Language Testing in Asia

RESEARCH Open Access

Understanding university English instructors' assessment literacy: a formative assessment perspective

Miaomiao Kang[1] and Ricky Lam[2*]

*Correspondence:
rickylam@hkbu.edu.hk

[1] College of Professional and Continuing Education, The Hong Kong Polytechnic University, PolyU West Kowloon Campus, 9 Hoi Ting Road, Yau Ma Tei, Kowloon, Hong Kong, China
[2] Department of Education Studies, Faculty of Arts and Social Sciences, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China

Abstract

Formative assessment is a powerful approach to inform, direct, and enhance teaching and learning. To deploy it in classrooms, teachers are expected to be assessment-capable. However, few studies have explored the extent to which teachers are literate in formative assessment applications and their levels of literacy, especially in Chinese tertiary settings. This study examined Chinese university English instructors' assessment literacy by investigating what three participants knew, thought about, and how they applied formative assessment. A knowledge-conception-practice framework was applied to explore assessment literacy in the four formative assessment processes (i.e., design, use, interpretation, and communication) to engage student learning. Individual interviews, classroom observations, and supplementary documents were adopted as instruments for data collection. Through iterative and comparative thematic analysis, findings revealed that notwithstanding variations, all three instructors possessed basic formative assessment knowledge and were willing to design and conduct formative assessment-related activities for students' learning engagement and enhancement. However, they held product-oriented rather than process-oriented mindsets when designing, using, interpreting, and communicating formative assessment. They also lacked sufficient assessment knowledge and strategies to effectively engage students in self-regulated learning through learner-centered task design, success criteria, and process-oriented feedback. Lastly, the study discussed theoretical and pedagogical implications regarding how to conceptualize and develop English instructors' formative assessment literacy in mainland China and beyond.

Keywords: English language assessment, Formative assessment, Assessment literacy, Chinese higher education

Introduction

The Chinese educational landscape, owing to its deep-seated exam-oriented culture, emphasizes assessment outcomes, diligence, and the teacher-student hierarchy rather than students' learning processes, abilities, and active agency in learning (Lam, 2022). In tertiary English teaching, although instructors have autonomy in pedagogical decision-making, they resort to adopting drillings of students' basic language skills without

encouraging students' active engagement in learning and development of higher-order thinking skills.

To transform this product-based pedagogy into a process-oriented one, formative assessment (hereafter FA) has been recommended in major Chinese educational curriculum guidelines or requirements since 2001 (e.g., *College English Curriculum Requirements*) (Ministry of Education 2020). Despite official recommendations, FA has not been widely or appropriately conducted in Chinese classrooms (Gan & Lam, 2020; Xiang et al., 2023). The test-oriented educational culture results in challenges of FA applications, such as teachers' fixation on grades (Chen et al., 2013; Huang & Luo, 2014), overemphasis on error correction instead of process-oriented knowledge construction (Lam, 2022), and unselected use of FA strategies which do not fit in student learning styles (Xiang et al., 2023). Teachers' capacities in the comprehension, beliefs, and operation of FA may be affected and even restricted by the product-oriented context and mindset (Yin & Buck, 2019).

Teachers' assessment ability was termed "assessment literacy" by Stiggins (1991), referring to the knowledge and competence of using assessment for student learning improvement. Thereafter, most researchers included assessment knowledge and practices in their definitions of assessment literacy (e.g., Boyles, 2006). Apart from assessment knowledge and practices, several researchers proposed that assessment conceptions are also an integral part of assessment literacy, because they mediate knowledge base and practices (Chan & Luk, 2022; Xu & Brown, 2016). In this paper, FA literacy refers to English language instructors' FA knowledge base, conceptions, and practices. It is regarded as a subset of the umbrella term "assessment literacy." FA literacy plays a dominant role in maximizing the benefits of FA on learning and pedagogy and effectively enhancing EFL teachers' overall capacity in instruction and assessment. Since most current studies on language assessment literacy are somewhat generic without a clear focus on teacher competence in performing classroom-based FA, this study focuses on this specific, yet much-needed aspect of language assessment literacy by exploring university instructors' FA literacy under the latest education reforms (Gu & Lam, 2023). It intends to understand university English instructors' FA literacy by qualitatively examining what they know, think about, and how they conduct FA tasks and further explore the interrelationships of these three components, namely knowledge, conceptions, and practices. To comprehend specific components of FA literacy and conceptualize the theoretical framework in this research, it is necessary to review the literature about FA definition, rationale, and processes and the three elements of FA literacy.

Literature review

Formative assessment: definition, rationale, and processes

Black and Wiliam (1998) proposed a definition of FA in their seminal review, namely "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (pp. 7-8). This definition subsequently served as the fundamental basis for the widespread adoption of FA in educational theory, policy, and implementation worldwide. Despite the discrepancy between different researchers' foci and conceptualization, FA is commonly regarded as an instructional strategy, either planned or

spontaneous, aiming to collate students' learning evidence, for enhancing future instruction and learning quality (Yin & Buck, 2019). FA rationale is twofold, namely self-regulated learning (SRL) and feedback. SRL refers to an autonomous and recursive process (see Fig. 1), including self-planning, where students examine the assignment, set learning goals, and build up a schedule; self-performing, where students monitor and control their learning behaviors, emotions, and cognition; and self-evaluating, where students review and refine their previous autonomous learning through self-reflection (Zimmerman, 1998). This regulatory process covers learners' three interrelated dimensions of learning engagement, namely behavioral, affective, and cognitive involvement (Andrade & Brookhart, 2020).

Engaging students in SRL is conducive to enhancing their cognitive thinking; boosting their intrinsic learning motivations; raising their sense of responsibility toward learning; and finally, cultivating them to be motivated, responsible, and effective learners, which is a fundamental goal of FA (Panadero et al., 2019). Effective FA applications promote students' SRL and autonomy through their interactive engagement in goal setting and progress monitoring (Xiao & Yang, 2019), open-ended task designs (Yin & Buck, 2019), transparent assessment criteria (Andrade & Brookhart, 2020), and external feedback as guidance (Chen, 2024).

Feedback is also a catalyst to identify students' strengths and limitations in learning, and scaffold them to monitor, reflect on, and improve their learning progress during their SRL process (Clark, 2012). To achieve the learning-supportive purpose, the timeliness and content of feedback are two crucial aspects to consider (McCallum & Milner, 2021). Process-oriented and task-specific comments are much more effective than product-focused error corrections in engaging students in SRL to support their language development (Panadero et al., 2019).

These two interconnected rationales (i.e., SRL and feedback) can only materialize through specific FA processes, namely design, use, interpretation, and communication

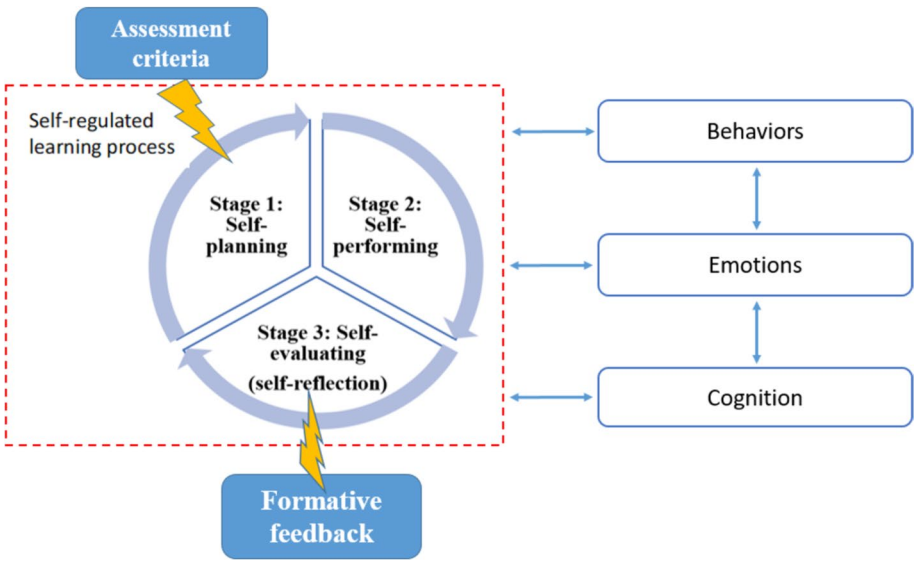

Fig. 1 Stages and dimensions of self-regulated learning

(Hill & McNamara, 2012). Although these four processes are separated owing to their unique attributes, they are inextricably interrelated in FA applications. For instance, the non-grade and student-centered FA design facilitates the use of FA data (i.e., for learning-supportive purpose). The success criteria designed in advance decide how FA data can be used, interpreted, and communicated. Instructors' communications with students through feedback relate to how instructors design FA procedures, use FA to facilitate student SRL, and interpret FA as their instructional input. Overall, the rationales for FA are represented by its specific processes, and instructors' knowledge, conceptions, and actions regarding the four FA processes determine the extent to which students are engaged in their SRL. In realism, instructors may have insufficient FA literacy to design, use, interpret, and communicate FA efficiently, especially when they are used to teaching to the test (Li & Gu, 2023).

