Vietnamese university EFL teachers' and students' beliefs and teachers' practices regarding classroom assessment

Hung Phu Bui

UEH University

Language Testing in Asia

RESEARCH Open Access

Vietnamese university EFL teachers' and students' beliefs and teachers' practices regarding classroom assessment

Hung Phu Bui*

*Correspondence:
hungbp@ueh.edu.vn

University of Economics Ho Chi Minh City (UEH), Ho Chi Minh City, Vietnam

Abstract

This study investigates Vietnamese university EFL teachers' and students' beliefs and teachers' practices regarding classroom assessment. Data were collected from a questionnaire administered to 455 participants (104 teachers and 351 students), interviews with nine teachers and twelve students, and 18 observations of 90-min lessons in five universities in Vietnam. In general, the teachers were aware of all assessment types in the current literature. They dominantly used interactive scaffolding and performance assessment. They also denoted some contextual factors as barriers to their assessment practices, leading to their seeming negligence of other classroom assessment types in their practices. The students preferred formative assessment and some assessment tasks (e.g., group work), while the teachers believed that summative assessment could engage students in learning and associated performance assessment with a summative assessment. Implications for L2 classroom assessment practices are discussed.

Keywords: Beliefs, Classroom assessment, Contextual factors, Practices, Teacher cognition

Introduction

Second language education research has been experiencing a growing interest in classroom assessment. Regarding the teacher's role in assessing students, classroom assessment or classroom-based assessment can be employed to give grades and certify students' academic records (summative) and improve teaching and learning (formative) (Shepard, 2019). While summative assessment is mainly used to measure students' achievements, formative assessment, known as the teacher's "use of artifacts and activities to understand learners to help them learn" (Gan et al., 2018), is well documented as to enhance the learning outcomes (Hao & Johnson, 2013; Shepard, 2019), motivate students to learn (Bui & Nguyen, 2022; Gan et al., 2018), and improve the teacher's performance (Stiggins, 2005). In the classroom, these two types of assessment can be integrated coherently at different stages of teaching to enhance students' achievements and "minimize the negative effects of grading on learning" (Shepard, 2019, p. 183).

From the researcher's knowledge, most, if not all, classroom assessment models have been developed in Western cultures. Simply adopting classroom assessment models

constructed in Western cultures and applying them in Asia, including Vietnam, are an example of overgeneralization as people in these different contexts may have different learning cultures. For instance, Thanh and Gillies (2010) found that Asian students hesitate to get involved in public discussions because having no answer or giving an incorrect answer can threaten face. Also, the specific characteristics of L2 (second for foreign language) assessment have motivated applied linguists to localize the classroom assessment framework. Gan et al. (2018) found that not all types of L2 classroom assessment are liked by Chinese students and motivate them to learn.

Research into teachers' and students' beliefs may be used to advance the classroom practices and justify classroom assessment framework. However, there exists in the current literature a dearth of research on Asian University EFL (English as a foreign language) teachers' beliefs and practices and students' beliefs and reactions to teachers' classroom assessment. This study is, therefore, timely and aims to shed light on how the L2 classroom assessment framework should be adapted concerning the contextual factors of Vietnam. It explores Vietnamese university EFL teachers' and students' beliefs and teachers' practices regarding classroom assessment.

Literature review

The renewed interest in L2 classroom assessment

The L2 classroom assessment construct has been recently renewed. The renewed interest, supported by jurisdictions and research in countries (Leung et al., 2018), conceptualizes classroom assessment as teachers' use of artifacts and activities to diagnose and know about learners, facilitate learning, and promote the learning outcomes (Hao & Johnson, 2013; Shepard, 2019). The underlying mechanism is that teachers should not conduct assessment of learning but assessment for learning. This paradigmatic movement is congruent with the sociocultural theory of learning that classroom assessment is viewed as an integral component of the teaching and learning process (Pourdana, 2022). As McMillan (2013) puts it, classroom assessment demonstrates a crucial role of teachers as it directly influences learners and learning in a cyclical process in which teachers actively, flexibly, and continuously collect and interpret evidence about students to decide what they should do and how they should do it to facilitate the learning process and maximize the learning outcomes. Other researchers (e.g., Shepard, 2019; Stiggins, 2005) propose that classroom assessment techniques employed by teachers may engage students in learning and affect students' behaviors and learning strategies, which, in turn, assists students in achieving the expected goals and objectives and becoming competent, responsible, and independent. In general, classroom assessment is contemporarily conceptualized as formative assessment (Davison & Leung, 2009). Yan et al. (2021) also confirm this renewed construct of classroom assessment, saying that classroom assessment is "used interchangeably with formative assessment, assessment for learning, and more recently learning-oriented assessment"; therefore, classroom assessment, including the summative component, "should be used formatively to facilitate student learning" (p. 542).

The recent reconceptualization of classroom assessment foregrounds and specifies what teachers should do in the classroom. For instance, Swaffield (2011) posits

that classroom assessment can be either formal or informal in which teachers can use oral feedback in an interactive classroom. Edwards (2014) postulates that the benefits of classroom assessment can be maximized when it is dialogic. Consequently, students' capacity can be fostered when they have opportunities to generate self-regulated learning and learning motivation. Also, students may learn more from actively participating in the assessment process (Berggren, 2019). The benefits of working as assessors and being assessed may outweigh those of solely having works assessed by others (Fernández, 2020). In learner-centered pedagogy, diagnostic assessment can be used by teachers to understand learners and learning. L2 teachers' attention to individual learners can provide information for teachers to support struggling learners (Alderson et al., 2015), for which L2 teachers can "scaffold the L2 learners when they are working on a complex task and mediate them individually to accomplish the task until their learning needs are fulfilled" (Rafi et al., 2022, p. 2).

