Asia Pacific Journal of Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/cape20

Exploring Chinese university English teachers' language assessment literacy: a mixed-method study

Lulu Zhang, Ying Zheng & Dongying Yang

To cite this article: Lulu Zhang, Ying Zheng & Dongying Yang (24 May 2024): Exploring Chinese university English teachers' language assessment literacy: a mixed-method study, Asia Pacific Journal of Education, DOI: 10.1080/02188791.2024.2354684

To link to this article: https://doi.org/10.1080/02188791.2024.2354684

Published online: 24 May 2024.

ASIA PACIFIC JOURNAL OF EDUCATION
https://doi.org/10.1080/02188791.2024.2354684

Routledge
Taylor & Francis Group

Exploring Chinese university English teachers' language assessment literacy: a mixed-method study

Lulu Zhang [a], Ying Zheng [a] and Dongying Yang [b]

[a]Faculty of Arts and Humanities, University of Southampton, Southampton, UK; [b]College of Foreign Languages, North China University of Science and Technology, Tangshan, China

ABSTRACT

As the development of language assessment literacy (LAL) has grown to be a highly regarded and much-researched field of teacher professional education in recent decades, this study seeks to investigate the needs of in-service university English teachers for LAL development, teachers' current level of LAL as well as factors that facilitate or impede teachers' development of LAL. The primary aim of the study is to identify areas in which teachers require further knowledge and training in language assessment and to devise effective strategies for enhancing their language assessment competency. By running Principal Component Analysis, four dimensions were identified: *assessment in language pedagogy; technical skills; language assessment for learning; and assessment principles and concepts*. 871 in-service English university teachers participated in the survey. Descriptive analysis was employed to compare teachers' assessment knowledge with their actual practice. Moreover, Chi-square test was used to examine the correlation between teacher backgrounds and their development of LAL. Results indicated that the disparities in LAL development needs between different teacher cohorts gradually diminished as teachers gained teaching experience. Teachers demonstrated their confidence in the understanding of LAL, but there was still a need to enhance their knowledge of assessment principles and concepts, as well as psychometric analysis of language assessment.

ARTICLE HISTORY
Received 16 October 2021
Accepted 1 May 2024

KEYWORDS
Language assessment literacy; Chinese university English teachers; needs for language assessment literacy development; background variables; mixed-method research

The development of language assessment literacy in China

In the past two decades, there has been an upsurging attention and interest placed on language teachers' assessment literacy. Numerous studies were conducted to delve into the framework of language assessment literacy (e.g., Brookhart, 2011; Fulcher, 2012; Kremmel & Harding, 2020), to investigate the level of language assessment literacy (e.g., Crusan et al., 2016; Volante & Fazio, 2007; Xu & Brown, 2017), to develop training for LAL (e.g., DeLuca & Klinger, 2010; Harding & Kremmel, 2016; Koh et al., 2018). Language assessment literacy is commonly viewed as a repertoire of knowledge and practical skills in language assessment (Davies, 2008). Language teachers need to be "conversant and competent in the principles and practices of language assessment" (Harding & Kremmel, 2016, p. 145).

Language testing and assessment in China underwent significant development in the late twentieth century. Professor Gui Shichun, known as "the brave pioneering explorer" in applied linguistics, championed the application of international mainstream language testing theories and

CONTACT Lulu Zhang lulu.zhang@soton.ac.uk Faculty of Arts and Humanities, University of Southampton, Building 65, Avenue Campus, Highfield, Southampton SO17 1BF, UK

the establishment of the nationwide English Proficiency Test (EPT) (He & Zhang, 2020). This marked a pivotal point, as language testing was officially recognized as an independent field of scientific inquiry in China. In recent decades, driven by the emphasis on assessment and testing reform, there has been a growing recognition of the need to enhance language teachers' assessment literacy both empirically and theoretically (e.g., Jin, 2010; Lam, 2019; Xu et al., 2016; Yan et al., 2018; Zhang, 2018). Language assessment literacy has emerged as a crucial concept in teacher professional development in China, particularly within the English as a Foreign Language (EFL) teaching programme (Jin, 2010). A nationwide survey conducted by Jin (2010) indicated that essential aspects of language testing theory and practice have been adequately integrated into the training of tertiary-level EFL teachers in China, either as compulsory or optional coursework. China is taking strides towards equipping teachers with the necessary assessment literacy to meet the demands of language teaching and the requirements of high-stakes exams.

Previous studies conducted in China examined the assessment competency of teachers at various levels, ranging from primary to university settings. Primary school English teachers had only an intuitive understanding of assessment (Liu & Li, 2020), and faced difficulties incorporating manipulation criteria into their assessment tasks (Koh et al., 2018). Teachers in primary schools lacked training in grading and communicating assessment results (Koh et al., 2018; Liu & Li, 2020). Similar challenges existed in secondary schools, where English teachers in Hong Kong struggled with marking students' speaking abilities during school-based assessments due to inadequate skills (Qian, 2014). Secondary-level EFL teachers in mainland China also expressed their insufficient knowledge of and interest in assessment theory, although these teachers demonstrated good assessment awareness and skills from practice and self-reflection (Yan et al., 2018). Xu et al. (2016) showed that university teachers (n = 249) had better skills in rating and demonstrated a better understanding of language assessment and effective management of enacting assessment. Whereas Sun and Zhang (2022) found EFL university teachers (n = 272) had better knowledge of teaching pedagogy and classroom assessment but lacked knowledge of language assessment theories and principles.

In general, teacher assessment literacy in China is displaying signs of improvement, although there remain disparities among teachers at various levels. English teachers in China did not receive sufficient training in language assessment. Additionally, a noticeable gap exists between practical assessment skills and theoretical knowledge among teachers at all educational levels in China. Most of these studies had a limited number of participants, potentially restricting the applicability of their conclusions. Therefore, there is a need for a large-scale study to examine the current level of LAL and the specific challenges faced by university English teachers in China within the domain of language assessment. Such a comprehensive investigation can offer invaluable insights for designing targeted assessment training programmes tailored to address the identified needs of these teachers.

Factors influencing teachers' needs for LAL development and their level of LAL

Language teachers, being the primary users of language assessment, hold a crucial position in the language assessment process as they are responsible for analysing assessment results to feed them into their teaching and scaffold student language learning effectively. Teachers are expected to possess a comprehensive understanding of language assessment. The development of LAL is an ongoing process and a context-sensitive product. As language pedagogy undergoes reform, teachers' perspectives, practices and needs regarding language assessment are likely to evolve accordingly.

Research suggests that teachers' needs in LAL training can vary depending on their backgrounds. Malone (2013) conducted a comparative analysis of language testing experts and language teachers and discovered notable differences in their understanding and use of assessment. The study revealed that language testing experts place more importance on possessing knowledge of assessment theories, whereas language teachers prioritize the acquisition of practical "how-to" components in the tutorial. A similar finding was summarized in Gullickson's study (1993), which indicated

that test experts exhibit a preference for training focused on test analysis, statistics, and application of standardized tests. Conversely, language teachers demonstrate a stronger interest in formative and summative uses of tests, as well as non-test evaluation. The discrepancy between teachers and test experts can be partly due to the differences in research interests and professional focus. Test experts tend to be more concerned with the process of test validation, while teachers are inclined to leverage tests to enhance their teaching practices and gain insights into their students' performance.

Teachers' career stages and teaching experience have been identified that impact teachers' need awareness in assessment training (Brown & Remesal, 2012). Novice teachers tend to exhibit more confidence in their assessment knowledge and assessment practices compared to experienced teachers. And they demonstrate a higher level of assessment knowledge than experienced teachers (Crusan et al., 2016). Similarly, the findings in the study by Xie and Tan (2019) also revealed that novice teachers express fewer assessment training needs compared to their experienced counterparts in areas such as language pedagogy, teaching policies and technical skills. Experienced teachers exhibit a disparity between their beliefs and their actual practices regarding assessment for learning (Brown & Remesal, 2012). Those studies indicated that the development of assessment literacy does not synchronize with an accumulation of teaching experience. The differences can be attributed to several factors, including the inclusion of assessment components in pre-service teacher education, insufficient in-service assessment training, and the reluctance of experienced teachers to participate in assessment training opportunities (Crusan et al., 2016).

