self-assess their written outputs before submitting to the instructor for grading and feedback.

Ryan viewed his assessment experiences positively, valuing the use of varied methods at different educational levels. He emphasised the importance of considering students' learning competence when selecting assessment methods. He narrated, "assessing them [students] in various ways lends itself to motivating students' learning autonomy and developing their learning competence" (Ryan-NF1, pos.15). Like Emma, Ryan believed that the objective of assessment was to promote students' learning autonomy.

Studying in the UK exposed Ryan to a different assessment experience. The education experience empowered him to equip with knowledge of formative assessment, as he recounted,

... I have learned assessment concepts when I was doing my master's in the UK. I wrote a term paper concerning school success. This field covered teachers' competencies, one component of which was teachers' assessment skills. The instructors in the UK also assessed me through formative assessment. Before that, I didn't know formative assessment, because you know, in our country, no formative assessment existed in primary or secondary schools. Besides, even the assessment course I took during the undergraduate period did not cover formative assessment at all ... (Ryan-IT1, pos.36&37)

This excerpt suggests three key insights regarding Ryan's assessment background. First, his undergraduate assessment course failed to help him build up a necessary assessment knowledge base. Second, his formative assessment experience in the UK strengthened his understanding of this assessment method. Third, compared with assessment experiences at the primary/secondary level, his tertiary-level experiences had a more profound impact on his assessment practices as an English instructor compared to those at the primary and secondary levels.

Pursuing an EdD also significantly impacted his teaching and assessment practices. The EdD programme accessed him to "a lot of materials from which I [he] found theories concerning learning input and output" (Ryan-IT1, pos.39), gearing his attention to the theoretical underpinnings of English teaching, learning, and assessment. He was conscious to justify his teaching and assessment procedures by using theories learned through extensive reading. This theoretical knowledge enabled him to rationalise his teaching and assessment procedures, boosting his confidence in implementing flexible approaches.

Apart from assessment courses, Ryan participated in two formal assessment training programmes, both focusing on item writing. Ryan was selected as an item writer for CET6 reading tasks by a language publisher in China, well known for providing a variety of language

training to university English instructors. In this project, Ryan was offered a series of training opportunities, which enabled him to acquire theoretical underpinnings of assessment practices and raised his great interest in language assessment. However, he regretted that the training concentrated mainly on item writing, neglecting formative assessment in the classroom.

Ryan also leveraged academic lectures and reading as alternative venues for learning about language assessment. A lecture on writing assessment enhanced his understanding of formative assessment for student writing, while assessment literature provided him with theoretical insights into the relationship between teaching, learning, and assessment.

... Books helped me shift from understanding nothing about assessment to distinguishing formative and summative assessment. I realise that summative scores are not the end of teaching and learning. Instead, the formative process is to bolster students' learning, and teaching is aimed at promoting learning through assessment. (Ryan-NF1, pos.21)

Overall, Ryan's assessment experience at different educational levels, assessment training, academic lectures, and book reading empowered him with high-level LAL initially. He had a good grip on conceptual knowledge of language assessment and familiarity with alternative assessment methods and implementation of formative and summative assessment. In addition, he had self-evident awareness of the purposes and roles of assessment in language teaching and learning. The sufficient knowledge authorised Ryan to be an experienced assessment practitioner.

6.4 An Experienced Assessment Practitioner

6.4.1 Understanding Assessment Requirements

While Ryan's prior assessment experiences provided him with a solid foundation in language assessment, the requirements set by his department facilitated the smooth implementation of assessment tasks. Like many departments, Ryan's emphasises a combination of formative

and summative assessment, prioritising formative assessment (60%) over summative assessment (40%). The department also specifies each assessment task, ensuring that instructors know how to conduct them, as detailed in Table [6.1](#).

Table 6.1 Assessment Requirements for College English by Ryan's Department

Assessment tasks	Weight (%)	Descriptions of the tasks	Guidelines for the assessment
Class attendance	10	Students are required to attend every class on time.	1. Personal and sick leave with a stamped leave note: 1 point deducted. 2. Personal and sick leave without a stamped leave note (two points deducted). 3. Leave without reasons (five points deducted). <i>Note:</i> Those who are absent from classes without reasons for three times will not be allowed to take the final exam.
Tasks on HaoCe (a reading app)	10	Students need to complete the reading of an English book in the app and submit a book report with no less than 150 words. They also need to start at least one discussion based on the reading topic and involve themselves at least three times in other students' discussions.	Instructors need to set the grading for HaoCe learning as: The final score = learning progress (70%) + discussion (15%) and book report (15%).
Mid-term test	20	This is an in-class test scheduled by the department.	Some instructors are specified to write items for the mid-term exam.
Quizzes	10	Quizzes can include dictations, translations, or vocabulary tests.	Instructors can decide and implement their own quizzes.
Assignments	20	Assignments include the unit tests in the textbook or translation exercises aligned with the teaching content.	Instructors should assign and check the homework 2–3 times each semester.
Module assessment	20	It refers to the assessment on the content after one module (the same theme) is completed.	Narrative, report, and argumentative writing are covered in the first semester. The previous two should be completed as a group and the latter one individually.

Assessment tasks	Weight (%)	Descriptions of the tasks	Guidelines for the assessment
Participation	10	Participation refers to students' involvement and engagement in the class, such as students' responses to questions, individual, or group presentation	Instructors can decide the way to mark students' participation on their own.

Ryan outlined all assessment requirements for students in the first class of the semester and specified his own tasks accordingly. For instance, he required students to take one vocabulary quiz once every two weeks and complete the translation of three sentences from Chinese to English every week.

Ryan highly commended these assessment requirements for two reasons. Firstly, they ensured fair assessments by guiding instructors to evaluate students consistently, thereby minimising score gaps among students at the same level so that “fairness will not become an issue in making various decisions such as granting scholarships” (Ryan-IT1, pos.68).

Secondly, these requirements could prevent instructors from perfunctory assessment. Ryan believed they directed instructors to assess based on criteria rather than cater to student expectations for higher scores. This structure encouraged instructors to take assessment seriously.

... Some instructors will give students very high scores to make them happy so as to get good evaluations of instructors at the end of the semester. Students love high scores certainly. Yet, this is not the authentic assessment on students' language performance ... (Ryan-IT1, pos.69)

6.4.2 Optimising the Use of Quizzes

Ryan recognised the crucial role of language assessment in College English teaching and learning. He believed that language assessment served two major functions: diagnosing students' learning weaknesses and helping adjust teaching. Assessment results allowed students to identify their deficiencies, monitor their progress, and focus on areas for improvement. Similarly, teachers can modify their teaching methods

to better address students' needs. These perspectives prompted Ryan to refine assessment activities to foster learning and student autonomy and to advocate for the use of quizzes as diagnostic tools.

Ryan regularly used quizzes for two reasons. First, positive experiences of being assessed with quizzes during his undergraduate studies encouraged him to utilise the same assessment method for his students. He expected that quizzes could enhance students' learning as they did to his own learning. "Without quizzes, I will not take learning seriously. With them, I will work hard to improve my learning" (Ryan-IT1, pos.43). Second, the shift to online classes during the COVID-19 pandemic pushed Ryan to increase quiz frequency to draw students' attention, because "it is more effective to motivate and urge them [students] to accumulate words on the daily basis rather than cram words at the end of the semester" (Ryan-IT1, para17).

The quizzes, being vocabulary-related, serve dual purposes. Firstly, they are summative and diagnostic in nature as they are administrated at the end of each unit to assess what students have learned. Secondly, they serve a formative purpose, as they provide learning opportunities for students when embedded into daily instructions.

To enhance the effectiveness of these quizzes, Ryan implemented a punishment-for-wrong-words strategy, tailoring quiz requirements to students' English proficiency levels. As detailed in Sect. [6.2](#), students at Ryan's university were categorised into three levels, leading to different assessment benchmarks for the quizzes: 80% for advanced, 70% for intermediate, and 60% for elementary classes. Failure to meet these targets resulted in the requirement to write a story by using the mistaken words.

The quizzes and punishment worked well for students' learning in Ryan' view. He was pleased to see students' advancement in their vocabulary learning.

... Students are consciously using the words covered in the quizzes. For instance, in a writing assignment, some students utilised the words "meticulous" or "meticulously" ... (Ryan-IT1, pos.19)

Writing a story as a punishment was not Ryan's intention. Instead, he expected students to understand and remember how the words were used contextually. He also expected the quizzes could direct students to expand their vocabulary skills daily.

Ryan's students also expressed appreciation for the vocabulary quizzes, stating that "quizzes are effective" and served as a "spur" to push them to memorise words (Ryan-student-FGI-2, pos.40&47). In addition, as Ryan would "recap quizzes and correct errors in the use of words", students would "pay special attention to the erroneous words" (Ryan-student-FGI-2, pos.54). Students took the vocabulary quizzes as a tool to motivate and monitor their learning, just as they did in the high school where learning was jammed and closely tracked. They assumed that the quizzes spurred them to acquire new vocabulary, an indicator of learning progress rather than a waste of time.

Students' positive feedback and Ryan's reflections on the quizzes' advantages encouraged him to use this assessment method throughout the three semesters. He continuously adjusted the types of quizzes, items, and content to relate to current social events. For instance, "In the spring of 2020, a lot of vocabulary items were COVID-19 related, covering the issues of Chinese government's attitudes towards the pandemic and China's measurements against COVID-19" (Ryan-IT5, pos.66).

6.4.3 Conducting Group Peer Assessment on Writing Essays

In addition to quizzes, Ryan regularly conducted peer assessment on writing. Impressed by peer assessment used in his MA and EdD programmes, Ryan adopted and modified this method in his class. His experiences as both a student and instructor enriched his ability to efficiently implement and enhance assessment practices.

The dual identity of being an EdD student and an English instructor prompted me to consider how to effectively assess students in group work. When I am an EdD student, I consider if the activities my instructors used on me can be used on my own students. These activities will be evaluated from my

perspectives of both a student and an instructor. (Ryan-IT2, pos.3)

Ryan believed that assessment should focus on writing outputs, asserting that reading serves as a catalyst for writing. He initiated a “synergy of reading-to-write and assessment” practice across various writing genres. The following observation episode showcased how he embedded group peer assessment into a narrative writing task.

Ryan played a video on how to describe and narrate with five senses after students read a text of narrative genre in the textbook.

Ryan: Okay, that’s the video. And you can try to use the words in the appearance description, Okay? Right. So, the next session of this lesson, we’re going to, first of all, talk something about how to write details, in a narrative task, and then you are going to do the peer evaluation. You all have got the writings and also the rubrics as well as the evaluation form.

In the following 10 minutes, Ryan instructed students how to write a narrative in detail by presenting and critiquing the problems in narrative samples of previous students.

Ryan: Okay, that’s all about five senses and I hope you can make for yourself the five senses to make your story more descriptive and now it’s time for you to do the peer evaluation. And you have got your other groups’ writing, right? You also got two papers here, right? One is the rubric. It is an assessment rubric. You can see ABCD and different types of things you need to evaluate, ok? You can also go to the QQ group to download this as well to read that more carefully. And this one is what you need to fill in. And after class, you need to give it to me. Let’s see whether we have enough time. If we do, you should give it to me directly and then you need to read ... write down the number written by the group. So, this “written by” means... eh... it’s not your own group, okay? So, it’s the story, which group wrote the story and wrote down the number. And then, you discuss

because I have detailed descriptions of the requirements here. Okay. First of all, you read the story definitely and then you can have a discussion. All right? Now if you have any questions, you can let me know. Now, please.

Six groups with five to six students for each group began to read together the narrative story written by another group.

Ryan (Walking among groups and looking around): You can actually take some notes or make some comments just directly on the paper if you think it necessary. And then, once you finish reading that, you can start to discuss it. I think you should be critical according to the criteria.

About six minutes later, students began to discuss with each other while Ryan walked through different groups and stopped to listen for a while in each group, offering his comments and discussing with students who needed help.

(Ryan-F07)

Several inferences regarding group peer assessment could be made. First, students learned to write narrative essays with a “show, not tell” technique which enabled them to critically assess their peers’ work. Second, essays were graded using an assessment rubric that guided students to evaluate narrative components step by step. The criteria in the rubric helped students grasp key elements of the narrative genre. Third, Ryan enhanced collaborative learning and improved students’ narrative writing skills through peer assessment.

To make the group assessment function well, Ryan made two attempts. First, he read books on narrative writing to identify essential elements and understand how to integrate them into writing. He selected key elements and incorporated them into the assessment rubric. Before designing the rubric, he also reviewed “Standards for the Assessment of Reading and Writing”, which provided insights into various assessment methods for L2 reading and writing. Second, Ryan developed a clear rubric for students to evaluate their peers’ writing, detailing assessment components and grading criteria (see Fig. [6.1](#)).

This rubric allowed students to “evaluate narrative essays confidently” (Ryan-IT1, pos.28). Feedback collected from peers (see Fig. 6.2) indicated that students provided specific comments on their group’s narratives, identifying strengths and weaknesses while suggesting improvements for each component.

<u>Narrative Writing Assessment Rubric</u>				
Item	A	B	C	D
Theme	It is clearly about helping the disadvantaged (the disabled, the elderly or other disadvantaged people) from the first-person point of view. The narrator is a college student.	It is not clear enough whether it's about helping the disadvantaged and it is narrated from the first-person point of view. The narrator is a college student.	It is unclear whether it's about helping the disadvantaged and it is not narrated from the first-person point of view. The narrator isn't a college student.	It is clearly not about helping the disadvantaged and it is not narrated from the first-person point of view. The narrator isn't a college student.
Structure	The structure “orientation-problem-solution” is clearly identified.	The structure “orientation-problem-solution” can be identified but with minor confusion.	The structure “orientation-problem-solution” cannot be identified easily.	The structure “orientation-problem-solution” cannot be identified at all.
Descriptive and Cohesive Language	The language in the narrative writing is properly descriptive and cohesive by using all the ways learnt.	The language in the narrative writing is moderately descriptive and cohesive by using some of the ways learnt.	The language in the narrative writing is not descriptive and cohesive enough.	The language in the narrative writing is not descriptive and cohesive at all.
Character Development	All the four ways of developing characters are properly applied.	All the four ways of developing characters are applied but not properly enough.	Not all the four ways of developing characters are applied.	None of the four ways of developing characters is applied.
Grammar and Spelling	There are no or few mistakes.	There are some mistakes without causing confusion.	There are many mistakes which causes confusion.	There are too many mistakes.
Word Count	800-900 words or slightly more	700-800 words	600-700 words	less than 600 words
Role-play (Video)	All the group members <u>play</u> and it is done in a vivid way.	All the group members play but it is not done vividly enough.	Only half of the group members play.	Only one third of the group members play.

Fig. 6.1 A Screenshot of Ryan’s Narrative Writing Assessment Rubric for Narrative Writing

Written by Group 3

Narrative Writing Peer Evaluation Form

Item	Grade	Comments (Mainly about suggestions on how to improve)
Theme	C	The narrator isn't a college student. The selected theme is positive but not clear. It is suggested to slightly delete the previous lengthy background to highlight the theme.
Structure	B	The structure is relatively clear, but the connection in some places is not smooth. For example, after Guo's parent nodded, "I" asked Guo which university he wanted to test.
Descriptive and Cohesive Language	B	There is less descriptive language, mostly in the first page, and most of the content is directly been told, instead of describing characters' facial expressions or movements. It's a bit hard to let readers continue reading with excitement and curiosity.
Character Development	C	We think the Character development is a little baffling. And the human setting is too planar and routine (平面和单调). We think you should describe more details to shape the characters.
Grammar and Spelling	B ⁻	There were a little spelling and grammatical mistakes which confused us, but overall they didn't affect reading. Hope you could check it more carefully and there are a lot of careless grammatical mistakes.
Word Count	A	

Fig. 6.2 A Sample of Peer Feedback Sheet

The development and implementation of assessment rubrics were heavily influenced by Ryan's educational and test-taking experiences. His studies in the UK, ongoing doctoral work at a Sino-UK university, and IELTS experience familiarised him with various assessment criteria and rubrics. Specific assessment criteria helped students establish visible and measurable learning goals.

... When I studied abroad in the UK, I was told how to obtain a score range, say, from 60 to 70. The doctoral courses I am taking currently also provide detailed descriptions for different levels in the course assessment ... So, I know that a rubric should contain clear criteria so that students understand what they should do to achieve the goal ... The similar rubric can also be found in IELTS writing. Because I have taken IELTS myself, I would instruct my students, by referring to the assessment criteria in IELTS writing ... I think it's very good practice [for writing instruction]. (Ryan-IT1, pos.31-33)

Ryan was satisfied with the effectiveness of group peer assessment in writing. First, students were highly engaged in this assessment activity through collaborative reading, discussing, exploring, and sharing. Second, students were enabled to appreciate and generate insightful peer feedback based on the specific assessment rubric. Third, students' written essays were improved through two rounds of revision according to peer and teacher feedback. In the process of group peer assessment, students achieved the development of their feedback literacy (Carless & Boud, [2018](#)).

... I noticed students in my class discuss heatedly with each other. I think they know how to make their stories achieve a higher level through retrospection. If I simply show them the rubric, I am afraid they won't read the criteria carefully. To the opposite, if I ask them to discuss in a group, they have to read the criteria one by one ... I also observe that while discussing, students are quite attentive. They write down clear comments in the assessment sheet ... They consider revising their writing even without getting feedback from peers, as I hear some students saying they also have the same problems in their writing when assessing their peers. You see, they know how to reflect on their writing when assessing peers. (Ryan-IT5, pos.73)

Ryan's students also benefited from group peer assessment in the writing class. They all believed assessing peers' writing products helped them "learn about classmates' ideas or good writing" (Ryan-student-FGI-3, pos.29) and to "identify their [peers'] problems in writing and, in turn, reflect on problems in my [their] own writing and think about how to avoid the problems" (Ditto, pos.30). They favoured this assessment practice, as it provided more "comprehensive feedback from different perspectives" (Ditto, pos.32) for their essays, and, thus, their writing could improve from these meaningful and diverse peer feedback.

Throughout the three semesters, Ryan consistently implemented reading-to-write activities aligned with departmental assessment requirements. He maintained this method for three reasons. Firstly, he aimed to test the effectiveness of group peer assessment across

different writing genres and student groups. His enhanced teaching experience equipped him to implement and trial assessment methods, as he noted, “with my increasing teaching experience, I expect to offer more help to students through this assessment method and to examine if this method is effective or not” (Ryan-IT4, pos.9). Secondly, positive student evaluations of his assessment methods boosted Ryan’s confidence. He shared, “Students have given me the highest score on my assessment practices since becoming an instructor” (Ryan-IT4, pos.11), reflecting the rewarding nature of his commitment to group peer assessment. Lastly, Ryan regularly reflected on his assessment practices to improve them. This reflective approach involved analysing assessment challenges, developing solutions, and applying them in subsequent practices. He regarded group peer assessment as a successful initiative but recognised the need for more time and opportunities for improvement.

6.4.4 Giving Feedback

Peer feedback is a vital tool for enhancing student writing. In the group peer assessment outlined in Sect. [6.4.3](#), Ryan effectively designed an assessment task that encouraged students to provide feedback on each other’s essays. He supported this process with a detailed rubric and oral comments, facilitating students’ learning with meaningful peer interactions.

Apart from peer feedback, Ryan provided his own critiques to students’ writing. His feedback (see Fig. [6.3](#)) adhered to several recommended principles. First, he pointed out weaknesses in students’ writing and acknowledged their strengths, offering emotional support. Second, Ryan’s feedback had coverage of writing components, including language, theme, content and structure, ensuring a well-rounded approach (Ferris, [2014](#)). Third, Ryan’s feedback was specific and clear, utilising softeners like “I suggest..., I think..., Maybe...” to avoid potential embarrassment. He also posed engaging questions such as “Is being remembered your purpose...?” to foster cognitive engagement and promote student autonomy (Ferris, [2014](#)). All in all, Ryan demonstrated a high level of LAL in designing, enabling and giving feedback.

Home visit for girls' education

This it was my second year teaching in a remote rural middle school. As a volunteer teacher who was born in urban area and still studied in the university, I was shocked by the learning environment of the students in remote areas. It was normal common for them to walk for two hours to go to school. And the most important thing was, because of the steep mountain roads, several pitiful students suffered accidents on the way to school. Many families couldn't afford the cost of the treatment and we didn't have some feasible measures to support them to go get them back to school. More often than not, they still chose to drop out.

Feeling worried about the situation, I made up my mind to help them. As far as I knew, the majority of the disabled students couldn't receive quality education and necessary care. Because their families believed it was useless for them to receive advanced education and also financial problems discouraged them from being educated. Although I knew my own strength was tiny, I still wanted to help them to go to school. If you ask me why I must do this, I will answer you that I want to help poor families shed the wrong perception that "education is useless". I hold the firm belief that every kid has the right to receive education equally, despite they are disabled people.

I have done much about it. For example, I once begged in the street for gaining some money from kind people. However, someone thought I was a liar and didn't want to assist me. But I didn't give up. I began to launch my wishes on the newspapers and asked help from local ministry of education. Although there was were many difficulties, ultimately, after a lot of perseverance, I finally raised funds for supporting the disabled students to go to school.

I think you need to change a title, as you have changed your story a lot.
24 October 2021, 18:42

Reply

I think you need to give more rationale for this. Why did you feel worried about the situation? Was it related to your own learning experience

Reply

Are you going to just focus on disabled students? It seems to me from the title that you are going to talk about girls' education.

Reply

How did you manage to do this?

Did you see her at her own home or at school? I was actually a little confused. In the later part, you mention that you went to her home and visit

Reply

What happened during this period of time?

Reply

You can make this part shorter, as it is not so important in this story.
24 October 2021, 18:22

Reply

Maybe Miss XX is better.

Reply

At the beginning I taught in this rural middle school, a little girl attracted my attention. Her name is Guo Tingting, who had petite figure with coarse brown hair and dark skin. Every time I met her, she always wore scanty clothes, staring at me with her big and glistening eyes. She was so shy and self-conscious that she seldom spoke in my class, but she was always conscientious with her big eyes which implied her eagerness to learn.

But the happy time didn't last long, she disappeared was then away from school for a period of time. It was that her parents came to ask for her leave. When I saw her again, she was on the walking stick. After that, she became more silent, less willing to communicate with others and often walked alone.

After a long time being together, the girl who was not good at communicating with others opened up her mind at last and talked about her story of disability.

That was a rainstorm and gloomy Tuesday with raindrops falling from the sky and the thunder rolled and deafened the ears. To arrive at school on time, she rushed to the

downpour without wearing a raincoat. Unluckily, because of the rough mountain road and the continuous storm, the road was undermined heavily. When she struggled to school, a huge stone suddenly rolled down. Although she escaped quickly, she didn't get rid of it and the stone pressed against her foot. She sweated with pain. Her face was blue around the gills. And she almost fainted. She tried her best to push the stone and cried helplessly. Because of the bad weather, there were almost no passers-by. After a long time, there was a villager from the same village passed by and sent her to the nearest hospital immediately.

Luckily, her legs needn't to be cut off. But there was a lifelong disability and she had to walk by walking stick for the rest of her life.

When she finished that, she smiled to me helplessly, which made me sorrowful from the bottom of my heart. I stared at her with my bloodshot eyes filled with sorrow and misery. Suddenly, she turned to me, with her big eyes still clear and full of desire and murmured, "teacher, please, I want to study as usual, I don't want to drop out." Without getting my response, she left away instantly. I stared at her thin body walking on wooden crutches and making a "twisted" sound on the ground. The slight wheezing from her every step was just like a dull bell which beat my heart heavily.

Fig. 6.3 A Sample of Ryan's Written Feedback

Despite the relief provided by peer assessment, Ryan found it challenging to give timely feedback to students' written essays due to his heavy teaching workload.