Context and L2 classroom assessment

Although the L2 classroom assessment literature has been shifted to be formative, classroom assessment in some contexts remains traditional. The study by Guo and Xu (2020) found that L2 teachers did not regard the learning objectives and rarely used peer assessment, self-assessment, and interactive assessment. According to Cheng and Sun (2015), this assessment practice does not provide much information about learners and cannot be used to improve teaching. Chen and de La Torre (2014) argue that this practice neglects students' active role in language learning. Zhou and Deneen (2016) also identified that L2 teachers in Asia mainly employ grades and scores to inform students of their capacity. These researchers criticized this practice for its lack of detailed feedback. It is teachers' specific feedback that helps students know what they should do in their learning process to achieve goals and objectives (Xu & Liu, 2009).

Classroom assessment in Asia is characterized by sociocultural variables, such as interpersonal communication and learning culture (Bui & Nguyen, 2022; Thanh & Gillies, 2010). Divaharan and Atputhasamy (2002) suggest that classroom assessment theory can be adjusted according to contextual factors. For example, Asian students may refuse to assess others openly. Carson and Nelson (1996) also found that interactive assessment in China reflect Chinese culture to a certain extent. For instance, students prefer self-assessment to peer assessment as they regard unsatisfactory comments as unfriendly comments, indicating its face-losing culture. According to Thanh and Gillies (2010), Asian students may hesitate to initiate class discussion because they are afraid that their opinions are incorrect or challenged in public, resulting in losing face. Thanh and Gillies (2010) also found that Vietnamese students interact more when they find that the environment is safe enough. Therefore, "Face-to-face discussion needs to be designed in a way that suits the learning culture of Vietnamese students." (p. 82).

Beliefs and practices regarding L2 classroom assessment

Beliefs, which mainly refers to individual cognition, describe what people believe to be true (Borg, 2017). Discrepancies between teachers' and students' beliefs may adversely affect students' behaviors (Bell, 2016), satisfaction, and learning outcomes and instructional practices (Bui, 2022; Ellis, 2008). Therefore, it is important

to investigate matches and mismatches between teachers' and students' beliefs in L2 education.

Recent studies are inconclusive about the relationship between teachers' beliefs and practices regarding classroom assessment. The studies by Zhou and Deneen (2016) and Wu et al. (2021) concluded that teachers' beliefs directly impact their instruction and lesson focus in EFL contexts. According to Yan et al. (2022), teachers' beliefs can be predictors of their practices of L2 classroom assessment. Nevertheless, the studies by Vattoy (2020) and Wang et al. (2020) indicated that L2 teachers' beliefs may not directly influence their practices of classroom assessment. The (mis)matches between teachers' beliefs and practices can be ascribed to sociocultural factors (e.g., students and policy), teachers' assessment literacy and workload (Narathakoon et al., 2020), and experience Vattoy (2020).

Differences between teachers' and students' beliefs are also well-documented in the literature. Investigations into (in)congruences between teachers' and students' beliefs may provide implications for language teaching and learning (Bell, 2016; Ellis, 2008). For instance, Ha et al. (2021) found some (mis)alignment between teachers' and students' perceptions of how teachers should give feedback and correct errors in Vietnamese secondary schools. Bui & Nguyen, (2022) study also showed some in(congruences) between teachers' and students' beliefs about some language teaching aspects, including error correction. In general, these studies concluded that it is important to explore teachers' and students' beliefs. It might be better for teachers to explicitly address any discrepancy in beliefs.

Although it is well-documented in the literature about the significance of investigations comparing teachers' and students' beliefs and teachers' practices regarding L2 classroom assessment, it is underexplored. Also, teaching English in Asia shows a growing demand, but little is known about English classroom assessment in this context. This study addresses the aforesaid gap proposing the research questions below:

> RQ1. What are Vietnamese university EFL teachers' practices regarding classroom assessment types and tasks? How do Vietnamese university EFL students react to their teachers' practices of classroom assessment?
> RQ2. Are there any incongruences between Vietnamese EFL teachers' and students' beliefs about classroom assessment?

Research methods

Research design

This study employed a mixed-methods design. To answer RQ1, the researcher first observed the classroom to examine the classroom assessment practices in universities in Vietnam. Right after observing each class, he administered a questionnaire in the absence of the teachers. Finally, teachers and students were invited to participate in in-depth interviews. Data collected from the questionnaire and in interviews were used to answer RQ2.

Participants and settings

The participants ($n = 467$) were randomly selected from five different universities in Vietnam (see Table 1). However, as answers of twelve participants (four teachers and eight students) were found invalid, the sample size reduced to 455. All 104 teachers (38 males and 66 females) had an experience of 3 to 15 years as EFL teachers. All of them had a master's degree or a doctorate degree in English Language Teaching or Applied Linguistics. All 351 students (149 males and 202 females) were sophomores and juniors. All of them learned English in the secondary level and in the first 2 years as required by their university curricula. They were willing to participate in this study, consented and were informed of the ethical considerations and their rights as participants. Their identity was kept confidential in the report.

The five universities where this study was conducted had a similar assessment system and principles. They set aside 40–50% of the total score for classroom assessment and the final test made up for the rest. This means that classroom assessment was considered to be both formative and summative. Teachers held an absolute control of this part; they could decide on test forms and content. The popular choices for the summative part were attendance, midterm tests, quizzes, and presentations. These universities did not have any test banks and suggested resources for teachers to make tests. A class size was from 40 to 50 students.

Instruments

This study employed a set of three instruments. The observation form was adapted from Guilloteaux and Dornyei (2008), which was designed to collect quantitative data. The researchers changed the observation form from a 5-point Likert-scale to categories to make notes on teachers' assessment and students' reactions during classroom observations.

