The Role of China's Standards of English Language Ability (CSE) Assessment Literacy in Promoting Teachers' Language Assessment Literacy Profiles: Evidence from Chinese University EFL Teachers

Tingting Yin[1] · **Xiucai Fang**[1]

Abstract Released in 2018, China's Standards of English Language Ability (CSE) set a benchmark for language education and assessment in China, integrating with international standards. This study investigates whether CSE assessment literacy (understanding and applying CSE standards in assessment practice) enhances Language Assessment Literacy (LAL) among 233 university EFL teachers in China. Using a self-designed questionnaire assessing current LAL and CSE assessment literacy, the findings indicate that the teachers' self-assessed LAL, classified into high, moderate, and low LAL profiles, was generally unsatisfactory. Awareness of the principles was highest in teachers with a moderate LAL profile and was minimal in teachers with a low profile. Proficiency in "CSE Background," "Competence Classification," and "Table Scales" was closely linked to elevated LAL profiles. Specifically, "CSE Background" improved LAL across low and moderate profiles, while "Competence Classification" and "Table Scales" mainly benefited the moderate profile, informing effective LAL and CSE training strategies.

Apply K-means clustering analysis to classify LAL profiles. Conduct logistic regression analysis to evaluate the impact of CSE assessment literacy on LAL profiles. Explore how CSE assessment literacy can enhance LAL profiles.

✉ Xiucai Fang
fangxiucai@hnnu.edu.cn

Tingting Yin
yintingting@hnnu.edu.cn

1 School of Foreign Languages, Huainan Normal University, Dongshan West Road, Huainan 232038, China

Introduction

Language Assessment Literacy (LAL) involves understanding and applying sound assessment practices for language learning evaluation in various contexts (Stiggins, 1991). Research has extensively examined the factors that improve teachers' LAL, revealing the interplay between contextual and experiential factors in the evolution of LAL (Yan et al., 2018). The China's Standards of English Language Ability (CSE) serve as a case study of these influences within China's unique language assessment landscape.

In 2018, China's Ministry of Education and National Language Commission introduced the CSE, the nation's inaugural national English proficiency framework. The release of the CSE aimed to standardize English proficiency measures, guide teaching and assessment, and ensure international compatibility of test scores (Liu, 2015). The initiative formed part of broader reforms to optimize examination and enrollment processes (Lin, 2016). This major policy shift motivated many EFL educators, especially those at Chinese universities, to adopt and research CSE-oriented assessment practices. A search conducted on "CNKI," China's comprehensive academic journal full-text database, found 554 articles dedicated to CSE were published from 2014 to April 2024. However, research on CSE peaked in 2019 and has waned gradually since then (see Fig. 1), indicating a potential decline in its appeal among Chinese scholars.

This period of declining interest has presented an important opportunity for educators to reflect on prior practices and adapt their assessment methods to the CSE standards.
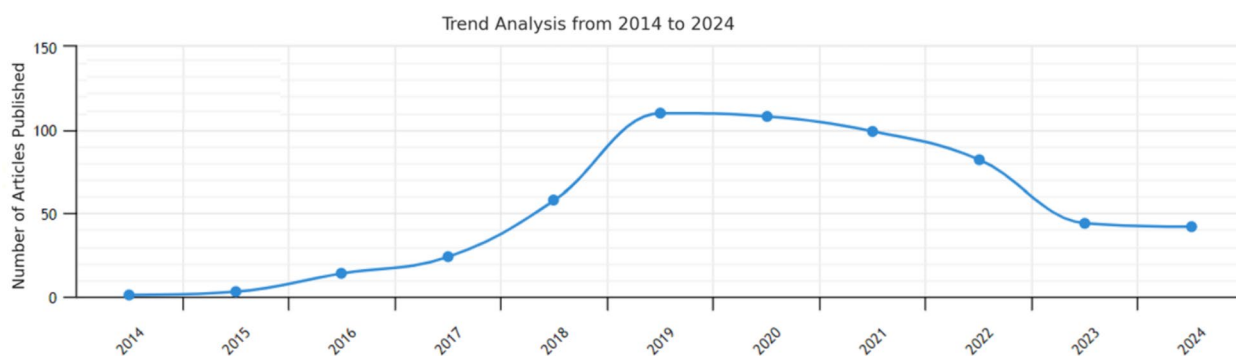
Fig. 1 Trend of articles counted by years (CNKI, 2024). A search on the subject of "China's Standards of English Language Ability" resulted in the retrieval of 554 articles from the CNKI database, as of April 28, 2024

This has led to proficiency in understanding and applying these standards to language assessment, known as CSE assessment literacy (Pan, 2020). However, the impact of this literacy on overall LAL is yet to be investigated. It is crucial to explore whether CSE assessment literacy can enhance general LAL because such research, conducted at this critical juncture, could uncover the CSE's ability to further language teachers' professional growth, inform CSE-based professional training, and potentially reignite interest in CSE-related studies.

Addressing the research gap in existing studies, this study firstly profiles Chinese university teachers' LAL at various developmental stages and then quantitatively examines the correlation between CSE assessment literacy and these LAL profiles.

Literature Review

Language Teachers' LAL Development

LAL has been conceptualized as a construct that encompasses various dimensions. Grounded in Davies' (2008) foundational components of *skills*, *knowledge*, and *principles*, research on LAL has incorporated sociocultural and sociopolitical dimensions to address its inherently contextual nature (e.g., Fulcher, 2012; Inbar-Lourie, 2008). In Fulcher's (2012) three-tier hierarchical LAL model, the language assessment *contexts* of history, society, politics, and philosophy are located in the top layer, with *principles* (processes, principles, and concepts) in the intermediate layer and *practice* (practical skills and knowledge) at its foundation. Taylor (2013) hypothesized the LAL profile of teachers, covering theoretical, technical, sociocultural, and decision-making domains. Complementary to this, Giraldo (2018) provided a detailed descriptor-based definition of language teachers' LAL.

