Check for updates

Teacher Classroom Questioning Practice and Assessment Literacy: Case Studies of Four English Language Teachers in Chinese Universities

Yan Jiang*

Basic Education College, Guangdong Teachers College of Foreign Language and Arts (GTCFLA), Guangzhou, China

This study examined classroom teacher questioning practice and explored how this process was affected by teacher assessment literacy and other mediating factors based on four case studies. Classroom observations were conducted to identify major patterns in teacher questioning practice, and semi-structured interviews were carried out to probe the participants' perception of classroom assessment, and how other factors, if there were any, along with teacher assessment literacy, impacted on teachers' questioning practices. The finding showed that different teacher assessment literacies resulted in variability in teacher questioning patterns. It also illustrated that teacher day-to-day classroom practice involved a complex interplay among factors at the teacher personal, institutional, and socio-cultural levels. This study highlights the legitimacy of teacher questioning as a sub-construct of teacher assessment literacy and calls for the establishment of a social and institutional culture aligning with assessment for learning principles.

Keywords: classroom questioning, teacher assessment literacy, formative assessment, mediating factors, student learning

INTRODUCTION

Classroom assessment has received a growing interest in recent decades due to its potential to enhance student learning. Classroom practices, however, heavily rely on teachers' knowledge and decision making, and thus, teacher assessment literacy becomes a fundamental factor that contributes to the effectiveness of classroom assessment (Leung, 2014). Despite its paramount importance, empirical studies on what teachers need to improve their assessment literacy in practice are still limited.

Teacher questioning has long been employed as an instructional technique. With the advances in educational assessment, it has been explored as a powerful assessment tool, which enables teachers to gather accurate information about learners and use the information to make better pedagogical decisions (Black et al., 2003). Regardless of its learning potential, teacher questioning, with specific reference to its use as an assessment strategy, remains relatively underexplored.

This study examines classroom questioning practice and explores potential factors impacting on teacher assessment literacy in practice. The study is part of a project investigating classroom assessment in Chinese educational context. This paper report shows teachers conducted questioning practices regarding what classroom questions were posed, how student responses were

collected, and what teacher feedback was provided. It also discusses how the whole questioning process was shaped by teacher assessment literacy and other mediating factors. It is hoped this classroom-based research contributes to our understanding of assessment knowledge and skills teachers may need to conduct effective questioning. It demonstrates the dynamic interaction among various components of teacher assessment literacy in practice and calls for joint efforts from government, university administrators, and students to build a culture aligning with assessment for learning principles.

LITERATURE REVIEW

Teacher Assessment Literacy

Students' achievement, as research evidence demonstrated, is positively correlated to their teachers' competence to design or choose high-quality classroom assessments and to use them productively to support learning (Stiggins, 2010). Teacher assessment literacy has thus become an emerging issue in education. The term assessment literacy (AL) traditionally refers to the possession of the knowledge about educational assessment and related skills to apply that knowledge to various measures of student achievement (Stiggins, 1991).

Worldwide, there has been an agreement to enhance teacher AL, as well as that of stakeholders (Taylor, 2009), in part, due to the concern that teachers may lack sufficient training in what educational assessment entails (Stiggins, 1991; Taylor, 2009). Yet, there is no common definition of AL (Fulcher, 2012), and various efforts have been made to describe what AL might comprise.

The earliest and a documentable contribution is the *Standards for Teacher Competence in Educational Assessment of Students* (hereafter the *Standards*) [(American Federation of Teachers, National Council on Measurement in Education, and National Education Association (AFT, NCME, & NEA), 1990)]. The *Standards* prescribe seven competences teachers should be skilled, applying to both internal classroom assessment and external large-scale assessment, i.e., choosing and developing assessment methods for the classroom, administering and scoring tests, interpreting assessment results and using it to aid instructional decisions, developing grading procedures, communicating results to stakeholders, and being aware of inappropriate and unethical uses of tests. Other contributions include *the 11 Principles*, an updated version of *Standards* (Brookhart, 2011), and *the Fundamental Assessment Principles*, a basic guide for assessment training and professional development of teachers and school administrators (McMillan, 2000). The former reflects the recent developments in formative assessment and teacher needs in the accountability context (Brookhart, 2011), and the latter takes a social constructivist perspective into account and acknowledges the role of classroom assessment to promote learning (Inbar-Lourie, 2008). Despite the differences, all these standards or principles seem to share the goal of establishing a comprehensive knowledge base for AL.