The questionnaire was adapted from Bui & Nguyen, (2022), Gan et al. (2018), and Yan et al. (2021). The pool of items was subject to careful examination. The researcher first categorized and removed overlapping items. The remaining 28 items, on a 5-point Likert-scale (1 = strongly disagree, 2 = disagree, 3 = undecided, 4 = agree, and 5 = strongly agree) were then tested with 209 participants ($n = 209$). These participants did not participate in the main study. The internal coefficient reliabilities (Cronbach's alpha) for the whole questionnaire and each factor were greater than .7. The researcher then randomized the items and added three fillers. The questionnaire used in the main study consisted of five main categories and 3 fillers (see Table 2).

The interview scheme was semi-structured. As the current study employed a sequential mixed-methods research design, the interview scheme and questionnaire had the

Table 1 Participant description

Participant group ($n = 467$)	Number	Male	Female
Teacher	104	38	66
Student	351	149	202

Table 2 Questionnaire items

Category	Item
Diagnostic assessment	1, 9, 15, 19, 25
Self-assessment	3, 8, 14, 20, 27
Interactive assessment	6, 10, 12, 18, 23, 29
Teacher scaffolding	5, 11, 16, 22, 26, 30
Performance assessment	2, 7, 13, 17, 24, 31
Fillers	4, 21, 28

same structure. The interview scheme contained items like "How" and "Why" to make insights into the participants' beliefs about classroom assessment.

Data collection

Data were collected from January to May of 2022 and consisted of three main phases. First, the researcher sent an email invitation to teachers at five universities based in Southern Vietnam. One hundred four teachers were willing to answer the questionnaire, and nine of them agreed to let the researcher observe their classes and survey their students' beliefs. The researcher did not provide the teachers with the specific aims of the study but informed them that the observations were to view the classroom performances. In classroom observations, the researcher sat at the back of the class and kept silent to avoid disturbing the class performances, making notes on the teachers' assessment practices. Classroom observations made up for a total of 1.620 min (180 min x 9 teachers). Each teacher was observed for two 90-min lessons because several observations of one teacher can minimize the potential impacts of the lesson characteristics on classroom assessment practices and provide rich information about classroom assessment practices. After each classroom observation, the questionnaire was administered to the students in the absence of the teacher. Although the questionnaire was written in Vietnamese, the researcher explained very single item to the teachers and students. The students were informed of the importance of their responses to the success of the study and their rights to refuse to answer the questionnaire. After the researcher screened the students' responses to the questionnaire, he randomly stratified the student sample and sent an email invitation to 90 students and the nine teachers for interviews. As all the teachers consented at the start of the study, they all responded and showed up for interviews. Nineteen out of the 90 selected students responded and showed up for interviews. The interviews took place within 7 days after each class observation and were conducted in the participants' native language (Vietnamese). The researcher started the interviews with lead-in questions and informed the ethics to which the study was committed. He used the prompt semi-structured interview strategy and asked guiding questions such as "Why do you think so?" and "How can classroom assessment be done?" to explore the teachers' and students' beliefs about L2 classroom assessment. The researcher confirmed and clarified teachers' and students' responses concerning the trustworthiness of their self-reports. The researcher redirected the participants and explained the main questions when he found that their responses were unfocused. Each interview

lasted for 25–35 min and was audio-recorded for analysis. For identity confidentiality, the teachers and students were anonymized as teacher 1–teacher 9 and student 1–student 19 in the data report.

Data analysis

The quantitative and qualitative data collected were analyzed through different processes. The researcher first cleaned the collected quantitative data. As a result, data from twelve participants (four teachers and eight students) were identified invalid (multiple choices for an item and unselected choices); therefore, they were excluded from data analysis. After cleaning data, the researcher projected the remaining data ($n = 455$) to SPSS 28.0 (IBM Corporation, Armonk, NY). First, to gain insights into the factor structure of the questionnaire, the resercher subjected all data ($n = 455$) to confirmatory factor analysis (CFA) by using principal axis factoring, Promax rotation, Kaiser's eigenvalues-greater-than-one option, scree plot, and suppression of small coefficients lower than .5. The pattern matrix showed that the loadings were greater .6, and 55.136% of the total variance were explained. The pattern matrix was then subjected to structural equation modeling through IBM Amos 28. As a result, the three fillers and two included items (SA1 and IA2) were removed because their loadings were smaller than 4. All the remaining items had loadings greater than 6. Kaiser-Meyer-Olkin measure of sampling adequacy was .862 (df = 325), and the significance level satisfied the standard ($p < .001$). Afterward, the scale of reliability (Cronbach's alpha) of each factor was examined ($\alpha > .7$). Subsequently, independent-samples *t*-test was used to compare EFL students' and teachers' beliefs about classroom assessment. Mean scores (M), standard deviation (SD), and significance level (p) were examined to figure out if the discrepancies between EFL students' and teachers' perceptions of observed factors and variables were significant.

The qualitative data collected from the classroom observations and interviews were analyzed thematically. The data analysis procedure was content-based and inductive (Nguyen & Hung, 2021; Creswell & Creswell, 2018). The researcher focused on the teachers' assessment behaviors and students' reactions documented from the classroom observations. Regarding the objectives of the study, he first categorized the assessment events into types: diagnostic assessment, peer assessment, self-assessment, interactive assessment, teacher scaffolding, and performance assessment. Subsequentially, the assessment events were scrutinized to identify themes which were then reread and refined on a cyclical basis. When analyzing the interview data, the researcher did not consider the respondents' language problems but their beliefs about classroom assessment.