The American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association (1990) formulated seven standards for teacher assessment development, focusing on skills and principles in creating, administering, scoring, and utilizing assessments. While these standards have been used globally in researching teachers' LAL development, there is a trend to distinctively distance LAL from assessment literacy (AL) and associate it with teacher professional growth and practice (e.g., Lan & Fan, 2019). Hence, it is essential to study teachers' LAL development anchored in the defined concept of LAL.

Contextual and Experiential Factors Mediate Teachers' LAL Development

Recent research has underscored that teachers' LAL development stems from a multifaceted interplay of factors. Contextual factors such as national and local assessment cultures, educational policies, institutional mandates, and infrastructures, along with experiential factors such as prior experiences, educational background, and teaching practices significantly influence LAL (Crusan et al., 2016). Yan et al. (2018) examined the integration of contextual and experiential factors in shaping teachers' LAL development, highlighting that these factors, through continuous self-reflection, guide the evolution of intuitive and principled assessment competencies in teachers. However, considering teachers' LAL as a developmental continuum (Pill & Harding, 2013), the interaction of these factors across various stages of LAL development warrants deeper exploration.

The Case of China: CSE Assessment Literacy Impacts Teachers' LAL Development

The Definition of CSE Assessment Literacy

CSE assessment literacy encapsulates essential LAL elements tailored to the CSE standards, focusing on grasping

and effectively applying these standards to language assessment (Jin, 2018; Pan, 2020). Originating from Jin's (2018) framework, it involves understanding the CSE's social context, theoretical foundations, and practical application. Pan (2020) extended this concept by introducing a model delineating CSE assessment literacy profiles for five key groups, incorporating dimensions such as social context, theoretical underpinnings, development methods, and application suggestions (see Fig. 2). Similar to Taylor's (2013) LAL profiles, this model visually represents varying CSE assessment literacy levels among different stakeholders. It is a conceptual tool that requires further empirical validation.

*The Practices of CSE Assessment Literacy*

Following its launch in 2018, the CSE has had a profound impact on research in English language teaching and assessment. This research has not only been prominent within China (see Fig. 1) but has also attracted attention in the international academic community (e.g., Peng et al., 2021) (see Fig. 3).

In summary, previous literature has suggested that the interaction of contextual and experiential factors can mediate teachers' LAL development, with continuous self-reflection during assessment practices acting as a catalyst for development. Within China's CSE context, although practitioners tailor their approaches to align with the CSE standards, thereby enhancing their CSE assessment literacy to varying extents, there is a noticeable gap in empirical research exploring the potential of CSE assessment literacy for fostering LAL advancement. A thorough comprehension of this dynamic could empower professional training entities to devise CSE-oriented programs addressing practitioners' specific needs in utilizing CSE for LAL enhancement.
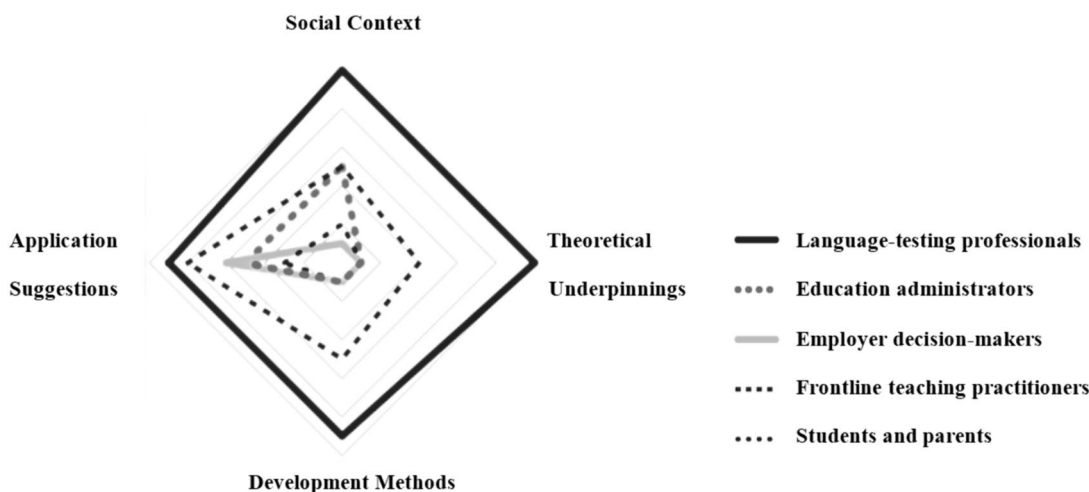
Fig. 2 CSE assessment literacy dimensions for different stakeholder groups (Pan, 2020)
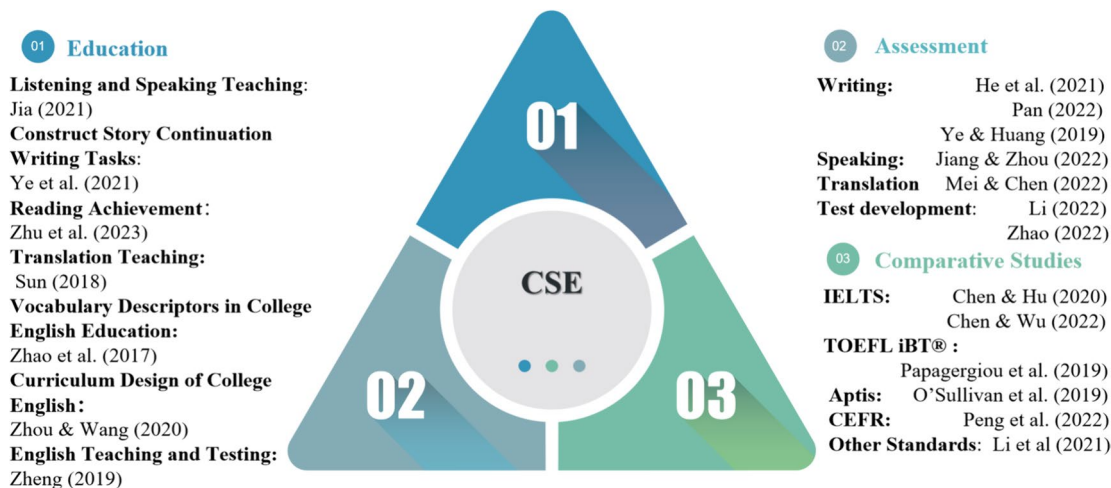
Fig. 3 Studies on CSE applications. See Appendix I for the references