I have six classes this semester, two for advanced levels and four for developing levels...Every class has six groups and then there are 12 poem writing scripts and 24 stories for narrative writing ... I don't have much time to write specific comments and return them back to students quickly. (Ryan-IT2, pos.88)

Ryan also struggled to give timely feedback due to his heavy non-teaching responsibilities. In the first semester, he was tasked with training students for national contests in English public speaking, reading, and writing. In the following semester, he prepared for a provincial teaching contest on behalf of his university, while working on his doctoral research proposal. These commitments left him with little time and energy.

Nonetheless, Ryan managed to solve the problems of giving feedback in several effective ways. He granted students the autonomy to submit their essays early for detailed feedback. Additionally, he utilised winter and summer holidays to review student essays, promising to return them with comprehensive feedback in the following semester. He noted, "I know this is not effective. Anyway, the feedback is detailed, though delayed" (Ryan-IT2, pos.88). Ryan acknowledged that the feedback, though delayed, provided sufficient information for students to improve their writing.

While Ryan found ways to manage feedback challenges, he acknowledged a gap in his knowledge of language testing and assessment. He had not developed a systematic understanding of language testing and assessment which left him without a solid theoretical foundation to support his classroom assessment practices.

I did not systematically learn language assessment, the same way I had learned language pedagogy in which I learned and practised. Although I took the course of language assessment, I just learned item writing skills instead of classroom-based assessment skills and methods ... Probably time is a problem. I

don't have time to learn language assessment sustainably. So, I think I am deficient in sort of theoretical stuff. When I conduct assessment activities such as quizzes, I evaluate their effectiveness based on my experience instead of theoretical underpinning. (Ryan-IT2, pos.94)

6.4.5 Grading

Although giving timely feedback posed a challenge, Ryan was very confident in and satisfied with his grading of students' performance on the seven assessment tasks mentioned in Sect. [6.4.1](#) each semester.

Personally, I think I grade students fairly. Students' final exam is marked by several instructors based on the criteria. I may not mark my own students' test papers at all. Therefore, in this part, the grading is both objective and fair..., I am sure my grading is unbiased, either. Even though some students may have a good relationship with me, I won't give them a higher score, if they do poorly in every task. I have made this point clear on the first day of the semester ... (Ryan-IT2, pos.146)

Ryan firmly believed that his grading method and principle could ensure fairness in both formative and summative assessment and that his grading would not be influenced by his relationship with students.

A record sheet (see Fig. [6.4](#)) of students' performance on assignments could also guarantee objectivity in his grading. As seen from the sheet, a clearly defined grading scheme was used to assess students' performance on assignments in each unit. Students could also earn a bonus by volunteering to do some extra assignments. Based on these grades, Ryan would give a holistic grade to students in the "assignments" task specified in Table [6.1](#). Regarding his grading, Ryan recounted,

I try my best to be fair and unbiased by using grades to mirror students' performances ... I will reward with a high score those students who make great efforts to perform well in English learning. Those who are poor in English but reluctant to make their efforts will surely get low grades. The department has set a

baseline grade on assessment. If students perform well, they will get a baseline score of 80. If students perform better, they get 90 or above. (Ryan-IT4, pos. 57&59)

This excerpt implied that students' performance was prioritised in Ryan's grading while students' effort was an important non-performance factor Ryan considered. This non-performance factor was not taken as an index for grade inflation. Instead, it was used as a bonus to spur students to learn harder and better.

姓名	平时成绩										期中	期末	总评	备注
	U1	Speech	U1	U1	U3	U3	U3	U3	U3	U3				
013	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
16	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
08	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
16	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
013	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
17	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
011	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
1+1+1+1	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
15	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
15	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
17	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
20	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
16	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
012	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
013	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
15	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
16	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
013	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
16	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
15	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
19	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
14	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
17	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A
012	✓	✓	A	✓	A	A	A	A	A	A	A	A	A	A

Students' Performance on Assessment Tasks												
Mid-term exam	U1 (Unit 1)	Speech	U1 B1	U1 B2	U3 A	U3 B	Mid-term	U4 A	U4 B	U6 A	U6 B	Recitation
013	✓	✓	A	✓	A	A	B	A	A	A	A	A ⁺ A ⁺ A ⁺
16	✓	✓	A	✓	A	A	A	A	A	A	A	
08	✓	✓	A	✓	A	A	A	A	A	A	A	
16	✓	✓	A	✓	A	A	A	A	A	A	A	A ⁺ AAAA
013	✓	✓	A	✓	A	A	A	A	A	A	A	AA
17	✓	✓	A	✓	A	A	A	A	A	A	A	
011	✓	✓	A	✓	A	A	A	A	A	A	A	
1+1+1+1	✓	✓	B	✓	A	A	A	A	A	B	A	AAA
15	✓	✓	A	✓	A	A	A	A	A	A	A	A ⁺ A
15	✓	✓	A	✓	A	A	A	A	A	A	A	A ⁺ A
17	✓	✓	A	✓	A	A	A	A	A	A	A	
20	✓	✓	B	✓	A	A	A	B	A	B	B	
16	✓	✓	B	✓	B	B	A	B	B	B	B	
012	✓	✓	A	✓	A	A	A	A	A	A	A	
013	✓	✓	A	✓	X	A	X	A	A	A	B	
15	✓	✓	A	✓	A	A	A	A	A	A	A	
16	✓	✓	A	✓	A	A	B	A	A	A	A	
013	✓	✓	A	✓	A	A	B	A	A	A		
16	✓	✓	A	✓	A	A	A	A	A	A	A	
15	✓	✓	B	✓	B	B	A	A	A	A	A	
19	✓	✓	A	✓	A	A	A	A	A	A	A	A ⁺ A ⁺ A
14	✓	✓	A	✓	A	A	A	A	A	A	A	
17	✓	✓	A	✓	A	A	A	A	A	A	B	
012	✓	✓	A	✓	A	A	B	A	A	A	A	

Fig. 6.4 A Record Sheet of Students' Performance on Assessment Tasks

Apart from the grading sheet, the university and department assessment policies also ensured fairness in Ryan's grading.

... Each teaching and research section of our department will unify an average score for the assessment of the whole class. It is not allowed that the average score in a class is very high or low. Otherwise, it is unfair to students from different departments and biased toward students from the same department which makes decisions about students' scholarships according to their scores of College English ...

Our university ... requires that scores in College English cannot be inflated ... If the number of students who get 90 or above is beyond the required proportion, the instructors have to provide supporting documents to explain... (Ryan-IT1, pos.66&70)

Seemingly, both Ryan's university and department had made efforts to rule out grade inflation currently common in higher education (Tucker & Courts, [2010](#)), to guarantee unbiased decisions with students' scores.

Ryan embraced the efforts the university and department made in grading. He explained to students the purposes of these efforts in the first class of each semester. That is, the university expects the assigned scores are able to represent students' authentic English proficiency and that students should make efforts to obtain satisfying scores. In this sense, students will become more competitive at home and abroad. "This is an effective way at the institutional level to protect those who learn hard" (Ryan-IT1, pos.71).

With respect to Ryan's grading, his students gave very positive responses. A great majority of students considered that Ryan's grading was "fair", "rational", "justifiable", and "unbiased" because of his "detailed assessment rubrics" and "clearly documenting learning performance and progress of every student" (Ryan-student-FG3, pos.105-112).

6.4.6 Preparing Students for CET 4/6

While completing required teaching tasks, in Ryan's department, instructors are expected to prepare students for CET4/6 through in-class lectures when the test approaches. The lectures can, "relieve students, especially first-year students, of their anxiety and worries about tests" (Ryan-POC7, pos.12). Moreover, the lectures can help students raise their CET4/6 scores, which are essential in critical decisions on various admissions. For instance, "scholarship is closely related to the CET4/6 scores, as a minimum score is set for the scholarship application ... CET4/6 scores are decisive if students expect to be recommended to become a postgraduate in higher-ranking universities" (Ryan-IT3, pos.23).

CET4/6 preparation in Ryan's class was translation-focused. Ryan maintained that Chinese students were good at receptive skills but not

productive skills, particularly in translation and writing, two important “indicators of students’ English proficiency” (Ryan-IT2, pos.139). Therefore, he placed heavy weight on translation, which he believed was conducive to writing as well. Accordingly, he regularly lectured three translation sentences from Chinese to English every week. The following was an episode of translation lectures.

Ryan: Now, let’s look at the third sentence “他們也用手機購物，查找信息，因為手機便於攜帶” (Ryan pointed to the sentence shown on the screen). What version do you think **I expect you** [emphasis added] to come up with? (Ryan walked towards students, smiling) (Students laughed). Of course, you can also use the words “because, so” to show the logic of this sentence. They are fine. However, it will be better if you can flexibly handle this sentence rather than translate immediately **when you feel excited to see the word “because”** [emphasis added] (Students burst into laughter). Another thing, you will use the verbal phrases to translate this part “他們也用手機購物，查找信息”，won’t you? Can we use the noun phrases? As I mentioned previously, a verb indicates dynamism, but we can also change the dynamism into a static state. This exchange can make your sentences stand out of all the test takers. The uniqueness of sentences can illustrate your competence in using English. This will add extra points to your translation or writing.

Okay. Here we don’t use “because” directly and the second part of this sentence can be nominalised. Then, the sentence can be translated like this: they also use mobile phones to shop online and search for information owing to the facility mobile enjoys. See, we don’t need to put the second part of the sentence into sentences like “because mobile phones can be carried into...” We can nominalise it, right?

Now, we’ll see if we can use the passive voice in this sentence. We don’t say “they use”. Instead, we say “mobile phones are used/utilised”. In fact, this sentence was excerpted from a CET4 past test paper and was translated by me from English into

Chinese. If you locate this sentence from the test paper, you'll find that the sentence is logically structured ... It is also correct to use the structure like "shop online and search information", but it is more appropriate to use the parallel structure "make online purchases and conduct information search ..."

(Ryan-F01)

In this episode, Ryan provided detailed translation strategies. He presented commonly used sentence structures in translation and affirmed their correctness. Then, he offered alternative sentence structures that were more appropriate and native-like. He also explained why sentences translated in this way were better from the perspective of English syntax. Meanwhile, Ryan lectured in a humorous tone to relieve students' anxiety and make learning strategies understood easily.

Ryan justified his focus on CET4 translation. He believed that the strategies for translation were similar to those for writing, as both language skills required the appropriate use of words and sentences. Lectures on translation could be targeted at "hitting two birds with one stone".

They [students] will take CET4 soon. I attempt to make them understand what should be noted in translating. For instance, students are inclined to use simple words, phrases, and sentences ... Besides, I think lecturing on translation is helpful for their writing. For example, the variety of words and sentence structures used in translation can also be applied to students' writing ... (Ryan-POC4, pos.5)

Ryan's lectures on translation were spoken highly of by the students who were aware of the benefits of these lectures to both translation and writing in CET4.

ZYJ: ... Our English instructor not only pointed out our mistakes but also provided some wonderful and native words or phrases. He would offer different options for sentence structures as well ... all these are really useful.

LYR: I find that the sentence structures he [Ryan] proposes are truly useful. They can be used not only in translation but also in writing. If they are used well, they may help to add marks to our CET4.

Lumos: ... I have a friend who took *Gaokao* this year. I sent her all the materials related to translation our English instructor covered ... You know what? She gained 10 more points in her writing this year! She praised that [Ryan's] translation strategies were amazingly helpful.

(Ryan-student-FGI- 2, pos.112,114 &116)

6.5 Participating in a Teaching Contest

In view of his excellent performance in a previous national English teaching contest, Ryan was recommended by his department to take part in a provincial teaching contest on behalf of his university in the first semester.

...I participated in a teaching contest in 2019 and took it again this year [2020] ... Taking the contest requires a lot of preparation. Although the preparation is tough, I am determined to do well, as this is what I like to do...I like to and expect to advance my teaching, including assessment through this contest...A lot of my colleagues don't understand why I still take the teaching contest ... (laugh), because I think I can get more help from my veteran colleagues. When I present my teaching to my colleagues, they will offer their suggestions for my improvement... (Ryan-IT2, pos.48)

This excerpt indicates three inferences. First, Ryan had the intrinsic motivation to take the teaching contest, which seemed to be an institutional demand. He was fond of teaching from the bottom of his heart and enjoyed the liberty of adopting various pedagogical instructions. Second, Ryan aspired to improve his teaching practices through the contest. The preparation for the contest was a learning

process from which he could develop his teaching competence. Third, joining the contest was a collaborative task. Ryan would frequently present his teaching practices to his colleagues who would provide constructive feedback and comments. Collegial responses and peer feedback could contribute to the advancement of every aspect of his teaching, including assessment.

Preparation for the teaching contest was a long-lasting and time-consuming process. When informed that the teaching contest was postponed to the second semester, Ryan did not slow down his preparation. Instead, he focused more on his classroom teaching and took it as a venue to polish his teaching design. To meliorate and facilitate his teaching in the contest, Ryan wrote a reflective note (see Fig. [6.5](#)) to review his teaching design. As highlighted in this note, assessment was included and explained, though not specifically, which suggested that Ryan knew well how to align assessment methods with his teaching.

Teacher's Reflective Notes	
Item	Specific Reflection
Fostering Virtue Through Education	In this period, it lies in students' comprehension of the theme: be positive about our life (from the perspective of Stevie's personal traits) and be altruistic in our society (from the perspective of those who helped Stevie).
Teaching Techniques	1. The preview tasks are intended to help students to be more involved in the classroom learning. My past teaching experience shows that students have found the vocabulary exercise useful in learning key words and expressions. As for the comprehension questions, they contribute to students' preliminary understanding of the text so that deep analyses are possible during the classroom teaching. 2. Appropriate examples in our daily life can be useful in reducing students' potential difficulty in understanding the concept of "altruism" that can motivate us to care for people around us. 3. QR codes grant students easy access to learning resources by merely using their mobile phones, justifying my use of technology in teaching. 4. Additional resources relevant to disability and altruism are provided for students to develop their autonomy in (English) learning, which is of lifelong use.
Teacher's Role as a Facilitator	1. During students' completion of their preview tasks and after-class assignments, I should be ready to offer professional and practical guidance when it is needed. 2. During the pair work and group discussion, I will walk around the classroom to ensure students are doing what they should do and to provide help if necessary.
Assessment	Students' self-assessment (the preview tasks and assignments whose answers are provided for students' own assessment) and Teacher's assessment (students are asked to share their own knowledge and understanding via teacher-student interaction during the class as well as do writing practice after class) are combined to enhance students' learning outcomes subsequent to this period of learning.

Fig. 6.5 A Screenshot of Ryan's Reflective Note of Teaching Design

In the middle of the second semester, Ryan completed the contest and was awarded the third runner-up. Although this result was not what the university administrators expected, Ryan accepted it happily, because "all achievements are part of my professional development ... [from which] I have gained a lot" (Ryan-IT4, pos.47).

Taking the teaching contest lent itself to Ryan's professional development in both language assessment and teaching. It raised his more attention to integrating assessment into his teaching and "facilitated more exploration in language assessment practices" (Ryan-NF2, pos.17). When designing an assessment activity, he would "consider what role I [he] play[s] in assessment, if assessment

facilitates students' learning and in what way students can be assessed" (Ryan-IT2, pos.47).

The contest also provided an opportunity for Ryan to have communication with his experienced colleagues. As Ryan expressed the theoretical and practical need in language assessment, he asserted that while assessment theories were accessible by reading books, assessment practices could be acquired and advanced through consultations with colleagues who "have a lot of experience in language teaching and assessment". Collegial feedback and comments provided him with "a lot of inspirations", as "their perspectives are totally different and so unique" (Ryan-IT2, pos.115).

Additionally, the teaching contest had opened a new window for Ryan to share his experience with peers within and beyond his own university. The contest publicly showed his competence in English teaching, which qualified him for different opportunities, either being academic or pedagogic, as he narrated:

This contest is very important for my professional development. Without this experience, as an inexperienced teacher, I am not qualified to join any teams to apply for some kind of provincial or national courses ...The experience of taking the teaching contest also exposed me to publishers that would invite me to cooperate with them, say, sharing my teaching experiences with peers. I think, when sharing, I can obtain peer feedback, which, in turn, is an evaluation of my teaching. (Ryan-IT4, pos.51)

6.6 Perceived Assessment Identity

6.6.1 "I Don't Have a Strong Identity as an Assessor"

Initial assessment training, prior assessment experiences, and Ryan's efforts in assessment practices seemed not to reinforce his identity as an assessor. Although Ryan was observed to effectively use both traditional and alternative assessment methods, he did not regard himself as an assessor, stating that "being an assessor means we simply expect an outcome rather than the learning or assessment process" (Ryan-IT2, pos.187).

I don't have a strong identity as an assessor. The main purpose of assessment is to promote students' learning, not simply to give them a score or a grade... Assessment, in my view, is a tool [to promote learning] rather than a result. Therefore, I think of myself as an assessment designer and facilitator more than an assessor. (Ryan-IT1, pos.78)

Like Emma, Ryan linked the identity as an assessor with merely assigning scores or grades. This perception resided in the traditional assessment culture that assessment was considered an outcome represented by numbers or letters (Shrestha & Roffey, [2018](#)). Ryan maintained that this outcome-oriented assessment was not conducive to students' learning, as this perceived identity would focus instructors on assessment itself and assessment outcomes only. As such, he despised the identity as an assessor and placed greater value on the assessment process.

In the second semester, Ryan held the same stance towards his assessor identity, despite his strengthened awareness of embedding assessment with teaching. He recounted,

I don't have a new interpretation about the assessor identity. My point is consistent [that I don't identify myself as an assessor] from the beginning, because I view myself more as an instructor. Doing assessment is only one part of being an instructor. I cannot say I am simply a marker or a feedback giver. The role of being an instructor should not be confined to giving scores or feedback. I can also be a friend of students ... (Ryan-IT3, pos.47)

This excerpt showed that Ryan recognised assessment as part of teaching and learning yet saw himself more as an instructor than an assessor. Ryan also subsumed the assessor identity under that of the instructor, separating the two, by confining assessment to the roles of score or feedback giver.

Ryan took assessment as one minimal part of his teaching, in which he was an instructor, a friend, a designer, a facilitator and finally an assessor. "I can't take it [assessment] as my principal job. That's why I place my assessor identity as the last in the list" (Ryan-IT3, pos.48).

Despite not identifying himself as an assessor, Ryan was conscious of the standards required for this identity. He adhered to a principle of “being fair and square towards students, without being influenced by personal ties with students” (Ryan-IT5, pos.136). This principle was also welcomed and approbated by his students, as he narrated,

Recently, I had a dinner chat with some students who graduated a few years ago. They recalled that I was not that type of teacher who raised students’ scores because of the good personal relationship with them. And I said, “you got me!” (Ryan-IT5, pos.136)

As the study progressed, Ryan reiterated that he did not identify himself an assessor. He highlighted that being an assessor seemed to “focus more on the outcomes” but he was actually “involved throughout the assessment process”. “Students’ learning outcomes result from the instructors’ involvement” (Ryan-IT6, pos.35). Consequently, he preferred to be identified as a facilitator.

6.6.2 “I Am a Designer and a Facilitator As Well”

From the start, Ryan identified himself as an assessment designer. He believed an instructor should design both effective assessment activities and specific methods tailored for students. This identity was evident in his teaching, learning, and assessment approaches. To design effective assessments, Ryan emphasised building rapport with students, which he felt was crucial for developing and administering assessments. He aimed for students to see him not just as a grade-giver but as an emotional supporter, facilitating a positive learning experience through shared stories.

...Being friends with students, I can know what students think of assessment. Meanwhile, I share my own experiences and understanding with them. In this way, I’m able to design and develop teaching and assessment activities that relate the topics of the coursebook to students’ own lives and learning. So, students can become fonder of my teaching and, hence, the ways and methods I used to assess them. Students can also feel that their English instructor is not a task-assigning robot but a

human being who likes to communicate with them. (Ryan-IT3, pos.47)

The approach of befriending with students proved to be successful. Students approved that although Ryan was “an authority” in the assessment, he did not assess their performance “in a superior manner” just like those teachers in high schools. Instead, he assessed like “a friend”, “give[giving] students a chance for equal and friendly communications” (Ryan-student-FGI- 2, pos.186).

Beyond being an assessment designer, Ryan, like Emma, preferably identified himself as an assessment facilitator who “facilitate[s] the implementation [of assessment activities] by all means” (Ryan-IT1, pos.76). In this way, he felt that “assessment is a process instead of an outcome, in which students’ learning is ignited by assessment methods” (Ditto). Ryan illustrated his view with his reflections on peer assessment in narrative writing.

... In this semester, I used group assessment to instruct students in narrative writing. I found that students’ writing was dry and meaningless at the beginning. So, I showed them videos and materials about how to write a narration. After several rounds of revisions based on their peers’ feedback and my feedback, they gradually came up with lots of meaningful ideas ... they could narrate a very interesting story of 800 to 900 words. This is a process I facilitate, which is good. (Ryan-IT1, pos.76).

Ryan facilitated students’ learning by guiding them throughout assessment activities. He provided input and feedback before and after assessment tasks. He highlighted the assessment process in which he provided emotional and practical scaffolding to students. For instance, one of his students reported that Ryan would “have a private talk with those students who fail in exams. He would encourage us to do better in next assessment” (Ryan-student-FGI-3, pos.106). In this sense, Ryan was also an emotional supporter of students.

In addition, as Ryan held that it was students who dominated the assessment, he strove to facilitate the assessment process by involving them in the group work. Ryan was aware that students’ engagement in

language learning and assessment could exert a positive impact on students' learning outcomes, as he commented,

... In this assessment practice, I did not expect my own comments but their [students'] feedback ... Yes, it is the students themselves who do the assessment. I didn't give any comments on their peer feedback. This feedback is their comments on their peers' writing from the perspective of an individual reader. I can't judge whether their feedback is right or wrong ... (Ryan-IT5, pos.75)

Ryan's identity as an assessment facilitator was reinforced in the second semester. He emphasised that conducting assessment was not for the sake of assessment itself but to facilitate students' English learning. His ultimate objective was to help students to gain advancement in learning.

I must assign them [students] an assessment task. Without the task, how can I know if they have learned what they need to learn? ... As I mentioned earlier, I have enabling activities in which students need to complete tasks. Then, I give feedback to their tasks. The tasks are used to assess students' learning. You see, assessment is to facilitate the entire teaching and learning process. (Ryan-IT4, pos.93)

To facilitate students' learning through assessment, Ryan reiterated the purposes of assessment and the long-term impact of assessment on students. He expected that his "assessment activities can help them [students] in learning instead of simply completing a task" (Ryan-IT3, pos.54). The most important thing was that he "need[s] to do homework to improve assessment competence so that the assessment can become more effective and attractive" (Ryan-IT3, pos.54). Apparently, Ryan understood the alignment of assessment with his identity as an assessment facilitator so that he could enhance his LAL to optimise his assessment effectiveness for students' learning.

In the third semester, Ryan iterated his identity of being a designer and facilitator in language assessment. He emphasised again that assessment should serve and improve students' learning through his

guidance as an instructor. “I shall prioritise students’ learning all the time when designing and implementing assessment practices” (Ryan-NF3, pos.34).

6.7 Ryan’s Case in Summary

6.7.1 LAL Profile

Ryan’s LAL profile is evident in his assessment-related narratives. He had a solid understanding of language assessment concepts and content due to his training to become an English teacher and his educational experience in the UK (principles & concepts). Familiar with both traditional and alternative assessment methods before becoming a university instructor, he applied these methods creatively to enhance student learning (language pedagogy). Well-versed in departmental assessment tasks (local practices), Ryan adapted these tasks, considering students’ weaknesses in English and the curriculum (knowledge of theory). For example, he used group peer assessment for different written genres, provided detailed rubrics, being aware that Chinese students need to improve productive language skills (language pedagogy, knowledge of theory). He also provided effective feedback on revised essays based on peer input (technical skills).