Results from analyzing the data collected from the observations were used to answer research question 1. To answer research question 2, quantitative (questionnaire) and qualitative (interviews) results were incorporated. As this study employed the explanatory sequential mixed-methods design to answer research question 2, the qualitative data were used to explain and supplement the quantitative data.

Results

Initial statistical analysis

The statistics from initial data analysis confirmed the relationship between the 26 items and five components of the questionnaire (Table 3). The statistics showed that the five-factor structure, confirmed by CFA, had a good model fit with CMIN = 339.006, CMIN/DF = 1.173, p = .001, GFI = .914, CFI = .990, TLI = .989, PCFI = .881, RMSEA = .032, and PCLOSE =.978. The internal reliability coefficient (Cronbach's alpha) for the entire questionnaire was .766. The Cronbach's alpha coefficients (α) were .886 for diagnostic assessment, .759 for self-assessment, .792 for interactive assessment, .876 for teacher scaffolding, and .919 for performance assessment.

Table 3 Initial statistical results

	Observed variable	Item corre.	α if item deleted	Mean	SD	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	Diagnostic assessment *(n = 455, α = .886, M = 4.63, SD = .46)*									
1	**DA1**	.689	.823	4.6747	.53892	.739				
9	**DA2**	.710	.824	4.6352	.56599	.762				
15	**DA3**	.754	.813	4.6088	.57166	.816				
19	**DA4**	.740	.825	4.6813	.52428	.799				
25	**DA5**	.735	.808	4.5670	.57424	.792				
	Self-assessment *(n = 455, α = .759, M = 4.62, SD = .45)*									
8	**SA2**	.602	.677	4.6022	.61729		.741			
14	**SA3**	.533	.717	4.6220	.61726		.621			
20	**SA4**	.515	.725	4.6549	.58774		.605			
27	**SA5**	.582	.691	4.6308	.55920		.697			
	Interactive assessment *(n = 455, α = .792, M = 4.56, SD = .45)*									
6	**IA1**	.609	.742	4.5736	.58859			.716		
12	**IA3**	.590	.747	4.5275	.61085			.679		
18	**IA4**	.546	.761	4.5516	.61647			.622		
23	**IA5**	.537	.765	4.5648	.62587			.618		
29	**IA6**	.578	.751	4.5824	.59487			.663		
	Teacher scaffolding *(n = 455, α = .876, M = 4.43, SD = .56)*									
5	**TS1**	.707	.850	4.4066	.72065				.773	
11	**TS2**	.674	.855	4.3978	.73460				.742	
16	**TS3**	.673	.855	4.4198	.72832				.721	
22	**TS4**	.644	.860	4.4813	.71113				.686	
26	**TS5**	.686	.853	4.4505	.70654				.741	
30	**TS6**	.691	.852	4.4505	.71583				.749	
	Performance assessment *(n = 455, α = .919, M = 4.48, SD = .54)*									
2	**PA1**	.752	.907	4.4659	.65910					.793
7	**PA2**	.766	.905	4.4484	.63755					.806
13	**PA3**	.735	.909	4.4681	.65249					.759
17	**PA4**	.776	.903	4.4703	.66596					.813
24	**PA5**	.779	.903	4.5209	.62183					.829
31	**PA6**	.810	.899	4.5077	.62214					.856

Item corre, corrected item-total correlation

Teachers' practices and students' reactions regarding classroom assessment

Analysis of data collected from classroom observations showed how Vietnamese EFL teachers' practiced types of classroom assessment. Most of the teachers used various classroom assessment techniques, among which they dominantly employed interactional scaffolding and performance assessment. Diagnostic assessment was used least often by the teachers. All of them started a lesson with warm-up activities and usually asked, "Do you understand?" to know if the students could make sense of the instruction or keep up with the process. In only one session, teacher 3 asked students about their prior experiences; nevertheless, her attempt was mainly inclined to activate students' existing knowledge for the target lesson rather than diagnosing students' problems.

All the teachers assessed students' performance using exercises, quizzes, and mini-tests. Teachers 1 and 6 made comments on the students' performance, but their comments were mainly restricted to letting students identify the right and wrong answers and appraising students' performance. That is, their feedback was general and did not direct students to what they should do for improvement. The teachers used scores and grades to indicate the teachers' comments instead of giving oral corrective feedback.

The teachers employed self-assessment and peer assessment marginally. Self-assessment and peer assessment accounted for less than 5 min in each session. After group work, the students were given 2 to 3 min to reconsider their discussion to present their opinions to the rest of the class. Sometimes, the students also had about 2 to 3 min to re-examine their answers in written exercises and mini-tests before they submitted their works.

The students showed their preference for interacting in small groups. Student-initiated interaction with the teachers was not detected during the teacher-fronted instruction time. They responded to the teachers when they were called on. However, in pair work and group work, they interacted quite actively. Some also started conversing with the teachers by asking about their queries and for hints individually when the teachers circulated the classroom during the students' group work. In other words, they preferred speaking in small groups to speaking to the whole class. Female students generally hesitated to talk without the microphone. Male students, by contrast, answered the teacher's questions immediately and seemingly more confidently than female students.

Most male students were more reserved than female students in interactive assessment and teacher scaffolding. Their utterances mainly contained the core information. They also demonstrated more engagement in individual work. They searched for information, looked up words in e-dictionaries, and completed tasks individually. However, many female students extended their answers by giving examples and elaborating their ideas further. They also dominated interaction in group work. They started group brainstorming, asking group members about their opinions, giving personal views, and summarizing the key points in their discussion.