Knowledge, conceptions, and practices

Teachers' knowledge base includes theoretical knowledge and practical skills. Specifically, teachers need to know when, why, and how to design, use, interpret, and communicate FA to support student learning involvement and improvement (Heritage, 2007). In the test-dominated educational context, Chinese instructors have insufficient knowledge, experience, and training about the process-oriented FA (Gan & Lam, 2020; Li & Gu, 2023). They usually possess basic knowledge about the nature and purpose of FA, but lack in-depth and contextualized expertise to effectively achieve its formative purpose through active student engagement, such as how to design and conduct peer- and self-assessment (Lam, 2019). Researchers revealed that teachers might lack FA expertise regarding how to engage student SRL effectively during FA process, such as why and how to frame FA as a non-graded, purely learning-supportive tool (Arrafii & Sumarni, 2018), design more productive questions and student-centered tasks (Yin & Buck, 2019), provide clarified and transparent assessment criteria (Arrafii & Sumarni, 2018; Li & Gu, 2023), and offer timely and high-quality feedback (McCallum & Milner, 2021).

Assessment conceptions embrace cognitive and affective parts, namely, beliefs and emotions (Lam, 2019; Xiang et al., 2023). Previous studies found that most instructors believed FA was necessary for learning improvement (e.g., Chen & Teo, 2020), yet they still hold a product-oriented mindset, prioritizing the assessment outcomes rather than the learning process (Chen et al., 2013). Some instructors still possessed grade-conscious mindsets, which demotivate students to pursue task-specific enhancement and inhibit their cognitive development (Andrade & Brookhart, 2020). Emotionally, instructors lacked confidence or motivation in designing and conducting FA (e.g., reluctance to provide timely and informative feedback or empowering students to self- or co-regulate their learning) owing to inadequate skills, individual challenges, or practical constraints (e.g., large class size) (Widiastuti et al., 2020; Xiang et al., 2023).

Only a few studies examined assessment conceptions, especially in higher education (Xu & Brown, 2016). Among the limited research, Samuelowicz and Bain (2002) and Postareff et al. (2012) regarded academics' assessment conceptions as a continuum, from the most reproductive (surface-level) to the most transformational (deep-level) in another extreme. These findings echoed Samuelowicz and Bain's (2001) research outcomes on college lecturers' beliefs concerning teaching and learning, ranging

from teaching-centered to learning-centered beliefs. However, all these three studies focused on cognitive attitude without covering the affective dimension, which is also a necessary part of assessment conceptions. Xiang et al. (2023) found that compared with cognitive beliefs, affective conceptions were more closely related to instructors' FA practices although their research was limited to the use of quantitative self-reported data.

Regarding FA practices, studies have suggested that instructors did not design, use, interpret, and communicate FA effectively in classrooms. They overemphasized summative grading (Chen et al., 2013), insufficiently clarified assessment goals and criteria (Levi & Inbar-Lourie, 2020; Li & Gu, 2023), highlighted error correction rather than the content and thoughts in feedback provision (Lam, 2022), and underused task-specific written feedback (Widiastuti et al., 2020). Studies unveiled that when implementing FA, instructors usually did better in collecting data on students' learning evidence through FA-based task design, while they did not perform well in data interpretation, use, and communication (e.g., providing effective feedback to facilitate student SRL) (Raffe & Loughland, 2021). Liu's (2024) recent research got much more positive findings, in which a teacher informant attempted to provide evidence-based, process-oriented, and individualized feedback, and her scaffolding was always adjusted to facilitate student self-regulated learning. However, this study only included one single case in the formative writing assessment field.

The abovementioned investigations were typically site-based and context-specific, exemplified by Lam's (2019) and Liu's (2024) study which concentrated on the secondary-level writing settings. Many adopted a large-scale quantitative research design, without generating detailed or in-depth insights (e.g., Xiang et al., 2023). For those qualitative studies, they mainly used self-reported data (e.g., interviews) from instructors' perspectives (e.g., Chan & Luk, 2022; Raffe & Loughland, 2021), which might lead to personal bias in their findings. Further, most studies failed to cover all three components of FA literacy (i.e., knowledge, conceptions, and practices).

Previous studies suggested that FA knowledge and/or conceptions are closely related to practices. Teachers with necessary knowledge and positive conceptions are likely to promote effective FA in classrooms (Liu, 2024; Widiastuti et al., 2020). Herman et al. (2015) found that teachers could effectively facilitate student learning through FA applications despite their limited knowledge base. Although this finding seems inconsistent with others, the knowledge measured in this study was only the content and pedagogical knowledge rather than a specific understanding of FA concepts and related topics (e.g., feedback). Studies also indicated that teachers' practices may not always align with their beliefs (Liu, 2024; Nimehchisalem & Abdalla, 2020). For example, some teachers believed in student-centered classrooms but adopted teacher-centered pedagogy (Kaymakamoglu, 2018). Their belief-practice misalignments may stem from misconceptions, fixed mindsets, or lack of skills to overcome challenges in classrooms (i.e., time constraints, large class size, heavy workloads, and students' reluctance to criticize their peers' work due to protection of face) (Raffe & Loughland, 2021). Given numerous constraints in FA practices, instructors have to strengthen their knowledge and conceptions and possess capabilities to handle contextualized issues, to facilitate student learning engagement and enhancement through FA processes.

From the abovementioned literature, FA knowledge base, conceptions, and practices are indispensable and interrelated components in determining whether teachers are assessment-capable in FA applications. First, few studies covered all three aspects concurrently. Most studies focused on one or two aspects, failing to take a holistic view of all three components and present FA literacy comprehensively. Second, inadequate studies investigated FA literacy in the Chinese context, where exam-oriented and teacher-centered mindsets have been dominating the teaching and learning landscape for thousands of years. Third, limited attention was paid to university English instructors, especially those who taught English majors, the group undertaking considerable English language learning and assessments. Fourth, few research obtained qualitative in-depth data from triangulated research sources. To fill the above research gaps, this study qualitatively explored FA literacy of Chinese university English instructors, who taught English majors, from three aspects, namely knowledge base, conceptions, and practices through multiple data sources. Three research questions were proposed to fulfill this research purpose:

1. To what extent do the instructors know about formative assessment?
2. What are the instructors' cognitive and affective beliefs of formative assessment?
3. How do instructors apply formative assessment as part of their pedagogical practices?

Aligning with the conceptualization of assessment literacy and FA literacy (see Sect. "Introduction"), we adopted a knowledge-conception-practice framework to examine FA literacy in Chinese university settings. Combining with FA rationale, processes, and components specifically, this framework entails instructors' knowledge, perceptions (beliefs and emotions), and experiences toward deploying the FA processes (i.e., design, use, interpretation, and communication) to support student learning engagement and enhancement (Hill & McNamara, 2012; Xu & Brown, 2016).