Ryan believed that instructor support could ease students’ anxiety about standardised tests, so he dedicated class time to CET4/6 preparation when the tests approached (personal beliefs/attitudes, technical skills). He acknowledged the importance of CET 4/6 certificates for education and employment (sociocultural values). Consequently, he emphasised objective grading by carefully documenting students’ learning processes (grading & decision-making). Aligning with university requirements, Ryan ensured that scores he gave accurately reflected students’ language proficiency, which could benefit their future prospects (sociocultural values). Ryan’s LAL profile over the three semesters is summarised in Fig. [6.6](#).

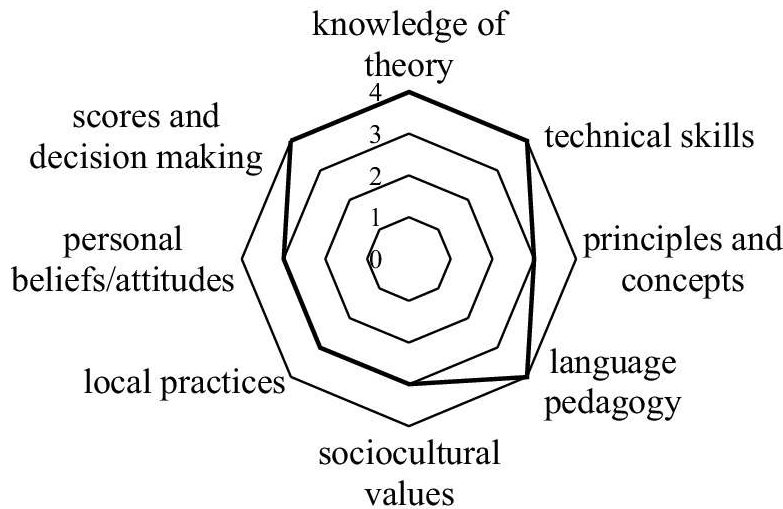


Fig. 6.6 Ryan's LAL Profile

6.7.2 LAL Developmental Trajectory

Ryan's initial LAL did not develop substantially over the three semesters, but the development was also featured with its individual trajectory. Being educated to be an English teacher, Ryan's well-received training in language testing and assessment and assessment experience in the UK enabled him to have a considerable LTA knowledge base and practices initially. This solid foundation placed him ahead of other participants in LAL before joining this study.

Ryan acquired extensive knowledge about standards of assessing writing through academic reading when exploring assessment criteria for peer assessment on narrative writing. To maximise the effectiveness of group peer assessment, Ryan turned to reading materials to obtain writing assessment skills. Informed by these materials, he developed a detailed assessment rubric for students to conduct group peer assessments. Additionally, his ongoing EdD programme renewed his understanding of alternative assessment methods experienced as both a doctoral student and an instructor.

Preparation for a teaching contest further heightened his awareness of integrating assessment with teaching practices. During this time, he reflected on his teaching and assessment weaknesses through discussions with veteran colleagues. These preparation activities prompted him to focus on designing assessments that facilitate student learning. As a result of his heightened assessment awareness, Ryan

successfully implemented effective practices like quizzes, peer assessments, and CET 4 translation preparation, which significantly boosted both his and his students' self-efficacy.

However, Ryan showed no desire to change his established assessment methods. Despite this inertia, Ryan became more confident with his assessment practices. His confidence was boosted along with accumulated teaching experience and iterative reflection. This summary of Ryan's LAL developmental trajectory is illustrated in Fig. 6.7.

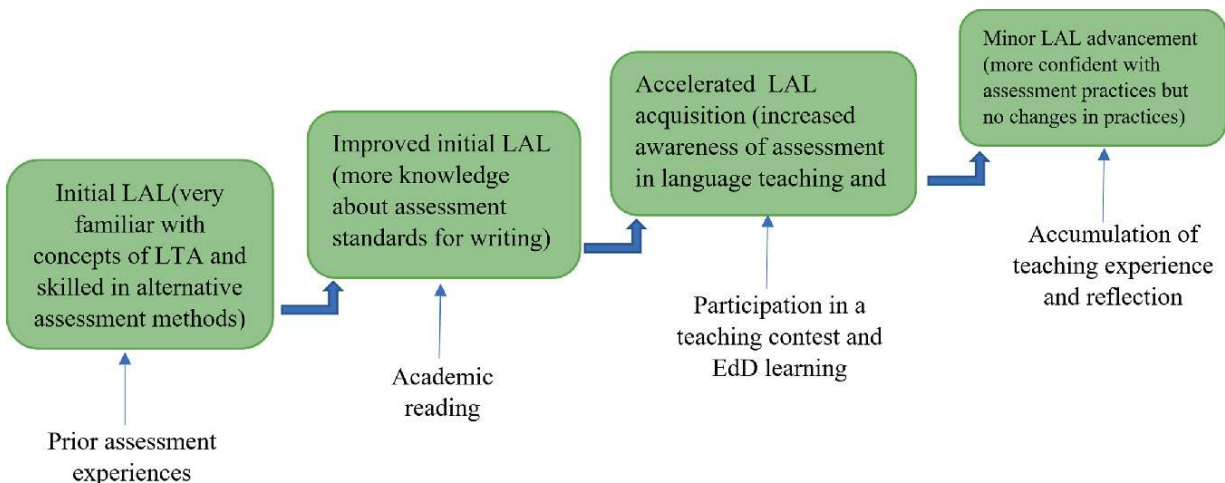


Fig. 6.7 A Trajectory of Ryan's LAL Development

6.7.3 Assessment Identity

Throughout the study, Ryan did not recognise his identity as an assessor. He associated this identity solely with scores or grades, believing that his role extended beyond mere evaluation to include fostering student learning through assessment. He maintained this non-assessor identity throughout the three semesters, even though he had advanced his assessment awareness and confidence in assessment practices. Viewing assessment as a part of his instruction, Ryan prioritised his responsibilities as an instructor who provides knowledge and skills, feedback and emotional support.

Although he did not identify as an assessor, Ryan was effectively a veteran assessment practitioner. He skilfully employed a range of assessment methods—traditional and alternative, formal and informal. Ryan designed effective rubrics for peer assessments in writing and

provided explicit, constructive written feedback on student essays. He meticulously graded each assessment task, efficiently tracking students' learning process and progress with an assessment sheet.

Ryan preferred to view himself as an assessment designer and facilitator. He designed effective assessment tasks that engaged students in learning through assessment by building rapport with students. Just like Emma, Ryan underscored his identity as a facilitator who used assessment to advance students' learning rather than for summative assessment only. He highlighted students' active role in the assessment process, seeking to accelerate their learning by accentuating the purposes and impact of assessment and providing necessary intellectual and emotional support.

6.7.4 Challenges and Opportunities for LAL Development

Ryan's rich experience in language assessment enabled him to handle classroom assessment practices with ease. He sought solutions to difficulties in designing and conducting assessment tasks. For instance, he read academic books to learn the standards of assessing narrative writing based on which he produced a rubric for group peer assessment. He also addressed the problem of providing timely feedback on students' writing despite his heavy workload.

However, Ryan faced a significant challenge in acquiring comprehensive theories of language testing and assessment. His busy teaching and extracurricular training schedules, preparation for the teaching contest and EdD programme deprived him of the time on formal learning in language testing and assessment. Consequently, his assessment practices were largely based on personal experience, lacking a solid theoretical foundation, which hindered the sustainability and effectiveness of his assessments.

Despite these challenges, Ryan found opportunities to develop his LAL over the three semesters. His participation in the provincial teaching contest created venues for him to discuss teaching and assessment issues with experienced colleagues. The collegial support was evinced in boosting his assessment awareness in language teaching. Despite an unexpected result of the contest, Ryan took it as an important part of his professional development both in language teaching and assessment.

References

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>

[Crossref]

Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing*, 19, 6–23.

[Crossref]

Shrestha, M., & Roffey, C. (2018). An alternative to traditional assessment: The debate showcase. *Journal of Education and Research*, 8(2), 5–31. <https://doi.org/10.3126/jer.v8i2.22888>

[Crossref]

Tucker, J., & Courts, B. (2010). Grade inflation in the college classroom. *Foresight*, 12(1), 45–53. <https://doi.org/10.1108/14636681011020155>

[Crossref]

7. The Case of Tim

Ling Gan¹✉ and Ricky Lam²✉

- (1) School of Languages and Communication, Beijing Technology and Business University, Beijing, China
- (2) Department of Education and Psychology, Hong Kong Baptist University, Hong Kong, China

✉ **Ling Gan (Corresponding author)**

Email: linggan@btbu.edu.cn

✉ **Ricky Lam (Corresponding author)**

Email: rickylam@hkbu.edu.hk

7.1 Overview

This chapter explores Tim's LAL development, his assessment identity construction and the difficulties he went through and the opportunities he embraced over three semesters. Unlike Rachel, Tim, as a novice English instructor, began with substantial assessment knowledge obtained from his comprehensive assessment education. However, he did not advance this knowledge over the semesters as he was not motivated to do so due to the influence of his family, institution and students. Despite his assessment knowledge, Tim's assessment activities were test-orientated, which led to his uncertainty about his identity as an assessor. He integrated self-and peer assessment into writing classes without seeking to improve their effectiveness. While he preferred giving individualised feedback, he felt powerless when feedback was not effectively utilised. Tim was concerned about assessment criteria and enthusiastically discussed with peers when taking a training programme. However, he did not take this opportunity

to enhance his LAL in this area. In addition, his participation in a teaching contest did not significantly boost his LAL. Over time, through reflecting on his responsibilities, Tim gradually acknowledged his identity as an assessor, but preferably considering himself to be a life mentor and acting as a lenient marker.

7.2 Background Information

Born in the early 1990s, Tim got his bachelor's and master's degrees from the same university in China, with his research area being corpus linguistics. After graduation, he was employed by LU as one of the only two recruited English instructors that year. When joining my study, he had been teaching non-English majors College English for four years at this university.

College English teaching at LU experienced a reform in the autumn of 2020. Students admitted in 2020 are no longer placed into A, B, and C levels through an in-house placement test. Instead, they are divided into A (average) and B (elementary) levels. While a great majority of students are placed into A level, those whose mother tongue is not Mandarin into B level.

Among A-level students, their English proficiency varies substantially owing to different entry requirements for majors. For instance, students majoring in economics, accounting, statistics, management, and law are admitted with higher English proficiency, while those of other majors have a lower English level. In the current study, Tim taught students majoring in animation in the first two semesters and accounting in the third semester.

English learning can be divided into two phases in LU. Students have General College English in one academic year. There are four lectures (45 minutes for each lecture) of classes per week. In the first year, students are required to learn the integrated skills of reading, writing, listening, speaking, and translation. Different language skills can be scheduled into four lectures each week. In the second year, students do not have General College English classes, but they can select courses of English for Specific or Academic Purposes including Western cultures, English literature, Academic English, or Business English, if they want to continue their English learning. All these courses are not compulsory

in College English teaching. Tim taught both General College English and English for Specific Purpose over three semesters, but he was only observed on the former in the study.

7.3 Prior Assessment Experiences

Tim, like other participants, experienced numerous paper tests in primary and secondary schools to prepare for high-stakes exams. Unlike others, he found university assessments were similarly reliant on traditional tests and he felt that important decisions based solely on scores were unfair. Nonetheless, Tim firmly believed that preparing students for tests in the Chinese context “meets students’ learning needs, in view of tests’ positive washback” (Tim-IT1, pos.10).

During his master’s programme, Tim took a compulsory course of language testing and assessment taught by a well-known language testing professor in China. Although the course immersed him in LTA theories, he struggled to apply this knowledge in teaching, mainly because “the course was not orientated to how to conduct assessment practices in the classroom but do research in language testing and assessment” (Tim-IT1, pos.12). Like Emma, the lack of practical application meant Tim “had not established an assessment knowledge base” (Tim-IT1, pos.26).

Despite this, Tim was exposed to the literature related to language testing and assessment, placing Bachman’s book series on his reading list. In the list, books and papers written by the professor lecturing the assessment course boosted his understanding of washback at most. More importantly, academic reading opened an avenue for him to access hot issues in assessment research, which in turn contributed to his research area.

After becoming an instructor, Tim attended training programmes on grading standardised tests like CET4/6 or National Graduate Entrance English Examination (NGEEE). The training, usually lasting one or two hours, focused on how to grade writing and translation with high efficiency. Surprisingly, Tim did not speak highly of these opportunities but complained that he was only “a rating machine”, as the training did “promote the rating efficiency”, but “contributed nothing to my [his] LAL development” (Tim-IT5, pos.98).

While formal learning gave Tim initial LAL, its lack of practical relevance limited his development to fundamental assessment concepts. Additionally, Tim’s understanding of test washback seemed slightly deviated from that of testing and assessment professionals, leading to his test-orientated practices in the classes, which could be demonstrated in the following sections.

7.4 A Confident Test-Orientated Practitioner

7.4.1 Understanding Assessment Requirements

Tim seemed not to care about his limited assessment know-what and know-how. Instead, he was confident enough to handle assessment issues, even without strictly following the assessment requirements (see Table [7.1](#)) for College English prescribed by his department.

Table 7.1 Assessment Requirements for College English by Tim’s Department

Component	Assessment task	Weight (%)	Description of the task
Formative assessment (40% of the total assessment)	Class attendance & in-class performances	30	Students are required to attend every class on time. They need to actively participate in the class activities like Q & A, oral presentations, group work, or written assignments
	After-class assignments & autonomous learning	30	Students need to complete homework out of classes, like projects, and reading assignments. They also need to complete tests presented on WeLearn, an autonomous learning app
	Oral testing	40	Instructors conduct a spoken test with different methods decided independently
Summative assessment (60% of the total assessment)	Final exam		This is an achievement test, consisting of listening, reading, translation, and writing

Three insights emerge from these assessment tasks. First, although there was an intended balance between formative and summative assessment, Tim’s university prioritised summative assessment. This meant that Tim’s evaluation of students’ language performance primarily depended on their final test scores. Second, while his

department allowed him considerable autonomy in planning and administering assessment tasks, including in-class and after-class assignments and speaking tests, this freedom might have led to his reduced emphasis on formative assessment. Third, the lack of specific guidance on assessments posed challenges for new instructors, as Tim recounted that he “had the least idea of what to do with formative assessment and summative assessment when I [he] started to teach in LU” (Tim-IT5, pos.62).

Despite the autonomy for assessment, Tim’s in-class assessment activities were dominantly CET4/6-orientated. As such, assessment in his class seemed not to be given due attention, as students reported, “the assessment methods our instructor used on us are too monotonous and few assessment activities are conducted to improve our learning” (Tim-student-FGI-3, pos.20).

7.4.2 CET4/6-Orientated Practices

Tim centred his teaching activities around CET4/6 by lecturing past test papers. This choice was largely influenced by his preconception of CET 4/6 positive washback on students’ learning. This choice, Tim argued, met students’ learning needs, as “they [students] don’t want flowery and useless activities” (Tim-IT1, pos.10).

Tim clearly understood passing CET 4/6 was his students’ major concern, a concern he had experienced as a student. When he shifted his identity from a student to an English instructor, he felt an obligation to help students prepare for standardised tests. He was confident that he could manage the preparation with his current assessment knowledge base and through modelling his teachers.

I chose to lecture CET 4 for students mainly because of my experience as a student ... I know what I want from assessment as a student ... I just copied what and how my teachers assessed or tested me. I copy them subconsciously and have never questioned the effectiveness of their assessment activities ... I feel comfortable with these activities. Similarly, I like the way I lecture on CET 4 for students, and I believe students also like these lectures. (Tim-IT2, pos.82)

Tim's teaching of language components was also CET4/6-orientated. For instance, writing assignments were narrowed down to argumentative writing, a default type in CET4/6 and all the writing topics stemmed from past CET4/6 test papers. For him, argumentative writing assignments enable students to learn how to write logically and formally, two abilities Chinese students need urgently. He also believed that these assignments could train students to better cope with the test.

... This is like a "training to tests". After all, they [students] need to take CET 4/6 and expect to get a good score in it. As a College English instructor, I have the responsibility and obligation to help them achieve the goal, right? ... (Tim-IT5, pos.21)

Tim's listening classes were also targeted at CET4/6. As listening tasks in the textbook were similar to those in CET4/6, Tim focused on test-taking strategies in listening. The following is an observation excerpt from a test-taking lecture in the listening class.

Tim played an audio file of a short conversation between a man and a woman about French learning. When the conversation ended, one question was asked.

Tim (sitting behind the platform and the microphone): Okay. Let's look at the four items. Item A is "she has a keen interest in French language culture ... Is it the answer? What does the word "k-e-e-n" mean? (No students responded)

Tim: It means "eager, enthusiastic", right? ... Now, let's listen one more time. (Tim played the audio again)

Tim: The woman says her grandfather can speak French. Grandfather believes that "truly cultured persons must know..." Okay. From this sentence, we can infer that her grandfather influenced her in French learning. So, can we say that item A sounds reasonable? Well, here, I would talk about the "over inference" in listening comprehension. Tell me, what can we infer from what the woman says? Can we infer that her grandfather's view has impacted her? Yes! That is to say, she

agrees with her grandfather, right? ... Then, we can infer that the girl thinks truly cultured persons must know French. **This is the first level of inference** [emphasis added]. Based on this inference, we may further infer that the girl is pretty interested in French. **This is the second level of inference** [emphasis added]. However, **in CET 4/6, you are not required to make two levels of inferences in the listening comprehension** [emphasis added]. If you need to do so, this item must be wrong. So, when listening, you must be cautious and ask yourself, **is it the information you get from the listening texts directly or do you need to infer indirectly?** [emphasis added] In addition, **do you notice there is a flaw in item A?** [emphasis added] (Tim looked up at students. No one replied) ...

(Tim-F01)

Two major inferences can be drawn from this excerpt. First, Tim demonstrated the ability to critically evaluate listening test items, a skill developed through his own experiences as a student in an exam-focused Chinese context. He learned to discern the intentions of item writers and assess the strengths and weaknesses of test items. He was quite proud of this capability.

... The multiple-choice items in the speaking textbook are poorly written. I mean, the distracting items are too evident that students can choose the correct answer without listening to the audio ... Obviously, the test difficulty has decreased drastically, although the editors of the textbook claim that the test difficulty of the listening materials is equal to that of CET4/6. I can claim that the item writers of the textbook do not design and write the items carefully. (Tim-POC1, pos.3&4)

Second, Tim lectured the listening module alone without examining whether students understood the listening materials. Mostly, he asked rhetorical questions to which he did not expect students to respond, as he seemed to be more comfortable with lectures alone.

Although the lack of interaction made CET4/6 lectures less engaging for students, Tim confidently defended his approach. He

justified it by emphasising the importance of these standardised tests each semester. On the first day of class, he detailed the tests students might face over the next four years, including CET4/6, NGEET, and TOEFL/IELTS, along with preparation and test-taking schedules. This discussion heightened students' awareness of the significance of these exams and the importance of adequate preparation.

In the second semester, Tim allocated one-third of class time to CET6 preparation, lecturing on past test papers and providing detailed test-taking strategies for each component. He acknowledged that these lectures slightly delayed his teaching schedule but deemed them worthwhile as they addressed students' needs and expectations. While recognising the challenges first-year students faced with CET6, he remained unconcerned about their acceptance of the difficulty. He remarked, "As grown-ups, students are quite aware of the difficulty of CET6. How can they give it up simply because of the difficulty they face?" (Tim-IT4, pos.60).

In the third semester, Tim employed the same approach with a new cohort of freshmen, dedicating four weeks to the listening module of CET6 despite a tight teaching schedule. When asked about lecturing on CET 6 for freshmen, he stated it was driven by his "conscience and morality" (Tim-IT6, pos.14). He was committed to supporting students in improving their CET6 scores, unbothered by the additional workload.

Regardless of Tim's professional responsibility, students' responses to the test-orientated practices were not positive.

WS: Because listening is excluded in *Gaokao* of my province, I did not have any listening training in high school. I totally can't understand CET6 listening recordings ... I force myself to listen, but I can't, because I really don't understand. I think lecturing on CET6 is useful for those who have practiced listening in high school but not for me...

Aurora: I did not have listening training in high school, either ... Therefore, it was painful for me when our English teacher gave CET6 listening lectures. As listening tasks in the textbook are much easier, I can understand. However, I can't understand CET

6 listening at all. Although the English instructor would play the same recording several times, I couldn't figure out every single word ... it was a waste of time ...

(Tim-student-FGI-3, pos.114&115)

Apparently, not all students appreciated Tim's CET6 listening preparation, as it proved challenging for those with limited listening skills. Additionally, focusing solely on CET4/6 test-taking strategies neglected skills that could enhance overall language learning.

Despite student complaints, Tim remained committed to his CET4/6-oriented practices. He believed these methods were the result of an extensive exploration into assessment practices that aligned with the goals and core competencies of university students. Consequently, he saw "no need for any change and improvement in assessment practices" (Tim-IT6, pos.18).

7.4.3 Using Self- and Peer Assessment in Writing Classes

Different from the first semester, Tim started to involve students in assessment activities in the second semester, as he found that "my (his) class is too monotonous and I (he) want something different" (Tim-POC4, pos.10). He regularly implemented self- and peer assessment in his writing class, enabling students to identify issues in their writing alongside teacher evaluations. This approach aimed to make the assessment of students' writing "more comprehensive and objective" (Tim-IT5, pos.26). The following assessment episode illustrated how self- and peer assessment was utilised in his writing class.

Tim: Well, today, we are going to assess and appreciate some written essays ... The five writing essays I present here are representative of the problems in your writing. As usual, you read the essay first and then assess it in three steps. First, if you were a teacher, what score would you give out of 15, a full score? Second, you need to identify the strengths and weaknesses of this essay. Third, you give suggestions on how to revise and improve the essay. Today, we focus our revision on the syntax ... we will work together to see how to change a simple sentence into one with better grammar and expressions ...

Tim presented the first essay on the screen. While he told the students that all essays were anonymous, students burst into laughter, as the name obviously appeared at the end of the essay. Hurriedly deleting the name, Tim asked students to read the essay.

Tim (Three minutes later): Okay. Time is up. Let's have a look at this essay ... Who wants to share your views? YD. Would you please tell us what you think of this essay? (Tim grins and students laugh loudly)

YD (Stand up, laughing): Generally, the structure is good. (Students laughed)

Tim (Walk towards YD): It's all right. What do you think of this fellow student's, aha, or your own essay? (Students burst into laughter). What problems does it have? How can you improve it?

YD: The weakness lies in the expressions. When I present the arguments, I find it hard to show my arguments in English ...

Tim: ... Anything else?

YD: I can come up with some advanced vocabulary items and I will avoid using simple words in my next writing.

Tim: Could you give some examples of what words can be advanced now?

YD: Sorry, I mean the words I used are my ...

Tim: You mean they have been the advanced words you can use, right?

YD (Grins): Yes, yes.

Tim: Ok. Can you give some tips to improve your Chinglish? ... In second language acquisition, the problem you mentioned just

now is called the transfer of the first language. Can you give one example to show how you are influenced by the mother tongue?

YD: Hmm ... I can use specific examples to illustrate my points in Chinese, but I can't in English. For example, ...

Tim: ... Okay. Good. Out of 15, what score will you give to your essay?

YD (In a low voice): 9?

Tim: Okay. Fine ... Thank you. Now, let's hear other students' opinions of this essay. (Tim puts his hand on the desk of a girl sitting in front of the screen)

(Tim-F04)

Three inferences can be made from this episode. First of all, Tim pinpointed the purposes and steps of self- and peer assessment for writing essays, although he did not explicitly inform students that he was using these methods. Second, unlike Rachel, Tim gave no training to students on self- and peer assessment, failing to present specific criteria for students to evaluate their own and peers' writing. Without assessment criteria, students would fail to generate effective feedback. This lack of criteria hindered students' ability to give effective feedback, which is essential for accelerating learning and enhancing metacognitive development (Deneen & Hoo, [2021](#); Zhao, [2014](#)). Third, Tim seemed to use these two assessment methods simply to make his class more interesting rather than motivate and promote students' learning by engaging them in assessment tasks.