Teachers' and students' beliefs about classroom assessment

Independent-samples *T*-test results indicated comparisons of the students' and teachers' beliefs about different aspects of classroom assessment (see Table 4). In general, the incongruence in the students' and teachers' beliefs was significant ($p < .05$) with a medium effect size (Cohen's $d = .391$). The teachers generally agreed more strongly upon

Table 4 Students' and teachers' beliefs about classroom assessment

Type	Equal variances	Student (n = 351)		Teacher (n = 104)		t	Mean diff.	Sig (2-tailed)	Cohen's d (f^2)
		Mean	SD	Mean	SD		S/T		
DA	Not assumed	4.62	.43	4.68	.59	− .791	− .058	.431	.116
SA	Assumed	4.62	.46	4.64	.42	− .308	− .018	.758	.045
IA	Assumed	4.57	.44	4.52	.48	.868	.051	.386	.108
TS	Not assumed	4.41	.58	4.56	.48	− 2.250	− .145*	.026	.281
PA	Assumed	4.44	.53	4.68	.55	− 3.445	− .240**	.001	.444
Total	Assumed	4.53	.23	4.62	.23	− 2.754	− .082*	.006	.391

*$p \leq .05$; **$p \leq .001$, *S/T* students over teachers

the concepts in classroom assessment ($M = 4.53$, SD = .23) than the students ($M = 4.62$, SD = .23).

There were found significant discrepancies between students' and teachers' beliefs. The difference between students' and teachers' beliefs about performance assessment was significant ($p = .001$), with a medium effect size (Cohen's $d = .444$). This means that the teachers generally rated this category quite moderately higher than the students. The incongruence between the two groups' beliefs about teacher scaffolding had a significance level of .026. However, the effect size was relatively small (Cohen's $d = .281$), suggesting that this difference should not be considered significant (Bonett, 2009; Cohen, 1988). Also, the differences between the students' and teachers' beliefs about diagnostic assessment, self-assessment, and interactive assessment were insignificant ($p >$. 05), and their effect sizes were small (Cohen's $d <$. 2). Thus, these incongruences should not be considered statistically significant.

Data collected from interviews generally confirmed the students' and teachers' self-reports provided in the questionnaire and provided insightful information about their beliefs. First, the students expected their teachers to diagnose their competences (e.g., lexical resources and knowledge) prior to a task assignment. However, as noted by the teachers, the workload did not allow them to spend much time doing it. Instead, they attempted to diagnose their students in the first session in each semester to have an overview of the students' competences. The class size, as reported by the teachers, was another barrier to the use of diagnostic assessment.

Contextual issues were found to be a concern in interactive assessment. Many students did not acknowledge the importance of classroom interaction, especially teacher-student interaction. They believed that it was a choice, reflecting their learning styles. Some students hesitated to interact with their teachers and peers as they were afraid of losing face in case they were corrected. Female students showed a stronger sense of small group talks than public talks. As responded by many female students, speaking loud in public (without the microphone) would lose their elegance. They had been educated by their family to build and maintain their charming image in public. The teachers were generally aware of the importance of interaction in the classroom. They sometimes interacted with their students to understand them and scaffold their learning. However, they did not take the contextual issues that interfered with classroom assessment as revealed by the students. They showed their

awareness of such behaviors but supposed that the students refused to get involved in interactive tasks as they were not interested in the topic of discussion.

Performance assessment was found to be the biggest concern. Most students believed that tasks, exercises, quizzes, and short tests should be used as tools for teachers to understand students and find appropriate remedies to support the learning process rather than to give scores and grades. In other words, they preferred classroom assessment to be prone to formative rather than summative assessment. They responded that the use of summative assessment sometimes made the classroom tense. Also, most of them did not think that the scores and grades reflected their competences. That is, they did not trust the reliability and validity of the summative assessment in the classroom. All of them realized that different teachers administered tests at different levels of difficulty even though they taught at the same level.

Unlike the students, most teachers believed that using summative assessment in class was essential. Summative assessment was supposed to require students to engage in learning. From their experience, removing the summative component of classroom assessment might increase the absenteeism rate. Scores could also be used to encourage students to do assigned tasks. In other words, they believed that the evaluation of students' efforts should be taken into their total scores.

Overall, data collected from the questionnaire and interviews showed Vietnamese university EFL teachers' beliefs compared to students regarding types of classroom assessment. The teachers believed that L2 classroom assessment should be both summative and formative, which led to their dominant use of exercises, quizzes, and mini-tests in their performance assessment. They perceived that summative assessment engaged students more in learning. However, the students preferred the classroom to be totally formative. Data collected from classroom observations showed the teachers' practices of classroom assessment types and students' reactions to their teachers' assessment. Accordingly, the teachers and students showed their preferences for some assessment tasks.

Discussion

The current study compared teachers' and students' beliefs and explored teachers' practices and students' reactions regarding classroom assessment. The classroom observations provided insights into the classroom assessment practices at five universities in Vietnam. A questionnaire and interviews delved into students' and teachers' beliefs. Collected data provided interesting socio-psychological factors in L2 classroom assessment in Vietnam.