Smith et al. (2014) looked into pre-service teachers' assessment beliefs and found that pre-service teachers' beliefs are framed by their personal experiences of assessments, rather than solely by what they have learned in universities. Teachers learn about assessment "on the job" and become more self-aware of their training needs (Vogt & Tsagari, 2014, p. 382). Teachers are more likely to develop their assessment skills through a combination of hands-on experience and training, leading to improved application of assessment processes (Reynolds-Keefers, 2010). Xu and Brown (2017), p. did not observe any significant effect of the assessment training experience on teachers' AL. It is notable that according to those studies, the more hands-on experiences in assessments, the higher level of AL teachers may have, which is contradictory to the findings above (Brown & Remesal, 2012; Crusan et al., 2016), experienced teachers did not demonstrate high level of LAL compared to novice teachers. With the improvement in workplace-based assessment training and institutional contexts in China, further investigation must be required to assess the influence of assessment training and experiences on teacher AL and examine the challenges in assessment training in China.

Sun and Zhang (2022) found that teachers' academic backgrounds significantly influence their knowledge of language assessment theories and principles, as well as language teaching pedagogy. Teachers with a background in linguistics tend to have a higher level of LAL compared to those from literature or other majors such as translation. However, the study did not find significant effects related to teaching experience or higher degrees. An analysis of LAL training courses offered by 86 course instructors in Chinese universities, as reported by Jin (2010), revealed that there are sufficient language testing and assessment courses available for pre-service teachers. However, certain topics related to educational and psychometric measurements, as well as student classroom practice, received less attention compared to the fundamental aspects of assessment theories and practices in language testing (Jin, 2010). These findings indicated that teacher training programmes in China lack support for psychometric aspects and fail to provide "representativeness and relevance of content in light of transformations in the assessment landscape (e.g., accountability system, conceptions of formative assessment)" (Gotch & French, 2014, p. 17).

Researchers have examined the impact of career stage, teaching experience, learning and training experience and academic backgrounds on teachers' needs for AL development and their assessment literacy. Wang et al. (2020) categorized these factors into three levels: micro-level (e.g., assessment training, teaching experience, student attribute), meso-level (e.g., class size, workload), and macro-level (e.g., assessment culture). Yan et al. (2018) also identified contextual (e.g., educational landscape, assessment policy) and experiential factors (e.g.,

teacher training, prior experience in item writing) that influence teachers' LAL need development. It is necessary to consider the diverse context-specific needs of teachers in developing LAL. Other mediating factors such as gender, age, professional status, target student, pre-service LAL learning, and test development experience need to be considered. Understanding the effects of these factors on LAL training needs and LAL development requires comprehensive analysis and integration.

Teachers usually learn about assessment with a cookie-cutter approach (Leung, 2014) and mechanically follow the instructions of assessment decontextualized (Lam, 2019). Xu and Brown (2017) identified three types of reasons for this insufficiency: a lack of assessment policies and professional standards; the absence of AL standards in recruitment criteria for university English teachers; and inadequate assessment training in pre-and in-service teacher education programmes. With the promotion of assessment policy in China particularly the introduction of formative assessment in the early 21st century (College English Curriculum Requirement in 2004) (Ministry of Education of the People's Republic of China, 2004) in China, it has become necessary to explore additional explanations for why language teachers may excel in certain areas but lack proficiency in others. "Assessment illiteracy is a global concern" (Xu & Brown, 2017, p. 151). It is a continuing process to determine what teachers need to learn in language assessment and how to efficiently enhance their language assessment competency. Two research questions were addressed in this study:

RQ1. To what extent and in what ways do the language assessment literacy of Chinese university English teachers differ empirically based on their diverse backgrounds?

RQ2. What are the factors influencing teachers' needs for LAL development?

Research methodology

Leech and Onwuegbuzie (2009, pp. 267–268) concluded a three-dimensional typology of mixed methods design: "(a) level of mixing (partially mixed versus fully mixed); (b) time orientation (concurrent versus sequential), and (c) emphasis of approaches (equal status versus dominant status)". Given the research aims and questions, this study adopts a partially mixed design with quantitative priority, using both a questionnaire and semi-structured interviews. The questionnaire investigates the knowledge and practices of language assessment among English language teachers in Chinese universities. The semi-structured interviews aim to explore language assessment contexts and factors that impede or promote the development of language assessment literacy among Chinese English university teachers.

Survey design and data collection

The questionnaire was constructed by drawing from two sources: the Language Assessment Literacy Survey ($n = 71$ items) developed by Kremmel and Harding (2020) and the Assessment Competencies of Teachers: A National Survey ($n = 35$ items) developed by Plake et al. (1993). The survey by Plake et al. (1993) comprises 35 items designed to assess teacher assessment competency in a general sense, regardless of the specific subject matter being taught. It identifies six broad skill areas in educational assessment, covering the selection, processing, and interpretation of assessment from the perspectives of teaching and learning, as well as ethical considerations. The questionnaire used in the Kremmel and Harding (2020) study introduces nine distinct components and focuses specifically on language assessment literacy among a wide range of stakeholders (e.g., language teachers, examiners): (1) developing and administering language assessments; (2) assessment in language

pedagogy; (3) assessment policy and local practices; (4) personal beliefs and attitudes; (5) statistical and research methods; (6) assessment principles and interpretation; (7) language structure, use and development; (8) washback and preparation; (9) scoring and rating.

One notable difference between the two questionnaires is their target participants. The 1993 questionnaire does not closely relate to language assessment, whereas the questionnaire by Kremmel and Harding (2020) specifically focuses on language assessment. It targets a wide range of stakeholders, including language teachers and examiners. Although the 1993 survey is considered outdated, the themes covered in the survey are broader, particularly on the selection of assessment and ethical consideration, compared to the Kremmel and Harding (2020) survey. The 2020 survey delves into LAL in detail, exploring technical language assessment skills and policy (e.g., CEFR). And it incorporates new concepts, including assessment for learning. The updated competencies in the more recent questionnaire reflect current perspectives on teaching, learning, and the social role of assessment (Inbar-Lourie, 2008).

Both questionnaires aim to assess people's knowledge about assessment, but there is a lack of alignment between teachers' assessment knowledge and their assessment practical skills. In this study, we examined both teachers' self-evaluation of their knowledge and practices in language assessment. To achieve this, we combined the 35-item test developed by Plake et al. (1993) with the Language Assessment Literacy Survey created by Kremmel and Harding (2020) in our questionnaire. The questionnaire items are divided into two subscales (1) language assessment knowledge and (2) language assessment practice. The questionnaire consists of three parts: (1) teacher demographic information; (2) 34 5-point Likert scale items (1 = not at all, 2=very little, 3=a moderate amount, 4=quite a lot, 5= a very great deal) to measure teachers' self-evaluation on their language assessment knowledge; (3) 27 5-point Likert scale items (1=never, 2=seldom, 3=sometimes, 4=often, 5=always) to measure language assessment practices. The items are designed using "have knowledge of" statements and "can do" statements. For example, item 3 "have knowledge of writing good quality multiple-choice questions" in the subscale of language assessment knowledge. "Can write good quality multiple-choice questions" in the subscale of language assessment practice. The "have knowledge of" and "can do" statements provide descriptions of what teachers are expected to be able to know and do. Additionally, it standardized measurement to assess respondents' knowledge and abilities. However, individual self-evaluation may lead to subjective responses, resulting in underestimation or overestimation of their performance in the tasks. Nevertheless, it also provides insights into individuals' confidence in their knowledge and competency in language assessment.

The questionnaire was administered to in-service university English language teachers in China through Wen Juan Xing. Snowball sampling was used to recruit participants, enabling contact with individuals from a diverse range of universities and mitigating bias. A pilot study was conducted in Hebei province to ensure internal consistency, as well as clarity, coherences, and usability of the survey for teachers. The questionnaire was then distributed to other provinces, excluding Hebei province during the main study to avoid data overlap. The collected data was analysed using SPSS (version 27). To provide further validation for the scale, the Principal Components Analysis (PCA) was used to identify the latent components and examine how the items load onto the latent dimensions. In the main study, the Chi-square test was carried out to examine the correlation between teachers' LAL development and their background variables.

A total of 204 respondents participated in the pilot study. For the main study, 669 questionnaires were completed and returned. After carefully vetting, 667 questionnaires were deemed valid for data analysis. Two questionnaires were excluded from the analysis due to the duplicate participant information. Altogether 871 English teachers were gathered from 12 provinces (Hebei, Henan, Yunnan, Inner Mongolia, Ningxia, Guangdong, Guizhou, Shanxi, Qinghai, Gansu, Sichuan, and Xinjiang), including 204 pilot study participants (see Table 1). Of those teacher participants, 86.7% are female ($n = 755$) and 13.3% are male ($n = 116$). Most teachers are between the ages of 31 and 40 with a master's degree ($n = 375$). 292 are Associate Professors and 43 are Professors. Approximately half of the respondents are lecturers, with 80.2% ($n = 366$) of all lecturers ($n = 456$) holding a master's

Table 1. Demographic information of the respondents by educational background.