Later, Tim clarified that he did present students with assessment criteria before the self- and peer assessment. These criteria were derived from the CET4/6 writing section rating standards, and his students confirmed that he displayed and explained them (see Table [7.2](#)). However, the criteria were too general to be helpful, which raised doubts about the students' ability to provide meaningful peer feedback based on this criteria sheet.

Table 7.2 A Criteria Sheet for Writing

Criteria for writing	
Content	Alignment with the topic
Position	Clearance and elaboration
Structure	Logic: <i>Topic sentence</i> <i>Main ideas</i> <i>Supporting points</i>
Sentence	Accuracy Diversity Complexity Pragmatics Formality

Overall, Tim was quite satisfied with the self- and peer assessment, finding that his students made progress, as “they can immediately identify what is wrong with the essays and know how to improve it” (Tim-POC4, pos. 7).

One month later, however, Tim changed his attitude towards the effectiveness of self- and peer assessment.

... I gave them [students] detailed feedback about how to write a good CET6 argumentative piece, but they repeated the same mistakes! ... I am even angry that some essays are ghost-written. This is not a matter of competence but a matter of learning attitude! I take my job seriously, but they [students] do not listen to me. I don't think there's a problem with my teaching and lecturing. There's a problem with their learning attitudes and motivation! Their negative responses to my teaching have posed a heavy blow to my devotion to College English.

Now, I feel that the so-called self- and peer assessment is a hollow exercise. It is useless for students' learning. You can hear students talk confidently about what problems the essays have and how the problems can be solved. It seems that they can improve their writing output. The reality is that I collected a pile

of trash! I am sorry to say that. I know it is rude, but their writing is trash indeed!

(Tim-IT4, pos. 22 & 25)

Tim talked furiously and indignantly during the interview when commenting on the effectiveness of his self- and peer assessment in the writing. He attributed the inefficiency of these methods to students' low motivation and negative attitudes rather than his teaching. Despite his efforts to enhance students' writing using these approaches, he was disappointed to see no improvement in their essays, which led to decreased confidence in his teaching practices.

In a subsequent interview, Tim clarified his changed attitudes.

It is a different story between witnessing students' progress and feeling disappointed with their writing. It is because I have high expectations of students. For instance, I expect that they can achieve 90 marks in writing, but they get 80 marks, actually. In this case, I think they are not good enough. However, I must admit that they have made progress because they have raised their marks from the initial 70 to 80. The progress is visible.

(Tim-IT5, pos. 64)

Although Tim was discontented with students' writing, he did not neglect the positive impact of self- and peer assessment on student learning. That is, they did promote students' writing and individuals' learning attitudes.

In the third semester, Tim conducted self- and peer assessment in a writing class and a speaking class in the same manner as he did in the previous semester, but on students majoring in accounting who had much higher English proficiency. Similarly, Tim spoke highly of the effectiveness of these two methods.

From my perspective, the effectiveness is self-explanatory ... For example, when a student self-assesses his own essay, he identifies his own problems immediately, ones he does not realise when he is writing ...

I also find that some students did very good jobs. For instance, they noticed no “introduction” at the beginning of the essay or “the arguments are irrelevant to the original topic”. By identifying peers’ writing errors, students themselves can develop their ability to make their writing more logical ...

(Tim-POC6, pos. 4&5)

Tim viewed the effectiveness of self- and peer assessment as students’ ability to identify problems in writing. This understanding was also imparted to his students, as seen in their responses.

WS: ... I think when we read our own writing, we can’t identify the problems at all. Instead, we can quickly locate problems in others’ writing. I will be over the moon if my own writing can be assessed by my fellow students. That will indicate I have more room for improvement ...

Aurora: ... Agreed. We are not aware of the problems in our own writing, but we can identify a lot of them in others’ writing. For instance, I can find there is a logic problem with a sentence, or the argument is not illustrated by examples. By evaluating others’ writing, we can reflect on our own writing: do I have the same problems in the writing?

KS: It is interesting to read peers’ writing. I am curious if my peers have new ideas or arguments in their writing and if I can learn from them ... This practice is awesome.

(Tim-student-FGI-3, pos. 65,66 &68)

When looking back on his assessment practice in the three semesters, Tim regarded the amalgamation of students’ self- and peer assessment with writing as an advancement in his LAL and an indicator of his professional development. He felt that he had become “more mature in English teaching” (Tim-IT6, pos. 22). However, owing to a lack of means to reflect on his use of self- and peer assessment, Tim failed to discern

deficiencies in this practice. As a result, the effectiveness of using self- and peer assessment in writing was undermined substantially.

7.4.4 Grading

Despite a combination of formative and summative assessment advocated by his department, Tim attached more importance to the latter. He did not trust the formative assessment of students' performance, as "Formative assessment could not differentiate students' English proficiency" (Tim-IT3, pos. 59). However, in order not to disadvantage students in an exam-driven culture which values students with "a good-looking GPA", he had to rely on formative assessment.

My department requires the different weights for formative and summative assessment, but I cannot follow this requirement strictly in reality. If I do, those students who have a poor performance in formative assessment will get a terrible score in the final. This is what I don't want to see. Besides, it is a time of GPA inflation. I mean, students' grades are inflated in universities. Under such a circumstance, I can't rigidly implement formative assessment ... I prefer summative assessment, as its scores are always objective and representative of students' English proficiency. Instead, formative assessment is comparatively subjective. I can raise the score of formative assessment a bit to give my students higher scores in the final, hence a "good-looking" GPA ... Implementing formative and summative assessment in College English teaching is conceptually good but is practically difficult. (Tim-IT2, pos.62)

The excerpt illustrated Tim's conceptions of formative and summative assessment. He perceived the former as complementary to the latter, believing it could be subjective and adjusted flexibly in tandem with summative assessment scores to help students achieve higher final grades. In addition, Tim prioritised the use of students' scores/grades, as their scores/grades impact significant decisions like scholarships, postgraduate recommendations and employment opportunities. Consequently, he graded generously, sometimes even deliberately

raising students' scores to reflect "care for students" and acknowledge "the social significance of scores for students" (Tim-IT2, pos. 64).

Tim's "care for students" became more self-evident in the third semester when he explained the final exam to his freshmen students.

Tim: That is to say, the 25 points come directly from the textbook. I think all of you should grab these 25 points. Then everyone can get a good score on the exam so that you can have a good time during the Spring Festival (Students burst into laughter) ... **If the failure in my course makes you sad throughout the Spring Festival, I, as an English instructor, feel sad and sorry for you, too** [emphasis added] (Students laugh out loud).

One student asked a question about the test difficulty.

Tim: ... I got no extra information about the test difficulty. I don't know the difficulty, either, as I'm not the item writer. Anyway, the test difficulty is not a problem. **I can ensure you that the test difficulty isn't decided by the test itself but by me. Do you understand?** [emphasis added]

Students: Yes! (The whole classroom rocks with laughter)

Tim: Once again, **you don't worry about the test difficulty, and I, will deal with it** [emphasis added] (Students laugh). No matter how difficult the test is, I can handle it. Can you understand me? ... **You take the test and leave the result to me. I'll take care of it** [emphasis added]. Don't worry. Ok, any other questions?

(Tim-F010)

As shown in this episode, Tim attempted to mitigate freshman students' anxiety by introducing the final achievement test. His humorous allusion to his control over the grading revealed that he had a great say in assessment. However, this allusion was seemingly unethical, as he should have told students that their exams were scored against the

criteria, but his emotional concerns for students overshadowed his commitment to objectivity and fairness in grading.

Tim's "hodgepodge" grading approach resembled a softer form of grade inflation, which would undermine students' learning motivation and enthusiasm in the long run (Arrafii, [2020](#)). His students even questioned this practice, stating that their scores "have no discrimination and are meaningless for learning" (Tim-student-FGI-3, pos. 134). Tim acknowledged students' inflated scores. While these scores allowed students to achieve their desired outcomes, he was also aware that this practice was unfair to those truly excellent students. However, he prioritised the perceived benefits for the majority over the fairness owed to a minority of high-achieving students.

Why do I inflate students' scores? My students need to have a good score sheet to apply for a job, or a position in the postgraduate programme both home and abroad. I know it is very hard for them [to get a high score]. So, I lend them a hand. Of course, this will result in GPA inflation. It is not weird in the education field around the world. I am clear that students' scores cannot represent their real English proficiency, but I have no way out and have to go with the flow. After all, I care a great deal for my students. (Tim-IT6, pos. 6)

In the meantime, Tim was sure that his grading on students' speaking was objective and fair since it was strictly criteria-based. "If students can meet the criteria, okay, they would be given a high score. If they don't, they would be given a baseline score" (Tim-IT4, pos. 44). Although Tim approbated his grading practice, some student interviewees questioned this practice on the speaking test.

Tutuanna: ... I feel that grading is subjective ... I don't think the score can reflect our students' real English proficiency. The paragraphs we draw lots to select have different difficulty levels. Dialogues are much easier than arguments. Luck counts in the test.

LYQ: Agree. The difficulty of reading paragraphs is completely different. It cannot say that the grading is 100% fair.

LZ: I believe that the grading for speaking is partially unfair. For instance, students who took the speaking test earlier could be given more time to read three paragraphs while those who took the test later read only one short paragraph. In addition, some paragraphs selected are easy to read while others are difficult.

(Tim-student-FGI-1, pos. 122,123 &124)

The interview excerpts implied that students doubted the validity of speaking test materials and the scheduling of the test administration. Variations in the difficulty of test materials and preparation time could lead to grading inconsistency. Consequently, the given scores could not be representative of students' authentic speaking proficiency.

Despite students' concerns, Tim remained confident in grading and ascribed this confidence to his consistent focus on assessment criteria. He was keen to know what criteria to include and how to weigh them, especially for speaking assessment, because establishing these criteria can be challenging for someone like him who judges student speeches or debates.

To improve his LAL in assessment criteria, Tim used to take several initiatives. He read journal articles on rubrics or scales and volunteered to be a judge for students' speaking competitions. This experience enabled him to practice using criteria to assess students' speaking ability and meanwhile provided him with opportunities to communicate with more experienced judges. However, he did not report any similar opportunities during the three semesters of the study.

Although Tim felt confident in his grading and saw no challenges in language assessment, he lacked the motivation to enhance his LAL through specialised training. For one thing, he was restrained by his family who needed his care and support. He felt "guilty" to leave his family to develop his own professionalism. For the other thing, the opportunities and time were limited. His department barely provided the opportunities to send instructors out to take assessment training and invite assessment experts in to give lectures or training.

Without formal learning, Tim deemed that it was beneficial to enhance LAL informally. He followed official accounts in WeChat to

obtain the latest information about linguistics and applied linguistics including language testing and assessment. In the second semester, he attended a *National Training Programme for in-service EFL Teachers in Tertiary Education* held in the summer holiday. He believed his LAL would grow along with “the increase of teaching experience and the implementation of teaching activities” (Tim-IT4, pos. 80). Additionally, the incremental scores from student evaluations also bolstered his confidence in his teaching and assessment practices.

... It is good to make friends with other instructors who join the programme. We discuss and share our teaching experiences. A good example is that this summer holiday, I discussed with an English instructor from University T about how to grade students accurately and efficiently. We had discussions during the programme. We even spared extra time after the programme to have further discussions ... (Tim-IT5, pos. 132)

7.4.5 Giving Individualised Feedback

Concerned with assessment criteria, Tim produced assessment criteria for students’ speaking by referring to *China’s Standards of English Language Ability* and IELTS speaking paper, with different focuses each semester. For instance, pronunciation and fluency were the two major criteria in the first semester. Tim expected that the first-year students could pay more attention to their pronunciation, because

Most students are speaking-deprived in their high school. Their English learning is exam-orientated ... they are not proficient in speaking. If I require them to have a conversation or give a description in the speaking test, I believe they find it extremely difficult to do that. Therefore, I give them one semester to practise and improve their pronunciation first ... (Tim-POC3, pos.3)

Tim well understood the weaknesses of Chinese students in English learning and the influence of the assessment culture on their English learning. Accordingly, he helped students build up their confidence in English speaking by concentrating on basic speaking skills.

From the second semester on, having referred to *China's Standards of English Language Ability*, Tim realised that students should move forward beyond pronunciation skills after one-semester of practice. He decided to align the speaking test with IELTS speaking, focusing on logic, authenticity, and articulation. With these criteria, he assessed students' speaking and offered them individualised feedback as well.

Therefore, the speaking test was not simply an occasion for Tim to assess students but one where he could communicate with students by providing emotional and academic scaffolding for English learning. This communication could be seen in the following field observation excerpt.

A girl chose a topic of her talk about someone she admired and talked about Audrey Hepburn in the speaking test.

Tim: Okay. It should be “sympathetic”, right? How does she influence you in your daily life?

S7: Hmm ... I look up a lot of books about her and I admire her hard work, and I think she... before acting she will look up a lot of books about the knowledge and have a better understanding of the movie. So, I think her experience influenced me a lot and I will use of her as my example to study hard.

Tim: Okay. Good. Very good. One suggestion. I think there are too many descriptions in your talk. Descriptions should be integrated with comments. You can describe what kind of person Audrey Hepburn is. Next, you can summarise her elegant appearance and her acting images. Then, you add your comments on these two aspects. For example, you can say “She’s already left marks on the entire history of movie making’. You can also comment on her from a historical perspective, say, “She has become generations’ memory in the movie history”, right? Your talk is well organised, but I think you need to improve it by giving some comments apart from descriptions. Okay. You get 98. Good job. Keep moving. By the way, do you have any difficulties with your English learning?

S7 (Think for a while): None so far. Thank you.

Tim: Okay. Keep doing your good job.

(Tim-F06)

Tim insisted that as an instructor, he had done his job perfectly to give students specific and individualised feedback in their speaking tests. This type of feedback was also welcomed by his students who believed Tim's direct feedback could help them focus on speaking problems and "speak in a native-like way" (Tim-student-FGI-2, pos. 90).

Students' positive responses to Tim's feedback in the end-of-semester teacher evaluation made him very proud of his feedback-giving skills. However, Tim himself was not satisfied with the effectiveness of his feedback on students' English learning.

Regarding its effectiveness, from my perspective, I haven't seen any obvious improvement in students' learning my feedback has brought about, ... I have begun to question the appropriateness of giving feedback to students. I could have skimmed through the feedback, but actually, I have put great effort in providing them with customised strategies for English learning. Yet, no improvement can be found, which really makes me sad. (Tim-IT4, pos. 38)

Tim anticipated that his feedback could lead students to their reflection on learning, ultimately leading to their learning improvement. The reality was that "they repeat the same errors. That means they did not make their efforts to use my feedback" (Tim-IT4, pos. 73). Students' failure to use feedback depressed Tim again, just as mentioned in Sect. [7.4.3](#). He blamed this failure on students' low learning motivation and agency. Tim emphasised again that he, as an instructor, had perfectly "doing my [his] job" in giving individualised feedback, as his practice was aligned with updated educational guidelines.

... Providing customised feedback satisfies me the most in terms of language assessment. It is aligned with a Confucian tenet that teaching should be tailored to individual unique needs.

Regarding giving feedback, I have successfully achieved individualisation by tailoring my teaching to all student's learning strengths and weaknesses ... (Tim-IT6, pos. 4)

7.5 Participating in Teaching Contests

Apart from his teaching, Tim was engaged in professional development activities. In the third semester, he enthusiastically entered teaching contests at both the department and university levels. When his department invited participation, Tim immediately joined, excelling at the department level and advancing to the university contest. There, he delivered a 20-minute College English session to judges from diverse academic backgrounds and earned third place in the university contest.

However, these teaching contests did not contribute much to Tim's LAL development. He complained that he received no effective feedback from the judges regarding his assessment. He joined the contests with a humble intention of obtaining suggestions and feedback from collegial peers, but these colleagues failed to give him constructive suggestions except for some "pointless" comments and compliments because of the "face issue" (Tim-IT6, pos. 30). He had hoped for valuable feedback from peers but was left dissatisfied.

Yet, the teaching contest prompted Tim to reflect on his teaching and assessment. Through reflections, he developed his unique understanding of language assessment. For instance, he identified improvement within an output-based teaching paradigm he adhered to, emphasising that student self- and peer assessments are output of learning— "output is assessment" (Tim-IT6, pos. 44) (Tim-IT6, pos. 44).

Tim also prided his enhancement in handling assessment-related issues through joining the teaching contests. For instance, he was able to "mitigate students' anxiety of tests promptly" and to "handle mishaps properly when students take tests" and to "interpret scores to students whenever they have queries" (Tim-IT6, pos. 69). However, this self-reported enhancement has not yet been demonstrated in field observations.

7.6 Perceived Assessment Identity

7.6.1 “Am I an Assessor?”

As Tim invested most of time to lecture alone and prepare students for CET4/6 by lecturing on past test papers, he questioned his identity as an assessor in the outset of the study.

“Am I an assessor? I am not sure. My interest in language testing and assessment is not to assess students. I am kind of more interested in assessment research ... In my class, I am inclined to lecture alone. I don’t care about much if students like my lecture or not Of course, students’ reactions to my assessment practices have disheartened me before. I feel frustrated with these practices ... (Tim-IT2, pos. 103)

Research on language assessment was more of Tim’s interest so that he could transfer methodologies into his own research areas. He also had a low self-efficacy toward assessment practices, feeling more comfortable with and confident in CET 4/6 lectures.

Tim’s low self-efficacy and negative emotions toward assessment were mainly enticed by students’ negative responses. For instance, in the second narrative frame, Tim recounted a story about a speaking test. Some students told him directly that they could not join the test because of their poor speaking proficiency. Tim was extremely angry when students said, “it is okay to give me whatever score.” Students’ perfunctory attitudes towards the speaking test flared him up. This and other similar unpleasant experiences discouraged him from performing assessment tasks to make himself “humiliated” (Tim-NF2, pos.21). As a result, without focusing much on assessment practices, Tim perceived himself not as much as an assessor.

Not identifying himself as an assessor, Tim had a clear understanding of the credo to be an ideal assessor.

If I am an assessor, first, I should be familiar with the assessment criteria. Second, I should be fair to all students. I should not give a high score or grade to students based on my close relationship with them or on my own preferences. (Tim-IT2, pos. 115)

For Tim, teachers' assessment should be criteria-referenced, objective and fair. An assessor should be able to prioritise objectivity and fairness in assessing students' performance, without being subjected to his/her personal dispositions. Yet, his grading practices shown in Sect. [7.4.4](#) revealed that his assessment failed to comply with his credo to be an ideal assessor.

At the beginning of the second semester, Tim's self-awareness of his identity as an assessor loomed from his reflections on assessment, although he was not completely certain about this identity yet.

... How do I ever see this identity? It is my obligation to give grades on students' performance. I mean, I give a score. I then will point out the advantages and disadvantages, okay, the beneficial factors or malefactors in students' performance. That's the stage number two. I explain again. Stage number one, grading, grading. Stage number two, okay, listing the advantages and distinct disadvantages and stage number three, making specific suggestions, giving very clear instructions, or improvement procedures for the future development of the students ... (Tim-IT3, pos. 73)

This excerpt has two indications. First, for Tim, being an assessor involves three responsibilities: grading, identifying strengths and weaknesses, and giving feedback. The three were placed on a continuous spectrum. Second, Tim was clearly aware of the role of language assessment in students' language learning. That is, he used assessment to diagnose students' learning problems, report their progress, and provide customised feedback based on assessment results. These roles could be confirmed and illustrated in Sect. 7.4.5. Tim's self-awareness of being an assessor, as he revealed, was the result of a gradual thinking and reflection rather than an "eureka" moment triggered by an event.

In the third semester, as he increasingly realised that what he was doing was part of the assessment, Tim had a clearer understanding of the assessor's identity. That is, an assessor needs to assess students' language performance based on designed assessment criteria.

Being an assessor is to score or grade students' performance or ability in language components by using criteria ... To assess is to express viewpoints, based on a framework. The framework is our critical thinking. This critical thinking is a reference or standard. With the help of the reference, we got our creative thinking ... This is the logic inside assessment. As an assessor, I need to change the framework into various assessment criteria targeted at students' language competence. (Tim-IT5, pos. 153)

Tim's understanding of the identity as an assessor was aligned with his assessment practices. He assessed students' speaking or writing in the class according to the criteria he produced, but he had "never thought about it [what he has done]". When he realised that he was actually doing an assessor's job, everything about being an assessor seemed to "make sense" (Tim-IT5, pos. 146).

7.6.2 A Life Mentor

Over the three semesters, Tim gradually recognised his identity as an assessor, but this identity was not as strong as the identity as "a life mentor". At the end of the second semester, Tim evidently expressed his identity as "a life mentor" more than as an assessor. A life mentor is "Mr Big" who "not only imparts knowledge but also cultivates morals and personalities by giving them an example to follow" (Tim-IT4, pos. 164). "The example" was considered exactly Tim himself, as he responded, I am not going to give students an outline or a sample to imitate, without any of those mandatory or compulsory versions of all these things like the criteria. I just intend to give them an example to follow. That example is myself. (Tim-IT5, pos. 169)

Tim maintained that this identity was more valuable than the identity of an instructor and an assessor. However, at that moment, he had no idea of how to be "a life mentor" for students in language testing and assessment.

An accident that took place in Tim's department at the end of the second semester enlightened him to think about what "a life mentor" should be. In the final exam of College English II held at the end of the second semester, items in the listening section were found to be

mismatched with the listening recording. As a result, there was a long-time interruption during the exam. What was worse, no grades were given to thousands of students when the semester ended. The department was under great pressure, confronted with complaints and blames from students, parents, and the university. Tim commented that this mishap was “a disaster caused by carelessness” and “an issue of an attitude rather than a competence” in language testing and assessment (Tim-IT5, pos. 107).

After this incident, Tim seemed to find out what “a life mentor” meant in LTA.

... I intend to make students understand that assessments on others should be fair, unbiased, and objective. Some students texted me privately to want me to give them high scores, because high scores can guarantee their success in applications for further study home and abroad. I would tell them that my grading is based on their performance in the class and the final exam. Being fair and objective is concerned with the ethics of language testing and assessment. By means of fair assessment, I showed my students that pulling some strings to obtain benefits does not work. (Tim-IT6, pos. 58)

Tim anticipated that he would be able to promote students’ morals through his own assessment practices. He intended to show that his assessment of students was ethical, so should his students’ behaviour be. This influence was more far-reaching than evaluating students’ language performance through scores, grades, or feedback. Fortunately, this influence was also felt and sensed by his students. “He [Tim] uses assessment to check whether his teaching is useful and helpful for us or not ... He often tells us his testing experiences. The experiences are so amusing that I feel he is just like ‘a life mentor’ who shares his life experience” (Tim-student-FGI-3, pos. 156).

7.7 Tim’s Case in Summary

7.7.1 LAL Profile

Tim's assessment practices over the three semesters explicitly illustrated his LAL profile. He learned fundamental assessment concepts during his master's programme and used his assessment experiences as a student to critique test items (principles & concepts). He was aware of the washback of CET4/6, but interpreted it narrowly, replacing regular English teaching with test practice (language pedagogy). This interpretation underpinned his CET 4/6-orientated practices over the three semesters (personal beliefs/attitudes).

Although aware of departmental assessment requirements, Tim prioritised summative tasks like speaking tests and final exams (local practices). Recognising Chinese students' weaknesses in English writing (knowledge of theory), he focused on writing instruction, though tasks were CET 4/6-based. He incorporated self- and peer assessments to support writing instruction (language pedagogy/technical skills), but these methods were not effectively implemented.

Considering the social significance of the scores in College English for students (sociocultural values), Tim marked students leniently, which led to students' inflated scores to some extent. However, he still adhered to his principles of being an assessor, grading students' speaking in accordance with criteria (grading & decision-making) to which he was willing to devote time. In addition, throughout the three semesters, Tim insisted on giving individualised feedback to students' speaking during speaking tests (technical skills), as he believed individualised feedback contributed to students' speaking and was aligned with the Confucian philosophy of learning and teaching (knowledge of theory). Tim's LAL profile is further detailed in Fig. [7.1](#), as referenced in Tables 2.2.