Research methodology

This study adopted a qualitative case study design. While this design might not generalize its findings, it has good potential in its transferability to other similar informants and settings (Drisko, 2024). This design facilitated a holistic and in-depth understanding of FA literacy through several bounded systems (i.e., instructors), via addressing "what" and "how" questions, namely, what instructors know, think about, and how they implement FA (Yin, 2018). Semi-structured interviews, classroom observations, and documents were instruments for data collection. Using various tools warranted the triangulation of the data sources, which strengthened the study's internal validity, namely ensuring that the research findings were a trustworthy reflection of reality (Merriam & Tisdell, 2016).

Research sites and participants

This study was conducted in two non-key universities in Central China. Key universities are mainly top-ranked universities involved in some national initiatives intending to build and develop world-class research universities, such as Project 985, Project 211, and Double First-Class Project (Xie & Teo, 2020). By contrast, non-key universities are

outsiders to these projects. Compared with key universities, non-key universities are predominant in number and average-performing in terms of students' academic standards and instructors' pedagogy. These two universities (University A & University B) could represent other non-key Chinese universities regarding program structure, student English performance, and instructors' qualifications, knowledge, and skills in FA. Their "typical" features enhanced the transferability of the research (Patton, 2015, p. 284).

Purposeful sampling was adopted to align the research purpose and obtain rich information. The selection criteria were threefold. Each participant should be a full-time English lecturer, teach English majors, and teach undergraduates at least four lessons per week (one and a half hours per lesson). Volunteers satisfying these criteria were further selected to maximize their variation in age, gender, English courses taught, educational background, and working experience. Finally, three volunteers were recruited. Their backgrounds are shown in Table 1.

Data collection

Data collection started after all participants signed the informed consent. Apart from the permission request for obtaining their documentary, interview, and classroom data, the three participants were also informed that their data would be used solely for research purposes, with personal information kept confidential by adopting anonymity. Besides, they could withdraw from the study at any time without penalty. Throughout the research process, we adhered to ethical principles, including respecting informants, maintaining honesty, protecting privacy, and minimizing researcher bias (Hays & Singh, 2012).

The original interview protocol is attached to Appendix 1 for reference. The interview format was semi-structured, which not only guaranteed covering all theoretical components (i.e., knowledge, conceptions, and practices) but also facilitated adjusting situations (e.g., changing the words chosen or adding follow-up questions) according to the interviewees' responses to capture emerging ideas or explore more details and examples (Merriam & Tisdell, 2016). For each participant, the first interview lasted between 1.5 h and 2 h. Then, three to four 30-min follow-up interviews were conducted to obtain

Table 1 Description of the participants' backgrounds

Name (pseudonyms)	Connie	Thomas	Fiona
Working institution	University A	University B	University A
Age	41	47	30
Gender	Female	Male	Female
English courses taught	English Linguistics; Extensive Reading; English Grammar	Intensive Reading; English Listening; English Phonetics	English Writing; English Listening; Translation; Intensive Reading
Qualifications	BA and MA in Foreign Linguistics and Applied Linguistics	BA in English Teaching; MA in English Language and Literature	BA and MA in English Language & Literature
Types of learning institutions	Comprehensive	Normal	Comprehensive
Locations of their learning institutions	Northern China; Southern China	Northern China	Northern China
Years of teaching	15	23	4

further clarification. All interviews were conducted face-to-face except for Connie's and Thomas's last interviews which were conducted online through WeChat video. Interviewees used their first language, i.e., Putonghua, to answer all item questions. The interviews were recorded, transcribed verbatim, and translated into English before analysis. All interview transcripts after translation were sent back to the interviewees for member checking. The three cases made minor corrections and expanded some of their ideas or examples during member checking. They were also asked to clarify some of their previous responses or ideas in the follow-up interviews to verify the interpretations of their previous data.

Nine classroom observations (3 times/case) were also conducted. Each one lasted for 1.5 h, namely, a whole lesson. The observation protocol is presented in Appendix 2. All observations were videotaped and documented through detailed field notes. They captured FA literacy in authentic contexts, reflecting instructors' practical FA skills and applications, and strengthening the validity of the related data obtained from other sources. Documents in this study consisted of the participants' course schedules, PowerPoint slides, reflective journals, and their students' quizzes/assignments for over half a year after the commencement of the research.

Data analysis

The data set included interview transcripts, observation field notes, and documents. These data were analyzed manually by the first author via two major steps. First, she identified any useful or potentially valuable segments or linkages from the data set. During this process, she started coding, namely, making notations such as labels, comments, questions, or reflections next to the unit of the data in the margins or with separate memos (Merriam & Tisdell, 2016).

After that, she moved on to the analytical coding process, namely, grouping or sorting similar open codes into generalized categories/subcategories based on her comprehension and reflection of the code meaning, and the comparisons between the different open codes (Richards, 2015). During this process, she constantly made comparisons between her subsequent notations and earlier ones, between different categories in one data source and various data sources, and between different cases. She also attempted to obtain more evidence, verification, or clarification from follow-up interviews or other data sources. Based on constant comparisons and iterative data collection and analysis, categories were subdivided, subsumed, or modified. Additional data were continuously collected during the data analysis process until data saturation was achieved. To maintain inter-rater reliability, the second author verified about 20 percent of the processed data twice. The refined categories were further conceptualized into themes. In this process, rounds of revisions were performed based on the second author's comments until a consensus on coding categorization was reached.

In the beginning, the data analysis was primarily inductive. As more data were collected and analyzed, a deductive strategy was added to check whether the previous categories were strong enough to take shape (Merriam & Tisdell, 2016). Because commonalities and varieties existed in the three cases, their subcategories and categories also showed similarities and differences. Finally, they were generalized into four salient themes: FA and Grades, FA Instruments, Assessment Criteria, and Feedback Literacy.

Each theme integrates participant knowledge, conceptions, and practices of FA based on our tripartite research frame. This thematic presentation of results facilitates the revelation of intricate relationships among the three components of FA literacy (knowledge, conceptions, and practices) within each theme.

Results

Formative assessment and grades

Both Connie and Fiona believed that FA should have summative values, and grades/scores were a must for FA. Therefore, they assigned every FA task a score toward students' final grades. Compared with Fiona, Connie's use of scores in FA tasks limited students' learning engagement. Connie has a grade-conscious mindset, so she overemphasized grades when implementing FA, namely, using grades for stimulating, ranking, and labeling.

Connie thought that these attempts could "always highly motivate" students (Connie-I-2). However, observational data indicated that her students were more enthusiastic about their scores than the task itself or their learning gaps. Some even cheated on the quiz to pursue high marks or neglected Connie's qualitative feedback (Connie-O-1; Connie-O-2). In her interviews, Connie also expressed her confusion about how to "persuade students to reveal their shortcomings in their self-reflection tasks" (Connie-I-1) and activate every student's involvement in collaborative tasks when "students tended to impose more, if not all, responsibilities on some well-performed group members for a better result" (Connie-I-3). With limited FA knowledge and a grade-conscious mindset, Connie failed to frame those FA-based tasks as merely learning-supportive tools for students to review, revise, and refine their learning developments, despite instructional endeavors invested in her FA applications (e.g., provision of specific feedback). Her emphasis on grades limited students' active learning engagement.

By contrast, Fiona realized that grades should not be highlighted in FA applications. She believed that "grades might distract student attention from learning and weaken the learning-supportive purpose of FA" (Fiona-I-1). Hence, she adopted two marking strategies in FA practices, namely delaying evaluation and offering sub-scores, to enhance students' engagement in self-regulated learning. For instance, in one assignment, students' assessment scores were designed to cover their performance in three stages: peer-assessment, self-reflection, and revisions. Fiona did not evaluate students' performance immediately after they had finished each phase. Instead, she offered scores after they had accomplished all three stages. This delayed marking aimed to attract students' attention to specific tasks and take full advantage of their peer feedback and self-reflection to revise and improve their written drafts.

Fiona also provided sub-scores for students' writing tasks to assess their levels in each aspect, such as content, organization, and language. She used sub-scores rather than an overall score to "inform students of their learning more specifically to facilitate their self-reflection and self-regulation" (Fiona-I-4). However, this intention might not have been well achieved because Fiona neither specified the scoring rubrics nor provided qualitative written feedback to illustrate how they had achieved these aspects. Her FA literacy

regarding assessment criteria and feedback are presented in Sects. "Assessment criteria" and "Feedback literacy", respectively.