The results from classroom observations showed that the teachers employed interactive scaffolding and performances the most frequently. According to Van de Pol et al. (2015, 2019), scaffolding can be used to support students' uptake in L2 classrooms. However, it is essential to note that the more the teacher scaffolds students, the more the students feel contingently controlled. In other words, teachers' use of scaffolding should depend on students' current competence. This finding suggests that the teacher should diagnose students to understand them and their learning before they use scaffolding in L2 teaching. However, the results from classroom observations showed that the teachers used diagnostic assessment marginally. As noted by Alderson et al. (2015),

diagnostic assessment can engage students in learning, providing teachers with information about students' learning problems from which they can suggest remedies. According to Pourdana (2022, p. 8), it is "the responsibility of the L2 teachers to diagnose their students' learning problems by constantly observing them, encouraging their learning efforts, and identifying the obstacles to their future learning progress". The results of classroom assessment also indicated that the teachers neglected the role of self-assessment and peer assessment. Rafi et al. (2022) noted that, as motivated by sociocultural theory in L2 education, self-assessment and peer assessment both may mediate the learning process. As Babaii et al. (2016, p. 414) put it, L2 self-assessment can be used to "promote(s) learning, establish(es) a goal-oriented activity, alleviate(s) the assessment burden on teachers, and finally continue(s) as a long-lasting experience." As a goal-oriented approach, L2 self-assessment helps students monitor their own learning and enhance their responsibility for their learning in the long run (Winke et al., 2022). Also, peer assessment assists students in exchanging opinions (Wen et al., 2006). It benefits both students who serve as assessors and students who are assessed (Fernández, 2020).

Data collected from the questionnaire and interviews gave insights into students' and teachers' ratings. Quantitative data showed some significant discrepancies between the beliefs of students ($n = 351$) and teachers ($n = 104$) about teacher scaffolding and performance assessment. However, the effect size (Cohen's d) was found to be medium, which may have been caused by the limited sample of teachers recruited in this study (Bonett, 2009; Cohen, 1988). Data collected from interviews could explain such differences. While students associated performance assessment with summative assessment due to their experiences with teachers' practices in the classroom and indicated their preferences for formative assessment, the teachers believed performance assessment to be a tool of summative assessment that could save time by engaging students in one test. This finding confirmed the result from classroom observations that teachers used performance assessment as a summative assessment tool. It is crucial to address that performance assessment can be conducted formatively. According to Gan et al. (2018), performance assessment can provide evidence of a wide range of skills teachers and students may use to plan remedial actions. The results also showed that the students preferred interacting in small groups to speaking to the whole class. As noted by van de Pol et al. (2015, 2019), peers can interactively support each other in learning. They may feel more confident when they interact with peers than they interact with teachers (Thanh & Gillies, 2010). Also, the students were concerned about their teachers' grading fairness. Fairness is a complex and salient issue in a summative component of classroom assessment as it reflects teachers' beliefs and assessment literacy; therefore, teachers' grading bias can derive from "factors beyond the scope of the test" (Camilli, 2006, p. 225). In the summative component of classroom assessment, it has been provoking ongoing debates (Kane, 2010). As classroom assessment is context-sensitive, L2 classroom assessment practices should take into account sociocultural factors. However, it should be used to engage students in learning rather than force them to learn. Appropriate use of classroom assessment may increase students' satisfaction and motivate them to learn; otherwise, it can cause anxiety in the classroom (Bui & Nguyen, 2022; Gan et al., 2018). As Asian students are generally emotionally charged, they may withdraw from the classroom activities when they feel demotivated or lose face (Thanh & Gillies, 2010). To

this end, it might be essential for institutions to have a test bank and grading criteria to ensure reliability and validity (Russell & Airasian, 2012) if they want to include summative assessment in classroom assessment. Also, while male students preferred self-regulated learning (self-assessment), female students preferred interactive assessment (Hu & Cheung, 2021). This finding provides a reference for Vietnam and similar contexts. Teachers may consider this finding to justify the classroom assessment framework. When teachers identify any incongruences between teachers' and students' beliefs, they can elicit such differences to students to improve students' behaviors in the classroom and promote the learning outcomes (Nguyen & Hung, 2021; Borg, 2017).

The findings suggest implications for classroom assessment practices in Vietnam and similar contexts. Regarding the sociocultural issues raised by the students, it might be necessary to adapt classroom assessment theories to suit the learning culture of Vietnam. The success of classroom assessment partly depends on students' emotions. They need to feel safe to receive and provide feedback, resulting in improvement in their competences. Rapports between teachers and students and between peers may prevent students from feeling criticized and annoyed when they receive less positive feedback. It is the student's attitude towards feedback that can contribute to their development of knowledge and skills. In other words, the more they find feedback constructive, the more willing they are to receive and provide feedback. Thus, it is important for language teachers to consider sociocultural factors to modify classroom assessment literature to suit the context.

Conclusions

The current study provides insightful information about teachers' and students' beliefs and teachers' practices regarding L2 classroom assessment in an Asian context. The teachers generally had a strong sense that summative assessment could engage students in learning, while the students generally preferred formative assessment. It might be necessary for the teachers to update the current literature on classroom assessment as summative assessment should not be simply understood as a tool to give scores and grades and force students to attend the class. Instead, summative assessment can be used formatively; teachers can understand learners and learning through summative assessment and use the assessment results to assist students in learning.

One important finding is that the teachers predominantly used interactive scaffolding and performance assessment. Previous studies (e.g., van de Pol et al., 2015, 2019) indicated the importance of these two assessment types in supporting students' learning. However, it may be essential for teachers to diagnose students' competences to scaffold students appropriately (Pourdana, 2022). That means scaffolding is dynamic in nature and should be used flexibly.

Also, incongruences between teachers' and students' beliefs were found in some assessment types. As incongruences between teachers' and students' beliefs may derive from the practices of classroom assessment that they experience, it might be relevant that teachers understand students' beliefs to improve their satisfaction, behaviors, performance, attitudes, and learning outcomes (Nguyen & Hung, 2021; Borg, 2017). They may justify the classroom assessment framework by considering contextual factors.

This study contributes to the classroom assessment literature. As classroom assessment is context-bound, it might be necessary for teachers to consider contextual factors to justify the classroom assessment framework. Summative assessment, if included in classroom assessment, can be performed formatively. That means teachers can use students' academic achievements to give feedback on what students should do and how they should do it to facilitate the learning process. In terms of the reliability and validity of summative assessment, institutions may need to have a test bank and educate teachers in assessment literacy.