		Undergraduate	Master	Doctorate	Total
Gender	Male	30	77	9	**116**
	Female	164	564	27	**755**
Age	20–30	2	78	2	82
	31–40	68	**375**	20	**463**
	41–50	78	157	14	249
	>50	46	31	0	77
Professional status	Teaching Assistant	3	77	0	80
	Lecturer	74	**366**	16	456
	Associate Professor	100	181	11	**292**
	Professor	17	17	9	**43**
Years of teaching experience	<3	2	63	4	69
	4–10	4	171	9	184
	11–20	102	315	15	432
	>20	86	92	8	186
Target student	English majors	51	182	24	257
	Non-English majors	143	459	12	614
Pre-service LAL learning	Yes	56	304	17	**377**
	No	138	337	19	494
Workplace LAL training	Yes	80	184	10	**274**
	No	114	457	26	597
Assessment/test research experience	Yes	30	95	10	135
	No	164	546	26	**736**
Test development experience	Yes	134	436	33	**603**
	No	60	205	3	268
LAL needs	Urgent need	28	117	9	154
	Need	116	420	18	554
	Neutral	34	72	5	111
	Not too much need	16	30	4	50
	No need	0	2	0	2

degree. 43.3% ($n = 377$) reported having received pre-service assessment learning during their university education. Although there are more master's degree teachers than doctorate teachers, teachers with a master's degree and teachers with a doctoral degree had a comparable chance (around 47%) of having pre-service assessment learning. 31.4% ($n = 274$) of all participants reported that they had workplace training. 603 out of 871 participants claimed that they had test development experiences for mid-term or end-term exams, indicating that the majority of university English teachers participate in test design. 736 respondents pointed out that they had never engaged in language testing-related research and 81.2 % ($n = 708$) of teachers highlighted they need to improve language assessment-related knowledge and skills.

Interview design and data collection

Following the questionnaire, semi-structured 30-minute interviews were conducted to delve deeper into the factors and challenges in the LAL development process. The selection of interviewees was based on convenience, aiming to include interviewees with diverse research interests and educational backgrounds. The interview guide consisted of five key points: (1) understanding of language assessment, (2) assessment policies, (3) role in the language assessment process, (4) challenges or difficulties in the process of practising language assessment, and (5) available language assessment learning resources (e.g., LAL training, pre-service LAL learning, assessment research, etc.). The interviews aim to find out how English teachers in Chinese universities perceive language assessment and their role in the process, as well as how teachers conduct low-stake language tests (e.g., term exams), and implement assessment for learning. The interviews also are interested in the assessment culture in China Higher Education which may facilitate or impede LAL development. The interviews were audiotaped, and the content was subsequently analysed to extract insights and themes.

Five university teachers from foreign language departments in five universities participated in this study. Each was identified with pseudonyms to protect their privacy. Two lecturers (Amy and Bella) are from two non-state-run colleges, specializing in applied linguistics in language teaching. Amy holds her master's degree obtained overseas and has been teaching in Henan province for six years. Bella gained her master's degree from a Chinese university and has been teaching in Henan province for five years. Both teachers teach different courses to English majors and non-English majors. Carnell holds a doctorate and has nine years of teaching experience. His research focuses on English as Lingua Franca. He teaches English majors at a state-run university. Dora has a background and research interests in English literature. Ella's research interest focuses on language assessment. Both Dora and Ella are associate professors with over 10 years of teaching experience.

Research findings

RQ1. To what extent and in what ways do the language assessment literacy of Chinese university English teachers differ empirically based on their diverse backgrounds?

To address the first research question, we conducted a principal component analysis (PCA) on pilot data ($n = 204$). This analysis aims to determine the underlying dimensions that can structure university teachers' knowledge and practices in language assessment. Additionally, we compared LAL among teachers ($n = 871$) with diverse backgrounds in the main study based on the identified dimensions to gain insights into the specific training needs of teachers in this domain.

The PCA analysis was conducted in SPSS (version 27). The language assessment knowledge scale was labelled as "K", and the language assessment practice scale was labelled as "P". Before the analysis, the internal consistency reliability of the questionnaire was computed by using Cronbach's alpha. The item-total correlation serves as a criterion for initial assessment and purification with the minimum criteria of 0.40 (Field, 2018). The results showed that most of the items in each subscale had appropriate item-total correlations except the items K2, K27, K31, and P2. Upon item interpretation, they were removed from the analysis. As a result, the assessment knowledge scale retained 31 items, while the assessment practice scale retained 22 items.

The component correlation matrix in both subscales exceeded the threshold of 0.32 (Tabachnick & Fidell, 2007, p. 646), indicating a need for oblimin rotation. The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's test of sphericity were calculated. Bartlett's test of sphericity was significant ($p < 0.001$). The KMO values were satisfactory, with a score of 0.952 on the knowledge scale and 0.952 on the practice scale, surpassing the threshold of 0.60. These results indicated that the data analysis could proceed. Communalities below 0.20 were sequentially removed to ensure data quality. Upon the observation of the initial solution, no communalities below 0.20 were found. Eigenvalues greater than 1.0 were retained.

Four factors were retained in the assessment knowledge scale, accounting for a cumulative variance of 70.9%. Component 1 (Eigenvalue = 17.1) explained approximately 55.3% of the variance, followed by Component 2 (Eigenvalue = 2.4) with 7.8%, and Component 3 (Eigenvalue = 1.3) with 4.3%, and Component 4 (Eigenvalue = 1.1) with 3.4% of the variance. Three factors were retained in the assessment practice scale and accounted for a cumulative variance of 67.2%. Component 1 (Eigenvalue = 14.1) explained approximately 56.6% of the variance, Component 2 (Eigenvalue = 1.5) with 6.1%, and Component 3 (Eigenvalue = 1.1) with 4.5% of the variance.

We proceeded to examine items with low factor loadings or those that cross-loaded on multiple factors. Items with factor loadings below 0.40 were removed one by one in reverse order of highest factor loading. Consequently, K1 (0.338) and P1 (0.308) were removed. Retained factors should have a minimum of three items with loadings above 0.40 and should not exhibit high cross-loading on other factors. Items that cross-loaded on more than one component with the highest ratio of loadings below 0.40 were excluded and K14 (0.315 & 0.346) was removed. Based on the consistency and qualitative interpretation, P14 was removed as well. K23 and P23, K33 were observed to be categorized into two different components. To evaluate if removing these items would enhance the

total variance, they were individually deleted to assess any improvement. The removal of K23, K33, and P23 led to an enhancement in the total variance, resulting in 72.4% for the knowledge scale and 69.7% for the practice scale, which is better than before deleting. As a result, the assessment knowledge scale retained 27 items (see Table 2), and the assessment practice scale retained 22 items (see Table 3).

A reliability analysis was run again for each scale. $\alpha_{KLangP} = 0.949$, $\alpha_{KPC} = 0.890$, $\alpha_{KTS} = 0.950$, $\alpha_{KAfL} = 0.877$ in the assessment knowledge scale; $\alpha_{PLangP} = 0.935$, $\alpha_{PTS} = 0.943$, $\alpha_{PAfL} = 0.885$ in the assessment practice scale (see Table 4). The inter-item correlations within each group were 0.676, 0.575, 0.682, 0.726 in the assessment knowledge scale; and 0.616, 0.647, 0.657 in the assessment practice scale. The Cronbach's alpha values exceeded 0.70, indicating good internal consistency, which supports the decision to retain the components for further analysis.

A total of 667 respondents returned the questionnaire in the main study, which provided a sufficiently large item-to-participant ratio, approximately 1:13 (Loewen & Gonulal, 2015). As there was no overlap between the pilot study data and the main study data, given that they were collected from different provinces, further analysis was conducted with a dataset of 871 respondents (204 + 667). The reliability of each dimension demonstrated good internal consistency with values ranging from 0.8 to 0.95 (see Table 5). The low standard deviation indicated that the values tended to be closely clustered around the mean of the dataset.