Thomas was the only case who knew FA has no relationship with or even conflicts with summative grades. According to his interview data, he grasped the core value of FA, namely, to promote students' learning process rather than measure their learning through marks. Thomas also evinced his dislike of the grade-conscious mindset resulting from the exam-oriented culture and his disagreement with using grades/scores as incentives for any assignments or quizzes. Aligning with his knowledge and beliefs, he tried to frame FA as an opportunity to strengthen student learning engagement and enhancement rather than an evaluative tool in practice. There were two dimensions to his efforts: lowering the stakes of FA tasks and frequently motivating students to focus on their learning progress rather than the results. First, "to raise students' learning involvement and motivation" (Thomas-I-1), Thomas ensured that his FA tasks were not evaluative by design (e.g., conducting non-grade tasks or mainly offering participation scores). Second, he frequently encouraged students to involve themselves in their learning process and avoid product-oriented mindsets. For example, he motivated them to "become the owners of themselves" and "not be afraid of failure" (Thomas-P; Thomas-S). In his class, he always appreciated students' efforts, courage, and any progress made. However, except for a "risk-free" design and frequent encouragements, Thomas did not provide systematic task-specific guidance through assessment criteria or feedback to enhance student deep learning and higher-level engagement.

Formative assessment instruments

All three informants recognized the learning-enhancing and process-oriented nature of FA. Therefore, they were willing to design different FA tasks during their teaching process to facilitate student learning involvement and improvement. Connie and Thomas reiterated that traditional teacher-dominant pedagogy without student active participation should be avoided. For example, Connie mentioned, "Students should not be passive receivers like machines" (Connie-I-1). Thomas expressed his dislike of teacher-centered instruction and emphasized that "students should always get actively involved during the instruction…instructors should let students be the owners of their learning" (Thomas-I-1).

Despite these beliefs, Connie and Thomas only carried out a watered-down version of FA-based activities. In Connie's case, closed-ended tasks were much more dominant in frequency and quantity than open-ended tasks. According to her self-reported and classroom data, she conducted a timed quiz involving multiple-choice, fill-in-the-blanks, and true or false items through Rain Classroom[1] at the beginning of every class to check on students' learning of the previous lesson. She frequently asked simple yes or no questions like "Understand?" to check on students' understanding during her lessons. However, she occasionally adopted open-ended questions to cognitively engage students. Connie realized that she needed to "use more open-ended assessment forms" (Connie-R-1). In her first interview, she mentioned using "some flexible grammar tasks

[1] Rain Classroom, a popular mobile application in China, was developed by Tsinghua University to facilitate teaching and learning in large classes.

like writing a short, interesting paragraph may activate student learning engagement" (Connie-I-1). However, she expressed her unwillingness to replace those closed-ended FA tasks with flexible and open-ended ones due to her lack of practical skills to handle a large class size with unbearable workloads.

Besides inadequate strategies to handle large class size, Connie's deep-rooted teacher-dominant mindset also hindered her willingness to implement some student-centered activities. For example, she never conducted peer-assessment of students' writing tasks because she thought "students are neither serious nor capable of assessing their peers' writing" (Connie-I-1).

Similarly, Thomas emphasized the importance of students' self-regulated learning and active engagement in his self-reported data, yet he did not empower students with sufficient responsibilities in practice. Compared with Connie and Fiona, Thomas equally possessed a teacher-dominant mindset to an even larger extent. He believed his "well-prepared slides are much worthier for showing and teaching" than students' (Thomas-I-3). Therefore, when observing that students could not prepare the topic sufficiently, he directly canceled their group presentation tasks to "save time and let them learn more about the topic efficiently" by using his slides (Thomas-I-3). Similarly, although Thomas designed many open-ended tasks, these activities did not fulfill their functions to actively engage students in providing prompt responses. After presenting the activity instructions through PowerPoint slides, he usually did not give students enough wait time to think and prepare the task, ignored offering guidance to trigger students' further discussions or thoughts, or even appropriated students' ideas to save time and catch up with his teaching schedule (Thomas-O-1; Thomas-O-2; Thomas-O-3). Besides, Thomas resisted adopting some student-centered tasks such as group discussions or activities in Rain Classroom although he did value them in student self-regulated learning.

Fiona designed and conducted the most open-ended and student-centered activities among the three informants. The vast majority (nearly 90%) of her FA activities were student-driven and open-ended. This might result from her "learning by doing" philosophy. As she highlighted, "Learning only occurs when students have chances to explore, organize, and construct their knowledge by themselves" (Fiona-I-1). Besides, with a lack of FA training and experience, she was enthusiastic about trying various student-centered tasks in different courses. For example, she applied role play in her Intensive Reading course to raise students' interest and deepen their understanding of the original test. She adopted peer-assessment and self-assessment in her writing course and attempted group oral presentation and translation debate in her translation course, aiming to involve students actively in their learning process. However, Fiona also expressed her helplessness after numerous attempts in different courses, "after I found the way to improve an activity for a course based on students' performance, I have no opportunity to try the same activity in the same course next semester or next year… some reflective adjustments are not suitable for the new course" (Fiona-I-4). This is because as a novice teacher, she had to comply with her department's randomly assigned teaching duties to resolve the manpower shortage problem. Fiona had no autonomy in choosing what she preferred to teach.

When conducting student-centered activities (e.g., peer-assessment and self-assessment), all three instructors mainly focused on students' behavioral and emotional

learning engagement but lacked awareness and strategies to enhance students' cognitive thinking, reflections, and learning regulation. Thomas's peer- and self-assessments were mainly closed-ended answer checking, depriving students of opportunities to interact, negotiate, and solve problems with their peers or reflect on their learning. Although most student-involved assessments in Connie's and Fiona's classrooms were open-ended, these assessments could not maximize students' co- and self-regulated learning due to insufficient guidance provided by instructors. For example, Connie and Fiona lacked practical skills to guide students to compose peer feedback. Facing the same challenges that their students only focused on the strengths of their peers' performance when providing peer feedback, Fiona felt "frustrated" and "unsatisfied" (Fiona-I-1). She even canceled some further attempts (e.g., the one after students' presentations in her Intensive Reading course), despite her initial passion for peer-assessment (F-I-4; F-O-3). Connie provided quantitative guidance (i.e., requiring students to identify two weaknesses in their peer feedback) and verbal encouragement. However, neither of them offered cognitive scaffolds or linguistic clues to improve students' feedback provision skills.

No matter what types of activities the three informants implemented, they lacked adequate literacy to provide systematic assessment criteria and effective feedback to engage students more actively in their learning process. Their FA literacy in assessment criteria and feedback will be presented in the following two sections respectively.

Assessment criteria

Connie, Thomas, and Fiona possessed insufficient knowledge and strategies, and inappropriate conceptions regarding when, why, and how to engage students in their self-regulation through use of assessment criteria. They misconceived that students "need more freedom" when approaching their tasks (Fiona-I-1) and "don't want to restrict students" from using detailed criteria (Connie-I-3; Thomas-I-1). They did not realize that students need well-constructed criteria to make effective plans when they start to regulate their learning to handle the task. With a traditional teacher-dominant mindset, all of them took it for granted to interpret students' performance with their own "implicit criteria" to a certain degree. Their subjective thinking was also revealed in the fact that none of the three instructors had the consciousness to involve students in constructing assessment criteria. They thought that "it is the instructor's responsibility to create success criteria" (Fiona-I-1; Thomas-I-2) and "only the instructor had the authority" to do so (Connie-I-1).