Some researchers (e.g., Stiggins, 1995; Boyles, 2005; Hoyt, 2005) choose to describe the characteristics of those considered to be assessment literate. Stiggins (1995) states that assessment-literate educators, regardless of whether they are teachers,

principals, curriculum directors, are clear about what they are assessing, why they are assessing, how best to assess different achievement targets, how to generate sound samples of student performance, what problems may arise from assessment procedure, and how to counter them with specific strategies that lead to sound assessment. Stiggins (2010) further points out that many teachers are engaged in assessment-related activities without sufficient training, and thus, "assessment illiteracy abounds" (Stiggins, 2010, p. 233).

In response to the dichotomy of literacy vs. illiteracy, other researchers prefer to view AL as a continuum. Pill and Harding (2013), for example, propose a frame including five levels of AL (i.e., illiteracy, nominal literacy, functional literacy, procedural and conceptual literacy, and multidimensional literacy) and argue strongly that different stakeholders in a wider assessment community are expected to meet different levels of AL. For non-practitioners like policy makers, a "functional level" might be sufficient for engaging with and drawing sound conclusion from tests. In contrast, practitioners like classroom teachers should reach advanced literacy at "procedural level" or even "multidimensional level," meaning "understanding central concepts of the field, and using knowledge in practice," and "knowledge extending beyond ordinary concepts including philosophical, historical, and social dimensions of assessment" (Pill and Harding, 2013, p. 383).

More recently, researchers believe that AL should be viewed as a social practice and better understood in specific context. A new framework of Teacher Assessment Literacy in Practice (TALiP) has been proposed, which includes six key components: the knowledge base, teacher conception of assessment, institutional and socio-cultural contexts, TALiP the core concept of the framework, teacher learning, and teacher identity (Xu and Brown, 2016). In this view, AL is neither a disembodied set of knowledge or skills, nor a static product. Instead, it is a dynamic and evolving process where factors at different levels interact with each other to impact on teacher assessment practices.

At the personal level, teachers' knowledge about assessment and their conception of assessment impact on the assessment practice. Teachers need knowledge and skills to implement classroom assessment effectively (Assessment Reform Group, 2002), and different levels of assessment expertise may result in different assessment practice (Leung, 2004). In Xu and Brown's (2016) framework, teachers are supposed to possess seven areas of knowledge: disciplinary knowledge and pedagogical content knowledge; knowledge of assessment purposes, content and methods; knowledge of grading; knowledge of feedback; knowledge of assessment interpretation and communication; knowledge of student involvement in assessment; and knowledge of assessment ethics. It is noted that the knowledge specified above is merely the key theoretical principles rather than the ready-made solutions to the problems happening in complex classroom assessment practice. What is more, teachers' mastery of this body of knowledge does not guarantee a smooth transfer of the knowledge to the classroom practice (Xu and Brown, 2016).

Teachers' conception of assessment represents their views of teaching, learning, and epistemological beliefs (Brown, 2008). Different teacher beliefs can result in a good deal

of variability in assessment (Leung, 2004). Roughly speaking, the more constructivism-oriented the teachers are, the more likely it is their assessment practices place the learner at the center of teaching; conversely, teachers who hold strong behaviorist views are likely to emphasize the transmission of content and ignore learner needs (Carless, 2011). In a similar vein, teachers' conception of assessment plays a decisive role in their interpretation of assessment knowledge and actual implementation of assessment practice (Brown, 2008). Cognitively, teachers are inclined to accept new knowledge and methods that are congruent with their conception of assessment, while resisting those that are not. Affectively, teachers' positive or negative emotions about assessment may cause effective or less effective learning about assessment knowledge, and successful or less successful implementation of new assessment policies.

At the institutional level, the immediate context within which teachers work exert an influence on TALiP. Some teachers may support assessment for learning principles but are constrained by institutional values (Yu, 2015). In a school culture where great emphasis is placed on curriculum or content coverage, teachers often feel pressured to complete the prescribed syllabus and are therefore less likely to conduct formative assessment activities because of the time constraint (Carless, 2011). Likewise, the stakeholder's needs, such as students' interests, are what teachers must accommodate in practice (Xu and Liu, 2009). In face of students with limited language competence, English as a Foreign/Second Language (EFL/ESL) teachers may adjust their questioning pattern by raising less demanding recall questions to engage learners in classroom interaction (Shomoossi, 2004). Student motivation, too, colors TALiP. For instance, formative assessment assumes students want to learn and requires them to engage in such activities as metacognition, self-evaluation, and peer assessment. In reality, not all students are sufficiently motivated to participate in these activities or to take greater ownership of their own learning process (Carless, 2011). As a result, teachers cannot conduct the assessment activities as they wish in actual practice.