In terms of methodology, Yan et al. (2018) utilized retrospective interviews with three teachers to explore the interaction between contextual and experiential factors in teachers' LAL, advocating for broader, more diverse sampling to better understand this influence across various LAL profiles. Our study was conducted using a self-assessment survey questionnaire, recognized for its ability to efficiently engage large participant groups and mitigate anxiety through self-evaluation (e.g., Sun & Zhang, 2022). While prior studies often point to suboptimal LAL levels among Chinese EFL teachers (e.g., Fan & Jin, 2020; Sun & Zhang, 2022), there are concerns regarding existing scales. These include reliance on generic AL frameworks rather than LAL-specific models (e.g., Lan & Fan, 2019), and a focus on limited aspects of LAL, such as knowledge and skills, without incorporating fundamental principles (e.g., Sun & Zhang, 2022).

Therefore, to address these issues, we administered a large-scale survey to reassess the current profiles of LAL among Chinese university EFL teachers in the CSE context and investigate the association between CSE assessment literacy and LAL profiles. Specifically, the study aimed to address two key research questions (RQ):

RQ1: What are the current profiles of LAL among Chinese university EFL teachers?
RQ2: Is there a relationship between their CSE assessment literacy and LAL profiles? If so, how does CSE assessment literacy impact their LAL profiles?

Methodology

This research used a quantitative method in developing the LAL & CSE questionnaire to investigate the LAL profiles and CSE assessment literacy of university EFL teachers in China. The detailed steps of the method are provided below.

Instrument Development

We initiated the development of a LAL & CSE scale that incorporated elements based on both LAL and CSE assessment literacy models. This involved a multi-stage development process, as shown in Fig. 4.

In developing the LAL section of our questionnaire, we initially adopted Davies' (2008) widely recognized LAL model and integrated Giraldo's (2018) descriptor-based definition into our framework. Giraldo's adaptation of Davies' model offered a detailed conceptual structure, highlighting essential *knowledge*, *skills*, and *principles* for language teachers (Puspawati, 2019). This structure featured eight dimensions with 66 unique descriptors, forming the basis of the LAL part of our questionnaire. The CSE section was based on Pan's (2020) model, encompassing four dimensions of CSE assessment literacy, and providing a comprehensive overview of the expected CSE assessment literacy. Thus, our initial questionnaire (Version 1.0) was created.

The initial questionnaire was adapted for China's higher education context, particularly considering the EFL teaching context during three internal reviews (Versions 2.0–2.2). After expert reviews, modifications were made to enhance the content and linguistic suitability, including considering the influence of Mandarin and regional dialects on English

Fig. 4 Overview of the instrument development process
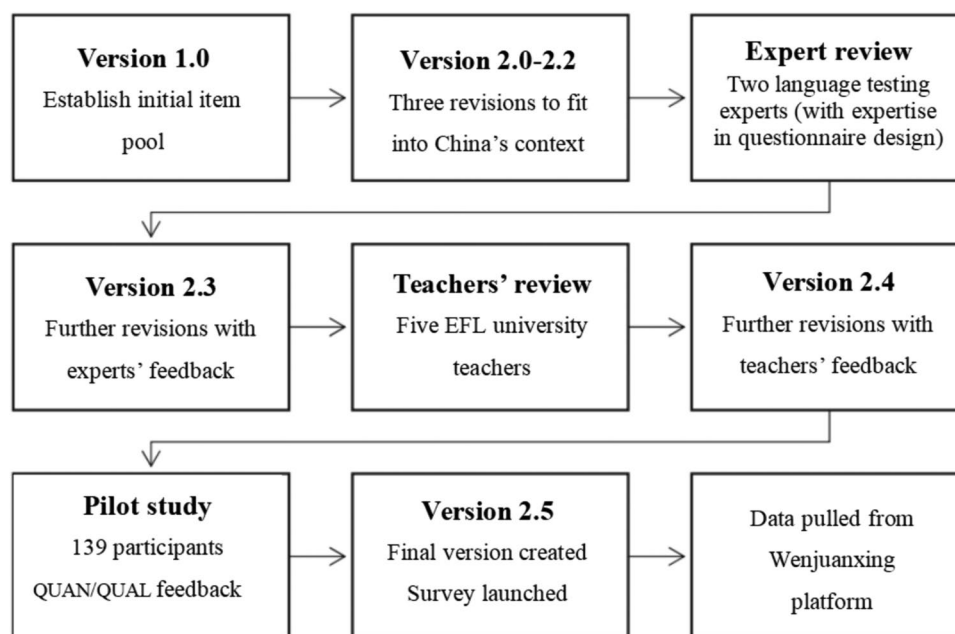
language assessment. This process led to Version 2.3 of the questionnaire, which was further refined to Version 2.4 based on feedback from five EFL university teachers.

A pilot study with 139 Chinese university EFL teachers assessed Version 2.4 of a questionnaire, with an attention check used to validate responses. Post-cleanup, 78 responses were analyzed, yielding high-reliability Cronbach alpha values: 0.975 for the LAL section, 0.836 for the CSE Assessment Literacy section, and 0.971 overall. Based on qualitative feedback, Chinese translations were added to the Likert scale options in Version 2.5, the final questionnaire.