When assigning tasks, Connie and Fiona offered generic assessment criteria to guide the major content of student assignments. For example, for a book report task, Connie asked students to include "a summary of the book, an introduction of its author, impressive phrases and sentences, and their reflections" in their reports (Connie-R-2). Apart from these brief criteria, neither of them offered systematic success rubrics to explicitly present what details or to what extent students should achieve under each criterion. Connie "intentionally delayed providing some specific criteria" until she offered students feedback after they finished their tasks, ignoring students' ongoing development of self-evaluative judgment when they independently or collaboratively prepared for the tasks (Connie-I-3). Fiona complained that sometimes "it was too difficult for me [her] to construct systematic assessment criteria when I [she]

designed various FA activities" owing to her lack of teaching and assessment experience to anticipate students' performance (Fiona-I-4). As mentioned in Sect. "Formative assessment instruments", Fiona's teaching duties varied year by year, so she "lacked chances to get enough evidence of students' performance and refine my [her] assessment standards in one fixed course" (Fiona-I-4).

Among the three informants, the assessment criteria provided by Thomas were the least effective for student learning engagement. These criteria focused on the language form rather than the content, without any in-depth pointers for students' task-specific cognitive involvement. For example, when conducting speech-related activities, he just highlighted the form-relevant items, such as "manner," "pronunciation," and "voice" without including any detailed description regarding the content of student speech production (Thomas-O-2). In his interviews, he mentioned that he intended to use those form-focused, flexible criteria to "encourage students' participation" and "boost their interest and motivation" (Thomas-I-4). However, the absence of instructions in assessment content usually made his assessment objectives too broad or ambiguous, making his students unable to engage in deep learning. For instance, in the only open-ended reflection task in his classroom, his students were observed to just pay attention to complete the sentence structure "I've learned…" provided by Thomas without any deep reflections on their learning strengths and weaknesses (Thomas-O-3). With insufficient FA expertise, Thomas did not realize that this kind of lower-level behavioral and motivational engagement was insufficient for students' effective learning regulation and enrichment.

Apart from generic assessment criteria, there was a lack of appropriate exemplars to enhance students' contextualized comprehension and internalization of those success criteria in the three instructors' FA applications. Connie never adopted exemplars to illustrate her assessment criteria because she was not cognizant of the necessity to do so. As she said, "It's easy to understand these simple guidelines…No need to use exemplars" (Connie-I-1). Fiona seldom provided exemplars (only once among all her FA practices). She just offered assessment exemplars in the argumentative writing task in her Writing II course. In that task, Fiona offered her self-generated examples to "show students what good work looks like" (F-R-2). However, these exemplars were provided not until all students completed their tasks in that writing cycle. This delayed provision of exemplars deprived students of opportunities to learn from good examples and better understand the quality criteria before they managed to self-regulate their writing tasks reflectively (Fiona-R-2; Fiona-I-3).

In contrast with Connie and Fiona, Thomas offered exemplars every time whenever assigning a task. His frequent exemplars, however, did not fulfill their functions of enhancing students' in-depth understanding of the success criteria or facilitating their higher-level learning engagement when they handled their tasks (Thomas-O-1; Thomas-O-2; Thomas-O-3). These exemplars predominantly served as simple motivational tools to support students' lower-level emotional and behavioral engagement due to the absence of his briefings to relate those exemplars to respective assessment criteria.

In summary, all three instructors lacked professional knowledge, appropriate conceptions, and practical strategies to use assessment criteria and tangible instructional scaffolds (e.g., exemplars) to facilitate students' goal-setting and evaluative judgment

when they engaged in self-regulated learning. Besides assessment criteria, feedback literacy is another salient theme to indicate instructors' FA literacy.

Feedback literacy

Connie, Thomas, and Fiona valued feedback and recognized its crucial role in students' learning improvement during their learning process. However, their understanding was partial regarding the process-oriented nature and the learning-centered function of feedback. To some extent, all of them had insufficient knowledge and misconceptions about when and how to provide constructive feedback to better engage students' learning.

The three instructors reiterated the importance of the timeliness of feedback. For example, Thomas stressed, "Feedback must be timely; otherwise, it's meaningless" (Thomas-I-1). They were also observed to offer students instant comments right after students' in-class assessments or feedback on their homework assignments within a week. However, Thomas' and Fiona's feedback was always provided after students finished their tasks rather than during their task preparation process. They misconceived that this kind of post-assessment feedback was "timely enough to improve students' learning" (Fiona-I-1; Thomas-R-2), without awareness to guide students during their self-regulated learning process when they are doing tasks. In practice, when students were preparing their in-class tasks, Thomas was observed just standing on the podium, without any interactions with students. In contrast, Fiona walked around to watch her students' preparation and communicated with them. However, her interactions mainly served a social rather than a pedagogical purpose. For instance, when students were preparing their role-play tasks in one of her Intensive Reading lessons, she observed their discussions and asked each group some questions, such as "Who will play the little boy?" or "Do you want to be the first group to perform the role-play?" (F-O-3). Fiona provided no profound task-specific comments or guidance to assist students' preparation process. Because of a lack of timely pre-assessment feedback, her students usually made some "terrible mistakes" in their formal performance, such as choosing inappropriate materials to translate, analyze, and share with their classmates (Fiona-O-1). Although it seemed that students in Thomas's classroom were never off-track without pre-assessment guidance, it might be due to the fact that Thomas did not have high expectations of his students' works in terms of task-specific and in-depth cognitive learning engagement.

Among the three instructors, only Connie had awareness and attempted process-oriented, pre-assessment feedback. For example, in oral presentation tasks, she asked students to send her their PowerPoint slides one week before formal presentations and gave them "face-to-face comments for further revisions and some presentation skills" (Connie-O-2; Connie-R-2). Through pre-assessment feedback, Connie intended to "guide students to reflect on and revise their PowerPoint slides" before students performed formal presentations in front of their classmates (Connie-I-4). However, Connie did not provide pre-assessment feedback in all tasks. In about 80% of her FA-based activities, such as self-reflection tasks and book reports, Connie only offered feedback near the end of the tasks due to "time limit" and "heavy workloads" (Connie-I-3; Connie-O-3). Furthermore, her pre-assessment comments mainly focused on general issues (e.g., "whether the presentation content is appropriate"), ensuring that students were not off task (Connie-I-1). She provided students with specific formative feedback only when they finished their

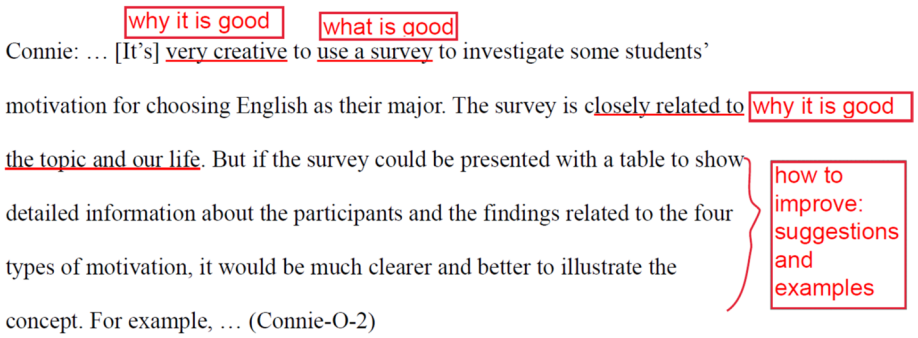

Fig. 2 An example of Connie's verbal feedback

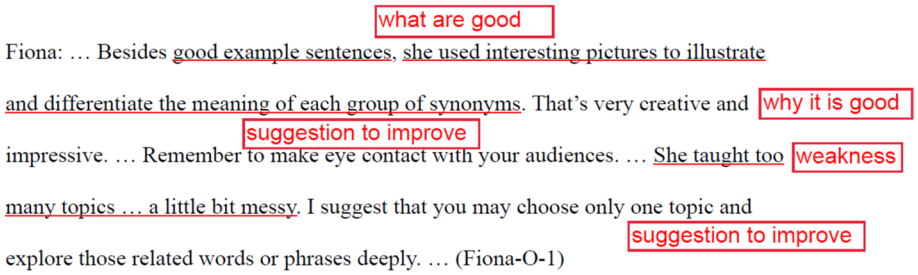

Fig. 3 An example of Fiona's verbal feedback

tasks, which was less likely to engage them in deep learning reflection and revision amid their self-regulated learning process (C-O-1; C-O-2; C-O-3).