Descriptive analysis was carried out to analyse LAL profiles of teachers based on their demographic characteristics including gender, age, educational background, professional status, teaching experience, target students, pre-service LAL learning, workplace LAL training, assessment research

Table 2. The results of principal component analysis of assessment knowledge.

	Component			
Item	1	2	3	4
K3. Have knowledge of writing good-quality multiple-choice questions.	.578			
K4. Have knowledge of writing good quality open-ended questions.	.802			
K5. Have knowledge of designing student self-assessment activities.	.766			
K6. Have knowledge of designing peer-assessment activities.	.821			
K7. Have knowledge of writing assessment criteria for open-ended questions.	.920			
K8. Have knowledge of giving students useful assessment feedback.	.831			
K9. Have knowledge of using assessment to diagnose learners' strengths and weaknesses.	.831			
K10. Have knowledge of adjusting test items to meet the needs of new teaching.	.838			
K11. Have knowledge of using specifications to develop test structures and items.			−.583	
K12. Have knowledge of organizing and adjusting classroom teaching on a basis of test results.	.667			
K13. Have knowledge of piloting/trying-out assessment to determine cut-scores.			−.526	
K15. Have knowledge of calculating and interpreting "Mean" and "Standard Deviation" of test results.			−.670	
K16. Have knowledge of using formative assessment to measure language learning progress.				.593
K17. Have knowledge of using summative assessment to measure language learning progress.				.725
K18. Have knowledge of assessing the proportion of project, test, and assignment when determining the rating scale.				.579
K19. Have knowledge of using statistics to analyse the quality of individual items/tasks.			−.870	
K20. Have knowledge of using statistics to analyse the difficulty of individual items/tasks.			−.830	
K21. Have knowledge of using different forms of alternative assessment (e.g., portfolio).			−.719	
K22. Have knowledge of using test results to determine teaching materials.			−.758	
K24. Have knowledge of communicating assessment results and decisions with other teachers or managers.			−.668	
K25. Have knowledge of using techniques other than statistics (e.g., questionnaire, interview, etc.) to examine the quality of test items and tasks.			−.508	
K26. Have knowledge of using assessment to motivate students to learn.				.489
K28. Have knowledge of the concept of reliability.		.854		
K29. Have knowledge of the concept of validity.		.823		
K30. Have knowledge of differentiating norm-referenced and criteria-referenced assessment approach.		.738		
K32. Have knowledge of aligning language assessment to language proficiency framework.		.745		
K34. Have knowledge of being aware of language assessment knowledge and language assessment literacy development approach.		.688		

Table 3. The results of principal component analysis of assessment practice.

Item	Component		
	1	2	3
P3. Can write good quality multiple-choice questions.	.557	.	
P4. Can write good quality open-ended questions.	.837		
P5. Can design student self-assessment activities.	.686		
P6. Can design peer-assessment activities.	.758		
P7. Can write assessment criteria for open-ended questions.	.913		
P8. Can give students useful assessment feedback.	.762		
P9. Can use assessment to diagnose learners' strengths and weaknesses.	.717		
P10. Can adjust test items to meet the needs of new teaching.	.829		
P11. Can use specifications to develop test structures and items.		−.573	
P12. Can organize and adjusting classroom teaching on a basis of test results.	.566		
P13. Can pilot/try-out assessment to determine cut-scores.		−.751	
P15. Can calculate and interpret "Mean" and "Standard Deviation" of test results.		−.830	
P16. Can use formative assessment to measure language learning progress.			.783
P17. Can use summative assessment to measure language learning progress.			.868
P18. Can assess the proportion of project, test, and assignment when determining the rating scale.			.671
P19. Can use statistics to analyse the quality of individual items/tasks.		−.914	
P20. Can use statistics to analyse the difficulty of individual items/tasks.		−.877	
P21. Can use different forms of alternative assessment (e.g., portfolio).		−.813	
P22. Can use test results to determine teaching materials.		−.853	
P24. Can communicate assessment results and decisions with other teachers or managers.		−.670	
P25. Can use techniques other than statistics (e.g., questionnaire, interview, etc.) to examine the quality of test items and tasks.		−.660	
P26. Can use assessment to motivate students to learn.			.476

Table 4. Reliability analyses and inter-item correlations in the pilot study ($n = 204$).

	Knowledge scale			Practice scale		
	Item	α	Inter-item correlations	Item	α	Inter-item correlations
Assessment in Language Pedagogy (LangP)	3, 4, 5, 6, 7, 8, 9, 10, 12	.949	.676	3, 4, 5, 6, 7, 8, 9, 10, 12	.935	.616
Testing Technical Skills (TS)	11, 13, 15, 19, 20, 21, 22, 24, 25	.950	.682	11, 13, 15, 19, 20, 21, 22, 24, 25	.943	.647
Language Assessment for Learning (AfL)	16, 17, 18, 26	.877	.726	16, 17, 18, 26	.885	.657
Principles and Concepts (PC)	28, 29, 30, 32, 34	.890	.575			

Table 5. Reliability analyses for all scales ($n = 871$).

Variable	Mean	SD	α	95% CI	
KLangP (Assessment in Language Pedagogy)	3.05	.88	.944	2.99	3.11
KTS (Testing Technical Skills)	2.81	.91	.942	2.75	2.88
KAfL (Language Assessment for Learning)	3.13	.99	.902	3.06	3.20
KPC (Principles and Concepts)	2.59	.83	.900	3.54	2.65
PLangP (Assessment in Language Pedagogy)	2.97	.86	.938	2.91	3.03
PTS (Testing Technical Skills)	2.69	.89	.938	2.64	2.76
PAfL (Language Assessment for Learning)	3.06	.98	.892	3.00	3.13

experience, and test development experience. The means between teachers' knowledge and teachers' practice of language assessment in terms of four dimensions derived from PCA were compared: *language assessment pedagogy* (KLangP, PLangP), *technical skills* (KTS, PTS), *language assessment for learning* (KAfL, PAfL), and *principles and concepts* (KPC) (See Table 6).

The descriptive analysis revealed that professors (KLangP = 3.47, PLangP = 3.36, KTS = 3.17, PTS = 2.87, KAfL = 3.73, PAfL = 3.62, KPC = 3.00) and teachers with assessment research experience (KLangP = 3.35, PLangP = 3.28, KTS = 3.20, PTS = 3.08, KAfL = 3.51, PAfL = 3.45, KPC = 3.11)

Table 6. Mean difference contrast of teachers' knowledge and practice of assessment.

		KLangP	PLangP	KTS	PTS	KAfL	PAfL	KPC
Overall average		**3.05**	**2.97**	**2.81**	**2.69**	**3.13**	**3.06**	**2.59**
Gender	Female ($n = 755$)	3.04	2.96	2.79	2.68	3.12	3.06	2.58
	Male ($n = 116$)	3.11	3.00	2.96	2.81	3.16	3.08	2.71
Age	<30 ($n = 82$)	3.07	3.01	3.00	2.94	3.12	3.09	2.81
	31–40 ($n = 463$)	3.00	2.90	2.76	2.64	3.05	**2.97**	2.53
	41–50 ($n = 249$)	3.15	3.07	2.86	2.73	3.29	3.22	2.68
	>50 ($n = 77$)	3.01	2.97	2.76	2.66	3.07	3.02	2.51
Educational Background	Undergraduate ($n = 194$)	2.97	2.88	2.76	2.63	3.09	3.00	2.48
	Master ($n = 641$)	3.05	2.98	2.81	2.70	3.12	3.06	2.61
	Doctoral($n = 36$)	3.34	3.27	3.14	2.96	3.42	3.31	2.96
Professional Status	Teaching assistant ($n = 80$)	3.05	2.95	3.01	2.91	3.09	**2.97**	2.75
	Lecturer ($n = 456$)	2.96	2.89	2.72	2.62	3.02	**2.98**	2.50
	Associate professor ($n = 292$)	3.12	3.03	2.84	2.73	3.21	3.13	2.64
	Professor ($n = 43$)	3.47	3.36	3.17	2.87	3.73	3.62	3.00
Teaching Experience	<3 ($n = 69$)	3.13	3.10	3.07	3.00	3.16	3.10	2.83
	4–10 ($n = 184$)	2.97	2.92	2.76	2.62	3.02	**2.92**	2.51
	11–20 ($n = 432$)	3.04	3.08	2.79	2.70	3.14	3.08	2.58
	>20 ($n = 186$)	3.12	3.12	2.79	2.65	3.19	3.12	2.63
Teaching Target	English majors ($n = 257$)	3.21	3.13	2.95	2.84	3.27	3.18	2.65
	Non-English majors ($n = 614$)	2.98	2.90	2.75	2.63	3.07	3.00	2.57
Pre-service Assessment Learning	Yes ($n = 377$)	3.18	3.08	3.01	2.88	3.28	3.21	2.85
	No ($n = 494$)	**2.95**	**2.88**	**2.66**	**2.55**	**3.01**	**2.94**	**2.40**
In-service LAL learning	Yes ($n = 274$)	3.23	3.15	3.05	2.94	3.37	3.32	2.87
	No ($n = 597$)	2.96	2.88	2.70	2.58	3.01	**2.94**	2.47
Assessment Research Experience	Yes ($n = 135$)	3.35	3.28	3.20	3.08	3.51	3.45	3.11
	No ($n = 736$)	2.99	2.91	2.74	2.62	3.06	**2.99**	2.50
Test Development Experience	Yes ($n = 603$)	3.11	3.03	2.84	2.73	3.17	3.10	2.61
	No ($n = 268$)	2.92	2.83	2.73	2.61	3.03	**2.97**	2.56