Instrument Format

The questionnaire consisted of 70 items in three sections: Section 1. Demographic Features, Section 2 LAL, and Section 3 CSE Assessment Literacy. The LAL section contained 54 items measured on a 5-point Likert scale: 5 (extremely knowledgeable), 4 (knowledgeable), 3 (moderately knowledgeable), 2 (slightly knowledgeable), and 1 (not at all knowledgeable). The CSE Literacy section included eight items measured on a 3-point Likert scale: 3 (extremely knowledgeable), 2 (generally knowledgeable), and 1 (not knowledgeable) (see Fig. 5).

Main Trial Sample

We used convenience and snowball sampling methods, utilizing Wenjuanxing, a leading online survey tool in China, to collect data from Chinese university EFL teachers. The survey, detailed on the introductory page with funding information, was broadly distributed via professional teacher groups to mitigate network biases and ensure varied participant representation. Initially disseminated through QQ and WeChat, targeting EFL teachers and educational competition participants, we encouraged sharing within their networks and promoted the survey via social media, influential individuals, and EFL teacher events at regional and national levels, thus significantly broadening our outreach.

Participants voluntarily completed an anonymous survey, assured their data was for research only. Over a month, we collected 440 responses, but after applying strict cleaning criteria (removing responses failing attention checks or incomplete for later data mining–based analysis), we obtained 233 valid ones: 107 from East China, 106 from Central China, and 20 from West China, detailed in Table 1.

Reliability and Validity of the Questionnaire

Analyzing data from 233 participants with SPSS 26.0 showed high reliability (Cronbach's alpha: 0.979 for LAL, 0.857 for CSE Assessment Literacy, 0.973 overall). The Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test indicated sampling adequacy for Exploratory Factor Analysis (EFA), with KMO values of 0.958 (LAL) and 0.835 (CSE), and both sections showing significant Bartlett's test results ($p < 0.00$).

The EFA for the LAL section initially suggested a 7-factor solution, later refined to 3 factors via screen plot analysis, accounting for 64.2% of variance. "Skills in Educational Measurement" and "Technological Skills" merged into "Knowledge," forming Factor 1: "Knowledge and Skills in Educational Measurement and Technology" (KSEMT), with Factor 2 as "Instructional and Language Assessment Design Skills" (ILADS), and Factor 3 as "Principles" (Appendix II). In the CSE section, EFA identified two factors from eight items, explaining 71.1% of variance, named "CSE Background with Theoretical Underpinnings and International Test Alignment" and "Yardstick for Language Education and CSE Practices" (Appendix III).

Fig. 5 LAL & CSE Assessment Literacy questionnaire
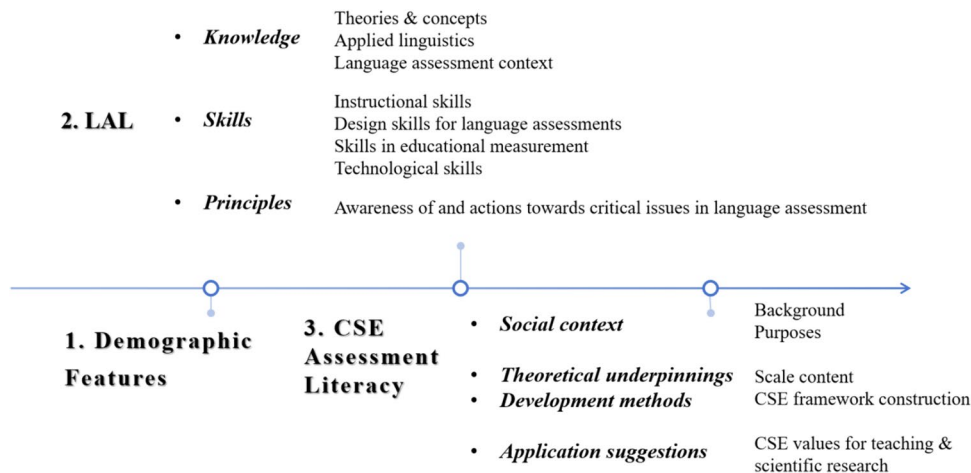
Table 1 Participants' demographic information (N = 233)

	Gender		Age				Degree				Geographical location		
	Male	Female	< 30	31–40	41–50	> 50	BA below	BA	MA	PhD	East	Central	West
N	57	176	15	66	113	39	1	38	156	38	107	106	20
%	24.5	75.5	6.4	28.3	48.5	16.7	0.4	16.3	67	16.3	45.9	45.5	8.6

	Teaching years				Test workshop participation experience		Test research participation experience		Test design experience	
	< 3	4–10	11–20	> 20	Yes	No	Yes	No	Yes	No
N	16	26	97	94	114	119	64	169	220	13
%	6.9	11.2	41.6	40.3	48.9	51.1	27.5	72.5	94.4	5.6

Data Analysis

To tackle RQ1, we conducted *K*-means cluster analysis with factor scores (Factor 1, 2, 3, and composite F) to categorize LAL profiles in university EFL teachers, supplemented by descriptive statistics and univariate ANOVAs for a general overview of LAL levels and a Chi-square test on Section 1 data for demographic differences. For RQ2, a Chi-square test assessed associations between LAL profiles and CSE assessment literacy, followed by a multinomial regression to examine the influence of CSE assessment literacy on LAL profiles.

Clustering organizes items into distinct groups based on high intra-cluster similarity and inter-cluster dissimilarity. *K*-means clustering, noted for its efficiency and effectiveness, often outperforms other methods for document data (Liang et al., 2012). In our study, *K*-means analysis involved calculating composite scores for Factors 1, 2, and 3, plus an overall LAL level score (F), using SPSS 26. Factor loadings were 28.389% (F1), 19.73% (F2), and 16.068% (F3), leading to a composite score formula (F = 0.28389*F1 + 0.1973*F2 + 0.16068*F3). The *K*-means clustering was executed in R 4.2.1, with detailed data in Appendix IV.

Results

RQ1: What Are the Current Profiles of LAL Among Chinese University EFL Teachers?