Regarding the feedback content, both Connie and Fiona were knowledgeable and showed their appropriate beliefs in their interviews and reflective journals. They recognized that effective feedback should point out students' learning "strengths," "weaknesses," and "ways to achieve enhancement" (Connie-I-1; Fiona-I-1). They also believed that feedback should focus on task-specific content, thinking, and skills rather than "mechanical language corrections" (Connie-R-3; Fiona-I-1). Connie even emphasized that she was enthusiastic about providing feedback on "specific content and ideas" rather than "mechanical issues like grammar" because the formal issues were especially beneficial for English majors who were "advanced level English learners" (Connie-I-2). Connie and Fiona's professional knowledge and positive conceptions were definitely mirrored in their oral feedback practices. Figures 2 and 3 illustrated two examples of their verbal feedback analyzed by the authors.

However, in contrast to their task-specific oral feedback, Connie's and Fiona's written feedback was always generic and language-focused (Connie-Q; Fiona-Q), which was a mismatch with their solid understanding and positive conceptions of effective feedback content. Based on their documentary data, they mainly highlighted students' language issues, especially those grammatical mistakes in their written feedback. Sometimes, they also gave a general evaluation like "Good!" without detailed and content-focused explanations. Regarding their generic written feedback, both of them expressed their helplessness caused by heavy workloads and time constraints in their interviews, demonstrating their inadequate pragmatic strategies to provide feedback in large classes.

Unlike Connie and Fiona, Thomas emphasized student emotions and language accuracy in both oral and written feedback forms. He usually offered person-referenced comments to appreciate students' attributes (e.g., courage, confidence, and diligence) and pointed out their language mistakes in grammar or pronunciation. Thomas thought that this kind of feedback could "encourage students" and "guide them to focus more on their learning improvement rather than how exactly well or poorly they perform" (Thomas-I-4). Without sufficient FA knowledge, Thomas did not realize that this kind of extrinsic motivation and lower-level error corrections were inadequate for students' higher-level learning engagement and enhancement. His disregard for cognitive guidance through feedback aligned with his form-focused assessment criteria mentioned in Sect. "Assessment criteria", reflecting his lack of awareness and strategies to scaffold students' higher-level thinking for deep learning.

Discussion

This section discusses how Connie, Thomas, and Fiona knew about, thought about, and engaged students in learning through FA and the interrelationships between their FA knowledge, conceptions, and practices.

Levels of formative assessment literacy

Overall, the three instructors had rudimentary FA knowledge, such as its learning-supportive purpose and its process-oriented nature. They possessed positive conceptions to design and conduct FA-based activities to improve student learning. Their favorable perceptions indicated their compliance with the recently published Chinese FA-oriented government mandates (Chen & Teo, 2020). However, the top-down policy mandates and instructor passive compliance could not be directly transformed into professional knowledge and effective practice. Teachers lacked adequate conceptual knowledge and contextualized practical skills and possessed teacher-dominant and product-focused mindsets when they engaged and enhanced students' learning through FA activities. Their insufficient FA literacy hindered students' self-regulated learning from behavioral, emotional, and cognitive perspectives. These findings generally echoed previous studies (e.g., Lam, 2019; Levi & Inbar-Lourie, 2020; Li & Gu, 2023) yet mismatched Liu's (2024) study in which the single case was much more proficient in adjusting her FA strategies to gradually facilitate student learning regulation and development.

Based on the degrees and features of the three informants' FA literacy, we extended Samuelowicz and Bain's (2001, 2002) and Postareff et al.'s (2012) views regarding instructors' assessment and pedagogical beliefs as a continuum (see Sect. "Knowledge, conceptions, and practices") to the whole picture of FA literacy (see Fig. 4). Similar to their belief continuums, this innovative FA literacy continuum also considers the most teacher-centered (reproductive) and the most student-centered (constructional) as the two ends. Instead of adopting categories to divide the continuum like Samuelowicz and Bain (2001, 2002) and Postareff et al. (2012) did, we divided this adapted FA literacy continuum into four-phase levels (i.e., elementary, lower-intermediate, intermediate, and advanced) and integrated them with the three dimensions of student engagement. These innovative attempts were intended to present the degrees of the three instructors' FA literacy in student learning engagement and enhancement and assist the discussion of

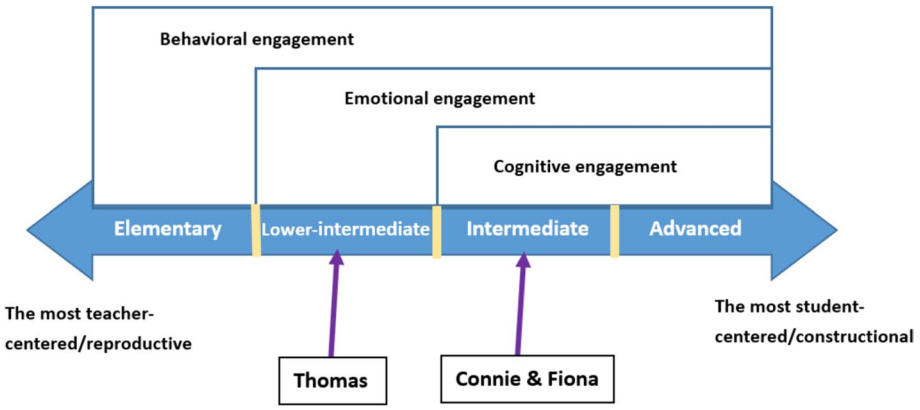

Fig. 4 FA literacy continuum

their FA literacy in specific FA processes (i.e., tasks design and operation, assessment criteria, and feedback).

Thomas's FA literacy was at the lower-intermediate level (see Fig. 4) because he mainly behaviorally and emotionally involved students in FA tasks from a more teacher-dominant perspective. He realized that students should actively participate in the lesson, so he integrated FA activities into his teaching, lowered their stakes, and frequently encouraged students to focus on their learning process. Yet, with a fixed teacher-dominant mindset and a lack of awareness and strategies to strengthen students' cognitive thinking, he restricted the activity types (e.g., closed-ended peer-assessment), appropriated student ideas, provided form-focused rather than content-focused assessment criteria, and offered person-referenced and language-oriented feedback. His lower level of FA literacy deprived students of opportunities to cognitively set effective task-specific learning goals, and further evaluate, manage, reflect on, and revise their self-regulated learning process (Andrade & Brookhart, 2020). Among the three informants, Thomas emphasized student emotional engagement most explicitly but at the lowest level. His frequent encouragement, use of motivational exemplars, and provision of person-referenced feedback were likely to engage his students emotionally in their learning and assessment process; however, these motivational scaffolds remained perfunctory, because they chiefly focused on extrinsic motivation. When designing and conducting FA activities, Thomas lacked knowledge, readiness, and practices to nurture students' intrinsic motivation through enhancing their learning autonomy and cognitive thinking, impeding students' higher-level affective involvement, such as a sense of achievement and responsibility initiated by active and deep regulatory learning (Panadero et al., 2019). Thomas's FA literacy was much lower than findings in previous research which indicated that despite insufficiency, instructors at least had awareness, skills, and attempts to cultivate student higher-level cognitive thinking and intrinsic motivation (e.g., Lam, 2019; Li & Gu, 2023). This misalignment can be explained by Thomas's limited learning about FA, his deep-rooted teacher-student hierarchy concept and his outcome-driven mindset shaped by Chinese traditional educational culture.

Both Connie's and Fiona's FA literacy were at the intermediate level (see Fig. 4), meaning that their capacities in students' learning engagement included all three dimensions. Yet, there was still much room for them to improve. Specifically, they empowered

students to self- and co-regulate their learning through open-ended student-centered activities, in which students were likely to have a sense of ownership, responsibility, and fulfillment when managing their learning independently and collaboratively. Thus, their emotional engagement became more intrinsic and active at a deeper level during their cognitive regulatory learning process (Panadero et al., 2019). With awareness of promoting students' cognitive thinking and learning, Connie and Fiona provided task-specific assessment criteria and feedback, raising students' motivation (self-efficacy), and facilitating their goal setting, learning management, reflections, and revisions (Andrade & Brookhart, 2016).