exhibited the highest level of LAL compared to other groups. The average scores for each dimension were higher than the overall average scores (KLangP = 3.05, PLangP = 2.97, KTS = 2.81, PTS = 2.69, KAfL = 3.13, PAfL = 3.06, KPC = 2.59) (Table 6). The LAL level was also high among teachers with doctoral degrees. In contrast, teachers without pre-service LAL learning experience exhibited the lowest level of LAL (see Table 6).

Table 6 shows that most teachers had a better understanding and practices in language assessment pedagogy, encompassing aspects such as student self-assessment, peer assessment, assessment feedback, and diagnostic assessment. Among all the participants, professors demonstrated the highest level of knowledge in language assessment pedagogy, with an average mean of 3.47. Although the average means of the practice of language assessment pedagogy were generally lower than the knowledge dimension, professors still showed a high level of practical implementation in this area, with a mean of 3.36. In contrast, teachers without pre-service LAL learning, workplace assessment training, assessment research experience, or test development experience exhibited less frequent engagement in designing assessment tasks or using alternative assessments. Their average means of PLangP range from 2.8 to 2.9.

Irrespective of their backgrounds, teachers generally displayed a lack of awareness regarding the use of technical skills in their language assessment. Both professors (Mean = 2.87) and teachers with a doctoral degree (Mean = 2.96) reported infrequent practice of technical skills in language assessment (see Table 6). In contrast, teachers who have assessment research experience (Mean = 3.08) demonstrated a higher level of awareness regarding the application of technical skills compared to other teacher groups.

All groups of teachers demonstrated a high level of understanding and implementation of assessment for learning (AfL). Most teachers felt they possessed a moderate understanding of the concept of AfL and could successfully apply it in their teaching, as reflected in the similar mean scores for AfL knowledge and AfL practice (Table 6). However, teachers between the ages of 31 – 40 showed slightly lower confidence in their ability to implement assessment for learning compared to

other groups. Specifically, the study revealed that teachers in the age range of 31 – 40 faced challenges in effectively utilizing formative assessment (Item P16[1] Mean = 2.86) (See Table 7). Both teaching assistants and lecturers demonstrated infrequent AfL practices. Teaching assistants also lacked practices of making use of summative assessment, as indicated by the mean score of item P17[2] (2.86). Teachers with 4 to 10 years of teaching experience did not implement AfL often as well ($M_{PAfL} = 2.92$). Teachers who lacked experience in pre-service LAL learning ($M_{PAfL} = 2.94$), workplace LAL training ($M_{PAfL} = 2.94$), assessment research experience ($M_{PAfL} = 2.99$), or test development experience ($M_{PAfL} = 2.97$) demonstrated weaker awareness of implementing AfL compared to their counterpart (See Table 6). However, all the teachers in China held a stronger awareness of using assessment to motivate students to learn (Table 7).

In general, teachers demonstrated a lack of sufficient knowledge of assessment principles and concepts, as evidenced by the mean score of KPC (2.59) (see Table 6). Table 6 highlights that only professors (Mean = 3.00) and teachers with assessment research experience (Mean = 3.11) expressed their confidence in their knowledge of KPC, with mean ratings above 3, surpassing the overall average. On the other hand, the remaining participants felt they had limited knowledge of assessment principles and concepts, with an average rating below 3.

RQ2. What are the factors influencing teachers' needs for LAL development?

The impact of teachers' demographic characteristics

In this sample, 81.3 % (n = 708) of university English teachers expressed a need for LAL development, while 6% (n = 52) did not agree and 12.7% (n = 111) of respondents remained neutral as some interviewees mentioned they did not benefit from or recognize the importance of LAL development. Chi-square tests revealed significant associations between gender and the needs for LAL development $\chi2$ (2, N = 871) = 11.56, p (0.003) < 0.05. In addition, the data demonstrated a significant relationship between the need for LAL development and educational background, $\chi2$ (4, N = 871) = 10.83, p (0.029) < 0.05, teaching experience, $\chi2$ (6, N = 871) = 18.82, p = 0.004, and workplace LAL training, $\chi2$ (2, N = 871) = 6.13, p = 0.047. However, no significant associations were

Table 7. Mean difference contrast of teachers' knowledge and practice of assessment for learning.

		K16	P16	K17	P17	K18	P18	K26	P26
Age	<30 (*n* = 82)	2.84	3.04	2.99	2.99	3.13	3.17	3.15	3.23
	31–40 (*n* = 463)	2.81	**2.86**	2.93	2.89	3.10	3.10	3.03	3.11
	41–50 (*n* = 249)	3.02	3.18	3.21	3.27	3.31	3.22	3.14	3.31
	>50 (*n* = 77)	2.68	2.88	3.00	3.04	3.13	3.10	3.10	3.09
Educational Background	Undergraduate (*n* = 194)	2.77	2.94	2.99	3.01	3.09	3.04	3.02	3.13
	Master (*n* = 641)	2.87	2.96	3.02	3.01	3.17	3.16	3.08	3.17
	Doctoral (*n* = 36)	**3.19**	**3.31**	**3.28**	**3.36**	**3.42**	**3.36**	**3.36**	**3.50**
Professional Status	Teaching assistant (*n* = 80)	2.89	2.91	2.93	**2.86**	3.11	2.99	3.28	3.16
	Lecturer (*n* = 456)	2.77	2.83	2.93	2.89	3.07	3.09	3.01	3.12
	Associate professor (*n* = 292)	2.91	3.10	3.09	3.16	3.23	3.21	3.09	3.20
	Professor (*n* = 43)	3.40	3.65	3.77	3.77	3.79	3.51	3.40	3.60
Teaching Experience	<3 (*n* = 69)	2.94	3.04	3.01	3.04	3.20	3.19	3.28	3.22
	4–10 (*n* = 184)	2.74	2.81	2.90	2.75	3.07	3.05	3.07	3.08
	11–20 (*n* = 432)	2.88	2.98	3.04	3.07	3.18	3.16	3.05	3.20
	>20 (*n* = 186)	2.89	3.07	3.10	3.17	3.22	3.18	3.10	3.18
Teaching Target	English majors (*n* = 257)	3.09	3.11	3.20	3.17	3.37	3.25	3.13	3.23
	Non-English majors (*n* = 614)	2.76	2.91	2.95	2.96	3.08	3.10	3.06	3.15
Pre-service Assessment Learning	Yes (*n* = 377)	3.06	3.14	3.20	3.18	3.29	3.31	3.25	3.33
	No (*n* = 494)	2.70	2.84	2.89	2.90	3.07	3.01	2.95	3.06
In-service LAL learning	Yes (*n* = 274)	3.13	3.32	3.32	3.35	3.38	3.35	3.22	3.34
	No (*n* = 597)	2.73	2.81	2.88	2.87	3.07	3.05	3.02	3.10
Assessment Research Experience	Yes (*n* = 135)	3.35	3.48	3.41	3.45	3.51	3.50	3.44	3.48
	No (*n* = 736)	2.77	2.88	2.95	2.94	3.10	3.08	3.01	3.12
Test Development Experience	Yes (*n* = 603)	2.91	3.01	3.10	3.07	3.21	3.16	3.14	3.19
	No (*n* = 268)	2.75	2.87	2.85	2.91	3.06	3.10	2.95	3.13

found between the needs for LAL development and age (p = 0.207), professional status (p = 0.167), target student (p = 0.280), pre-service LAL learning experience (p = 0.242), assessment research experience (p = 0.135) and test development experiences (p = 0.166), as *p* values were greater than 0.05 (see Table 8).