Cluster Results

A three-cluster model best fit the data (Fig. 6). Cluster 1 categorized a subset of 116 university EFL teachers, while Cluster 2 comprised 86 EFL teachers and Cluster 3 included 31 teachers. The clustering results showed imbalanced data distributions. To investigate the CSE assessment literacy in different LAL clusters, we applied the hard *K*-means clustering algorithm to categorize each teacher into each cluster, which is more likely to create imbalanced data distribution than the fuzzy *K*-means clustering, allowing multiple memberships across all clusters. This asymmetrical distribution

Fig. 6 Cluster categories

Cluster 2: *n*=86

36.91%

Cluster 1: *n*=116

49.79%

Cluster 3: *n*=31

13.30%
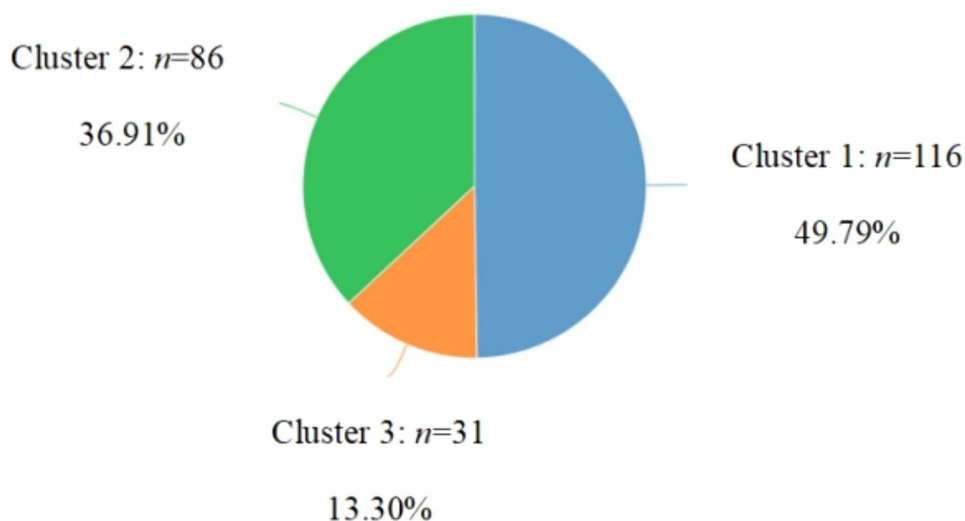
of data may reflect the intrinsic nature of the dataset in the real world (e.g., risk management) (e.g., Liang et al., 2012).

Descriptive Statistics and Univariate ANOVAs of LAL Profiles

As shown in Table 2, the participants' LAL levels were not very satisfactory (M = 3.39, SD = 0.35). Additionally, Chinese university EFL teachers rated KSEMT the lowest (M = 3.17, SD = 022) and "Principles" the highest (M = 4, SD = 0.08).

Based on the "Overall LAL Levels," we categorized Cluster 1 as the high LAL profile (LAL(H)), indicating advanced LAL development. Cluster 2 was termed the moderate LAL profile (LAL(M)), and Cluster 3 as the low LAL profile (LAL(L)). The mean scores of each LAL component within these profiles generally aligned with their respective developmental stages, aligning with their designated labels. However, the only exception was "Principles" as teachers in LAL(M) were rated higher (M = 4.35; CD = 0.11) than those in LAL(H) (M = 4.28; SD = 0.06). Additionally, teachers in LAL(L) displayed exceptionally low levels in "Principles" (M = 1.98; SD = 0.05).

Comparison of Background Characteristics in the Three LAL Profiles

The Chi-square test results indicated statistical significance for "Teaching Years" and "Test Research Participation Experience," while other variables showed no significant differences (see Table 3). These two variables served as covariables in the subsequent regression analysis.

Is There a Relationship Between Their CSE Assessment Literacy and LAL Profiles? If so, How Does CSE Assessment Literacy Impact Their LAL Profiles?

A Chi-square test was applied to find out whether LAL profiles were associated with CSE assessment literacy. Subsequently, multinomial logistic regression was employed to identify factors influencing LAL profiles.

Table 2 Descriptive statistics and univariate ANOVAs of LAL profiles

LAL profiles	KSEMT (N=29)		ILADS (N=17)		Principles (N=8)		Overall LAL levels		F	P
	M	SD	M	SD	M	SD	M	SD		
Cluster 1 (H)	3.76	0.25	3.96	0.18	**4.28**	0.06	3.90	0.27	9.815	0.000**
Cluster 2 (M)	2.68	0.31	3.24	0.31	**4.35**	0.11	3.10	0.65	23.437	0.000**
Cluster 3 (L)	2.32	0.10	2.39	0.08	**1.98**	0.05	2.29	0.16	18.734	0.000**
Overall LAL profiles	**3.17**	0.22	**3.49**	0.20	**4.00**	0.08	**3.39**	0.35		

KSEMT Knowledge and Skills in Educational Measurement and Technology; *ILADS* Instructional and Language Assessment Design Skills

[*]p < 0.05

[**]p < 0.01

Table 3 Comparison of background characteristics in the three LAL profiles

Variable		Cluster 1 LAL (H) (n=116)	Cluster 2 LAL (M) (n=86)	Cluster 3 LAL (L) (n=31)	Total	χ^2	p
Teaching years	<3	7 (6.03)	7 (8.14)	2 (6.45)	16 (6.87)	13.418	0.037*
	4–10	11 (9.48)	15 (17.44)	0 (0.00)	26 (11.16)		
	11–19	46 (39.66)	31 (36.05)	20 (64.52)	97 (41.63)		
	>20	52 (44.83)	33 (38.37)	9 (29.03)	94 (40.34)		
Total		116	86	31	233		
Test research participation experience	Yes	46 (39.66)	12 (13.95)	6 (19.35)	64 (27.47)	17.556	0.000**
	No	70 (60.34)	74 (86.05)	25 (80.65)	169 (72.53)		
Total		116	86	31	233		

[*]p < 0.05

[**]p < 0.01

Association of LAL Profiles with CSE Assessment Literacy

We conducted a Chi-square test with the data of the previous cluster results as independent variables and the data of Section 3 CSE Assessment Literacy (Items 63–70) as dependent variables. The results are shown in Table 4.