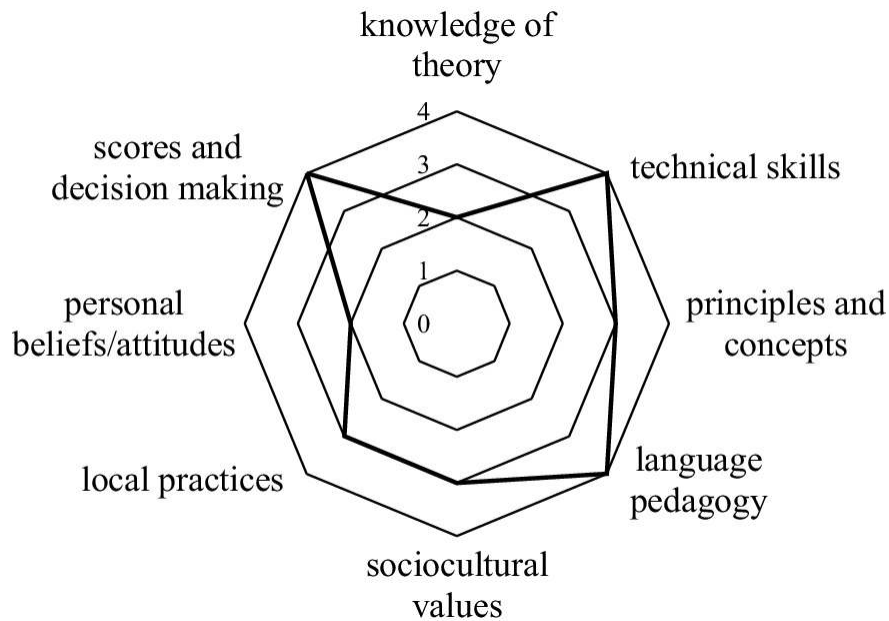


Fig. 7.1 Tim's LAL Profile

7.7.2 LAL Developmental Trajectory

Tim's LAL development was most tepid and incidental among all the participants in the study. Just like other participants, Tim acquired his initial LAL from his prior assessment learning experience and the experience of being assessed as a student. He understood the fundamental concepts of language testing and assessment at the beginning of this study. However, he was not seen to enrich his conceptual knowledge of language assessment. Similarly, Tim was aware of the prominent role assessment plays in promoting students' learning, but this assessment awareness did not grow over time, for his assessment practices kept orientated to CET4/6 lecturing throughout the three semesters.

Although self-reflection facilitated his use of alternative assessment, Tim did not take action to improve the effectiveness of alternative assessment on students' learning. For instance, he integrated self- and peer assessment with writing and speaking in the second semester. Unlike Rachel, he did not explore these two methods before the implementation, nor did he reflect on the problems after the implementation and actively seek solutions to improve their effectiveness. Instead, he blamed the ineffectiveness on students' low language proficiency and learning motivation. As a result, despite the

increasing use of alternative assessment methods in his class, Tim's understanding of these methods was not significantly enhanced. Nonetheless, Tim increased his concerns with assessment criteria when he joined a national training programme where he had in-depth discussions about grading students' speaking with peers from different institutions and universities.

In addition, although the participation in teaching contests prompted him to reflect on his assessment practices and developed his own understanding of “output is assessment”, minimal communication with colleagues during the contests made him believe his participation contributed little to his LAL advancement. Informed by the aforementioned discussion, Tim's LAL developmental trajectory is presented as follows (Fig. 7.2).

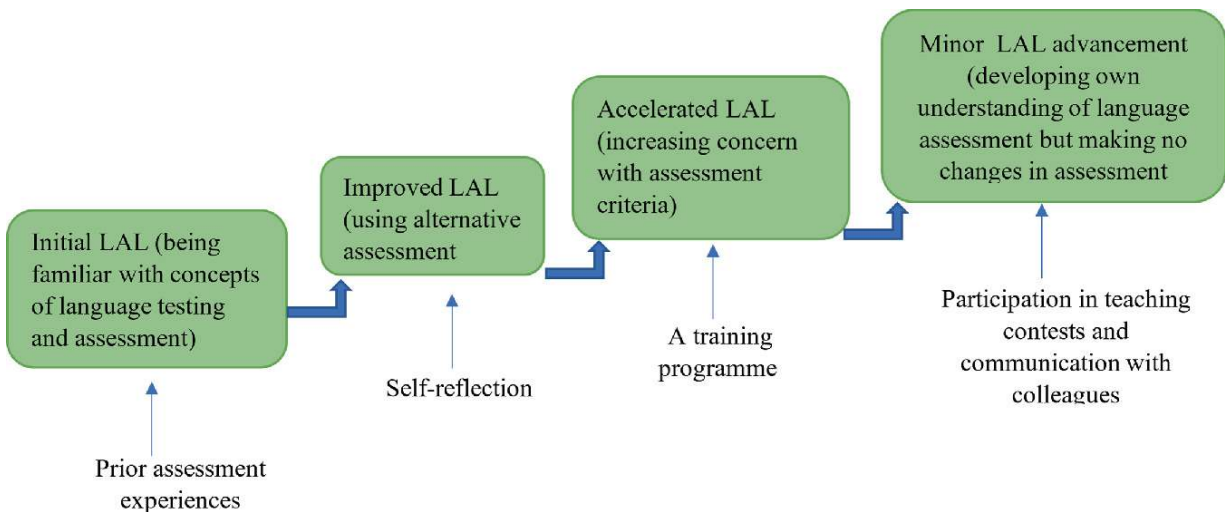


Fig. 7.2 A Trajectory of Tim's LAL Development

7.7.3 Assessment Identity

At the beginning of the study, Tim did not perceive himself as an assessor. He was hesitant about his assessor identity partly because initially he seldom conducted assessment practices and partly because he did not understand what being an assessor meant.

As our study proceeded, Tim began to develop a self-awareness of his assessor identity. Probably because the research interviews triggered Tim to think over his assessment practices, in the second and third semesters, he asked students to self- and peer-assess their own and fellow students' writing based on given criteria. With the

increasing use of assessment criteria in these assessment practices, Tim realised that he was actually “doing an assessor’s job”. The development of self-awareness of being an assessor was more ascribed to his reflection on the responsibilities of being an assessor.

Without a strong recognition of his assessor identity, Tim was actually a confident marker and feedback giver. He produced speaking criteria, against which scores on students’ speaking performance were given globally. In the meantime, he took the speaking tests as an opportunity to provide students with individualised feedback. While he pointed out individual students’ strengths and weaknesses in their speaking performance, he also offered suggestions on how to improve their speaking skills.

Tim was also a lenient marker. Despite a claim that his grading was criteria-based, objective, and fair, he graded students’ language performance generously and raised students’ formative assessment scores purposively. Being aware that his score-inflating practice was unfair to the higher achievers, he still compromised to the social consequence of score use over the ethical grading. This practice, however, was against his self-proclaimed identity as “a life mentor” who aimed to develop students’ morals in language testing and assessment by setting good examples.

7.7.4 Challenges and Opportunities for LAL Development

As all assessment activities Tim had conducted posed no challenges to his assessment knowledge base, he had no urgent need for assessment learning or training to improve his LAL. Accordingly, he viewed no challenges to his LAL development. Tim was content and comfortable with his CET4/6-orientated assessment practices and confident that he had been competent in grading and giving students individualised feedback.

Despite reduced assessment time caused by a large class size, slashed teaching hours and students’ negative responses, Tim seemed not to be disturbed too much by the low self-efficacy towards assessment activities. Instead, he insisted on the self-righteous assessment practices to “do his job” as an instructor and an assessor. Overall, Tim believed that he had solved all the assessment problems

through self-exploration and reflection and that his LAL could be enhanced with an increasing accumulation of teaching experience.

Tim was also given affordances for his professional development in language testing and assessment. For instance, he was recommended to take teaching contests, as they enabled him to have direct communications with experienced colleagues, despite few useful comments on testing and assessment. Additionally, he was sent to attend a training programme by the department in the summer holidays. The programme offered him an opportunity to communicate with peers from different universities sharing good practices regarding grading and assessment criteria. Nonetheless, Tim did not agentively take this opportunity to enhance his LAL in these two areas after the programme.

References

- Arrafii, A. M. (2020). Grades and grade inflation: Exploring teachers' grading practices in Indonesian EFL secondary school classrooms. *Pedagogy, Culture & Society*, 28(3), 477–499. <https://doi.org/10.1080/14681366.2019.1663246> [Crossref]
- Deneen, C. C., & Hoo, H. T. (2021). Connecting teacher and student assessment literacy with self-evaluation and peer feedback. *Assessment & Evaluation in Higher Education*, 1–13. <https://doi.org/10.1080/02602938.2021.1967284>
- Zhao, H. (2014). Investigating teacher-supported peer assessment for EFL writing. *ELT Journal*, 68(2), 155–168. <https://doi.org/10.1093/elt/cct068> [Crossref]

8. Discussion

Ling Gan¹✉ and Ricky Lam²✉

- (1) School of Languages and Communication, Beijing Technology and Business University, Beijing, China
- (2) Department of Education and Psychology, Hong Kong Baptist University, Hong Kong, China

✉ **Ling Gan**

Email: linggan@btbu.edu.cn

✉ **Ricky Lam (Corresponding author)**

Email: rickylam@hkbu.edu.hk

8.1 Overview

In this chapter, we synthesise and discuss the four university English instructors' LAL needs and profiles. University English instructors have four major LAL needs: knowledge of language assessment, language assessment skills, principles of language assessment and assessment identity. Different from Taylor's LAL profile for language teachers, the LAL profile for university English instructors we proposed embraces eight components with suggested levels. We revisit their LAL development trajectories and address research questions by discussing the complexities and the process of LAL development within specific work contexts. We also analyse the four instructors' assessment identities and the changes in their perceptions of these identities, identifying four relationships between assessment identity and LAL development. It is also found that institutional and classroom contexts, assessment experiences and instructors' conceptions of language

assessment posed both challenges and opportunities for instructors' LAL development and assessment identity construction.

8.2 LAL Needs for University English Instructors

Drawing on the findings and referring to the previous operationalisation of the eight components in Taylor's LAL, the current study discovered that university English instructors' LAL needs encompass components delineated as follows. These components intertwine with those in Taylor's profile but focus on specific assessment-related skills.

8.2.1 Knowledge of Language Assessment

Among a wide spectrum of LAL knowledge teachers are expected to possess, knowledge of language pedagogy and theory is the most essential for university English instructors in the Chinese context. This construct entails "language pedagogy" and "knowledge of theory" from Taylor's LAL profile. It pinpoints that language teachers should know how language teaching, learning, and assessment interact in specific contexts. They should understand the effect of assessment on teaching and learning, which is one of the LAL needs required for language teachers (Taylor, [2013](#)).

Teachers' language pedagogy guides their assessment focus. In this study, the instructors showed a strong understanding of how Chinese students learn English, selecting methods and creating opportunities to assess language abilities. For example, Emma emphasised speaking, viewing it as central to language proficiency, while Rachel experimented with speaking test methods to align with students' abilities. Ryan focused on writing and translation as key proficiency indicators, involving students in various writing genres and weekly translation tasks. Tim prioritised lectures on CET 4/6 past papers to help students pass the standardised test, addressing their needs directly.

Understanding language assessment theory, including concepts like validity and reliability, is another critical component. This component,

“language assessment theory”, is subsumed with the “concepts” under the component “principles and concepts” in Taylor’s LAL profile. All the participants in the study could cite a few basic concepts of language assessment with varying degrees of nuance, due to their prior experiences of being trained and assessed. However, participants’ LAL in this regard differed in assessment practices. For example, Emma and Ryan effectively used peer assessment, while Rachel initially struggled but improved over three semesters. Tim was well-versed in assessment theory but did not apply it in practice.

Knowledge of purposes and contexts of language assessment is another LAL component university English instructors need to acquire. It covers the understanding of curriculum/course requirements and institutional assessment policies, requirements, mandates, etc., which are usually linked with different purposes of assessment. This component overlaps with “local practices” in Taylor’s LAL profile. That is to say, language teachers need to know well local and institutional curriculum/course and assessment requirements or policies because these contextual factors could direct their assessment understanding and practices to a large extent (Scarino, [2013](#)). In the current study, all participants except Tim adhered to departmental assessment tasks, with Tim focusing more on CET 4-orientated lectures than formative assessments.

8.2.2 Language Assessment Skills

All participants demonstrated assessment skills, though Ryan sought more theoretical grounding and Tim showed less interest. This finding confirms the urgent need for assessment skills in language teachers’ LAL (Fulcher, [2012](#); Vogt & Tsagari, [2014](#)). This is unsurprising, as university English instructors in the study were observed to spend most of their time employing assessment skills in the classes. These skills encompass technical proficiency in item writing, statistics, and test analysis (Davies, [2008](#)), as well as classroom assessment via alternative methods, feedback, grading, and standardised test preparation.

Using various assessment methods, traditional or alternative, formal or informal, to assess students’ language skills was observed in all the participants’ classes, with differing frequency of use. For

instance, quizzes, presentations, Q&A, role plays, group discussions, and peer assessment were frequently employed to assess students' language performance. Whereas informal and traditional assessment methods were more often used, alternative assessment methods were rarely included as assessment routines, although peer assessment has been regularly used in Emma's and Ryan's classes. This finding corroborated Gan and Lam ([2020](#)) study that few alternative assessment methods were used in university English instructors' classes owing to their lack of training in this regard. While instructors did not avoid alternative assessments, they were not routinely applied (Babaii & Asadnia, [2019](#); Sultana, [2019](#)). As such, a combination of structured (e.g., pen-and-pencil tests or oral/written tests) and alternative ways of assessment is proposed to ensure a positive atmosphere and attitude towards language teaching and learning among students (Vogt, [2018](#)).

Giving feedback is one of the most needed skills in language teachers' LAL. Student interviews revealed a need for instructors' timely, informative, focused and descriptive feedback to guide improvement in learning. For example, Emma excelled at providing oral hints, but not at giving written feedback to students' portfolios. Conversely, Ryan was quite skilled in offering detailed written feedback on their writing performance, though often delayed. Tim, a novice instructor, did quite well in delivering individualised feedback on speaking, while Rachel, also a novice instructor, initially found it difficult to give feedback, but later, learned to provide informative oral feedback over time.

Grading is another indispensable skill required by university English instructors to make value judgement on students' language performance and achievement. In the study, participants graded assignments, participation, speaking, and final exams, supported by criteria. A grading sheet and assessment criteria ensured their grading practice was complete and comprehensive, although assessment criteria were not always transparent to students. Notably, all the instructors graded students by considering non-achievement factors, including students' efforts and attitudes. These factors, called academic enablers, were one of the challenging issues in grading (McMillan, [2008](#)), since they may lead to grade inflation. The inflated grades

highlighted university English instructors' unethical assessment practice, as their emotional concerns with students, instead of assessing students' performance against the criteria, might override instructors' knowledge of quality grading (Edwards, [2017](#)). However, this inflation practice was justified by the participants, because it was seen as beneficial and fair to students. This finding is in line with two grading factors, namely, appeal to "what is fair" and "what is beneficial for students", identified by Sun and Cheng ([2014](#), p. 326).

While the aforementioned components of LAL are commonly needed by all language teachers, preparing students for standardised tests, particularly CET 4/6, is specific to Chinese context LAL needs. Despite being low-stakes, all instructors in the study devoted a considerable amount of their time to preparing students for CET 4/6. For instance, Emma provided students with CET 4/6 listening note-taking strategies. Rachel's assessment activities were CET 4/6-orientated (e.g., CET 4 vocabulary tests, CET 4/6 writing topics). Differently, Ryan and Tim used part of their instruction time to lecture CET 4/6 past test papers. Except for Emma, all the participants believed that it was their primary responsibility to prepare students for CET 4/6. Test preparation was not expected to merely increase students' scores on the test. Instead, it was anticipated to motivate students in English learning, an instance of positive washback on College English instruction. Besides, test preparation contributes to enhancing students' language performance and promoting their self-regulated learning (Lam, [2013](#)).

Most of the existing literature regarding test preparation documented its negative washback on language teaching and learning, as it narrowed curricula, caused "teaching-to-tests", and brought about pressure on teachers and students (Gebril & Eid, [2017](#); Winke & Lim, [2017](#)). However, when the stakes of standardised tests decrease, the accountability requirements of these tests on teachers are no longer in tension with students' language learning. In this case, test preparation is taken as a means to boost students' learning motivations by considerably exposing students to language learning materials. Therefore, preparing students for standardised tests is crucial for university English instructors' LAL if they can understand the difference between these large-scale, standardised tests and classroom

assessments, and how to adapt the principles of the former and use them in the latter (Berger, [2020](#)).

Additionally, university English instructors' LAL needs should entail two important assessment skills, although little data about these two skills were revealed. They were writing items and analysing test results. The former referred to writing test items for various purposes and the latter meant using descriptive statistics to analyse the results of tests. These two skills were not considered by all participants in the study, possibly because they were unaware of them, as writing test items was not an assigned task for every instructor and analysing test results was usually done automatically by computer software. As these two skills are commonly embedded into university English instructors' assessment practices (Jiang, [2020](#)), they are indispensable components in instructors' LAL.

8.2.3 Principles of Language Assessment

Apart from assessment knowledge and skills, principles of language assessment could not be neglected in language teachers' LAL profiles. This component, similar to principles proposed by Davies ([2008](#)), refers to ethical assessment practices among university English instructors. They need to understand the consequences of language assessment on students, consider fairness in their assessment practices, and align assessment practices with local contexts or values.

In the current study, all four instructors experienced ethical assessment practices. For instance, both Ryan and Tim were aware of the impact of students' scores or grades on their future employment, scholarships or further study, but they chose different ways of grading. Ryan marked students' language performance by strictly following grading guidelines without considering non-achievement factors. He insisted that the given scores/grades should honestly represent students' authentic language proficiency, which would benefit them in society. Different from Ryan, Tim graded leniently by inflating students' scores, as he asserted that high scores helped students stand out in the prospective employment market. Inferably, Ryan and Tim aligned their grading practices with the social values of score use, although the way of alignment varied because of their differing understanding of ethics in assessment. In addition, both Emma and Rachel had to make a

compromise when their principles of grading were in conflict with departments or universities' requirements. Taken together, it seemed that principles of language assessment cover “principles”, “sociocultural values”, and “personal beliefs/attitudes” in Taylor’s LAL profile.

8.2.4 Assessment Identity

In addition to the three aforementioned components for LAL needs (e.g., Davies, [2008](#)), teachers’ self-awareness of assessment identity, not applicable in Taylor’s LAL profile, is added to the LAL profile for university English instructors. As mentioned in Sect. [2.4.3](#), teachers’ self-awareness of their assessment identity is viewed as part of LAL. Empirically, the present study discovered that when instructors gradually realised their identity as an assessor, they would increase their assessment frequency (Emma), strengthen their understanding of assessment, teaching and learning (Rachel and Tim) and attempt to improve the effectiveness of assessment (Emma and Rachel). Detailed discussions about the connection between university English instructors’ assessment identity and their LAL development can be seen in the following Sect. [8.4](#).

Informed by the discussions, Table [8.1](#) summarises the LAL needs of university English instructors and relates them to the components in Taylor’s LAL profile.

Table 8.1 LAL Needs of University English Instructors in the Chinese Context

Category subcategory		Description	Evidence related to Taylor’s LAL components
Knowledge of language assessment	1. Knowledge of language pedagogy and theory	Understanding how language is taught and learned in specific educational contexts Understanding the relationship between language teaching, learning, and assessment	<i>Language pedagogy/ knowledge of theory</i>
	2. Knowledge of language assessment theory	Understanding content, concepts and methods of language assessment	<i>Principles and concepts</i>

Category subcategory		Description	Evidence related to Taylor's LAL components
	3. Knowledge of purposes and contexts of language assessment	Understanding purposes of assessment, curriculum/course requirements and institutional assessment policies, requirements and mandates, etc.	<i>Local practices</i>
Language assessment skills	1. Using assessment methods	Using various assessment methods (traditional/alternative, formal/informal)	<i>Technical skills</i>
	2. Giving feedback	Providing written or oral comments on students' language performance	<i>Technical skills</i>
	3. Preparing students for standardised tests	Providing students with test-taking strategies	<i>Technical skills</i>
	4. Grading	Giving scores or grades to students' language performance	<i>Grading and decision-making</i>
	5. Writing items	Writing test items for various purposes	<i>Technical skills</i>
	6. Analysing test results	Using descriptive statistics to analyse results of tests	<i>Technical skills</i>
Principles of language assessment	1. Impact of assessment	Understanding influences or consequences of language assessment or one's beliefs/attitudes towards language assessment on students, teachers and universities	<i>Principles and concepts, personal beliefs/attitudes</i>
	2. Fairness	Being fair, objective, ethical and bias-free in assessment practices	<i>Principles and concepts</i>
	3. Alignment of assessment practices	Determining if assessment practices align with local, educational assessment contexts, policies, requirements and mandates Determining what impact social and cultural values have on the use of assessment in a given context	<i>Sociocultural values</i>
Assessment identity	Self-awareness of assessment identity	Being aware of the identity as an assessor and other identities in the assessment	<i>N/A</i>

8.2.5 An LAL Profile for University English Instructors

Informed by the LAL needs of the four participants, an LAL profile was proposed for university English instructors in the Chinese context (see Fig. 8.1), in contrast to that for classroom teachers stipulated by Taylor (2013). These two LAL profiles differ in the following ways.

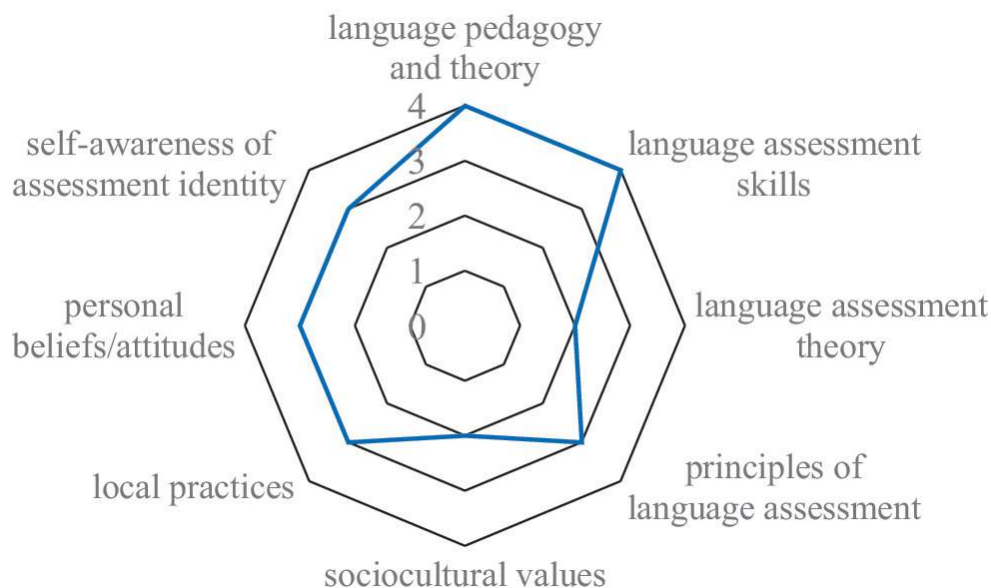


Fig. 8.1 An LAL Profile for University English Instructors in the Chinese Context

- *Knowledge of theory* and *language pedagogy* have been merged into *language pedagogy and theory* to indicate both procedural knowledge and conceptual knowledge of language pedagogy. This component has been required as the highest level (i.e., Level 4) as it is in Taylor's profile.
- *Principles and concepts* are separated, with *concepts* subsumed into *language assessment theory* to represent a focus on the theoretical and conceptual knowledge of language assessment and *principles* to *principles of language assessment*, which indicates the ethical aspects of language assessment. The former component remains at Level 2 as required in Taylor's profile, while the latter has been adjusted one level up (i.e., Level 3), compared with that in Taylor's profile.
- *Grading and decision-making* and *technical skills* are blended together as *language assessment skills* to indicate the process, and the procedural nature of skills related to the design, development and administration of language assessment. This component has been elevated to the highest level (i.e., Level 4).