For Tables 2 by 2 and larger 2 by 2, the value to report is Cramer's V. For *R*-1 (row variable) equal to 1, the criteria to use is: small = 0.01, medium = 0.30, large = 0.50; for *R*-1 equal to 2, the criteria to use is small = 0.07, medium = 0.21, large = 0.35; for *R*-1 equal to 3, the criteria to use is small = 0.06, medium = 0.17, large = 0.29 (Pallant, 2010, p. 220). Based on the criteria, for a contingency table with *R*-1 (row variable) equal to 1, the Cramer's V values of 0.115, and 0.084 fall into the small effect size category, which suggests that there is a weak, but still detectable association between the gender and LAL needs and between workplace LAL training and LAL needs. Due to the row being 2 in educational background, the criteria that equals 2 was applied and the results show that educational background (Cramer's V = 0.079) has a weak effect on teachers' LAL development needs. For

Table 8. Baseline characteristics * needs for LAL development crosstabulation.

Characteristics	Recognition of need n(%)	Neutral n (%)	No recognition of needs n(%)	Overall samples	Chi-square tests of independence
Age n(%)					χ2 (6) = 8.45
<30	71(86.6%)	6(7.3%)	5(6.1%)	82(9.4%)	p (0.207)>0.05
31–40	382(82.5%)	61(13.2%)	20(4.3%)	463(53.2%)	n = 871
41–50	197(79.1%)	32(12.9%)	20(8.0%)	249(28.6%)	Cramer's V = .070
>50	58(75.3%)	12(15.6%)	7(9.1%)	77(8.8%)	
Professional status n(%)					χ2 (6) = 9.12
Teaching Assistant	71(88.7%)	5(6.3%)	4(5.0%)	80(9.2%)	p (0.167)>0.05
Lecturer	375(82.3%)	59(12.9%)	22(4.8%)	456(52.4%)	n = 871
Associate Professor	228(78.1%)	43(14.7%)	21(7.2%)	292(33.5%)	Cramer's V = .072
Professor	34(79.1%)	4(9.3%)	5(11.6%)	43(4.9%)	
Target student n(%)					χ^2 (2) = 2.55
English major	202(78.6%)	35(13.6%)	20(7.8%)	257(29.5%)	p (0.280)>0.05
Non-English major	506(82.4%)	76(12.4%)	32(5.2%)	614(70.5%)	n = 871
					Cramer'sV=.054
Pre-service LAL learning experience n(%)					χ^2 (2) = 2.84
Yes	316(83.8%)	41(10.9%)	20(5.3%)	377(43.2%)	p (0.242)>0.05
No	392(79.4%)	70(14.2%)	32(6.4%)	494(56.8%)	n = 871
					Cramer's V = .057
Assessment research experience n(%)					χ2 (2) = 4.00
Yes	118(87.5%)	11(8.1%)	6(4.4%)	135(15.5%)	p (0.135)>0.05
No	590(80.2%)	100(13.6%)	46(6.2%)	736(84.5%)	n = 871
					Cramer's V = .068
Test development experience n(%)					χ2 (2) = 3.59
					p (0.166)>0.05
Yes	497(82.4%)	76(12.6%)	30(5.0%)	603(69.2%)	n = 871
No	211(78.7%)	35(13.1%)	22(8.2%)	268(30.8%)	Cramer's V = .064
Gender n(%)					χ2 (2) = 11.56
Female	**627(83.0%)**	87(11.5%)	41(5.5%)	755(86.7%)	p (0.003)<0.05
					n = 871
Male	81(69.8%)	24(20.7%)	11(9.5%)	116(13.3%)	Cramer's **V=.115**
Educational background n(%)					χ2 (4) = 10.83
Undergraduate	144(74.3%)	34(17.5%)	16(8.2%)	194(22.3%)	p (0.029)<0.05
Master	**537(83.8%)**	72(11.2%)	32(5.0%)	641(73.6%)	n = 871
Doctoral Degree	27(75.0%)	5(13.9%)	4(11.1%)	36(4.1%)	Cramer's **V=.079**
Teaching experience n(%)					χ2 (6) = 18.82
<3	59(85.6%)	5(7.2%)	5(7.2%)	69(7.9%)	p (0.004)<0.05
4–10	162(88.1%)	19(10.3%)	3(1.6%)	184(21.1%)	n = 871
11–20	**348(80.6%)**	**60(13.9%)**	24(5.5%)	432(49.6%)	Cramer's **V=.104**
>20	139(74.7%)	27(14.5%)	20(10.8%)	186(21.4%)	
Workplace LAL Training (%)					χ2 (2) = 6.13
					p (0.047)<0.05
Yes	235(85.8%)	24(8.8%)	15(5.4%)	274(31.5%)	n = 871
No	**473(79.2%)**	87(14.6%)	37(6.2%)	597(68.5%)	Cramer's **V=.084**

a contingency table with *R*-1 equal to 3, the Cramer's V value of 0.104 almost reaches the medium effect size, which suggests that there is a better association between teaching experience and teachers' LAL development needs compared to other groups.

A majority of survey participants in this study are female, and 83.0% of females recognized the need for LAL development (See Table 8). Among the participants who voiced the need for LAL development, the majority held master's degrees ($n = 537$). The number of teachers without workplace LAL training ($n = 597$) was double that of teachers with workplace LAL training ($n = 274$). Among the cohort of teachers with LAL training experience, 85.8% recognized the importance of LAL development, representing a considerable proportion of the total sample size. This suggests that LAL training in the workplace enhances teacher LAL development.

Teachers with 11 – 20 years of teaching experience were a large cohort of those aware of the needs for LAL development, accounting for 80.6%. However, the ratio was lower than the group of 4–10 (88.1%) and less than 3 years of teaching experience (85.6%). Teachers with less than 20 years of teaching experience have strong demands for developing LAL. The demands for assessment skills can be partially attributed to governmental policies. As the interviewees in this study mentioned, "*the policies encourage schools and teachers to carry out the formative assessment in class and after class*". Additionally, teachers in their career-developing stage perceive LAL development as a new direction for future academic research. Ella shared her research interests, "*more and more university teachers, especially in applied linguistics, concerned about assessment and testing and classroom assessment. The publications related to language assessment and classroom assessment are increasing, and universities set up language assessment modules for master students or even for undergraduate students*". The interviewees felt they need to engage in more assessment research and broaden their understanding of assessment, as they believe that practising assessment research enhances their assessment competence and encourages them to think and act as assessors. For teachers with less than 20 years of teaching experience, a potential research field can open doors to more publications, which in turn can contribute to job promotion.

The demand for quality assurance in language assessment

The interviewees highlighted the disconnect between assessment literacy and the criteria for selecting professional teachers, which has an impact on the development of LAL. The current teacher qualification test emphasizes knowledge of teaching pedagogy, classroom management, and ethical issues but does not emphasize assessment knowledge. "*Language assessment courses are optional in some universities or rarely included in the undergraduate teacher development programme*" (Ella). The absence of comprehensive knowledge in teacher qualification reflects the lack of quality assurance for teachers' assessment practices in China. Teachers who are pursuing a certificate of teacher qualification will not receive training on language assessment if it is not a mandatory requirement for the qualification.

In addition, teachers' awareness of responsibility impacts LAL development as well. Amy who had studied language assessment, expressed concern about the responsibility of teachers. Most teachers lack knowledge of language assessment in evaluating student performance. She stated, "*I believe that many diagnostic assessments and tests used in the classroom or elsewhere are unreliable and unscientific. They have a regulated format for writing up a test. Teachers need to fill in the format with selected test materials. From my understanding, most teachers design tests without considering test validity, reliability, or conducting pilot assessment*". Teachers are required to write local tests to assess students' language performance but seldom are required to test the validity, usefulness, and reliability of the local test. Typically, teachers in China adhere to the prescribed test structure provided by the institution, wherein they include questions or key points from the entire book to assess students' comprehension of the learning material and their ability to provide critical analysis. Teachers comply with regulations without being fully aware of their responsibility to consider the nature and the characteristics of the local test.