The results showed that CSE assessment literacy was associated with LAL profiles. Four CSE assessment literacy items were found to have statistically significant differences between the three different LAL profiles. They were Item 63 of the "CSE Background," Item 66 of "Competence Classification," Item 69 of "Table Scales," and Item 70 of "Teaching & Research Values." Other items were not statistically significant (P > 0.05).

Differences in CSE Assessment Literacy Influence LAL Profiles

A multinomial logistic regression analysis, with the LAL clusters as the dependent variable, the CSE assessment literacy of Item 63, Item 66, Item 69, and Item 70 as independent variables, and Item 5 of "Teaching Years" and Item 7 of "Test Research Participation Experience" as covariates, was applied to identify how CSE assessment literacy impacted LAL profiles (see Table 5).

When Item 63 was scored as 1 (Not knowledgeable), indicating teachers' lack of awareness of the "CSE Background," for each unit increase in this lack of awareness, the odds of being in Cluster 3 (LAL(L)) increased (a positive coefficient of 1.877) 6.535 times (Exp(B) = 6.535) that of the teachers who were extremely knowledgeable about this background (Item 63 = 3), compared to Cluster 1 (LAL(H)). Conversely, when Item 63 was scored as 2 (Generally knowledgeable), reflecting teachers' general awareness of "CSE Background," for each unit increase, the odds of being in Cluster 2 (LAL(M)) decreased (a negative Coefficient of −1.117) significantly by 67.30% (0.327–1.0 = −0.673) compared to LAL(H). This means that the odds of being in Cluster 1 (LAL(H) increased by 30.58% (1/0.327 = 3.058). Namely, when teachers did not know the "CSE Background," they were less likely to be in the LAL(H) profile. In contrast, when teachers had general knowledge of this aspect, they were more likely to be in the LAL(H) profile. Thus, knowing the "CSE Background" can promote a higher LAL profile.

In this sense, teachers with a deep understanding of the "Competence Classification" and "Table Scales" were found

Table 4 Comparison of CSE assessment literacy in the three different LAL profiles

Items	LAL (H)	LAL (M)	LAL (L)	Total	$\chi 2$	P
Item 63	China's Standards of English Language Ability (CSE) were officially published by the Ministry of Education of the People's Republic of China and the National Language and Literature Commission in 2018 (the CSE background).					
3	51 (43.97)	32 (37.21)	7 (22.58)	90 (38.63)	13.059	0.011*
2	42 (36.21)	27 (31.40)	8 (25.81)	77 (33.05)		
1	23 (19.83)	27 (31.40)	16 (51.61)	66 (28.33)		
Item 66	The CSE classified learners' English competence into three stages and nine levels: Elementary (Levels 1–3), Intermediate (Levels 4–6), and Advanced (Levels 7–9), in alignment with elementary years (Levels 1–2), junior and senior secondary years (Levels 3–4), non-English majors' undergraduate years (Levels 5–6), English majors' undergraduate years (Level 7), professional advanced English users (Levels 8–9).					
3	73 (62.93)	29 (33.72)	13 (41.94)	115 (49.36)	18.628	0.001**
2	20(17.24)	31 (36.05)	8 (25.81)	59 (25.32)		
1	23 (19.83)	26 (30.23)	10 (32.26)	59 (25.32)		
Item 69	The CSE descriptive framework consists of 86 scale and subscale tables involving language abilities of comprehension, expression, and mediation (namely, translation and interpretation), sub-abilities of listening, speaking, reading, writing, translation, and interpreting, and even learning strategies and self-assessment.					
3	26 (22.41)	8 (9.30)	3 (9.68)	37 (15.88)	12.572	0.014*
2	43 (37.07)	24 (27.91)	12 (38.71)	79 (33.91)		
1	47 (40.52)	54 (62.79)	16 (51.61)	117 (50.21)		
Item 70	Being familiar with the CSE (background, purposes, content tables, teaching application) is valuable for teaching and scientific research activities.					
3	70 (60.34)	44 (51.16)	14 (45.16)	128 (54.94)	15.956	0.003**
2	34 (29.31)	37 (43.02)	8 (25.81)	79 (33.91)		
1	12 (10.34)	5 (5.81)	9 (29.03)	26 (11.16)		

*p < 0.05
**p < 0.01

Table 5 Logistic regression analysis of the impact of CSE assessment literacy on LAL

Clusters		B	S.E	Wald	df	P	Exp(B)	95% C.I. for Exp(B)	
								Lower	Upper
3	Intercept	− 2.392	0.782	9.358	1.000	0.002			
	[Item 63 = 1]	1.877	0.520	13.042	1.000	**0.000**	**6.535**	2.359	18.100
	[Item 70 = 1]	1.027	0.526	3.817	1.000	**0.051**	**2.793**	0.997	7.827
2	Intercept	− 2.025	0.566	12.820	1.000	0.000			
	Item 5	− 0.307	0.116	7.011	1.000	0.008	0.735	0.586	0.923
	Item 7	1.216	0.250	23.760	1.000	0.000	3.375	2.070	5.505
	[Item 63 = 2]	**− 1.117**	0.302	13.644	1.000	**0.000**	**0.327**	0.181	0.592
	[Item 66 = 2]	1.425	0.317	20.166	1.000	**0.000**	**4.159**	2.233	7.747
	[Item 69 = 1]	0.752	0.356	4.450	1.000	**0.035**	**2.121**	1.055	4.266

The reference cluster is 1. The table only shows the items with P < 0.051. The reference item cluster is 3 (extremely knowledgeable)

to be more likely to be in the LAL(H) profile, while those with only a general or no knowledge were less likely to be in LAL(H). However, "Teaching & Research Values" had a borderline P of 0.051 and was still considered nonsignificant due to the typical significance threshold of 0.05. To summarize, three factors of "CSE Background," "Competence Classification," and "Table Scales" were identified as factors to promote LAL profiles. Additionally, "Teaching Years" and "Test Research Participation Experience" were not significant in differentiating low and high LAL levels but were significant between moderate and high levels.