At the wider social–cultural level, national policies, social norms, and behavioral expectations all exert influence on how assessment is conducted (Pryor and Crossouard, 2008; Gu, 2014). Although it is not politically correct to generalize about any culture, patterns do emerge as results of empirical research (Pierson, 1996). For example, texts and textbooks played a significant role in Chinese tradition (Carless, 2011). In contemporary classrooms, teachers put particular emphasis on covering textbook contents (Adamson and Morris, 1998) and favor text-oriented recall questions (Tan, 2007). For another instance, collectivism is highly valued in Chinese society (Hofstede, 2001), and within the classroom, this would mean a preference for a whole-class teaching and a discouragement of individual initiatives. In Hong Kong educational system, the class is reported to proceed together toward common goals, whereas individual differences are hardly taken into account (Cheng, 1997). This seems incompatible with assessment for learning principles that teachers should make effort to meet individual learner's need and enable all learners to achieve their best (Assessment Reform Group, 2002).

In short, TALiP is a complex entity with the interplay of teachers' assessment knowledge, their conceptions of assessment, and macro- and micro-contextual factors. Although research instruments have been developed to measure teacher AL levels, e.g., *Teacher Assessment Literacy Questionnaire* (Plake et al., 1993), *Assessment Literacy Inventory* (Campbell et al., 2002), and *Classroom Assessment Literacy Inventory* (Mertler, 2004), studies directly observing teacher AL in their classroom practice are sparse. Further research is warranted to examine how AL is enacted when different factors interrelate with each other to impact on classroom assessment practices.

Teacher Questioning

Questioning has long been used as a teaching technique to motivate student interest, facilitate teacher instruction, and evaluate learning achievement (Sanders, 1966). Although the dialogic process of classroom interaction is complex, the basic iteration of questioning involves three stages: the teacher initiates a question (Initiation), the student responds to that question (Response), and the teacher provides feedback (Feedback) to or makes evaluation (Evaluation) of the student response (Mehan, 1979). The IRF/E process has been extensively investigated in the field of classroom interaction, especially second language teacher talk.

Regarding teacher question types, diverse criteria have been proposed. For example, Bloom's (1956) taxonomy places questions in an ascending order of cognitive demand, including knowledge, comprehension, application, analysis, synthesis, and evaluation. Long and Sato's (1983) categories are based on whether the information elicited is known by the questioner: display questions request known, while referential ones request unknown, information. Richards and Lockhart (1994) divide teacher questions into procedural and convergent/divergent types. Procedural questions have to do with classroom routines and management. Convergent and divergent questions engage students in the content of learning; the former eliciting similar responses and focusing on low cognitive recall of previously presented information, the latter expecting diverse responses and requiring higher-level thinking (Richards and Lockhart, 1994).

Research evidences show that convergent and divergent questions, or low and high cognitive questions, have different educational merits. Convergent ones could test factual knowledge, arouse curiosity, and make students aware of the data they will require to answer higher-order thinking questions (Musumeci, 1996). They could better involve language beginners in classroom interaction due to their less intimidating nature (Shomoossi, 2004). Nevertheless, arbitrary facts solicited by recall questions are quickly forgotten (Sanders, 1966), and teachers' emphasis on this type of questions risks engaging learners in rote learning and discouraging their development of critical thinking skills (Tan, 2007). Divergent questions, in contrast, are more likely to simulate the development of students' thinking and cultivate such skills as synthesis and evaluation because they require students not only to recall facts but also manipulate information mentally to create an answer (Gall, 1970). Although posing higher cognitive questions does not promise the occurrence of higher-order thinking,

it is believed that when students are actively engaged in thinking and reflection, learning opportunities are maximized (Black et al., 2003). In second language classrooms, an extra benefit is that these questions effectively elicit longer and more syntactically complex responses from students (Wintergerst, 1994). Previous studies, nevertheless, consistently showed that convergent questions far outnumbered divergent questions in both content (Brown and Ngan, 2010) and language classrooms (David, 2007), with the implication that teachers should strike a balance between different types of questions to achieve various educational goals.