This study showed two main limitations. First, driven by the research questions, this study mainly investigated teachers' and students' beliefs and practices concerning L2 classroom assessment. Further studies can explore the effects of classroom assessment types and tasks on students' academic achievements. Second, because the current study was confined to the context of Vietnam, the findings mainly reported beliefs and practices of classroom assessment at the tertiary level. Future research can investigate how classroom assessment is conducted at other education levels and in other contexts.

Acknowledgements
I would like to send sincere thanks to University of Economics Ho Chi Minh City for funding this study. I also send great thanks to the participants. Without them, there would have been no chance for this study to complete.

Author's contributions
This is a single-authored research article. There are no other contributing authors. The author(s) read and approved the final manuscript.

Funding
This study was funded by University of Economics Ho Chi Minh City, Vietnam. Number: UEH2020_16.

Availability of data and materials
All data generated or analyzed during this study are included in this published article, and raw data will be available on request.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 3 January 2023 Accepted: 22 January 2023
Published online: 13 February 2023

References

Alderson, J. C., Haapakangas, E.-L., Huhta, A., Nieminen, L., & Ullakonoja, R. (2015). *The diagnosis of reading in a second or foreign language*. Routledge.

Bui, H. P., & Nguyen, T. T. T. (2022). Classroom assessment and learning motivation: insights from secondary school EFL classrooms. International Review of Applied Linguistics in Language Teaching. https://doi.org/10.1515/iral-2022-0020. (In Press)

Bui, H. P. (2022). Students' and teachers' perceptions of effective ESP teaching. Heliyon, 8(9), e10628. https://doi.org/10.1016/j.heliyon.2022.e10628

Babaii, E., Taghaddomi, S., & Pashmforoosh, R. (2016). Speaking self-assessment: mismatches between learners' and teachers'criteria. *Language Testing*, *33*(3), 411–437. https://doi.org/10.1177/0265532215590847.

Bell, D. (2016). Practitioners, pedagogies, professionalism in English for academic purposes: the development of contested field. Unpublished PhD Dissertation, The University of Nottingham, UK.

Berggren, J. (2019). *Writing, reviewing and revising: Peer feedback in lower secondary EFL classrooms*. Stockholm University.

Bonett, D. G. (2009). Estimating standardized linear contrasts of means with desired precision. *Psychological Methods*, *14*(1), 1–5. https://doi.org/10.1037/a0014270 PMID 19271844.

Borg, S. (2017). Teachers' beliefs and classroom practices. In P. Garrett, & J. M. Cots (Eds.), *The Routledge handbook of language awareness*, (pp. 93–109). Routledge.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement*, (4th ed., pp. 221–256). Praeger Publishers.

Carson, J., & Nelson, G. (1996). Chinese students' perceptions of ESL peer response group interaction. *Journal of Second Language Writing*, *5*(1), 1–19. https://doi.org/10.1016/S1060-3743(96)90012-0.