Some interviewees expressed the belief that learning about technical assessment skills in particular psychometric analysis is unnecessary, as their focus is primarily on preparing for teaching, which already consumes a significant portion of their working hours. Dora argued, "*It is just an end-term school-level exam, not Gaokao or College English Test, which is widespread nationwide*". Usually, teachers are required to provide test analysis at the end of the term, with ready-to-use Excel forms and predefined formulas for score calculations. Bella explained, "*It is a compulsory task with organised forms provided beforehand*". Consequently, many teachers do not invest substantial time in interpreting test analysis results. However, teachers with assessment research experience ($M_{KTS} = 3.20$, $M_{PTS} = 3.08$), like Ella, demonstrated better practice of technical assessment skills compared to their peers ($M_{KTS} = 2.74$, $M_{PTS} = 2.62$) (Table 6).

The absence of LAL in the teacher qualification, their lack of responsibilities in language assessment and teachers' focus and workload on teaching lead to the lack of quality assurance in language assessment, resulting in overestimating the development of LAL. There is a need for a robust quality assurance system to ensure the LAL training and the alignment of assessment knowledge with practices. Efforts should be made to foster a culture that values the effective use of assessment results to enhance learning and teaching.

The development of assessment culture

All five interviewees believed that teachers should have basic knowledge of language assessment. "*We should be aware of the purpose for assessment, the fitness of the assessment tools, and make use of assessment results*" (Carnell). Amy and Bella pointed out their master's degree helped shift their teaching perspectives and develop a greater awareness of LAL development through their studies. However, when it comes to testing technical skills, most university teachers find it impractical and less useful. Not all teachers share the same level of interest or knowledge in language assessment, as observed in the case of Dora, who specializes in English literature. Dora does not prioritize formative assessment and has limited knowledge compared to teachers who study linguistics-related programmes. Also, she did not receive any formal training in language assessment during her studies. The extent of pre-service teacher language assessment training largely depends on the nature of the study programme. Teachers who have studied linguistics-related programmes are more likely to possess a better understanding of language assessment principles and practices and are more inclined to learn assessment concepts that are relevant to language pedagogy, such as assessment for learning.

The implementation of assessment policies, particularly formative assessment, has significantly influenced the teaching practices of university English teachers and their development of LAL. Amy highlighted that the strict national and institutional requirements for assessment at the undergraduate level compelled teachers to adapt their teaching methods. "*The government emphasises the widespread use of formative assessment and the enhancement of student self-learning*" (Amy). Teachers were encouraged to provide feedback to help students improve their academic performance and to integrate various assessment formats, such as presentation, group work, and essay writing, into summative assessments. They are also aware of the value of student self-assessment and peer assessment during teaching and learning. As Ella explained, "*Recent official documents from the Ministry of Education have encouraged teachers to make use of assessment for student learning. Additionally, numerous studies on formative assessment have contributed to the dissemination of the assessment concept*." As a result, the use of presentation as a mid-term assessment has gained popularity in undergraduate courses. "*I include the presentation as one of my assessment tools*". Mentioned by Amy, Carnell and Ella. They highlighted that mid-term presentations inform formatively on student learning progress.

For most interviewees, the increasing workload with a growing number of students and research pressure, have jeopardized their enthusiasm for learning about language assessment. During the interviews, the significance of pre-service training was emphasized very often. Pre-service training in

their education can facilitate a smoother transition into language assessment and provide alternative pathways for exploration. Teachers who have prior learning experience in language assessment are more likely to apply the principles and concepts of language assessment and further develop their LAL. However, the lack of professional instruction during teaching acts as a deterrent to LAL development. As Ella said, "*Continuity is a key to the LAL development*".

In summary, teachers' linguistics background, assessment policies and pre-service training have an impact on teachers' practices in language assessment. Assessment approaches linked to language pedagogy are readily acknowledged and employed by university teachers in China to effectively support learners in their dynamic learning. It is important to develop a continuous process for LAL development. The process of constructing an assessment culture can foster teachers' LAL and their ability to produce effective language assessments.

Discussion and conclusion

In this study, we examined the level of in-service university English teachers' LAL, their needs for LAL development, and the factors that facilitate or impede the development of LAL. Teachers demonstrate a solid understanding of LAL, but as noted in previous studies by Jin (2010) and Sun and Zhang (2022), there is still a need for teachers to learn assessment principles and concepts and improve their technical assessment skills. Integration of knowledge with assessment practices remains a challenge for teachers, as highlighted in the study by Koh et al. (2018).

Among different groups of teachers, professors and those with assessment research experience show the highest level of LAL in the quantitative analysis across various dimensions, including language pedagogy, assessment principles and concepts, and assessment for learning. Additionally, teachers with assessment research experience exhibit a high level of knowledge and practice in technical skills. Interestingly, the LAL level of teachers with lecturer status was not superior to that of teaching assistants. However, most participants in the interviews, including all those professors and associate professors, were not familiar with technical assessment skills. This deficiency may stem from the regulated format of test reporting, which often overlooks the nuances of assessment techniques. Interviewees with abundant research experience in language assessment displayed a better language assessment competency than those who focused on language teaching (Malone, 2013). It is worth noting that teaching experience alone does not guarantee high levels of language assessment literacy. A keen research interest in language assessment does spark teachers' strong awareness of consolidating their knowledge of technical assessment skills.

Government and institutional policies serve as crucial drivers of teacher knowledge and practices in assessment for learning. The implementation of assessment for learning is widely embraced by university teachers in China and supported by government initiatives and academic research (Liu & Xu, 2017). The Chinese government has introduced an assessment reform that requires teachers to integrate AfL strategies into university EFL classes. "Grading should not only focus on student's academic achievement but also discover and support the development of students' multifaceted potential" (Ministry of Education of the People's Republic of China, 2001, p. 3). This policy support plays a crucial role in enhancing teachers' language assessment literacy in China. This study revealed that surveyed teacher participants demonstrated a statistically high level of understanding and competence in applying assessment for learning in their teaching. As confirmed by interview participants, their assessment practices are largely influenced by government and institutional policies. The policy establishes guidelines and requirements, shaping the standards and expectations in educational settings. Teachers are suggested adhering to regulations while designing and implementing assessments and teaching.

Regarding teachers' needs for LAL development, this study found that teaching experience has a slightly medium impact, consistent with the previous studies (e.g., Smith et al., 2014; Vogt & Tsagari, 2014). Contrary to the findings of existing studies (Brown & Remesal, 2012; Xie & Tan, 2019), our study revealed that teachers with less than 20 years of teaching experience exhibit a stronger

desire for LAL development. As highlighted in the interviews, LAL has emerged as a new focal point for academic research. The insights gained from LAL research not only assist teachers in comprehending the nuances of LAL but also facilitate their self-development of competence in language assessment. The gaps in LAL development needs between groups gradually narrow down from novice teachers to experienced teachers. Moreover, in China, teachers who have undergone workplace LAL training tend to emphasize the importance of LAL development, highlighting the positive impact of in-service training on fostering teachers' awareness of LAL.

Similar to the findings by Xu and Brown (2017), this study also identified in the interviews that the absent requirement for LAL in the teacher qualification and the insufficient emphasis on teachers' responsibility in language assessment contribute to teachers' perceived need for LAL. Without mandatory requirements regarding LAL, teachers may not prioritize the development of LAL, potentially hindering their ability to effectively implement assessment practices in their teaching. Although assessment research experience influences teachers' competency in language assessment, this study shows that there is no direct link between the need for LAL and assessment research experience, which indicates that assessment research experience may not translate into a perceived need for further development in LAL among teachers. Furthermore, this study found teachers with a linguistics background possess better awareness of LAL than teachers with expertise in English literature or translation.

Pre-service training has an impact on LAL development (Jin, 2010; Xu & Brown, 2017). The importance of LAL has been recognized and integrated into the teacher education programme for pre-service English teachers in Chinese universities, particularly in formative assessment, as indicated by the interview results demonstrated in this study. The strong demand for LAL development observed among teachers in their career-developing stage raises concerns about the effectiveness of in-service training programmes. As pointed out by Jin (2010) a decade ago, the current LAL training programmes still fall short in terms of practicality and sufficiency to enhance teachers' LAL competence. Rather than focusing solely on the infusion of theoretical assessment concepts, it is preferable for training programmes to provide opportunities for teachers to practice assessment and receive guidance on how to make effective changes (Yan et al., 2018).