- *Self-awareness of assessment identity*, as a new LAL component, is added to the profile, with a required Level 3.
- The names of *sociocultural values*, *local practices* and *personal beliefs/attitudes* remain unchanged, but all of them showcase broader procedural knowledge as they are operationalised in Table 2.2. The required level for the *sociocultural* component is suggested one level down (i.e., Level 2), while the levels for the last two components remain the same as those in Taylor’s LAL profile (i.e., Level 3).

This updated LAL profile is not intended to give a prescription of the components of LAL but is expected to make a step towards attesting and refining Taylor’s LAL profile for language teachers in practice. It attempts to call for a more data-driven conceptualisation of LAL, with clearly delineated components aligned with local contexts. Therefore, the components and levels of LAL for university English instructors are not permanent. Instead, they are suggested according to the findings of the present longitudinal study and our experiences of being a former university English instructor, similar to LAL profiles for novice and experienced language teachers in Xie and Tan’s (2019) study.

8.3 LAL Developmental Trajectories

The proposed LAL profile appears static, but its development among university English instructors is a dynamic process. This study reveals that instructors advance their LAL through a bottom-up, iterative approach, echoing findings from Yan and Fan (2021). It was also found that four instructors acquired fuller knowledge, more skillsets, and increasing awareness of the synergy of teaching and assessment by building on their prior and present experiences. Similar to previous studies (e.g., Xu & Liu, 2009; Yan & Fan, 2021), the study discovered that prior experiences of assessment training, course learning, or being assessed as a student empowered instructors with initial but limited LAL. As Lee et al. (2021) suggested, no one entered the teaching career with illiteracy in language assessment, since language teachers observed and followed their former teachers’ assessment practices to conduct assessment on their students. As such, LAL acquisition experiences in this period were considered an “apprenticeship of

observation” (Lortie, [1975](#), p. 61) and, thus, apprenticeship-based (Fulcher, [2020](#); Yan & Fan, [2021](#)).

The initial LAL of university English instructors varied substantially upon their entry into the teaching profession. For instance, Emma, Ryan and Tim all had basic theoretical knowledge of LTA and were all confident with their classroom-based assessment practices because of their accumulated teaching experience or prior assessment learning/training experiences. Among them, Ryan had more considerable LAL than the other two owing to his assessment learning in BA and MA programmes, rich experiences of different educational systems home and abroad as well as his in-progress doctoral learning. In contrast, Rachel had little and “superficial” understanding of LTA due to her paucity of assessment training and learning experiences. Regardless of the diverse initial LAL, whether and how initial LAL is enhanced and advanced depends on the ongoing dialogical interactions between the instructors and the contexts where they work (Inbar-Lourie, [2017](#)).

Interacting with their work contexts, instructors improved their LAL over time. Two patterns emerged. First, all the participants seemed to heighten their awareness of promoting students’ English learning through assessment. Second, they all increased student engagement in assessment practices (e.g., the increasing use of peer assessment). These patterns illustrate that language teachers’ LAL development is evidenced both propositionally and procedurally (Baker & Riches, [2018](#)).

Despite similar patterns, LAL development is unique to each instructor, which concurs with the views of Edwards ([2017](#)) and Yan and Fan ([2021](#)). Firstly, the extent of LAL development of university English instructors differed. Among the instructors, Rachel showed the most significant LAL growth. She had more exposure to assessment knowledge (e.g., critical reading assessment, *China’s Standards of English Language Ability*), gained more skills (e.g., giving feedback, using peer assessment) and catered more for students’ needs in assessment practices. Tim was also observed to develop his LAL in assessment skills (e.g., using self- and peer assessment methods), but his development seemed to be minor, tepid, and incidental. Different from Rachel and Tim, both Emma and Ryan, already having

comparatively high LAL at the beginning, showed steady awareness increases but little change in practice. This finding suggests that experienced English teachers sometimes are adamant about maintaining their signature assessment practices, even though they are subject to a new assessment milieu (Pu & Xu, [2021](#)).

The differing extent of LAL development demonstrates that instructors' LAL development is not necessarily all-rounded, foregrounding both strengths and weaknesses of language assessment knowledge, skills, and practices. This finding is consistent with Taylor's ([2013](#)) view that not everyone develops a comprehensive LAL profile. Teachers may gain development in one aspect of assessment but still remain novice in another (Adie, [2013](#)). For instance, despite an increase in conceptual knowledge of language testing and assessment, Rachel was not confident in evaluating the effectiveness of assessment methods. There was also much room for her to improve her expertise in giving feedback on students' language performance. Similarly, Tim, though versed in feedback, did not significantly enhance theoretical knowledge.

Moreover, instructors' LAL development remains incomplete, with areas for improvement. In this study, Emma enhanced her nuanced understanding of peer assessment, but she focused on students' involvement rather than peer feedback. Likewise, Rachel fostered assessment skills by giving feedback, but the feedback was not specific and informative enough for students' learning. She also advanced her LAL in using peer assessment, but did not care much about the quality of peer feedback. Similarly, Tim integrated self- and peer assessment into writing and speaking but failed to use them effectively. He neither instructed students how to peer assess, nor required students to generate feedback. Seemingly, despite LAL enhancement, these instructors were unconscious of deficiencies in their assessment skills.

Instructors' uneven LAL development in assessment practices could be explained by two aspects. First, as Hill ([2017](#)) asserted, teachers were short of the means to reflect on their assessment practices regularly. Although instructors reported reflecting on their assessment practices, it was challenging for them to detect deficiencies in their practices without specific guidelines that could be obtained from professional training or learning. Second, as most instructors learn

assessment practices on the job, which lacks theoretical assessment underpinnings (Vogt & Tsagari, [2014](#)), there seems to be no grounds for truly comprehensive improvement in their assessment practices. These two explanations underscore the importance of professional guidance and continuous assessment training/learning for teachers' LAL development. With specific guidance and theoretical learning, teachers' LAL tends to grow more comprehensively over time (Harding & Brunfaut, [2020](#)).

Secondly, university English instructors went through individualised LAL development routes (see Sects. [4.7.2](#), [5.8.2](#), [6.7.2](#), and [7.7.2](#)). This finding is not unexpected, as teachers in the same instructional contexts sharing the same assessment requirements or policies would have their own self-developed and individualised assessment practices (Poehner & Inbar-Lourie, [2020](#)), not to mention those in different institutional cultures where delivery of course content and designated assessment tasks vary considerably. Recognising unique institutional cultures is to situate learning and assessment within classroom contexts, thus creating a locally appropriate assessment culture (Inbar-Lourie & Levi, [2020](#)).

In the current study, individual instructors' LAL development usually started with their initial but limited LAL acquired from their past assessment experiences. Some advanced their initial LAL through academic reading to meet their or students' assessment needs (Rachel and Ryan). Some attended seminars, lectures, or conferences to gain new concepts or content of language assessment (Emma and Rachel). Some self-reflected on their assessment practices to acquire assessment skills (Rachel and Tim) or participated in teaching contests to accumulate assessment experience (Rachel and Ryan). These instructors made use of handy resources to enhance their initial LAL by interacting with their work contexts (Yan & Fan, [2021](#)), treading their unique paths of LAL development.

The different routes of LAL development also led to idiosyncratic rates of LAL development. In this study, Rachel gained the fastest and the most substantial development in LAL over time among the four participants, i.e., increasing assessment awareness, knowledge, and skills. She was also the most willing one to experience and demonstrate more changes in assessment. This finding partially corroborates H. Xu's

([2017](#)) claim that novice teachers tend to experience more fluctuations in professional development than their veteran counterparts. Rachel's progressive LAL development could be, at least, partly ascribed to the agency she exercised and the self-efficacy she gained from her assessment practices which in turn galvanised her agency to LAL enhancement. The other novice instructor, Tim, in contrast, because of his low self-efficacy towards assessment practices, seemed not to be so concerned with his LAL, which resulted in a tepid LAL acquisition. These findings emphasised the influence of agency on teachers' LAL development (Mansouri et al., [2021](#)), aligning with Xu and Liu ([2009](#))'s finding that teachers' agency was crucial for assessment knowledge construction.

Differently, experienced instructors, Emma and Ryan, were adept in using assessment skills and conducting assessment practices by building on their accumulated teaching experience, as DeLuca et al. ([2018](#)) revealed. Nonetheless, they tended not to innovate their assessment practices. It was not because they resisted changes but because it was not a good time for changes. As they had built up regular assessment practices through iterative trialling, they did not feel a need to change these off-the-shelf practices. Interestingly, the mindset of the experienced instructors in the current study differed from that of those who were going to experience drastic assessment reform in which they would undergo substantial changes in assessment practices (Carless, [2005](#)).

LAL development was accelerated through agentic activities and institutional mandates. The former includes conducting assessment-related research (Emma), communicating with collegial peers (Rachel and Ryan), attending seminars, lectures or conferences (Emma and Rachel) and engaging in academic reading (Rachel and Ryan). The latter covers the department's demands or recommendations on taking teaching contests (Rachel, Ryan and Tim) or joining training programmes (Emma, Rachel and Tim). However, some interactions, like Tim's teaching contests and Emma's grading discussions, offered limited LAL enhancement. Overall, all the instructors benefited from their reflections on assessment practices and accumulated teaching experiences to advance their LAL both in assessment awareness and

certain assessment practices. In view of these discussions, Fig. 8.2 illustrates a common LAL developmental trajectory for the instructors.

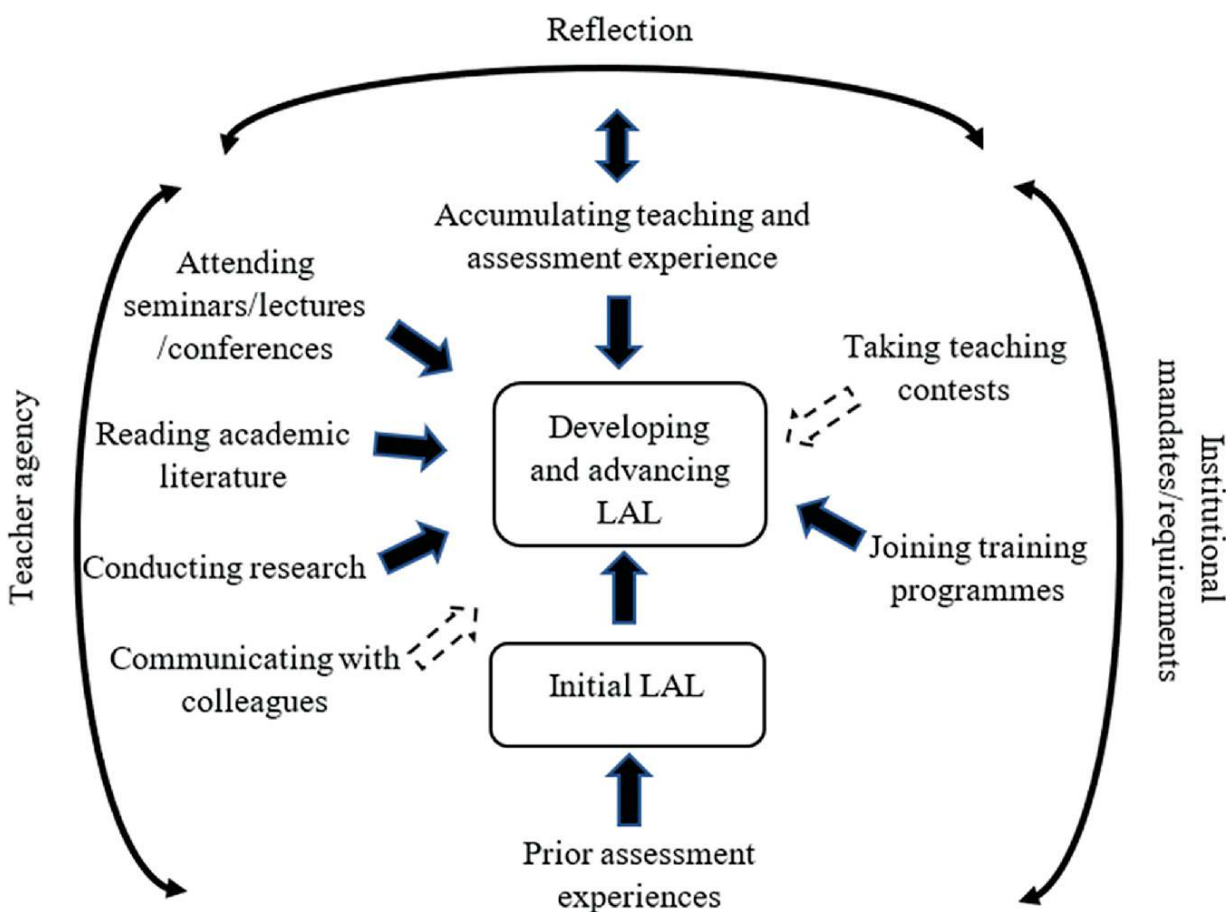


Fig. 8.2 A Trajectory of LAL Development for University English Instructors (Note ➡ indicates “contribute to” -.-➡ means “may or may not contribute to”)

The study emphasises caution in claiming language teachers lack assessment literacy due to limited formal training. As Vogt et al. (2020) noted, language teachers gained assessment-relevant experiences even without formal assessment training. This study shows that language instructors can develop their LAL to various degrees through continual practice and reflection on assessment practices. This has confirmed the findings from Levi and Inbar-Lourie (2019) and Yan and Fan (2021), who suggest that repeated practice and reflection foster LAL development. Self-reflection can reveal instructors’ conceptions of language assessment, reposition the relationship between assessment and instruction (Scarino, 2013) and enable instructors to become

“more empowered to have their own agency in the language assessment process” (Babaii & Asadnia, [2019](#), p. 758).

Overall, language teachers are not completely assessment illiterate as perceived. Instead, they have been bestowed with initial LAL from their prior assessment experiences (Yan & Fan, [2021](#)). However, becoming fully assessment literate requires sustainable professional guidance and support from language assessment professionals. Language teachers also need to adopt the role of “a testing citizen” (Fulcher & Davidson, [2007](#), p. 367), adhering to ethical practices as they integrate their assessor identity into their work.

8.4 Construction of Assessment Identity

As mentioned previously, self-awareness of assessment identity should be part of LAL needs for university English instructors, yet our study shows that instructors build their assessment identity alongside LAL development. Extant studies often assume that teachers are naturally labelled as assessors immediately when they conduct assessment activities and develop their LAL (e.g., Berry et al., [2019](#); Dao, [2021](#)). Few studies look into how teachers feel about being assessors, how they view their identity as assessors, and what assessment identities they undertake other than assessors. It is a good sign that there has been a call that due emphasis should be given to how teachers “feel” to be assessors (Estaji & Ghiasvand, [2021](#); Moqaddam et al., [2021](#)). This call is now responded to in the current study.

Initially, the four participants did not fully see themselves as assessors. Being an assessor is only one of the identities in assessment practices and there are interpersonal differences in the extent to which they identified themselves as assessors. These variations might partly come from their perceived role of a language instructor. They believed that their jobs automatically entailed assessment activities and that an instructor was supposed to help students rather than judge students only. It seems that if teachers focus much on assessment in their teaching, they would be torn between their roles as “a facilitator or monitor of language development and that of an assessor or a judge of language performance and achievement” (Rea-Dickins, [2004](#), p. 253)

and “they may feel ethically conflicted about their role as an assessor” (Sheehan & Munro, [2019](#), p. 6).

All instructors noted that their role as an assessor related primarily to scores or grades, namely, outcomes of assessment, reflecting their intuitive understanding of language testing and assessment largely shaped by their exam-orientated backgrounds in which testing and assessment are mainly related to summative paper-and-pencil tests whose scores or grades are the decisive criteria for students’ learning performance (Cheng & Fox, [2017](#)). This perspective separates assessment from everyday teaching and confines the assessor role to testing, not supporting learning. That is, being “assessors” is alienated from being “instructors”. Consequently, instructors do not realise that their engagement in assessment also develops their assessment identity, which in turn, consolidates their identity as an instructor (Adie, [2013](#)). Becoming an assessor requires an understanding update in assessment knowledge and practices (Xu & Brown, [2016](#)).

Instructors’ perceived assessment identity might or might not change when they became engaged in assessment practices from which they enhanced and advanced their LAL. This is partially consistent with Scarino’s ([2013](#)) claim that language teachers’ self-awareness of being an assessor would gradually develop along with their LAL enhancement. In the current study, Emma, initially unaware of her assessment identity, grew aware through peer assessment and research, yet found it strained her student rapport, prompting a reduced assessor role. Rachel, too, began as being unaware, but as her assessment skills and knowledge grew, she embraced her novice assessor identity. Tim doubted his role of being an assessor initially, but reflection made him recognise his assessment activities and hence his assessor identity. Conversely, Ryan never identified himself as an assessor, seeing it separate from his instructor role.

These findings align with previous studies (e.g., Sheehan & Munro, [2019](#); Xu, [2019a](#)), showing assessment identity as a dynamic, iterative process. In this process, instructors’ understanding of their assessment identity can change or stay the same with LAL development, highlighting its non-linear nature (Dao, [2021](#)). Thus, four quadrants of instructors’ assessment identity and LAL development are proposed in Fig. [8.3](#).

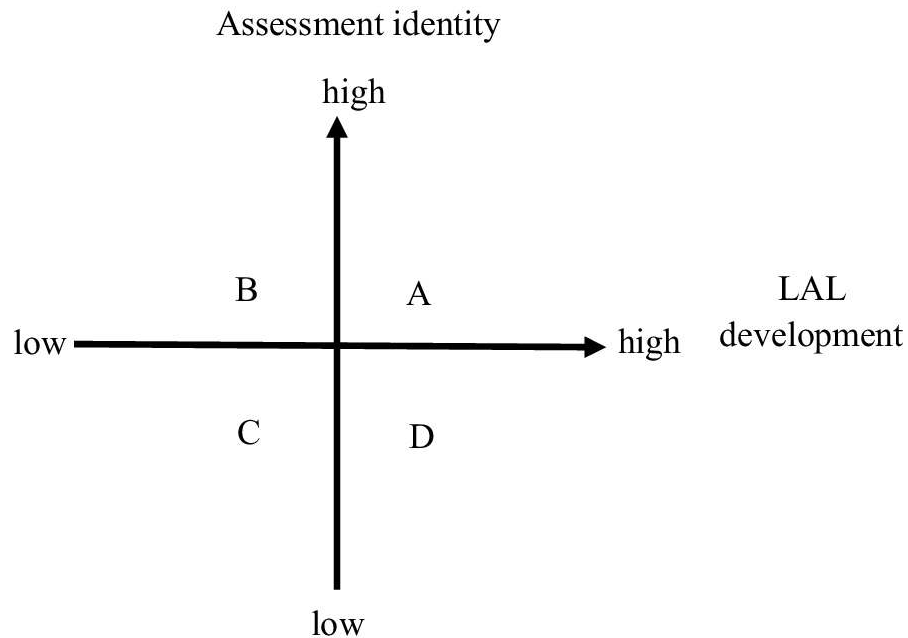


Fig. 8.3 Four Quadrants of Instructors' Assessment Identity and LAL Development

In quadrant A, instructors' assessment identity is incrementally constructed along with their LAL development. Instructors in this quadrant develop a strong self-awareness of being an assessor to promote students' learning through assessment tasks when they advance their LAL. In the meantime, the construction of the assessor identity lends itself to instructors' LAL development. By contrast, instructors in quadrant B are aware of their assessor identity and the attached responsibilities, but are reluctant or clueless to enhance their LAL, even though they realise LAL development is beneficial for their instructions and assessment.

Similar to those in quadrant B, instructors in quadrant C often have low LAL development. They also have a weak identity as an assessor. They might downplay the importance and benefits of language assessment in language teaching and learning. They might also just begin their teaching profession with little or even no learning or training in language assessment. However, instructors in the last quadrant (quadrant D) develop, refine, and advance their LAL consciously or unconsciously by improving their assessment practices. Yet, their self-awareness of the identity as an assessor is still weak and vague. For them, being an assessor only means grading and judging students instead of enhancing student learning through assessment.

Here it should be noted that the connection between LAL and assessment identity may not be as distinct as or limited to what the four quadrants present. How assessment identity and LAL are connected requires further exploration and explanation in future research.

Other than being an assessor, the four university English instructors viewed themselves as taking on multiple identities “when they became immersed in practices” (Adie, [2012](#), p. 55) and worked for different assessment purposes. They focused on an identity comfortable for them in specific contexts (McDowell & Harman, [2008](#)) and often felt obliged to select an identity to stick to (Rea-Dickins, [2004](#)), be it an assessment facilitator, a guide, a feedback giver, a score/grade giver, or an emotional supporter. These identities were attuned to instructors’ understanding of the multiple roles assessment plays in teaching and learning.

Instructors’ multiple identities coexisted throughout the assessment, with one or two identities foregrounded in an instruction situation. For instance, both Emma and Ryan identified themselves as assessment facilitators who designed and implemented assessment activities catered to students. They anticipated students’ learning could be regulated and promoted through assessment. Meanwhile, they underscored the identity as an emotional supporter. As an instructor, they cared more about students’ emotional distress caused by assessment and mitigated the negative impact of assessment on students. Yet, it should be noted that sometimes, when instructors are stuck with assessment dilemmas, they are torn apart by their identities as an instructor and a facilitator of students’ language development and as an assessor of students’ language performance (e.g., Emma and Rachel). Therefore, these multiple identities, as Moqaddam et al. ([2021](#)) asserted, required instructors to achieve different levels of LAL to enact in assessment practices.

Apart from their narrated identities (e.g., an assessment facilitator, a feedback giver, etc.), all instructors unconsciously constructed their assessment identity through day-to-day assessment practices. This construction was juxtaposed with their LAL development, contrasting Xu and Brown ([2016](#))’s assertion that becoming an assessor is the pinnacle of teachers’ assessment literacy development. For example,

Emma became an eager assessment researcher by immersing herself in research on peer assessment. Rachel formed her identity as an assessment explorer by trialling assessment methods. Ryan maintained his identity as an experienced assessment practitioner by optimising his assessment activities and Tim empowered himself to be a confident test-orientated practitioner with CET 4/6 instructions. These differing identities revealed that instructors enacted different aspects of their assessment identity as part of their assessment practices, depending on the assessment needs and demands they have to follow (Leonardsen et al., [2022](#)).

The construction of assessment identity among university English instructors shows it is dynamic and constantly evolving (Looney et al., [2018](#)), involving changes in perceptions of assessment's role. This finding corroborates that of Cowie et al.'s ([2014](#)) study in which pre-service teachers (re) shaped their assessment identity, often mediated by diverse factors in their assessment practices. Figure [8.4](#) outlines the trajectory of identity construction among the instructors.

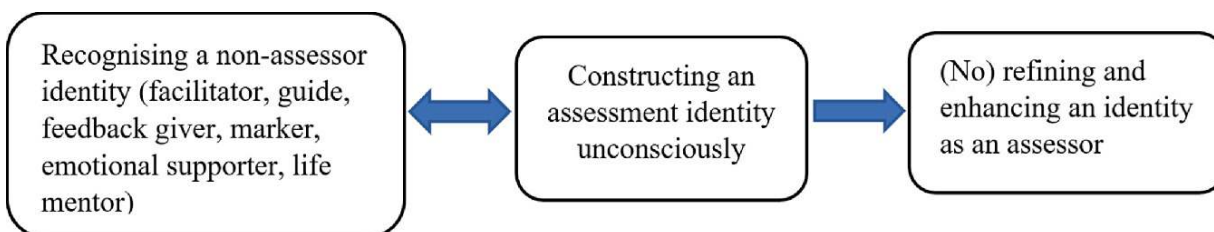


Fig. 8.4 A Trajectory of Instructors' Construction of Assessment Identity

Overall, constructing teachers' assessment identities is valuable yet challenging. It is valuable because the success of language testing and assessment is often ascribed to the objectives assessors expect to achieve (Hidri, [2021](#)). It is challenging because language teachers' intuitive understanding of language testing and assessment has often led to their misconceptions of the identity as an assessor. In the exam-driven Chinese context, despite increasing opportunities, language assessment remains marginalised in English teaching at the tertiary level. To promote language assessment in teaching and learning, one objective of professional training and learning for language teachers' LAL is to build an integrated approach towards the assessment identity and to understand the responsibilities and practices of an assessor (Lan

& Fan, [2019](#)). The driving force for the training and learning is that language teachers should be aware that their LAL is always developing and evolving (DeLuca & Braund, [2019](#)).