Chen, J., & de La Torre, J. (2014). A Procedure for diagnostically modeling extant large-scale assessment data: the case of the programme for international student assessment in reading. *Psychology*, *5*(18), 1967–1978. https://doi.org/10.4236/psych.2014.518200.
Cheng, L., & Sun, Y. (2015). Teachers' grading decision making: multiple influencing factors and methods. *Language Assessment Quarterly*, *12*(2), 213–233. https://doi.org/10.1080/15434303.2015.1010726.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.
Creswell, J. W., & Creswell, J. D. (2018). *Research design: qualitative, quantitative, and mixed methods approaches*. SAGE Publishing.
Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, *43*(3), 393–415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x.
Divaharan, S., & Atputhasamy, L. (2002). An attempt to enhance the quality of cooperative learning through peer assessment. *Journal of Educational Enquiry*, *3*(2), 72–79.
Edwards, J. G. H. (2014). Peer assessment in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment*, (pp. 730–750). Wiley-Blackwell. https://doi.org/10.1002/9781118411360.wbcla002.
Ellis, R. (2008). Explicit knowledge and second language learning and pedagogy. In N. Hornberger (Ed.), *Encyclopedia of language education, 6*, (pp. 1901–1911). Springer.
Fernández, D. A. (2020). Collaborative writing in mixed classes: what do heritage and second language learners think? *Foreign Language Annals*, *53*(1), 48–68. https://doi.org/10.1111/flan.12446.
Gan, Z., Leung, C., He, J., & Nang, H. (2018). Classroom assessment practices and learning motivation: a case study of Chinese EFL students. *TESOL Quarterly*, *53*(2), 514–529. https://doi.org/10.1002/tesq.476.
Guilloteaux, M., & Dornyei, Z. (2008). Motivating language learners: a classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly*, *42*(1), 55–77. https://doi.org/10.1002/j.1545-7249.2008.tb00207.x.
Guo, Q., & Xu, Y. (2020). Formative assessment use in university EFL writing instruction: a survey report from China. *Asia Pacific Journal of Education.* https://doi.org/10.1080/02188791.2020.1798737.
Ha, X. V., Nguyen, L. T., & Hung, B. P. (2021). Oral corrective feedback in English as a foreign language classrooms: a teaching and learning perspective. *Heliyon*, *7*(7), E07550. https://doi.org/10.1016/j.heliyon.2021.e0755.
Hao, S., & Johnson, R. L. (2013). Teachers' classroom assessment practices and fourth-graders' reading literacy achievements: an international study. *Teaching and Teacher Education*, *29*(1), 53–63. https://doi.org/10.1016/j.tate.2012.08.010.
Hu, J., & Cheung, C. K. (2021). Gender difference in the effect of cultural distance on academic performance among cross-border students in China. *Psicologia: Reflexão e Crítica.*, *34*, 33. https://doi.org/10.1186/s41155-021-00199-4.
Kane, M. (2010). Validity and fairness. *Language Testing*, *27*(2), 177–182. https://doi.org/10.1177/0265532209349467.
Leung, C., Davison, C., Hamp-Lyons, L., East, M., Evans, M., & Liu., Y.-C., & Purpura, J. E. (2018). Using assessment to promote learning: Clarifying constructs, theories, and practices. In J. M. Davis, J. M. Norris, M. E. Malone, T. H. McKay, & Y.-A. Son (Eds.), *Useful assessment and evaluation in language education*. Georgetown University Press.
McMillan, J. H. (Ed.) (2013). *Sage handbook of research on classroom assessment*. SAGE Publications.
Narathakoon, A., Sapsirin, S., & Subphadoongchone, P. (2020). Beliefs and classroom assessment practices of english teachers in primary schools in Thailand. *International Journal of Instruction*, *13*(3), 137–156. https://doi.org/10.29333/iji.2020.13310a.
Nguyen, L. T. & Hung, B. P. (2021). Communicative pronunciation teaching: Insights from the Vietnamese tertiary EFL classroom. System, 101, 102573. https://doi.org/10.1016/j.system.2021.102573
Pourdana, N. (2022). Impacts of computer-assisted diagnostic assessment on sustainability of L2 learners' collaborative writing improvement and their engagement modes. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(11), 1–21. https://doi.org/10.1186/s40862-022-00139-4.
Rafi, F., Pourdana, N., & Ghaemi, F. (2022). Computer-mediated diagnostic assessment of mixed-ability EFL learners' performance on tiered tasks: differentiating mediation on Google Meet™. *Journal of Modern Research in English Language Studies*, *9*(2), 1–26. https://doi.org/10.30479/jmrels.2021.16118.1950.
Russell, M., & Airasian, P. (2012). *Classroom assessment: concepts and applications*, (7th ed.,). McGraw-Hill.
Shepard, S. (2019). Classroom assessment to support teaching and learning. *The ANNALS of the American Academy of Political and Social Science*, *683*(1). https://doi.org/10.1177/0002716219843818.
Stiggins, R. J. (2005). *Student-involved assessment FOR learning*. Pearson Prentice Hall.
Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice*, *18*(4), 433–449. https://doi.org/10.1080/0969594X.2011.582838.
Thanh, P. T. H., & Gillies, R. (2010). Designing a culturally appropriate format of formative peer assessment for Asian students: *the case of Vietnamese students*. *International Journal of Educational Reform*, *19*(2), 72–85.
van de Pol, J., Mercer, N., & Volman, M. (2019). Scaffolding student understanding in small-group work: students' uptake of teacher support in subsequent small-group interaction. *Journal of the Learning Sciences*, *28*(2), 206–239. https://doi.org/10.1080/10508406.2018.1522258.
van de Pol, J., Volman, M., Oort, F., & Beishuizen, J. (2015). The effects of scaffolding in the classroom: support contingency and student independent working time in relation to student achievement, task effort and appreciation of support. *Instruction Science*, *43*(5), 615–641. https://doi.org/10.1007/s11251-015-9351-z.
Vattoy, K. D. (2020). Teachers' beliefs about feedback practice as related to student self-regulation, self-efficacy, and language skills in teaching English as a foreign language. *Stud. Educ. Eval.*, *64*, 100828. https://doi.org/10.1016/j.stueduc.2019.100828.
Wang, L., Lee, I., & Park, M. (2020). Chinese university EFL teachers' beliefs and practices of classroom writing assessment. *Stud. Educ. Eval.*, *66*, 100890. https://doi.org/10.1016/j.stueduc.2020.100890.
Wen, M. L., Tsai, C., & Chang, C. (2006). Attitudes toward peer assessment: a comparison of the perspectives of pre-service and in-service teachers. *Innovations in Education and Teaching International*, *43*(1), 83–92. https://doi.org/10.1080/14703290500467640.

Winke, P., Zhang, X., & Pierce, S. J. (2022). A closer look at a marginalized test method: Self-assessment as a measure of speaking proficiency. *Studies in Second Language Acquisition* (in press. https://doi.org/10.1017/S0272263122000079.
Wu, X. M., Zhang, L. J., & Dixon, H. R. (2021). Implementing assessment for learning (AfL) in Chinese university EFL classes: teachers' values and practices. *System*, *101*, 102589. https://doi.org/10.1016/j.system.2021.102589.
Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: a narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, *43*(3), 492–513. https://doi.org/10.1002/j.1545-7249.2009.tb00246.x.
Yan, Q., Zhang, L. J., & Cheng, X. (2021). Implementing classroom-based assessment for young EFL learners in the Chinese context: a case study. *Asia-Pacific Educational Research*, *30*(6), 541–552. https://doi.org/10.1007/s40299-021-00602-9.
Yan, Q., Zhang, L. J., & Dixon, H. R. (2022). Exploring classroom-based assessment for young EFL learners in the Chinese context: teachers' beliefs and practices. *Frontiers in Psychology*, *13*, 1051728. https://doi.org/10.3389/fpsyg.2022.1051728.
Zhou, J., & Deneen, C. C. (2016). Chinese award-winning tutors' perceptions and practices of classroom-based assessment. *Assessment and Evaluation in Higher Education*, *41*(8), 1144–1158. https://doi.org/10.1080/02602938.2015.1066306.

Publisher's Note