Classroom responses are assumed to be important evidence of students' current state of learning since they represent the externalization of individual thinking, coded in language (Leung and Mohan, 2004). Research reports from classrooms, nonetheless, suggest that the commonly found responses, i.e., student individual answer, student choral answer, student no answer, and teacher self-answer, do not always well represent students' thinking. First, when teachers pose convergent questions that are usually associated with a single correct answer, an individual student may provide various "trial" responses attempting to meet the teachers' expectations rather than demonstrate their own understanding (Mehan, 1979). Second, when questions are replied to by students as a whole class, teachers may collect very limited information about individual student, yet this pattern has been identified worldwide (e.g., Smith et al., 2004) and most often reported in Asian countries (e.g., Shomoossi, 2004; Tan, 2007). Third, student reticence is a problem encountered by many language teachers (Jackson, 2002; Peng, 2012). If questions repeatedly go unanswered, how much can one really understand about student learning, let alone what remedial measures could be taken? The fourth problem, and related to the third, is teachers' tendency to self-answer. Excessive teacher self-answering deprives students of the opportunity to exhibit their thoughts and may result in heavy teacher-dependent learners (Hu et al., 2004).

Teacher feedback is critical to the development of student learning, though feedback that moves learning forward is hard to engineer effectively. In a model differentiating teacher feedback at four levels (i.e., self level, task level, process level, and self-regulation level), Hattie and Timperley (2007) maintain that feedback at process and self-regulation levels promote learning more effectively. Studies on teachers' actual practices, though, suggest that classroom feedback is commonly operated at the less effective self and task levels. Other problems reported include teachers rejecting unexpected answers and failing to create further learning opportunities for students (Black and Wiliam, 1998), teachers merely indicating whether students' answers are correct or not without follow-up interventions to enhance learning (Leung and Mohan, 2004).

Recent research suggests questioning could be one of the key formative assessment strategies to promote learning when teachers use questions to elicit student understanding, interpret the information gathered, and act on student responses to achieve learning goals (Black et al., 2003; Jiang, 2014). Formative Assessment (FA) is defined as a set of procedures where "evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited" (Black and Wiliam, 2009, p. 9).

Although researchers do agree that the priority of FA is to improve student learning, the term FA itself is open to different interpretations, and variations of FA are perceived essential for different teaching settings. For example, in East Asian context where teachers are greatly revered, the idea of adopting a more student-centered approach to FA may not be easily accepted (Leong et al., 2018). Taking these culturally specific factors into consideration, Carless (2011) suggests that FA could be seen as a continuum ranging from "restricted" to "extended" forms. "Restricted" FA is a relatively pragmatic version and is largely teacher directed, while "extended" FA is a more ambitious version involving more self-direction by the student.

When questioning is used as an FA strategy, it follows FA procedure of eliciting, interpreting, and using the evidence, but it must go beyond its basic IRF/E process to represent learning intentions (Jiang, 2014). First, the questions teachers raised should contribute to student cognitive development (Black et al., 2003). Second, the responses teachers collected should reflect students' understanding and facilitate teachers' subsequent decision making (Hodgen and Webb, 2008). Third, the evaluation teachers made should move the learners toward the learning goals (Hill and McNamara, 2012). In other words, if teachers merely raise questions to develop interest, for example, without collecting useful information about students (eliciting evidence), or without acting on students' responses to facilitate learning (interpreting and using evidence), questioning can only be labeled as an instructional tool, rather than an FA tool. In brief, questioning serves the learning purpose at each stage of initiation, response, and evaluation when it functions as an FA strategy.

In the research field of assessment, Black et al. (2003) have demonstrated how to use questioning as an FA tool. Before initiating questions, the teachers in their study were encouraged to spend much time framing quality questions to engage students in active thinking; when collecting responses, the teachers deliberately extended wait time or conducted group discussions to allow every student to think and articulate his/her ideas; during evaluation, students shouldered the responsibility for informing the teachers of their learning strengths and weaknesses, and incorrect answers were used to create further learning opportunities. Evidences from other research corroborate this finding that quality questioning makes both teaching and learning more effective (e.g., Ruiz-Primo and Furtak, 2006, 2007; Jiang, 2014), though these studies are limited in number. Moreover, direct investigations into assessment knowledge and skills teachers required to conduct effective questioning are generally lacking, and further research is needed to explore specific teacher assessment literacy in classroom questioning.

Put simply, though the paramount importance of teacher AL has been highlighted in literature, empirical studies examining how teachers enact AL in practice are still limited. Existing studies on teacher questioning have heavily investigated teacher practices in Initiation–Response–Evaluation stages, questioning itself, with specific reference to its use as an assessment tool, has

remained relatively underexplored. This study bridges the two strands by examining teacher questioning from assessment angle and exploring potential factors impacting on teacher assessment literacy in questioning practice.

METHODS

The study aims to address the following research questions:

RQ1: How do teachers conduct their classroom questioning regarding what questions are posed, what responses are collected, and what feedback is provided?

RQ2: What factors influence teacher assessment literacy in questioning practice?