8.5 Challenges and Opportunities for LAL Development

Idiosyncratic LAL development and observed assessment identities in this study were possibly contingent on the assessment contexts (e.g., assessment requirements, institutional support in assessment), assessment experiences, and instructors' personal dispositions and beliefs (e.g., agency, self-efficacy, awareness of being an assessor). These factors may pose challenges to constrain or provide opportunities to facilitate individual instructors' LAL development with varied degrees. The following sections focus on contextual, experiential, and personal factors that may engender challenges and opportunities for instructors' LAL development.

8.5.1 Institutional and Classroom Contexts

Similar to previous studies (e.g., Vogt et al., [2020](#); Xu, [2019a](#)), this study discovered that institutional (meso-) and classroom (micro-) factors significantly affect language teachers' LAL development. The former included institutional assessment requirements and mandates, teaching workloads, and collegial communication. The latter covered students' responses towards instructors' assessment practices. These contextual factors either constrained or facilitated the extent to which instructors' LAL developed.

Departments' assessment requirements directed instructors' attention to assessment practices and supported or constrained instructors' commitment to language assessment and LAL development. Specifically, mandated assessment tasks ensure alignment between teaching and assessment. Detailed assessment plans and tasks with specific guidelines could provide resources and support for instructors to conduct effective assessments. For example, Rachel benefited from departmental guidance, enhancing her LAL with institutional support. Such support helps instructors, especially

novices, feel valued within the language assessment community, easing LAL development.

Institutional mandates can prompt instructors to adjust assessment practices but may also force ethical compromises, such as inflating grades per institutional demands, as Emma and Rachel experienced. Conversely, Ryan adhered to strict grading policies, illustrating how differing institutional cultures influence assessment practices. Institutions focusing on learning foster a learning-orientated assessment ecosystem and encourage professional growth in assessment (Inbar-Lourie & Levi, [2020](#)).

Heavy workloads impact instructors' dedication to assessment, resulting in mixed LAL development. This finding supports Yan et al. ([2018](#)), as instructors often prioritise instruction preparation over assessment (Rachel), reduce assessment time for lecturing (Tim), and struggle with administrative tasks (Emma) and extracurriculars (Ryan). This limits the time for giving feedback and reading academic papers, hindering assessment knowledge growth. However, instructors did not negatively view assessment due to workload, differing from Crusan et al. ([2016](#))'s finding.

Instructors rely on external resources for professional development, with collegial communication providing valuable feedback. Ryan joined teaching contests for feedback, while Rachel sought collegial help for assessment challenges. Tim engaged in cross-institutional discussions, albeit with little impact on his practices. Emma was encouraged by colleagues to prepare students for CET 4. These interactions confirm findings that collaborations (e.g., with colleagues, students, and researchers) facilitate language teachers' LAL development (Baker & Riches, [2018](#); Lee et al., [2021](#); Xu & Liu, [2009](#)).

Students' responses to language assessment also may shape instructors' self-efficacy towards assessment practices, which, in turn, influences instructors' commitment to LAL development. Rachel and Tim's experiences illustrate this. Both felt offended when their students refused to take speaking tests. While Rachel proactively decreased test difficulty to engage students in speaking tests, Tim took no actions and thus had low efficacy towards his assessment practices. Rachel's strategy worked in the speaking test, bringing her high self-efficacy towards assessment practices. This heightened self-efficacy stimulated

Rachel's willingness to access more assessment knowledge and skills and take risks in assessment methods. In contrast, Tim, due to his low self-efficacy, conducted fewer assessment activities and recognised less his identity as an assessor.

8.5.2 Assessment Experiences

Despite enhancing their LAL in their teaching contexts, the four university English instructors faced significant challenges owing to the lack of systematic and professional assessment training.

Comprehensive assessment training is crucial for sustainable LAL development as it provides essential knowledge, skills, theories, and practices (Malone, [2017](#)). Both Rachel and Tim expressed concern over the absence of assessment training for novice instructors, especially those without a background in assessments. Without early training or guidance, new instructors struggle with formative assessment implementation. Fortunately, Rachel benefited from extensive departmental support through detailed guidelines and examples. This finding confirms Gardner et al. ([2008](#)) claim that guidance can effectively develop teachers' professional understanding of assessment.

Instructors also enhanced their LAL through institutional and self-directed learning opportunities. Both Rachel and Tim were given the opportunity to attend national seminars organised and operated regularly by official organisations. In their summer holidays, they both joined *National Training Programmes for in-service EFL Teachers in Tertiary Education* aiming to strengthen teaching professionalisation. These programmes, while not focused on assessment, provided platforms for peer communication and observation of demo classes.

Nowadays, university English instructors increasingly access assessment seminars, webinars, and conferences via social media, which facilitates the accumulation of assessment experience and hones "assessment intuitions" about theoretical assessment content and concepts (Yan et al., [2018](#), p. 165). Such academic exchanges emphasise the need for continuous learning to inform assessment practices.

The continuous teacher learning can be provided by professional training. Without professional assessment training or learning, instructors' LAL development was not full-fledged (see Sect. [8.2](#)). Instructors require training to effectively reflect on their assessment

practices and enhance LAL. However, many current in-service teacher education programmes in China fail to address classroom-based assessments or instructors' specific needs (Jin, [2010](#)). For instance, Emma found that departmental IELTS training had little impact on her speaking assessment practices. Therefore, assessment training programmes should be content- and needs-specific (Fulcher, [2020](#)).

Besides assessment training and learning, participating in teaching contests is another valuable way for university English instructors to enhance their LAL. These contests, held at various levels in response to a national call for high-quality and innovative teaching, aim to develop excellent language teachers suited to the Chinese context. In this study, Rachel, Ryan, and Tim all took part in teaching contests at different levels. Although Emma did not compete during the study period, she supported a colleague preparing for a contest and had previously placed third in a national English teaching contest, which aided her career advancement.

Teaching contests offer a pathway for instructors to develop their LAL by presenting demo classes and engaging with veteran colleagues. These contests, whether institutional, regional, or national, have received significant attention and support from departments or universities, which form panels of experienced instructors to assist participants. Rachel and Ryan presented demo classes multiple times to veteran colleagues, finding these dialogues valuable for reflecting on instructional and assessment practices. Similarly, Emma's involvement in preparing a colleague for a contest prompted her to reevaluate her assessment knowledge and skills. These experiences heightened Rachel's, Emma's and Ryan's focus on integrating assessment with teaching and then promoting student learning.

Notably, the quality of collegial conversations is crucial for developing teachers' assessment capabilities (Livingston & Hutchinson, [2017](#)). Tim downplayed the role of such conversations, as their feedback, Tim believed, all flattery and complimentary, was unhelpful with both his instruction and assessment. This underscores the importance of supportive, meaningful, and critical dialogues for enhancing LAL.

Although preparing for teaching contests consumes time and energy, potentially diverting focus from classroom assessment and

learning, these contests foster a community of practice within institutions that support and facilitate language teachers' LAL development through interactions and collaborations (Ho & Yan, [2021](#)).

8.5.3 Teachers' Conceptions of Language Assessment

The four case stories revealed that teachers' conceptions of language assessment, as a most salient personal factor, could pose both a challenge and an opportunity for English instructors' LAL development. These conceptions, context-dependent and subject to individual experiences, often shape their assessment practices. This corroborates Lin et al.'s ([2021](#)) view that teachers' conceptions of assessment are a key indicator of teachers' LAL due to their strong link with assessment practices.

Instructors' prior assessment experiences shape their assessment conceptions, which in turn form their assessment practices. Rachel and Ryan, for example, retained traditional methods like quizzes, believing they would benefit their students as they had themselves. Emma, despite disliking paper-and-pencil tests, used a mix of traditional and alternative assessments such as portfolios for formative assessment. Likewise, Tim valued language tests for their positive washback, focusing on CET 4/6 in his teaching. These examples demonstrated that instructors prioritised assessment activities aligned with their assessment conceptions, regardless of their past negative assessment experiences. This is different from Xu's ([2019b](#)) claim.

Teachers' conceptions also affect LAL development pace and extent. Rachel's belief in assessment's role in promoting learning accelerated her assessment learning. Therefore, her LAL development was the fastest and most extended among the four instructors (see Sect. [5.4](#)). In contrast, as Tim held that language tests had positive washback on students' learning, he orientated his teaching towards CET 4/6 test papers (see Sect. [7.4.2](#)). Since he had no difficulty with CET4/6 lectures, Tim was not keen on LAL development. His LAL development was, therefore, comparatively slow, tepid and incidental. This finding echoes those of studies showing conceptions of assessment do influence the scope of LAL development (e.g., Xu, [2019a](#)).

To some extent, instructors' conceptions of language assessment can limit their construction of identity as assessors. Influenced by

exam-orientated cultures, the instructors associated assessment with grading. They did not see themselves as assessors, believing they did not constantly judge students. However, as Xu and Brown ([2016](#)) argued, without shifting their conception of assessment identity, teaching and assessment practices may stagnate. This “uncontested adoption” of existing practice may handicap the adoption of innovative assessment ideas (Tzagari & Vogt, [2017](#), p. 54) and hence, undermine their LAL development.

8.5.4 A Recap of Challenges and Opportunities for LAL Development

The challenges and opportunities mediated by contextual, experiential, and personal factors serve to be filters and amplifiers in instructors’ LAL development and assessment identity construction. Without a nuanced understanding of the challenges university English instructors face in implementing assessment tasks, predetermined professional assessment training programmes and courses that neglect teachers’ specific assessment needs may fail to cultivate their LAL. In this study, the four instructors have become agentic to varying extents to create and draw on available opportunities and/or resources to progress their assessment knowledge, skills, and practices. In this way, they reflect on the difficulties in conducting assessment practices and revamping their assessment repertoire. These reflections can potentially provoke them to shift their identity from an instructor to an assessor (Lee, [2010](#)) and, thus, the shift of the identity would enable the instructors to exercise their agency to self-direct their assessment learning or training (Xu & Brown, [2016](#); Zhang, [2020](#)).

8.6 Summary

In this chapter, we outlined the LAL needs of four university English instructors from four aspects, namely, knowledge of language assessment, language assessment skills, principles of language assessment, and assessment identity. Informed by these needs, we proposed an LAL profile for university English instructors in the Chinese context. This context-specific LAL profile refined and amalgamated the eight components in Taylor’s profile with suggested

LAL levels, attesting to Taylor's LAL profile in language assessment practice. In the remaining sections, we delineated features of four university English instructors' LAL development and portrayed their variable LAL development trajectories. We also explained the four instructors' perceived assessor identity and their construction process. Finally, we presented the challenges and opportunities in the process of LAL development and assessment identity construction, and how the four informants mobilised pertinent constraints and affordances to develop LAL and their assessment identities.

References

- Adie, L. (2012). Learning as identity and practice through involvement in online moderation. *Educational Assessment, Evaluation and Accountability*, 24(1), 43–56. <https://doi.org/10.1007/s11092-011-9132-4>
[Crossref]
- Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, 20(1), 91–106. <https://doi.org/10.1080/0969594X.2011.650150>
[Crossref]
- 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>
[Crossref]
- Baker, B. A., & Riches, C. (2018). 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>
[Crossref]
- Berger, A. (2020). Creating language-assessment literacy theory and practice in EFL teacher education. In B. M.-L. Julia Hüttner & S. R. A. B. Schiftner (Eds.), *Theory and practice in EFL teacher education: Bridging the gap* (pp. 57–82). Multilingual Matters.
- 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>
[Crossref]
- Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 12(1), 39–54. <https://doi.org/10.1080/0969594042000333904>
[Crossref]
- Cheng, L. Y., & Fox, J. (2017). *Assessment in the language classroom: Teachers supporting student*

learning. Macmillan.

Cowie, B., Cooper, B., & Ussher, B. (2014). Developing an identity as a teacher-assessor: Three student teacher case studies. *Assessment Matters*, 7, 64–89. <https://doi.org/10.18296/am.0128>

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>
[Crossref]

Dao, T. B. N. (2021). A narrative inquiry of a Vietnamese university EFL teacher's assessment identity. *International Journal of TESOL & Education*, 1(3), 260–277. <http://i-jte.org/index.php/journal/article/view/109>

Davies, A. (2008). Textbooks in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
[Crossref]

DeLuca, C., & Braund, H. (2019). Preparing assessment literate teachers. *Oxford Research Encyclopaedia of Education*, 1–21. <https://doi.org/10.1093/acrefore/9780190264093.013.487>

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>
[Crossref]

Edwards, F. (2017). *The development of summative assessment literacy: An exploration of the experiences of beginner secondary science teachers in New Zealand* [Doctoral dissertation, The University of Waikato]. The University of Waikato Research Commons. <https://researchcommons.waikato.ac.nz/bitstream/handle/10289/11567/thesis.pdf?sequence=4&isAllowed=y>

Estaji, M., & Ghiasvand, F. (2021). Assessment perceptions and practices in academic domain: The design and validation of an assessment identity questionnaire (TAIQ) for EFL teachers. *International Journal of Language Testing*, 11(1), 103–131. <https://files.eric.ed.gov/fulltext/EJ1291317.pdf>

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
[Crossref]

Fulcher, G. (2020). Operationalising language assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: From theory to practice* (pp. 8–28). Cambridge Scholars Publishing.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge.

Gan, L., & Lam, R. (2020). Understanding university English instructors' assessment training needs in the Chinese context. *Language Testing in Asia*, 10(11), 1–18. <https://doi.org/10.1186/s40468-020-00109-y>

Gebril, A., & Eid, M. (2017). Test preparation beliefs and practices in a high-stakes context: A teacher's perspective. *Language Assessment Quarterly*, 14(4), 360–379. <https://doi.org/10.1080/>

[15434303.2017.1353607](#)

[Crossref]

Gardner, J., Harlen, W., Hayward, L., & Stobart, G. (2008). *Changing assessment practice: Process, principles and standards*. https://www.aaia.org.uk/storage/medialibrary/o_1d94rld7sht0o83k3e3u22128.pdf

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 Nature Switzerland AG. https://doi.org/10.1007/978-3-030-35081-9_4

Hidri, S. (2021). Language assessment literacy: Where to go? In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for improved student learning* (pp. 3–12). Routledge.

Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment*, 6(1), 1–17. http://www.altanz.org/uploads/5/9/0/8/5908292/3.si1hill_final_formatted_proofed.pdf

Ho, E. C., & Yan, X. (2021). Using community of practice to characterize collaborative essay prompt writing and its role in developing language assessment literacy for pre-service language teachers. *System*, 101. <https://doi.org/10.1016/j.system.2021.102569>

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 257–270). Springer International Publishing AG.

Inbar-Lourie, O., & Levi, T. (2020). Assessment literacy as Praxis: Mediating teacher knowledge of assessment-for-learning practices. In *Educational linguistics* (Vol. 41, pp. 241–259). https://doi.org/10.1007/978-3-030-35081-9_12

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>

[Crossref]

Jiang, J. (2020). Gaoxiao yingyu jiaoshi ceping suyang xianzhuang jiqi yingxiang yinsu yanjiu [Study of college English teachers' assessment literacy and its influencing factors]. *Foreign Language World*, 6, 18–26.

Lam, R. (2013). Formative use of summative tests: Using test preparation to promote performance and self-regulation. *The Asia-Pacific Education Researcher*, 22(1), 69–78. <https://doi.org/10.1007/s40299-012-0026-0>

[Crossref]

Lan, C. S., & Fan, S. Y. (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>

[Crossref]

Lee, I. (2010). Writing teacher education and teacher learning: Testimonies of four EFL teachers. *Journal of Second Language Writing*, 19(3), 143–157. <https://doi.org/10.1016/j.jslw.2010.05.001>

[Crossref]

Lee, J., Butler, Y. G., & Peng, X. (2021). Multiple stakeholder interaction to enhance preservice teachers' language assessment literacy. *Languages*, 6(4), 213. <https://www.mdpi.com/2226-471X/6/4/213>

Leonardsen, J. K., Utvær, B. K. S., & Fjørtoft, H. (2022). The five faces of an assessor: Conceptualizing the enactment of teacher assessment identity in vocational education and training. *Educational Assessment, Advanced Online Publication*. <https://doi.org/10.1080/10627197.2022.2106967>
[Crossref]

Levi, T., & Inbar-Lourie, O. (2019). 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>
[Crossref]

Lin, L., Li, G. Y., & Guo, X. (2021). Pre-service Chinese language teachers' conceptions of assessment: A person-centered perspective. *Language Teaching Research*, Advanced online publication. <https://doi.org/10.1177/1362168821996529>

Livingston, K., & Hutchinson, C. (2017). Developing teachers' capacities in assessment through career-long professional learning. *Assessment in Education: Principles, Policy & Practice*, 24(2), 290–307. <https://doi.org/10.1080/0969594X.2016.1223016>
[Crossref]

Looney, A., Cumming, J., Keij, F. V. D., & 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>
[Crossref]

Lortie, D. C. (1975). *School teacher: A sociological study*. The University of Chicago Press.

Malone, M. E. (2017). Training in language assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment, encyclopaedia of language and education* (pp. 225–239). Springer International Publishing AG.
[Crossref]

Mansouri, B., Molana, K., & Nazari, M. (2021). The interconnection between second language teachers' language assessment literacy and professional agency: The mediating role of institutional policies. *System*, 103. <https://doi.org/10.1016/j.system.2021.102674>

McDowell, L., & Harman, K. (2008). 'The designer', 'the teacher' and 'the assessor': Changing academic identities. In *The SRHE Annual Conference: Valuing Higher Education*.

Moqaddam, Q. J., Khodabakhshzadeh, H., Motallebzadeh, K., & Khajavy, G. H. (2021). Measuring EFL teachers' assessment identity: Development and validation of a questionnaire. *Journal of Language and Translation*, 11(1), 29–46. https://journals.iau.ir/article_679444_6270eb6f508688975ac0f2bec03a2081.pdf

McMillan, J. (2008). *Assessment essentials for standards-based education* (2nd ed.). Sage.

Poehner, M. E., & Inbar-Lourie, O. (2020). An epistemology of action for understanding and change

in L2 classroom assessment: The case for praxis. In M. E. Poehner & O. Inbar-Lourie (Eds.), *Toward a reconceptualization of second language classroom assessment: Praxis and researcher-teacher partnership* (pp. 1–20). Springer Nature Switzerland AG.

[[Crossref](#)]

Pu, S., & Xu, H. (2021). Examining changing assessment practices in online teaching: A multiple-case study of EFL school teachers in China. *The Asia-Pacific Education Researcher*, 30(6), 553–561. <https://doi.org/10.1007/s40299-021-00605-6>

[[Crossref](#)]

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249–258. <https://doi.org/10.1191/0265532204lt283ed>

[[Crossref](#)]

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>

[[Crossref](#)]

Sheehan, S., & Munro, S. (2019). *Classroom assessment: The development of teachers' cognition*. https://www.teachingenglish.org.uk/sites/teacheng/files/J175%20ELT%20Research%20Papers%20-%20teacher%20cognition%20and%20assessment_A4_FINAL%20v2_Web.pdf

Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9, Article 1. <https://doi.org/10.1186/s40468-019-0077-8>

Sun, Y., & Cheng, L. (2014). Teachers' grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326–343. <https://doi.org/10.1080/0969594X.2013.768207>

[[Crossref](#)]

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>

[[Crossref](#)]

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. http://www.altaanz.org/uploads/5/9/0/8/5908292/5.si3tsagarivogt_final_formatted_proofed.pdf

Vogt, K. (2018). *Needs analysis report*. https://taleproject.eu/pluginfile.php/2127/mod_page/content/2/TALE%20Needs%20analysis%20report.pdf

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>

[[Crossref](#)]

Vogt, K., Tsagari, D., & Spanoudis, G. (2020). What do teachers think they want? A comparative study of in-service language teachers' beliefs on LAL training needs. *Language Assessment*

Quarterly, 17(4), 410–433. <https://doi.org/10.1080/15434303.2020.1781128>
[Crossref]

Winke, P., & Lim, H. (2017). The effects of test preparation on second-language listening test performance. *Language Assessment Quarterly*, 14(4), 380–397. <https://doi.org/10.1080/15434303.2017.1399396>
[Crossref]

Xie, Q., & Tan, S. Y. (2019). Preparing primary English teachers in Hong Kong: Focusing on language assessment literacy. *Journal of AsiaTEFL*, 16(2), 653–673. <https://doi.org/10.18823/asiatefl.2019.16.2.14.653>

Xu, H. (2017). Exploring novice EFL teachers' classroom assessment literacy development: A three-year longitudinal study. *The Asia-Pacific Education Researcher*, 26(3), 219–226. <https://doi.org/10.1007/s40299-017-0342-5>
[Crossref]

Xu, Y. T. (2019a). English language teacher assessment literacy in practice. In X. S. Gao (Ed.), *Second handbook of English language teaching* (pp. 517–539). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-02899-2_30

Xu, Y. T. (2019b). How teacher conception of assessment mediate assessment literacy: A case study of university English teacher in China. In S. Papageorgiou & K. M. Bailey (Eds.), *Global perspectives on language assessment* (pp. 197–211). Routledge.
[Crossref]

Xu, Y. T., & Liu, Y. C. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 493–513. <https://doi.org/10.1002/j.1545-7249.2009.tb00246.x>

Xu, Y. T., & Brown, G. T. L. (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>
[Crossref]

Yan, X., & Fan, J. (2021). “Am I qualified to be a language tester?”: Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246. <https://doi.org/10.1177/0265532220929924>
[Crossref]

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>
[Crossref]

Zhang, X. (2020). Assessment for learning in constrained contexts: How does the teacher's self-directed development play out? *Studies in Educational Evaluation*, 66,. <https://doi.org/10.1016/j.stueduc.2020.100909>

9. Conclusion

Ling Gan¹✉ and Ricky Lam²✉

- (1) School of Languages and Communication, Beijing Technology and Business University, Beijing, China
- (2) Department of Education and Psychology, Hong Kong Baptist University, Hong Kong, China

✉ **Ling Gan (Corresponding author)**

Email: linggan@btbu.edu.cn

✉ **Ricky Lam (Corresponding author)**

Email: rickylam@hkbu.edu.hk

9.1 Overview

Chapters [4](#), [5](#), [6](#), and [7](#) delineate the different LAL development experiences of four university English instructors in their classroom assessment practices. This chapter begins with a summary of the findings regarding instructors' LAL development. We further discuss implications the study provides for university English instructors, university administrators, and assessment trainers in an effort to enhance instructors' LAL development. We finally discuss possible limitations of the study and accordingly present suggestions for future research in terms of research participants, timespan and themes.

9.2 Key Findings

The study addressed the first research question, "How do university English instructors develop their LAL over time?" by examining their LAL development and assessment identity construction across three

semesters. Firstly, university English instructors' LAL was an amalgam of assessment knowledge, skills, principles, and practices, all reflected in their day-to-day assessment practices. In assessment practices, each instructor developed his/her LAL with an individualised route and to varying extents based on their initial LAL influenced by previous assessment experiences. They enhanced this foundation through daily practices and academic engagements such as reading, conferences, and lectures. They also optimised their assessment practices by agentively reflecting on and (re)adjusting their assessment approaches to cater for students' learning needs. Influenced by various contextual, experiential, and personal factors, their LAL development followed idiosyncratic trajectories and was not full-fledged.