Despite awareness, willingness, and attempts to engage students at higher emotional and cognitive levels, neither Connie nor Fiona had sufficient FA literacy to involve students in their self-regulated learning. For example, they did not know how to coach students' peer-assessment skills (e.g., through cognitive scaffolds for guiding their peer feedback), so they restricted students' evaluative judgment and reflections during their collective learning process. When Connie and Fiona assigned tasks, their generic assessment criteria, albeit content-focused rather than form-focused, were not comprehensive enough to trigger students' higher-level task-specific cognitive thinking (Andrade & Brookhart, 2016). The absence of tangible scaffolders in Connie's case and the delayed exemplars in Fiona's case also impeded students' deep learning engagement due to students' lack of contextualized comprehension of the success criteria when they engaged in their self-planning stage (see Fig. 1). Furthermore, although Connie and Fiona had a clear understanding of how to provide task-specific feedback, they needed coping strategies to provide content-focused rather than language-oriented written feedback for large classes. The effectiveness of their descriptive verbal feedback was largely reduced owing to its timeliness. They possessed inadequate knowledge, inappropriate conceptions, and ineffective practices of providing process-oriented, pre-assessment feedback when students still had opportunities to review and revise their learning process. Their feedback literacy also supported McCallum and Milner's (2021) research, which suggested that both timeliness and task-specific content of feedback were crucial factors to impact feedback effectiveness on student learning engagement. In other words, the absence of either reason (i.e., timely yet generic feedback in Connie's case or delayed descriptive feedback in both Connie's and Fiona's cases) was likely to inhibit students' learning engagement and enhancement. Some previous studies have already proved that instructors had insufficient literacy in these FA processes (e.g., guiding peer- and self-assessment, sharing assessment criteria, and providing feedback), but they did not explore the details due to their quantitative research design (e.g., Arrafii & Sumarni, 2018; Xiang et al., 2023). However, this current research provided more specific, nuanced, and comprehensive evidence owing to its qualitative, in-depth nature by exploring all three literacy components, namely knowledge, conceptions, and practices.

Interrelationships between FA knowledge, conceptions, and practices

The three instructors' FA knowledge, conceptions, and practices were always closely interrelated. Their fundamental knowledge and enthusiasm triggered their FA applications as parts of their pedagogies. However, their FA practices usually could not meet their expectations owing to insufficient knowledge and strategies. For example, students

were unwilling to play the role of a critique in peer-assessment as in Connie's and Fiona's cases and performed unsatisfying group presentations as in Thomas's case. Without mastery of related knowledge and skills, Connie and Fiona found their assessment innovations futile and accordingly diminished their passion to further try out FA in a continuous manner (e.g., peer-assessment in Fiona's Intensive Reading course and group presentation in Thomas's case).

Besides the knowledge-conception-practice alignment, there were still mismatches between the informants' FA conceptions and practices, echoing previous research (e.g., Nimehchisalem & Abdalla, 2020; Widiastuti et al., 2020). Specifically in this research, although Connie, Thomas, and Fiona expressed opposition to the traditional teacher-dominant classrooms, they remained somewhat teacher-centered when designing and conducting FA activities (Kaymakamoglu, 2018). For instance, they all created assessment criteria by themselves or even judged students' performance with their subjective "implicit standards," without involving students in constructing, internalizing, and negotiating the success criteria. Connie and Thomas limited student-centered FA types (e.g., using closed-ended tasks) in practice. Thomas usually appropriated students' thoughts and rushed to finish his own PowerPoint slides. While they valued the process-oriented nature of FA, they remained product-focused in FA practices. A case in point is that Connie and Fiona still allocated marks as summative judgments that impacted student GPAs for every task. Connie overemphasized grades through frequent ranking, labeling, and comparisons. All three instructors did not provide effective diagnostic feedback to guide students' self-regulated learning process promptly. They still focused on students' error corrections rather than their process-oriented learning development when providing written feedback.

These conception-practice discrepancies may result from two reasons. First, instructors lacked theoretical knowledge and practical strategies about FA. They did not know what, why, and how issues concerning engaging students more actively and cognitively during FA processes (e.g., activity design, use of assessment criteria, and feedback provision). Lacking sufficient knowledge, they had misconceptions about what a real student-centered classroom entails. For instance, they misconceived that giving students more freedom through generic assessment criteria ensured "student-centeredness," failing to realize that a real student-centered and learning-oriented classroom required instructors' "critical presence" through systematic guidelines to foster students' knowledge construction and learning regulation (Ontai, 2021, p. 2). Compared with Connie and Fiona, Thomas's knowledge base was lower, without including students' higher-level cognitive involvement.

Second, the three informants still possessed deep-seated teacher-centered and product-focused mindsets shaped by the Chinese exam-oriented culture, which highlighted the teacher-student hierarchy and assessment outcomes (Gan & Lam, 2020). These fixed mindsets led to their dominance in assessment design and interpretation, their (especially Thomas's) emphasis on language accuracy rather than critical thinking and learning improvement, and their delayed feedback which restricted students' learning engagement and development. Compared with Thomas and Connie, Fiona seemed to be more passionate, reflective, student-centered, and open-minded. However, owing to her lack of experience and uncertain teaching assignments, her student-centered reflection, task design, and task redesign had no opportunities to be materialized in her FA practices.

Implications and conclusion

This study explored the three components (i.e., knowledge, conceptions, and practices) of FA literacy in the Chinese university context. The findings broadened the assessment literacy research by considering all three components simultaneously, specifying the process of FA (i.e., design, use, interpretation, and feedback), and adding valuable research insights into university English instructors' FA literacy. They enhanced the theoretical understanding of the nature and constituents of FA literacy, so as to reach a validated conceptual consensus on what teacher assessment literacy in FA applications is defined. This evidence-based definition, albeit small-scale, sheds new light on future research on FA literacy in China and beyond. The current study further extended and modified the previous belief continuums regarding assessment, teaching, and learning and proposed a renewed FA literacy continuum. This updated continuum combined instructors' FA literacy with the three perspectives of student engagement, facilitating researchers' holistic understanding of Chinese instructors' levels of literacy in achieving the ultimate purpose of FA, namely, by cultivating self-regulated learners. Practically, this research shows the directions to enhance instructors' FA literacy, namely, by equipping instructors with sufficient theoretical knowledge and practical skills and changing their teacher-centered and product-oriented mindsets regarding FA-based activity design and operation, assessment criteria, and feedback. These areas pave the way for the design of FA training courses for serving university English instructors. Service providers should offer human, technical, and financial support to establish systematic, collaborative, and reflective teacher training programs, such as creating a community of practice, conducting formative peer observations, running sharing sessions on good practices, encouraging micro-teaching of FA skills with reflection, and subscribing useful tools (e.g., apps and websites) to facilitate FA applications in large classes.

Notwithstanding its implications, this research had limitations in its data sources and samples. First, owing to the Chinese COVID-19 prevention policy, the classroom observations were restricted to nine times. Second, this study did not involve student data, which is important for examining instructors' FA applications (Hill & McNamara, 2012). To guard against teacher bias and enhance trustworthiness, observational data (e.g., classroom observation and student assignments) were collected alongside instructors' self-reported data. Third, the research sites were restricted to two universities in Central China with three participants, but these informants remain representative of typical English instructors in non-key universities. Further research may conduct more classroom observations, involve student data, and select more instructors from different parts of China to understand instructors' FA literacy more comprehensively.

Abbreviations

FA Formative assessment
SRL Self-regulated learning

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40468-024-00323-y.

Supplementary Material 1.

Acknowledgements
Not applicable

Authors' contributions
MK was responsible for conceptualization, data collection and analysis, and drafting/revising the manuscript. RL was responsible for conceptualization, data analysis, and editing the manuscript.

Funding
Not applicable.

Data availability
The data sets will be available upon request. Interesting parties can write to the first author to get access to the data.