Participants

The study was conducted in two tertiary institutions in China, one private foreign language college (UA) and the other state-supported foreign language university (UB). The participants were four EFL teachers and 31 first-year undergraduates. Maximum variation sampling helped to select teacher participants from different educational backgrounds (Merriam, 1998). Fiona and Lily (pseudonym) from UA were female teachers with 20 and 12 years of English teaching experience, whereas Max and Rick from UB were male teachers with 18 and 5 years of teaching experience. Fiona, Lily, and Rick held Master's degrees as their highest qualification, while Max had a doctorate degree in English language education. Student participants were from different academic levels, and identification codes were assigned accordingly. C1, C2, C3, and C4 represent the classes of Fiona, Lily, Max, and Rick, and H, A, and L suggest the students' academic levels. For example, C1–H1 indicates a high achieving student from Fiona's class, whereas C4–L1 refers to a low achiever from Rick's class.

Data Collection

Classroom observations and semi-structured interviews were adopted to collect data. Each participant teacher was observed for eight sessions on a consecutive basis, which produced a total length of 48 h of observational data. The observations were audio recorded, and field notes were also taken following an observational protocol.

The four teachers and 12 students were individually interviewed, and the rest of the students were invited to the focus group interviews, yielding 37 h of interview data. The schedules for teacher (see **Appendix**) and student interviews were designed based on the existing literature and developed from the research questions. The interviews were carried out in Mandarin, the mother tongue of the participants to facilitate natural communication. Written and informed consent for publication of the participant verbatim quotes was obtained.

Data Analysis

Observational data were analyzed quantitatively to identify major patterns in teacher questioning practice. This included an initial analysis that consisted of going through the recordings, sorting out episodes involving question–answer interactions, transcribing the interactions verbatim, and the

subsequent coding of question types, response types, and teacher feedback types. Teacher question and feedback types were coded based on Richards and Lockhart's (1994) typology and adapted Hattie and Timperley's (2007) classification system. Classroom responses were coded according to a self-generated categorization since existing classifications in the literature did not seem to accommodate the responses in the current study. Observational data were revisited at regular intervals to maintain coding consistency. Measures to ensure trustworthiness of the categories, including checking by an "external" coder, in this instance a PhD student, were also undertaken. The inter-rater reliability between the two researchers was 87%, and the small number of initial differences in coding was resolved through discussion.

Interview data were analyzed following an inductive coding procedure adapted from the qualitative analysis protocols established by Miles and Huberman (1994). The analysis was an ongoing iterative process including reading and coding, interpreting and categorizing, making inferences, and developing models. The following example serves as an illustration. During the reading of interview transcripts, both teacher and student participants repeatedly mentioned that students' English proficiency was limited, which inhibited students from responding to high cognitive questions. The code "limited language competence" was therefore assigned to the relevant texts. The code was then grouped under the category of "student factor" with other inter-related concepts such as "a lack of motivation." Major themes emerged from recurring regularities and factors affecting classroom questioning practice were identified under three categories: "student factors," "teacher factors," and "contextual factors." These categories were further linked to each other in various relationships to build a logical chain of evidence (Merriam, 1998), i.e., factors influencing teacher assessment literacy in questioning practices were placed at three layers: the first layer referred to teacher factors at the teacher personal level; the second layer to student factors and contextual factors at the institutional level; and the third layer to contextual factors at the wider socio-cultural level. To enhance the credibility of the study, the researcher's interpretations were continually checked by the participants during the interviews, and key findings were shared with participants for verification.

FINDING

Findings presented in this paper are drawn from a wider research project investigating classroom assessment practices in two Chinese universities. Detailed findings pertinent to questioning practice have been reported in Jiang (2014), which highlights good practice of one experienced teacher and illustrates the learning potential of effective questioning. The focus of this study is to examine teacher questioning practice and explore how various factors may interact with each other in shaping teacher assessment literacy in questioning practice.

Teacher Questioning Practice

Table 1 below shows the average number of questions teachers posed in each session, **Table 2** presents how these questions

TABLE 1 | Types of questions.

Teacher	Question/session		Procedural		Convergent		Divergent	
	n.	%	n.	%	n.	%	n.	%
Fiona	67	100	6	9	59	88	2	3
Lily	137	100	20	15	108	79	9	6
Max	66	100	6	9	44	67	16	24
Rick	73	100	2	3	66	90	5	7
Average	86	100	9	11	69	80	8	9

The data was drawn from Jiang (2014).

TABLE 2 | Types of responses.