Secondly, university English instructors unconsciously constructed their identity as assessors during LAL development. None perceived themselves as an assessor at the outset. However, as their LAL progressed, some instructors changed their perceptions and recognised this identity while maintaining their identities as facilitators, guides, feedback givers, or emotional supporters. These identities were considered highlighting instructors' main roles in assessment practices. Thus four quadrants were proposed to elucidate the relationship between instructors' assessor identity and their LAL development.

Addressing the second research question, the study identified challenges and opportunities in LAL development and assessment identity formation, mediated by contextual, experiential, and personal factors. Contextual factors included institutional requirements and mandates, teaching workloads, collegial communications, and students' responses to instructors' assessment practices. Assessment experiences including assessment training and learning or participation in teaching contests acted as amplifiers or filters to impact instructors' LAL development and their assessment identity. Personal factors mainly referred to instructors' conceptions of language assessment which covered their beliefs, attitudes, and emotional responses to language assessment. These factors interplayed with each other, either exacerbating challenges or creating opportunities that influenced the route, rate, and extent of instructors' LAL development and assessment identity construction.

9.3 Implications of the Study

In this section, we summarise the practical implications for LAL development from the perspectives of university English instructors, institutional administrators, and assessment trainers.

9.3.1 Implication for University English Instructors

Understanding the importance of promoting learning through assessment, university English instructors need to be aware that they require considerable theoretical knowledge about language testing and assessment and become more familiar with effective assessment practices in English teaching and learning. It is essential that they purposefully develop their knowledge and skills in assessment practices. They also need to strengthen their awareness of a range of contextual, experiential, and personal factors that may amplify and impede their LAL development. With these factors in mind, they could be well prepared for the challenges they encounter in language assessment practices.

We suggest that university English instructors be more concerned about their emotional or affective responses in language assessment mediated by students' expectations or their expectations on students. Teachers' emotions are profoundly connected with assessment practices and teachers' LAL development (Nazari et al., [2023](#)). Realising LAL development is a complex, variable, and dynamic process, they understand that it is very common if they cannot develop sufficient LAL to meet assessment challenges. With this understanding, they would not have low self-efficacy towards their ineffective assessment practices. Instead, they would focus on what they should and can improve via agentive actions or regular self-reflection, which has been proven to be effective and efficient in fostering teachers' LAL (Babaii & Asadnia, [2019](#); Tian et al., [2021](#)).

Self-directed development also plays out in university English instructors' LAL enhancement. They are committed to sustaining their assessment expertise by actively using available affordances and creating new opportunities when external supports are constrained. The affordances or opportunities can be the assessment-relevant resources accessible through social media or provided by institutions

(e.g., the series of *Studies in Language Testing* published by Cambridge University Press). To make good use of these resources, instructors need to galvanise their self-agency to update their assessment knowledge base and skillsets and synergise existing knowledge with external resources (Zhang, [2020](#)). This self-directed development definitely involves instructors' continuous reflections on their assessment practices from which they accumulate assessment experience and construct assessment identities.

Becoming a reflective assessor is highly recommended for university English instructors' sustainable LAL development. The instructors can often consider what the identity as an assessor means. They may reflect on such questions as “what is an assessor” and “what responsibilities should an assessor shoulder?”. Reflections like this would help instructors have a better understanding of the nature of language assessment and thus take agentive actions to acquire new assessment knowledge and skills. They can accumulate and self-reflect assessment experience by involving themselves in assessment practices (X. Yan, personal communication, June 25, 2025).

Finally, university English instructors are recommended to join or create an assessment professional community to develop their LAL through “active and collective conversations with colleagues” (Smith et al., [2014](#), p. 259). The community should include those colleagues with assessment and other applied linguistic backgrounds. The dialogic collaboration among colleagues can provide a supportive environment where meaningful interaction contributes to the shared understanding of problems or challenges in assessment practices (Ho & Yan, [2021](#)). This mutual understanding will be, in turn, good for developing a new knowledge base and adjusting the existing LAL (Gan & Lam, [2022](#)).

9.3.2 Implications for University/Department Administrators

Informed by the findings of the study, university/department administrators (e.g., course coordinators, and heads of teaching sections) are suggested to provide more specific guidance in assessment practices, especially for novice instructors. For example, they introduce the institutional assessment culture, specify detailed assessment requirements, and present more situated assessment

examples for novice instructors to follow (e.g., Rachel's case). They may establish an expert-and-novice partnership where the novice instructors can gain assistance or support from the experienced one when facing problems in assessment or teaching.

Besides, they need to be mindful of the challenges university English instructors experience in developing their LAL and create more opportunities or resources for them. For instance, the department or university administrators may invite external assessment experts or internal academic staff with language testing and assessment backgrounds to regularly deliver assessment workshops, lectures or seminars, especially on the topics of classroom-based assessment. Moreover, the workshops, lectures or seminars should be orientated to classroom assessment practices rather than language testing and assessment research. This scaffolding may help instructors address assessment problems and mitigate their negative emotions when confronting these problems. With the help of the administrators, instructors can sense that they are not alone when they are struggling with professional development in language assessment.

It is also recommended that university/department administrators give instructors more autonomy to make assessment-related decisions in a supportive manner instead of controlling every aspect of assessment. For instance, instructors have the autonomy to use whatever assessment materials, tasks or activities that they consider necessary to facilitate learners' learning. In this manner, instructors are empowered to explore assessment methods that work best for their students, hence, developing their LAL agentively in their assessment practices.

9.3.3 Implications for Assessment Trainers/Training Providers

Although varied LAL enhancement among the participants was identified, it should be noted that their LAL development could not be sustainable without ongoing formal training and learning in language testing and assessment. It is recommended that assessment trainers/training providers offer training programmes to instructors to develop a robust assessment knowledge base and skill kit, particularly a classroom-based one and help instructors to "scrutinise their

professional identity as assessors within classroom contexts” (Pastore & Andrade, [2019](#), p. 134). Assessment trainers or training providers need to draw on the proposed LAL profile to understand the LAL needs of university English instructors before designing and developing the training module. With this in mind, they may vary the depth of training on different assessment topics by prioritising instructors’ LAL needs.

To improve the effectiveness of training, the pedagogical elements of assessment training require more attention (Gan & Lam, [2022](#)). Assessment trainers/training providers should also innovate instructional materials and the modes of delivering the training, as the traditional top-down approaches to training (e.g., through textbooks or courses), usually decontextualised, fail to help instructors internalise and apply LAL concepts (Kao, [2023](#)). As such, contextualised instructions such as concept-based instruction, apprenticeship and synergies (e.g., interaction between students and instructors) have been suggested to innovate LAL training for language teachers (Fulcher, [2020](#); Kao, [2023](#); Ukrayinska, [2024](#)).

More importantly, the assessment training needs to be embedded into university English instructors’ continuous professional development. For example, it should be a part of the *National Training Programme for In-service EFL Teachers in Tertiary Education*, which can invite a great majority of university English instructors to this professional development programme. The assessment training can be produced into a MOOC (Massive Open Online Course) one where the instructors can access learning content and materials without time and space constraints (Westbrook & Spiby, [2024](#)). Instructors can even enjoy LAL MOOC courses or training free of charge for some durations. They can also find solutions to the problems they encounter in their language assessment practices through open discussion forums afforded by MOOC. Without time and space constraints, English instructors can be motivated to invest more efforts in their ongoing LAL development.

9.4 Limitations of the Study and Suggestions for Future Research

Although this study provides theoretical insights into teachers' LAL development, it has limitations due to its qualitative nature. First, despite the discussion of our roles and stances in the study in Chap. 3, we have brought the inevitable influence on the interpretations of the data and, thus, the findings, given our multiple identities as an instructor and researcher. The first author's experience as a former university English instructor and a current assessment researcher may affect the conversations with the participants because of the so-called power differential (Gibson & Hua, 2016), as they might have thought the first author was taking advantage of hearing their stories. Moreover, although we built rapport with the participants, we were rejected to enter some critical observation fields as an "outsider", such as Ryan's speaking tests for students and informants' discussions with their colleagues, as our presence might have posed a threat to students' testing performance and sensitive conversations between the informants and their colleagues. The absence of these classroom episodes may have led to inconclusive data collection and may have had an impact on data triangulation.

Second, the interview questions may have prompted the participants to articulate expected responses. For instance, the interview questions provoked Emma, Rachel, and Tim to realise their responsibilities as an assessor, which could have possibly changed their perceptions of their identity as an assessor. While interview prompts can cause participants to reflect on and change their perceptions and practices accordingly (Merriam & Tisdell, 2016), this influence would not substantially make a difference to the trustworthiness of the study (Gibson & Hua, 2016).

Third, because of the constraint of sample selection, the study did not provide a comprehensive view of English instructors' LAL development in various types of universities located in different areas of China. For example, three participants in the study worked in the key universities while only one worked in a non-key university and all these universities are located in Wuhan, a city in the central part of China. Although the key participants are representative informants, data would have been more diverse and conclusive if more participants from non-key universities or Sino-foreign universities had been included. In these universities, educational policies, assessment requirements, and

demands on teachers' professional development are drastically different from those in key universities. Data would also have been more comprehensive if participants from universities located in the western and eastern areas of China had joined the study, as the socio-economic differences in these areas may influence the quality of instructors and policies of teacher professional development in the universities. Differences caused by the types and locations of universities may have resulted in idiosyncratic LAL development among university English instructors.

Despite the above limitations, at least five areas could be extended from the current study for future research. Firstly, more participants from different tiers of universities in diverse regions of the Chinese context (e.g., universities located in western or eastern parts of China) can be involved to explore their LAL developmental trajectories with more foci on multiple data sources (e.g., field observations, interviews, reflection journals, think-aloud, documents, etc.). In this way, the chasm between participants' self-reported beliefs, perceptions, and observed behaviours can be discerned.

Secondly, longitudinal research into instructors' LAL development needs to be improved. It would be useful to extend the study to probe the *after* moment of LAL development over even longer periods (e.g., how instructors' enhanced LAL would affect their own teaching and their students' learning outcomes in one way or another), since there should be repercussions of LAL enhancement in classrooms (Y. Xu, [2017](#)). Thirdly, it is worth exploring how the agency of university English instructors has contributed to their LAL development and assessment identity construction over time. The current study discovered that instructors' LAL development involved their agency (e.g., Emma, Rachel, and Ryan), as both LAL development and teacher agency are mediated within a sociocultural paradigm (Mansouri et al., [2021](#)). The alignment of teacher agency and LAL development deserves further exploration.

Likewise, university English instructors' construction of assessment identity would be another valuable research agenda. The current study elucidated how university English instructors constructed their assessment identity along with LAL development, whereas it has not yet examined how instructors' assessment identity construction

contributes to their LAL development. Although assessor identity formation is considered conducive to LAL development (Xu & Brown, [2016](#)), no empirical studies have been conducted to attest to this claim.

Additionally, LAL rubrics for language teachers based on empirical studies need to be developed and validated to probe teachers' LAL realities. LAL needs for language teachers generated from specific educational contexts can provide a prerequisite for researchers to develop and validate an analytical LAL rubric. This rubric can be used to document the progressive LAL development from being a "novice" to being "an expert". It is also helpful for assessment trainers to identify teachers' LAL needs, diagnose the strengths and weaknesses in teachers' LAL, notice shifts in their LAL development, and adjust assessment training practices accordingly.

9.5 Conclusion

Our study investigated the LAL development of four university English instructors in their teaching settings in the Chinese context. An LAL profile for Chinese university English instructors was outlined by redefining LAL components and refining the levels of these components in Taylor's LAL profile. This proposed profile is one initiative to attest to Taylor's LAL profile in a specific educational context.

LAL development among the four participants was featured with a bottom-up process in which the routes and rates of LAL development varied significantly because of the extent to which the participants reflected and exercised their agency. Different trajectories of LAL development were also influenced by instructors' prior assessment experiences, institutional requirements, and their personal dispositions including assessment efficacy and understanding of assessment identity. These factors either posed challenges to or created opportunities for LAL development. Additionally, the four instructors constructed their diverse assessment identities, building up relationships between their assessment identities and LAL development. These relationships could extend instructors' LAL research beyond knowledge, skills, and practices to a larger scope.

References

- 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> [Crossref]
- Fulcher, G. (2020). Operationalising language assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: From theory to practice* (pp. 8–28). Cambridge Scholars Publishing.
- Gan, L., & Lam, R. (2022). A review on language assessment literacy: Trends, foci and contributions. *Language Assessment Quarterly*, 19(5), 503–525. <https://doi.org/10.1080/15434303.2022.2128802> [Crossref]
- Gibson, B., & Hua, Z. (2016). Interviews. In Z. Hua (Ed.), *Research methods in intercultural communication: A practical guide* (pp. 181–195). John Wiley & Sons.
- Ho, E. C., & Yan, X. (2021). Using community of practice to characterize collaborative essay prompt writing and its role in developing language assessment literacy for pre-service language teachers. *System*, 101. <https://doi.org/10.1016/j.system.2021.102569>
- Kao, Y. T. (2023). Empowering preservice English teachers with language assessment literacy concepts and practices: Application of Vygotskian concept-based language instruction. *The Modern Language Journal*, 107(S1), 68–94. <https://doi.org/10.1111/modl.12818> [Crossref]
- Mansouri, B., Molana, K., & Nazari, M. (2021). The interconnection between second language teachers' language assessment literacy and professional agency: The mediating role of institutional policies. *System*, 103. <https://doi.org/10.1016/j.system.2021.102674>
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.
- Nazari, M., Karimpour, S., & Ranjbar, M. (2023). The ecology of language teacher assessment literacy: Insights from emotional critical incidents. *Studies in Educational Evaluation*, 78, 101285. <https://doi.org/10.1016/j.stueduc.2023.101285> [Crossref]
- 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> [Crossref]
- 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 (Eds.), *Designing assessment for quality learning* (pp. 303–324). Springer.
- Tian, W., Louw, S., & Khan, M. K. (2021). Covid-19 as a critical incident: Reflection on language assessment literacy and the need for radical changes. *System*, 103. <https://doi.org/10.1016/j.system.2021.102682>

Ukrayinska, O. (2024). Synergies in developing pre-service teachers' language assessment literacy in Ukrainian universities. *Education Sciences*, 14(3), 223. <https://www.mdpi.com/2227-7102/14/3/223> [Crossref]

Westbrook, C., & Spiby, R. (2024). Providing continuous professional development through a language assessment literacy MOOC. In B. Baker & L. Taylor (Eds.), *Studies in language testing—language assessment literacy and competence, volume 2: Case studies from around the world*. Cambridge University Press & Assessment.

Xu, Y. (2017). *Assessment literacy of university English instructors in China: a mixed-methods study* [Unpublished Doctoral Thesis, University of Hong Kong].

Xu, Y. T., & Brown, G. T. L. (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> [Crossref]

Zhang, X. (2020). Assessment for learning in constrained contexts: How does the teacher's self-directed development play out? *Studies in Educational Evaluation*, 66. <https://doi.org/10.1016/j.stueduc.2020.100909>

Appendix 1: Narrative Frame 1

Dear Teacher,

Thank you very much for your participation in this study. This is a narrative frame focusing on your assessment experiences as a university English instructor. In this frame, you will look back on your past experience and give an account of how you conduct assessment in your classroom. Please note that all the information you provide will be kept confidential and used only for research purposes.

Please share your experience of language testing and assessment in your English learning and teaching. You may write about:

1. how you were tested or assessed in your English learning.
2. how you test or assess your students in English teaching.
3. any assessment training experiences you are exposed to.
4. any books or materials related to language testing and assessment you are exposed to.
5. your own understanding of language testing and assessment.
6. the assessment policies or requirements of your department or university.
7. your own understanding of the assessment policies or requirements.
8. your understanding of the identity as assessors

1. I am a _____ (gender) teacher working in _____ (University). I have been teaching _____ (course) at this university. I obtained _____ (degree) in (year) majored in _____ (or I am doing my _____ (degree) (professional title) in my university.

2. Please describe in detail how you were tested or assessed in your English learning and how you test or assess your students. _____

3. I have received _____ (no/ a little/some) assessment training when I work in this university. (Fill in the blank if you have received no assessment training because _____ assessment training in _____ (assessment contents). It lasted _____ (hours or days). I have received _____ (frequency) so far. I think the training _____

(Your opinions of the training effectiveness, relationship with your teaching and your suggestions for the training)

4. I (have/have not) read books or materials related to language testing and assessment. (Fill in the following if you have read any books or materials related to language testing and assessment, because _____ language testing and assessment when I _____ (pre/in-service time). These books or materials are _____ (books or materials). From these books/materials I have learned _____ (benefits of reading for your teaching or assessment practices).

5. I think language testing and assessment is _____ (relevance or contribution) to/for English teaching/learning. _____

6. The assessment requirements of my university for assessment practice are _____ requirements _____. (Your opinions)

7. I think I am _____ (your identity/identities) for my students in terms of testing and assessment because _____ (reason for your choice of the identity/identities). I have _____ (need to pay for the identity you choose)

Appendix 2: Narrative Frames 2-4

Dear Teacher,

Thank you very much for your effort in this study. This is a narrative frame focusing on your assessment experiences as a university English instructor. Please note that all the information you provide will be kept confidential and used only for research purposes.

You may write about:

1. your reflection on your assessment practices this semester.
 2. your needs of knowledge, skills or competence related to language testing and assessment.
 3. events or incidents that motivate, facilitate, or impede learning, training or practices of language testing and assessment.
 4. difficulties or challenges and opportunities related to language testing and assessment.
 5. strategies to cope with difficulties/challenges and decisions of opportunities.
 6. perceptions of these difficulties/challenges, opportunities, strategies, and decisions.
 7. your understanding of the identity as assessors.
-
1. My own reflection on assessment practices this semester (assessment activities, evaluation of these activities, improvement, etc.)

 2. I need to improve _____ (specific knowledge, skills, or competence of language testing and assessment). For instance, _____
 3. This semester I have experienced/encountered _____
(Some events that impact your opinions, learning, or practices related to language testing and assessment)
 4. I have found some difficulties/challenges in _____. (Specific assessment practices in classroom reading/understanding language testing and assessment). For example, _____ (why). As for these difficulties, _____ (strategies)
 5. I have been given some opportunities in _____ (specific assessment practices in classrooms/assessment language testing and assessment). For example, _____ where, who, what, and why). I _____. (Your decisions to deal with these opportunities)
 6. I think these difficulties/challenges and opportunities are _____
(opinions).
 7. I think I am _____ (your identity/identities) for my students in terms of testing and assessment because _____ (reason for your choice of the identity/identities). I have _____ pay for the identity you choose).

Appendix 3: Interview Protocol

Interview Protocol Project: Language assessment literacy development of university English instructors in China: An ethnographic case study.

Date and Time:	Place
Interviewer:	Interviewee:

Sample questions for the first interview

1. Could you please tell me how you assess your students in your class? Could you give some specific examples to tell me how you make decisions of what and how to assess?
2. How did you know to assess students? Were you trained or did you learn to assess students in your pre-service or in-service programme? If so, how?
3. How do you think the COVID-19 pandemic influences the way you assess your students?
4. Now we date back to the days when you were a student. How were you assessed? What do you think of this experience?
5. What do you know about language testing and assessment?
6. What do you want to improve in terms of language testing and assessment?
7. What roles do you think language testing and assessment plays in your teaching?
8. What assessment policies or requirements do you know about? (national or institutional level)? What do you think of these policies or requirements?
9. What roles do you think you play when you are assessing students? How do you feel like the roles?

10. Do you have any learning or training experience concerning language testing and assessment? If you do, could you give me more details about the experience?
11. How do you view your identity as an assessor?
12. Do you think your learning or training experiences concerning language testing and assessment may influence your identity as an assessor? If you do, could you give some examples of the influence?

Sample questions for the second interview with instructors

1. What challenges or difficulties have you encountered in assessing students?
2. What have you done to deal with these challenges or difficulties?
3. Can you share me some difficulties or challenges you have experienced in developing your knowledge, skills, or practices related to language testing and assessment?
4. Have you got any opportunities to learn about language testing and assessment? If any, what are they? Did you take the opportunities? Why or why not?
5. How do you comment on the assessment practices you have conducted in this semester?
6. What impact do you think your tutorials with students and discussion with colleagues have on your understanding of language testing and assessment?
7. How do you feel about the assessment decisions you have made in this semester?
8. What do you think may trigger, facilitate, or impede your assessment practices?

9. What do you think you need to know about language testing and assessment to help you assess students?
10. What have you done to improve your knowledge, skills, or understanding related to language testing and assessment in this semester?
11. What identities do you think you take on when you are assessing students? How do you feel like these identities?
12. How do you feel like your identity as an assessor this semester?

Sample questions for the third interview with instructors

1. What assessment knowledge/skills do you think you have learned in this semester?
2. What assessment activities are you going to conduct in this semester?
3. How do you perceive the assessment schedules for this semester?
4. What challenges or difficulties do you think you may run into when implementing the assessment activities?
5. What do you plan to do to deal with these challenges or difficulties?
6. What assessment skills, knowledge, theories, or practices do you think you need to implement the assessment activities?
7. What do you think you can do to improve your assessment skills, knowledge, theories, or practices?
8. Do you have any opportunities to learn about language testing and assessment? If any, what are they? If no, what opportunities do you expect to have?

Sample questions for the fourth interview with instructors

1. How do you comment on the assessment practices you have conducted in this semester?
2. What impact have the assessment activities you conducted had on your teaching and students' learning?
3. What do you think of your grading this semester?
4. Do you have any opportunities to develop your assessment knowledge, skills, theories or practices? If any, what are they? If not, why not?
5. What challenges or difficulties have you encountered when assessing students? How did you deal with these challenges or difficulties?
6. What challenges or difficulties have you encountered when enhancing your assessment knowledge, skills, theories or practices?
7. How did you feel like your identity as an assessor this semester?
8. Have you made any efforts to construct your identity as an assessor or other identities? If yes, what are they?
9. Do you have any changes in your views of language teaching, learning, and assessment?

Sample questions for the fifth interview with instructors

1. How do you comment on your assessment practices in the past one year (What do you learn from this practice? Do you think you need to improve it in this semester? In what way?)
2. What changes will there be in your understanding/knowledge of language assessment in this semester?
3. What changes will there be in your assessment practices in this semester? (Do you have any expectations of changes? Why or why

not?)

4. Do you think your identity as an instructor and identity as an assessor are easily separated?
5. Do you see any differences between how you perceive your assessment identity and how you enact your assessment identity in the class?

Sample questions for the sixth interview with instructors

1. How do you comment on your assessment practices and your understanding of language assessment in the past three semesters?
2. What have you learned from your assessment practices in the past three semesters?
3. What comments would you like to make about your identity in the assessment?
4. Have you heard the term “language assessment literacy” before? How do you understand it?
5. How do you evaluate your own LAL in the past three semesters?
6. How do you expect to improve your LAL in the future?

Sample questions for focus group interviews with students

1. How does your instructor assess you usually in the class?
2. What do you think of the assessment methods used by your instructor?
3. What influence do you think these assessment methods have on your learning?
4. In your opinion, what are the effective assessment methods?

5. What feedback is usually given by your instructor? What do you think of the feedback?
6. What information can you get about high-stake tests (CET4/6, IETLS, TOFEL, GRE, etc.) from your instructor? What do you think of the information?
7. What do you think of grades you get from your instructor? Do you think the grades given by your instructor can reflect your learning performance?
8. What suggestion can you give to your instructor in terms of language testing and assessment?
9. What identities do you think your instructor develop in assessment practices?

Appendix 4: Observation Protocol

Date and Time	
Participant	
Location	
Length of activity	
Observed notes	Reflective notes
<u>Settings:</u> Mode of interaction Classroom dynamic <u>Activities:</u> Lectures Pair work Self-assessment, etc <u>Students:</u> Responses to the instructor and to the activities Interaction between peers and with the teacher <u>Participants:</u> Responses to students Feedback to students' activities	(Personal reflections, insights, ideas, confusions, hunches, initial interpretations, or breakthroughs)