Declarations

Ethics approval and consent to participate
Prior to the start of the study, ethical approval was granted by HKBU REC (Ref. no. SOSC-EDUC-2019-20_015). Informed consent was obtained from all individual participants included in the study.

Consent for publication
Consent to publish has been obtained from all participants to report their individual data.

Competing interests
Not applicable.

Declaration of Generative AI and AI-assisted technologies in the writing process
None.

Received: 24 June 2024 Accepted: 11 October 2024
Published online: 31 October 2024

References

Andrade, H. L., & Brookhart, S. M. (2016). The role of classroom assessment in supporting self-regulated learning. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 293–309). Springer. https://doi.org/10.1007/978-3-319-39211-0_17

Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment as the co-regulation of learning. *Assessment in Education: Principles, Policy & Practice,27*(4), 350–372. https://doi.org/10.1080/0969594X.2019.1571992

Arrafii, M. A., & Sumarni, B. (2018). Teachers' understanding of formative assessment. *Lingua Cultura,12*(1), 45–52. https://doi.org/10.21512/lc.v12i1.2113

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan,80*(2), 139–148.

Boyles, P. (2006). Assessment literacy. In M. H. Rosenbusch (Ed.), *New visions in action: National assessment summit papers* (pp. 11–15). Iowa State University.

Chan, C. K. Y., & Luk, L. Y. Y. (2022). A four-dimensional framework for teacher assessment literacy in holistic competencies. *Assessment and Evaluation in Higher Education,47*(5), 755–769. https://doi.org/10.1080/02602938.2021.1962806

Chen, J. (2024). Effects of learning analytics-based feedback on students' self-regulated learning and academic achievement in a blended EFL course. *System,124*, 103388. https://doi.org/10.1016/j.system.2024.103388

Chen, J., & Teo, T. (2020). Chinese school teachers' conceptions of high-stakes and low-stakes assessments: an invariance analysis. *Educational Studies,46*(4), 458–475. https://doi.org/10.1080/03055698.2019.1599823

Chen, Q., Kettle, M., Klenowski, V., & May, L. (2013). Interpretations of formative assessment in the teaching of English at two Chinese universities: a sociocultural perspective. *Assessment & Evaluation in Higher Education,38*(7), 831–846. https://doi.org/10.1080/02602938.2012.726963

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review,24*(2), 205–249. https://doi.org/10.1007/s10648-011-9191-6

Drisko, J. W. (2024). Transferability and generalization in qualitative research. *Research on Social Work Practice* (pp. 1–9). https://doi.org/10.1177/10497315241256560

Gan, L., & Lam, R. (2020). Understanding university English instructors' assessment training needs in the Chinese context. *Language Testing in Asia, 10*(1), 13. https://doi.org/10.1186/s40468-020-00113-2

Gu, Y., & Lam, R. (2023). Developing assessment literacy for classroom-based formative assessment. *Chinese Journal of Applied Linguistics, 46*(2), 155–161. https://doi.org/10.1515/CJAL-2023-0201

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

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan,89*(2), 140–145. https://doi.org/10.1177/003172170708900210

Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., & Timms, M. (2015). Investigating the dynamics of formative assessment: Relationships between teacher knowledge, assessment practice and learning. *Assessment in Education: Principles, Policy & Practice,22*(3), 344–367. https://doi.org/10.1080/0969594X.2015.1006521

Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing,29*(3), 395–420. https://doi.org/10.1177/0265532211428317

Huang, J., & Luo, S. Q. (2014). Formative assessment in L2 classroom in China: The current situation, predicament and future. *Indonesian Journal of Applied Linguistics,3*(2), 18–34. https://doi.org/10.17509/ijal.v3i2.266
Kaymakamoglu, S. E. (2018). Teachers' beliefs, perceived practice and actual classroom practice in relation to traditional (teacher-centered) and constructivist (learner-centered) teaching. *Journal of Education and Learning,7*(1), 29–37. https://doi.org/10.5539/jel.v7n1p29
Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System, 81*, 78–89. https://doi.org/10.1016/j.system.2019.01.006
Lam, R. (2022). Assessment training in the use of portfolios: Voices from writing teachers. In L. Hamp-Lyons & Y. Jin (Eds.), *Assessing the English language writing of Chinese learners of English* (pp. 235–252). Springer.
Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly,17*(2), 168–182. https://doi.org/10.1080/15434303.2019.1692347
Li, J., & Gu, Y. P. (2023). Developing classroom-based formative assessment literacy: an EFL teacher's journey. *Chinese Journal of Applied Linguistics,46*(2), 198–218. https://doi.org/10.1515/CJAL-2023-0204
Liu, X. (2024). Formative writing assessment: An EFL teacher's beliefs and practices. *Changing English,31*(2), 200–210. https://doi.org/10.1080/1358684X.2024.2328163
McCallum, S., & Milner, M. M. (2021). The effectiveness of formative assessment: student views and staff reflections. *Assessment & Evaluation in Higher Education,46*(1), 1–16. https://doi.org/10.1080/02602938.2020.1754761
Merriam, S., & Tisdell, E. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.
Ministry of Education. (2020). *College English curriculum requirements. Higher.* Higher Education Press.
Nimehchisalem, V., & Abdalla, M. M. (2020). Assessment beliefs and practices of Libyan secondary school English language teachers. *International Journal of Applied Linguistics & English Literature,9*(6), 99–109. https://doi.org/10.7575/aiac.ijalel.v.9n.6p.99
Ontai, G. (2021). Reflections on teacher's role in SRL. *Academia Letters* (pp. 1–4). https://doi.org/10.20935/AL248
Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using formative assessment to influence self- and co-regulated learning: The role of evaluative judgement. *European Journal of Psychology of Education,34*(3), 535–557. https://doi.org/10.1007/s10212-018-0407-8
Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Sage.
Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation,38*(3–4), 84–92. https://doi.org/10.1016/j.stueduc.2012.06.003
Raffe, C. P., & Loughland, T. (2021). "We're not data analysts": Teachers' perspectives on factors impacting their use of student assessment data. *Issues in Educational Research, 31*(1), 224–240. http://www.iier.org.au/iier31/raffe.pdf.
Richards, L. (2015). *Handling qualitative data* (3rd ed.). Sage.
Samuelowicz, K., & Bain, J. (2001). Revisiting academics' beliefs about teaching and learning. *Higher Education,41*(3), 299–325.
Samuelowicz, K., & Bain, J. (2002). Identifying academics' orientations to assessment practice. *Higher Education,43*(2), 173–201.
Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan,72*(7), 534–539.
Widiastuti, I. A. M. S., Mukminatien, N., Prayogo, J. A., & Irawati, E. (2020). Dissonances between teachers' beliefs and practices of formative assessment in EFL classes. *International Journal of Instruction,13*(1), 71–84. https://doi.org/10.29333/iji.2020.1315a
Xiang, X., Yum, S., & Lian, R. (2023). Effect of cognitive attitude and affective attitude on formative assessment practice: Intention as a mediator. *Social Behavior and Personality: An International Journal,51*(9), 1–16. https://doi.org/10.2224/sbp.12484
Xiao, Y., & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. *System,81*, 39–49. https://doi.org/10.1016/j.system.2019.01.004
Xie, C., & Teo, P. (2020). Institutional self-promotion: A comparative study of appraisal resources used by top- and second-tier universities in China and America. *Higher Education,80*(2), 353–371. https://doi.org/10.1007/s10734-019-00483-4
Xu, Y. T., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: a reconceptualization. *Teaching and Teacher Education,58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010
Yin, X., & Buck, G. A. (2019). Using a collaborative action research approach to negotiate an understanding of formative assessment in an era of accountability testing. *Teaching and Teacher Education,80*, 27–38. https://doi.org/10.1016/j.tate.2018.12.018
Yin, R. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.
Zimmerman, B. (1998). Academic studying and the development of personal skill: a self-regulatory perspective. *Educational Psychologist,33*(2–3), 73–86. https://doi.org/10.1080/00461520.1998.9653292

Publisher's Note