Teacher	Question/session		Choral answer		Individual answer		Teacher answer		No answer	
	n.	%	n.	%	n.	%	n.	%	n.	%
Fiona	67	100	35	52	26	39	3	4.5	3	4.5
Lily	137	100	61	45	64	47	7	5	5	3
Max	66	100	8	12	39	59	8	12	11	17
Rick	73	100	37	51	25	34	9	12	2	3
Average	86	100	36	42	39	45	6	7	5	6

The data was drawn from Jiang (2014).

were answered, and **Table 3** summarizes at which levels teacher feedback was operated. The general trends are as follows: (1) the majority of the teacher questions were convergent (80%), followed by procedural (11%) and finally, divergent questions (9%); (2) a large proportion of these questions were responded to with student individual (45%) and choral answers (42%), and a small proportion with teacher answers (7%) and no answer (6%), respectively; (3) upon receiving student responses, teachers provided their feedback mainly at task level (92%), with a small proportion at self and task combined level (5%), process level (2%), and self-regulation level (1%).

Questions

Teacher questions were coded according to Richards and Lockhart's (1994) classification, and definitions and examples are provided below in **Table 4**.

A further analysis of these questions revealed two notable trends. First, convergent questions dominated in Fiona, Lily, and Rick's classes (representing 88, 79, and 90%, respectively) and were largely text oriented, meaning requiring students to recall basic information about the text. These questions, according to the students, "consolidated what we previously learned" (C1–A2), "laid foundation for future learning" (C4–A3) and facilitated "a sense of accomplishment" (C1–A1) when a correct answer was managed to offer. Nevertheless, they were "boring in nature, and associated with fixed single answers which stifled students' opinions" (C2–A3) and encouraged "learning everything by rote" (C1–L1).

Second, the most divergent (24%) and least convergent questions (67%) were raised by Max, whose convergent questions were both text oriented (58%) and world-knowledge related

(42%) (refer to **Table 5** for examples). Students seemed to appreciate the value of both world-knowledge convergent questions and divergent questions. As they commented, "world-knowledge questions provide a supplement to what we learn in the course" (C3–A1), especially when "most of us merely focus on the textbook and know very little about the world" (C3–A3). Divergent questions, due to their open nature, ensured psychological safety as "I can say whatever I like without feeling too much pressure in front of my peers" (C3–L1), and they also "provide us with opportunities to articulate our opinions, which helps teachers to know what we are thinking" (C3–A1).

Responses

Classroom responses were categorized based on a self-developed coding scheme. The definitions and examples are listed in **Table 6** for illustration.

At response stage, the practices of Fiona, Lily, and Rick shared similarities that student choral and individual answers represented the two largest portions of responses, whereas in Max's class, individual answers and no answer were the two most common responses. In what follows, findings related to the three teachers' practices and that of Max's are presented consecutively.

Fiona, Lily, and Rick seemed to favor choral answers (representing 52, 45, and 51%, respectively), and the following classroom scenario is representative.

(Lily—Session 2) Lily commenced the lesson with a review of a listening material entitled "Britain's solo sailor."
Lily (at 3:33): "Today we will go on with the topic traveling and we will still talk about Ellen. You remember her, right? What is she?"

TABLE 3 | Types of feedback.

Teacher	Question/session		FT		FST		FP		FR	
	n.	%	n.	%	n.	%	n.	%	n.	%
Fiona	58	100	48	83	9	15	1	2	0	0
Lily	111	100	104	94	7	6	0	0	0	0
Max	52	100	48	92	0	0	2	4	2	4
Rick	62	100	61	98	0	0	1	2	0	0
Average	71	100	65.5	92	4	5	1	2	0.5	1

FT, feedback about task; FST, feedback about self and task combined; FP, feedback about process of the task; FR, feedback about self-regulation.

TABLE 4 | Question types and examples.

Type	Definition	Example
Procedural question	It is associated with classroom procedures and routines, and classroom management (Richards and Lockhart, 1994).	"Have you got the answers now?" (Lily—Session 1)
Convergent question	It has to do with content of learning. It encourages similar and short responses and focuses on the recall of previously presented information (Richards and Lockhart, 1994).	"Cathy wrote the letter for two reasons, the first is directly to show what?" (Fiona—Session 2)
Divergent question	It has to do with content of learning. It encourages diverse responses and requires higher-level thinking (Richards and Lockhart, 1994).	"What are the advantages of smoking?" (Max—Session 2)

Ss: "A sailor."

Lily: "Yeah, she is a sailor. Can you spell sailor?"

Ss: "S-A-I-L-O-R."

Lily: "Yes. OK, sail, sailor. (Lily wrote "sailor" on the blackboard) Where is she from?"