Discussion

LAL Profiles Among Chinese University EFL Teachers

Overall LAL Profiles

In this study, the overall LAL levels were deemed unsatisfactory, with Chinese university EFL teachers averaging a 67.8% LAL score, despite 49.7% of participants being categorized in the high LAL profile. This finding corroborates recent studies (Fan & Jin, 2020; Sun & Zhang, 2022) that similarly underscored the insufficient LAL levels among Chinese EFL teachers.

Overall, participants exhibited the highest literacy in "Principles" and the lowest in "Knowledge and Skills in Educational Measurement and Technology," aligning with Puspawati's (2019) observation of university teachers' superior understanding of assessment principles over their knowledge and skills. In contrast, Sultana's (2019) findings indicated a lack of awareness among teachers about the fairness and impact of standardized tests. The variation in findings could stem from the degree of assessment autonomy, as university teachers in Puspawati's study had greater freedom compared to their secondary-level peers in Sultana's study, who were subject to stricter, government-regulated assessment guidelines.

The Varied "Principles" in Different LAL Profiles

Within different LAL profiles, the literacy regarding "Principles" was complicated. The teachers in the low LAL profile exhibited an exceptionally low level of "Principles," despite the general high levels in the other two profiles. This might be explained by teachers' LAL developmental stages. The development of deeper, more theoretical, and principle-based insights necessitates a robust foundation of practical knowledge as a basis for LAL development (Fulcher, 2012). Moreover, the "Principles" scored the highest in the moderate LAL profile rather than in the high LAL profile, suggesting that teachers with a high LAL profile might not uniformly excel in all dimensions. Alternatively, teachers with a high LAL profile might underestimate their confidence in adhering to strict ethical standards or fairness as they gain deeper insights into assessment knowledge, skills, and principles.

Additionally, the statistically significant association of "Teaching Years" and "Test Research Participation Experience" with the three LAL profiles in the Chi-square test (see Table 3) does not imply an influencing impact on LAL levels. The role of these variables as potential factors influencing LAL profiles requires further exploration in regression analysis, which is the focus of the next section.

Impact of CSE Assessment Literacy on LAL Profiles

Relationship between CSE Assessment Literacy and LAL Profiles

The Chi-square test revealed significant associations between the three LAL profiles and CSE assessment literacy, focusing

on "CSE Background," "Competence Classification," "Scale Tables," and "Teaching & Research Values." Subsequent multinomial logistic regression analysis was applied to further identify the first three items of "CSE Background," "Competence Classification," and "Scale Tables" as factors positively influencing the LAL profiles, while "Teaching & Research Values" was found to be nonsignificant.

The significance of "Teaching & Research Values" varied between the Chi-square and multinomial regression analyses. A variable significant in a Chi-square test, showing an overall association, may not retain its significance in multinomial regression, where the emphasis is on distinguishing between specific groups defined by the dependent variable (e.g., Bayaga, 2010; Sun & Zhang, 2022). Its borderline P value of 0.051 may suggest that while "Teaching & Research Values" is important, the "CSE Background," "Competence Classification," and "Table Scales" may hold greater significance in affecting the development of the LAL profiles in the more detailed, group-focused contexts.

The Promoting Role of CSE Assessment Literacy on LAL Profiles

Teachers with extensive CSE background knowledge from both low and moderate LAL profiles could progress to the high LAL profile. However, only those in the moderate LAL profile with proficiency in CSE competence classification and table scales were able to advance to the high LAL profile. Thus, CSE background seems to more broadly facilitate LAL advancement than the other two factors.

CSE Background

Understanding CSE background could promote teachers' LAL development. This can be explained by the significance of contextual factors in assessment practices (e.g., Crusan et al., 2016; Fulcher, 2012). Comprehending the CSE background enables EFL teachers to familiarize themselves with China's reform in the foreign language landscape, aiming to bridge gaps in goals, standards, and assessments, and to synchronize with global benchmarks (Liu, 2015). Prompted by the broader educational context, EFL teachers could gain a better understanding of the significance and urgency of the CSE framework, thus motivating themselves to align their daily assessment practices with the CSE standards and improve their LAL through practicing and engaging in self-reflection.

This impact of contextual factors on teachers' LAL development through practices has been confirmed in empirical studies (e.g., Yan et al., 2018). Within contexts that promoted an alignment between teaching and testing (e.g., in the CSE context), the simple task of writing test items could make teachers attentive to the validity of assessment content, ensuring that student assessment is aligned with the standardized curriculum. This assessment practice of writing standardized test items also generates opportunities to encourage teachers to improve their test item writing skills, thereby deepening their insight into the core assessment knowledge, skills, and principles.

Competence Classification and Table Scales

Contrary to the influence of CSE background, a profound mastery of CSE competence classification and table scales would effectively aid teachers in progressing from the moderate to high LAL profile. This advancement requires a foundational moderate LAL level. This could be explained by their theoretical association with LAL knowledge. The development of CSE competence classification and table scales drew upon three major theories: Bloom's Revised Taxonomy, the Communicative Language Ability Model, and the functional linguistic model (Pan & Xiao, 2022). A deep understanding of the two CSE factors demands a familiarization of these applied linguistic theories, integral to assessment knowledge in LAL models (Davies, 2008). Thus, in this process, teachers' LAL could be advanced.

Furthermore, when EFL teachers apply the CSE competence classification and table scales in classroom assessment practices or standardized testing, their LAL can be developed inductively by hands-on experience and reflections on those experiences, reflecting a constructivist view of LAL development (e.g., Inbar-Lourie, 2008). Sang's (2023) research on English teachers' perceptions of the CSE in Chinese universities corroborates this impact. Participants noted that the CSE table scales could clearly define student behaviors, furnishing a direct, quantifiable method for appraising learning outcomes and teaching efficacy, thereby offering a viable framework to gauge both teaching success and student progression.