Despite its exploratory nature, this study offers some insights into the strategies that can be adopted to enhance teacher assessment competence and boost teachers' confidence as assessors. Given the power and significant influence of governmental and institutional policies in promoting assessment for learning and the prescribed test structure for test writing, this study suggests constructing a shared language assessment community to improve teachers' LAL and develop teacher identity as an assessor. Creating a collaborative learning environment among teachers can facilitate the sharing of new information and assessment-related expertise. In our view, the pinnacle of internalizing the "essence" of assessment literacy is to construct a teacher's identity as an assessor. Teachers' cognition of their perceived role as language assessors play a vital role in the evaluation of their effectiveness in the field of language assessment (Looney et al., 2018).

To create such an assessment community and awareness of being an assessor, it is important to bring individuals together and foster a sense of shared purpose and understanding. Regular communication among teachers is necessary to ensure that everyone is informed about updated practices in terms of language assessment and to promote collaboration and teamwork. Clear and shared goals of language assessment should be established, and teachers should be encouraged to share their expertise, knowledge, and resources through discussions and activities. In addition, encourages teachers with assessment expertise to work together with teachers who lack assessment knowledge. Necessary reflection on teacher assessment literacy is needed as well. "Teachers can become more in tune with their sense of self and gain a deeper understanding of how this self fits in with larger contexts involving others" (Xu & Brown, 2016, p. 158). Continued efforts from the government and assessment researchers are needed to make assessment knowledge more accessible to teachers.

This study contributes to an understanding of the level of teacher assessment literacy within the researched context and therefore benefits teacher professional development. Understanding what teachers know and how they perceive their role could enable them to construct purposefully responsive teacher professional assessment education that responds not only to current testing and assessment circumstances in China but also to teachers' learning needs.

Although this study has revealed some mediating factors that can influence teachers' LAL based on interviews with five in-service English university teachers, there is insufficient evidence to show how those influences occur and the effect size of those factors. Further study on how those factors (e.g., LAL training, policy, etc.) affect the teacher language assessment literacy development could make a valuable contribution in this domain. Due to the lack of detailed questions on LAL needs, the analysis of the impact of teachers' perceived LAL needs on teachers' perceived LAL competence could be our further study as well. Investigating through the lens of teachers' perceptions may result in subjective evaluation and not reflect their actual knowledge and practices. Further study could incorporate classroom observations on teachers' actual assessment practices in the classroom to triangulate the conclusions drawn from this study.

Notes

1. P16. Using formative assessment to measure language learning progress.
2. P17. Using summative assessment to measure language learning progress.

Acknowledgements

We would like to thank the teachers who helped us to collect the survey data and all the respondents for their participation. We would also like to thank the interviewees for their valuable feedback and participation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Lulu Zhang is a PhD student in Modern Languages and Linguistics at the University of Southampton. She received her MA (2014) in Applied Linguistics in Language Teaching from the University of Southampton. Her research focuses on foreign language teaching, language assessment and teacher education.

Ying Zheng is an associate professor in Faculty of Humanities at the University of Southampton, UK. She obtained her MEd (2005) and PhD (2009) degrees in Cognitive Studies from Queen's University, Canada, specializing in Second Language Testing and Assessment.

Dongying Yang is a professor in College of Foreign Languages at North China University of Science and Technology. Her research interests include English language teaching and English literature.

ORCID

Lulu Zhang http://orcid.org/0000-0002-0544-9187
Ying Zheng http://orcid.org/0000-0003-2574-0358

References

Brookhart, S.M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practices*, *30*(1), 3–12. https://doi.org/10.1111/j.1745-3992.2010.00195.x

Brown, G.T., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology*, *15*(1), 75–89. https://doi.org/10.5209/rev_sjop.2012.v15.n1.37286

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

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, *25*(3), 327–347. https://doi.org/10.1177/0265532208090156

DeLuca, C., & Klinger, D.A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education Principles, Policy & Practice*, *17*(4), 419–438. https://doi.org/10.1080/0969594X.2010.516643

Field, A. (2018). *Discovering statistics using SPSS* (5th ed.). Psychology Press.

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

Gotch, C.M., & French, B.F. (2014). A systematic review of assessment literacy measures. *Educational Measurement Issues & Practice*, *33*(2), 14–18. https://doi.org/10.1111/emip.12030

Gullickson, A.R. (1993). Matching measurement instruction to classroom-based evaluation: Perceived discrepancies, needs, and challenges. In S.L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 1–25). Buros Institute of Mental Measurements, University of Nebraska-Lincoln. http://digitalcommons.unl.edu/burosteachertraining/3

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). De Gruyter.

He, L., & Zhang, J. (2020). Origin and development of language testing in China. *Journal of Zhejiang University (Humanities and Social Science*, *5*(6), 29–38. https://doi.org/10.3785/j.issn.1008-942X.CN33-6000/C.2019.08.101

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

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

Koh, K., Burke, L.E.C.A., Luke, A., Gong, W., & 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

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

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

Leech, N.L., & Onwuegbuzie, A.J. (2009). A typology of mixed methods research designs. *Quality & Quantity: International Journal of Methodology*, *43*(2), 265–275. https://doi.org/10.1007/s11135-007-9105-3

Leung, C. (2014). Classroom-based assessment issues for language teacher education. In A.J. Kunnan (Ed.), *The companion to language assessment* (pp. 1510–1519). Wiley Blackwell.

Liu, J., & Li, X. (2020). Assessing young English learners: Language assessment literacy of Chinese Primary School English teachers. *International Journal of TESOL Studies*, *2*(4), 36+. https://link.gale.com/apps/doc/A653407914/AONE?u=anon~e5c60557&sid=bookmark-AONE&xid=762ae917

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. 17–37). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-60092-9_2

Loewen, S., & Gonulal, T. (2015). Exploratory factor analysis and principal components analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp.182–212). Routledge.

Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education Principles, Policy & Practice*, *25*(5), 442–467. https://doi.org/10.1080/0969594X.2016.1268090

Malone, M. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, *30*(3), 329–344. https://doi.org/10.1177/0265532213480129

Ministry of Education of the People's Republic of China. (2001). *基础教育课程改革纲要(试行)* [Compendium of basic education curriculum reform (trial implementation)]. http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_309/200412/4672.html

Ministry of Education of the People's Republic of China. (2004). *College English curriculum requirements (for trial implementation)*. Shanghai Foreign Language Education Press.

Pallant, J. (2010). *SPSS survival manual*. McGraw-Hill Education (UK).

Plake, B.S., Impara, J.C., & Fager, J.J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement Issues & Practice*, *12*(4), 10–12. https://doi.org/10.1111/j.1745-3992.1993.tb00548.x

Qian, D.D. (2014). School-based English language assessment as a high-stakes examination component in Hong Kong: Insights of frontline assessors. *Assessment in Education Principles, Policy & Practice*, *21*(3), 251–270. https://doi.org/10.1080/0969594X.2014.915207

Reynolds-Keefers, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment, Research & Evaluation*, *15*(8), 1–9. https://doi.org/10.7275/psk5-mf68

Smith, L.F., Hill, M.F., Cowie, B., & Gilmore, A. (2014). Preparing teachers to use the enabling power of assessment. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 303–323). Springer.

Sun, H.Y., & Zhang, J. (2022). Assessment literacy of college EFL teachers in China: Status quo and mediating factors. *Studies in Educational Evaluation*, *74*, 1–10. https://doi.org/10.1016/j.stueduc.2022.101157

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Pearson Education.

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

Volante, L., & Fazio, X. (2007). Exploring teacher candidates'assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, *30*(93), 749–770. https://doi.org/10.2307/20466661

Wang, L., Lee, I., & Park, M. (2020). Chinese university EFL teachers' beliefs and practices of classroom writing assessment. *Studies in Educational Evaluation*, *66*, 1–11. https://doi.org/10.1016/j.stueduc.2020.100890

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

Xu, Y.T., & Brown, G. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, *6*, 133–158. http://hdl.handle.net/10722/245352

Xu, Y.T., & Brown, G.T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

Xu, Y., Han, S., & Chen, Y. (2016). 大学英语教师语言评估素养调查报告[College English teachers' language assessment literacy: A survey report]. *中国外语教育[Foreign language education in China]*, *9*(4), 60–69. https://kns.cnki.net/kcms/detail/detail.aspx?FileName=WYUJ201604008&DbName=CJFN2016

Yan, X., Zhang, C., & Fan, 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

Zhang, L.L. (2018). 我国教师测评素养研究述评[Exploration of research on Teacher assessment literacy in China]. *中国考试[China examinations]*, *2018*(8), 69–73. https://doi.org/10.19360/j.cnki.11-3303/g4.2018.08.012