Ss: "Britain."

Lily: "Yeah, she is from Britain, OK? What the boat she used is called?"

Ss: (student reticence).

Lily: "She used a kind of very special ship, right? A very special boat."

Ss: "Yacht."

Lily: "Yacht, right? Yes. Can you spell yacht?"

S1: (Many students shouted the answer at the same time, and S1's voice was the most audible): "Y-O-C-H-I."

Lily: (wrote "yochi" on the blackboard) "No, I don't think so."

S2 (voluntarily provided an answer): "Y-A-C-H-T."

Lily: "Yes, that is yacht. And do you still remember the picture I gave you? We know a yacht is composed of three parts, right? Remember? The first part is a?"

Ss: "Mast."

Lily: "Yes, a mast. So, what is a mast?"

Ss: "桅杆." (The Chinese counterpart of "mast")

Lily: "Yes, it is used to support the yacht. Can you spell?"

Ss: "M-A-S-T."

Lily: "OK. What's another part?"

Ss: "Sail."

Lily: "Sail. OK. The last part is?"

Ss: "Hull." (5:33).

Student choral responses above (see underlined portions) shared some characteristics. First, all were elicited by convergent questions. Second, all were short utterances (usually limited to a single word or the spelling of that word) relating to the basic information in the text. In other words, student contribution was limited to factual recall, seemingly curtailing students' thinking. Third, the interval between the responses was notably brief: during the 2-min interaction (3:33–5:33), 10 questions were put forward and answered, meaning that each question–answer session was completed within 12 s on average. This, on the one hand, represented smooth classroom interaction; it indicated, on the other hand, that Lily neither expected thoughtful answers from her students, nor intended to test learning at a more challenging level.

Though the teachers welcomed choral answers, the students preferred individual answers instead.

C1–H1: Choral answers could not reflect individual student's problems, and teachers thus ignored them. Individual answers are more effective to check student mastery of the knowledge.

C4–L2: I don't like choral answers. When Rick checks answers as a whole class, he goes them through very quickly, it is hard for me to follow, and I always have some problems unsolved, which made me a bit frustrated.

C4–A1: Individual answers are certainly better. When we shout the answer as a whole class, very often is the case that we follow those strong students without our own thinking. Individual responses allow more students to think independently.

In brief, collecting choral responses could be time saving in classroom teaching, nevertheless, choral responses were less likely to reveal valuable information about individual students or for teachers to identify learning difficulties and make informed

Convergent	Definition	Example
Text oriented	It generally requires factual recall, focuses on the content, structure and language of the texts, and the answers could be found in the text (Jiang, 2014).	"How many smokers are there in high income countries?" (Max—Session 2)
World-knowledge related	It relates to the text in some way, but demands world knowledge not specified in the text (Jiang, 2014).	"What is the largest tobacco company in China?" (Max—Session 3)

Type	Definition	Example
Student individual answer	It refers to the answer offered by individual student to teacher questions (Jiang, 2014).	Fiona: "Suppose…what could you do to make your friend feel better?" S1: "My desk mate Anita…." (Fiona—Session 3)
Student choral answer	It refers to the answer provided by students as a group/whole class to teacher questions (Jiang, 2014).	Rick: "To do this dish, first of all, what should we do?" Ss: "Cut the chicken into pieces." (Rick—Session 3)
Teacher answer	It refers to the answer offered by the teacher himself/herself (Jiang, 2014).	Lily: "Do you know where Madrid is?" Lily: "You don't know? 马德里 (Chinese counterpart for Madrid) is in Spain." (Lily—Session 3)
No answer	It refers to student reticence in response to teacher questions (Jiang, 2014).	Lily: "What the boat she used is called?" Ss: [student reticence] (Lily—Session 2)

S1, an individual student; Ss, students as a group/whole class.

decision to serve learning. Individual responses, on the other hand, required every student to provide an answer and, thus, might encourage independent thinking. Additionally, individual answers had the potential of foregrounding learner strengths and difficulties and, therefore, facilitated the teacher's next step in teaching.

In Max's class, students more often than not responded to his questions with silence. A closer look at these unanswered questions revealed that approximately two thirds were convergent. Since this type of questions is less cognitively challenging and therefore relatively easier to be answered, why did students repeatedly fail to offer answers? To expose the reasons, classroom data was re-examined.

(Max—Session 4) Some acronyms were listed in the course-book, and Max asked his students, "What is acronym? What is abbreviation? And what is initialism?" The whole class was quiet. Waiting for a few seconds, he continued, "To satisfy your curiosity, find subtle differences between them. And you can tell me, rather than wait for me to tell you." Max left these questions as homework and moved on to a reading task.