"Teaching Years" and "Test Research Participation Experience" as Covariables to Affect LAL Profiles

The covariates "Teaching Years" and "Test Research Participation Experience" were not significant in differentiating low from high LAL profiles but were significant between moderate and high profiles, suggesting that the impact of these covariates on LAL development is contingent on reaching a certain threshold of LAL proficiency. This discovery provides fresh perspectives on resolving debates about the effectiveness of certain LAL-influencing factors, such as years of teaching (e.g., Sun & Zhang, 2022).

Conclusion

The study found that Chinese university EFL teachers' LAL was mostly unsatisfactory, distributed across high, moderate, and low profiles. Awareness of the principles was highest in the moderate LAL profile and lowest in the low profile, outperforming the other LAL components. Mastery of "CSE Background" enhanced LAL across low and moderate profiles, while "Competence Classification" and "Table Scales" benefited only the moderate profile. Therefore, proficiency in CSE assessment literacy in terms of background, competence classification, and table scales promoted the LAL profiles of Chinese university EFL teachers.

This study has several limitations that suggest directions for future research. First, the reliance on self-report scales may limit the generalizability of findings, recommending the use of interpretative methods like teacher interviews and classroom observations to deepen the data (Giraldo, 2020). Second, while "Teaching Years" and "Test Research Participation Experience" did not distinguish between low and high LAL profiles, they were significant in differentiating moderate from high profiles. This highlights their impact at higher LAL proficiency levels and underscores the need for longitudinal studies to examine LAL development across career stages.

This study carries theoretical and practical implications for researchers, policymakers, and EFL teachers. Theoretically, this study investigates the integration of contextual and experiential factors in LAL development using China's CSE as a focal point. Employing *K*-means clustering analysis, it identifies diverse LAL profiles among Chinese EFL university teachers, responding to Yan et al.'s (2018) call to examine LAL developmental trajectories. This study is the first large-scale empirical research on the role of CSE assessment literacy in enhancing LAL profiles, thus enriching the existing literature on CSE and LAL. It also clarifies the impact of factors such as teaching experience on LAL, suggesting the existence of a threshold at which these factors become significant.

Practically, the findings will inform LAL enhancement strategies and international policymaking. Educational leaders should provide diverse LAL development programs tailored to teachers' LAL profiles. Teachers can thoroughly understand assessment contexts, effectively integrate theory with practice, and actively engage in self-reflection (Yan, et al., 2018). Continuous training will help them stay abreast of educational trends and support their professional development. Additionally, the standardization of China's CSE aligns local educational practices with global standards, helping policymakers balance global uniformity with local specificities. This is crucial for the diverse and multifaceted educational systems in the Asia–Pacific, offering a roadmap for integrating international standards with local educational cultures and languages.

References

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice, 9*(4), 30–32. https://doi.org/10.1111/j.1745-3992.1990.tb00391.x

Bayaga, A. (2010). Multinomial logistic regression: Usage and application in risk analysis. *Journal of Applied Quantitative Methods, 5*(2), 288–297.

CNKI (2024) *Trend of articles counted by years* [Figure]. CNKI. https://kns.cnki.net/kvisual8/article/center?language=CHS&uniplatform=NZKPT

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

Fan, J., & Jin, Y. (2020). Standards for language assessment: Demystifying university-level english placement testing in China. *Asia*

Pacific Journal of Education, 40(3), 386–400. https://doi.org/10.1080/02188791.2019.1706445

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

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

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

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. (2018). Defining and developing the assessment literacy of foreign language teachers. *Foreign Language Education in China, 1*(2), 65–93.

Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation, 61*, 112–122. https://doi.org/10.1016/j.stueduc.2019.03.003

Liang, J., Bai, L., Dang, C., & Cao, F. (2012). The K-means-type algorithms versus imbalanced data distributions. *IEEE Transactions on Fuzzy Systems, 20*(4), 728–745. https://doi.org/10.1109/TFUZZ.2011.2182354

Lin, H. (2016). Establish a national foreign language assessment system to improve national language proficiency. *China Examinations, 12*, 3–4. https://doi.org/10.19360/j.cnki.11-3303/g4.2016.12.001

Liu, J. (2015). Some thoughts on developing China common framework for english language proficiency. *China Examinations, 1*, 7–11. https://doi.org/10.19360/j.cnki.11-3303/g4.2015.01.002

Pan, M. (2020). Revisiting language assessment literacy: A perspective based on China's standards of english language ability. *Journal Beijing International Studies University, 42*(5), 46–56. https://doi.org/10.19360/j.cnki.11-3303/g4.2020.07.005

Pan, M., & Xiao, Y. (2022). *Research into EFL teachers' language assessment literacy*. Shanghai Jiao Tong University Press.

Peng, C., Liu, J., & Cai, H. (2021). Aligning China's standards of english language ability with the common European framework of reference for languages. *The Asia-Pacific Education Researcher, 31*(6), 667–677. https://doi.org/10.1007/s40299-021-00617-2

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

Puspawati I (2019) Understanding teachers' knowledge, skills, and principles on language assessment: A survey on teachers' language assessment literacy. In *Humanity, Education and Social Sciences* Third International Conference on Sustainable Innovation (pp. 70–75). Atlantis Press. https://doi.org/10.2991/icosihess-19.2019.11

Sang, Y. (2023). China's standards of English language ability: Voice from english teachers at Chinese universities. *SAGE Open, 13*(4), 1–12. https://doi.org/10.1177/21582440231205434

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan, 72*(7), 534–539.

Sultana, N. (2019). Language assessment literacy: An uncharted area for the english language teachers in Bangladesh. *Language Testing in Asia, 9*(1), 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*, 101157. https://doi.org/10.1016/j.stueduc.2022.101157

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

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