In the above episode, the question "what is acronym, abbreviation, and initialism" demanded little mental processing, meaning only if they retrieved relevant information from the memory, students should have been able to answer it. Nevertheless, a scrutinization of the course-book showed that the required information was not specified; nor had it been taught previously as suggested in student interviews. Accordingly, a lack of knowledge was probably the reason for the silence.

When Max was asked about this, he stated that he had largely anticipated students' lack of responses, and the reason that he put forward the questions was to inform the students of his expectations and to prompt them to search for the answer independently. In other words, Max deliberately put forward certain questions to elicit no answers, to expose students' knowledge deficiencies in some areas, and to give them directions for further effort.

When invited to comment on student individual responses, the most common responses in his class, Max stated:

Max: When students were less responsive, I had to nominate one of them. …For those open questions, I usually asked them to discuss as a group and then invite group representative to report. Everyone may approach the topic from one perspective, and group discussions give students different ideas and make them think more thoroughly.

Students seemed to favor this question–discussion–response sequence, as they remarked:

C3–H2: Group discussions give us more time to formulate the answer. We need time to organize our ideas and express them in a logical way.

C3–A2: I'm reluctant to offer an answer in class for fear that my answer is stupid. Group discussions help test my ideas and make me less anxious.

As illustrated above, individual answers elicited in Max's practice had the potential of reflecting student thinking for the following reasons. First, when more time was provided to process the question and group discussions were offered to trigger each other's thinking, student responses were more likely to represent deep rather than superficial thinking. Second, when a micro-community was established where ideas could be shared without fear of derision, students became more willing to

voice their real thoughts. Third, from assessment perspective, when group ideas were collected after discussion, a better-rounded picture of both individual and group learners' thinking was obtained.

Evaluation

Analysis at this stage excluded procedural questions and teacher self-answers, meaning it looked solely at how teachers reacted to student choral, individual, and no answers. The results showed that all the teachers tended to provide feedback on tasks (on average 92%), informing students how well a task was accomplished. Nevertheless, variances were also noted.

Fiona, Lily, and Rick preferred positive feedback, Fiona's comments being encouragement based, and Lily and Rick's being more specific. One classroom episode is presented for illustration.

(Fiona—Session 3) Students were required to apply some reporting verbs they had learned, such as "warn," "encourage," "suggest," "recommend," in an imaginary situation. Fiona said "Suppose one of your best friends is desperately depressed. What could you do to make him/her feel better?" She made students discuss for around six minutes and then invited answers.
Fiona: Ellis, would you please share your ideas?
S1: My desk mate Anita complained that she was under the stress these days, so I encourage her to go to Karaoke to sing out loudly and release her stress. And I advise, she told me she want to change her life style, but I persuade her not to go, not to change her life style because her style is very OK. And I want her don't do the thing she will regret in the future.
Fiona: OK, that's all? Yeah, that's right. OK, Joanna, do you have something to share with us?
S2: My friend complaint that her life is meaningless, and she don't know what the sense of life. I warned her don't always be in the passive condition. And I persuade, I encourage her go traveling, so she will see the life is meaningful.
Fiona: That's OK. Thank you and sit down please. Jane, do you have some ideas?
S3: My partner Daphne complained that she has failed an interview, so I suggested her to do something she likes to release, to relax, because I think she is just nervous and stressful. And I recommended her doing some sports to release out the stress.
Fiona: Is that all? Suggest doing something. Right, good.

In this scenario, the purpose of the task was to practice the reporting words. In responding to her students, Fiona provided comments like "that's right" "that's OK" "Good" (see the underlined portions). Despite various grammatical errors in students' utterance, especially those related to the reporting words ("I warned her don't…," "I encourage her go traveling"), only one correction was made by Fiona ("suggest doing something"). The observation implied that there was a gap between the student current knowledge and the desired learning target, which was regretfully not pointed out by the teacher. Also, Fiona's feedback included no explanation concerning in what aspect students had done well or in which area they could make more effort to.

Although teachers like Fiona emphasized encouragement-based feedback, students believed encouragement, itself, did not suffice, as embedded in the following quotes:

C1–H1: I hope Fiona could point out the errors I made so that I have a direction for further effort. If these problems were not identified, how can I make an improvement? I may make the same mistake again.

C2–A1: I expect my teacher to tell me my problems, though praise is also necessary.

Distinct from Fiona, Lily, or Rick, Max was the only participant teacher who provided feedback at more effective levels of process and self-regulation. His practice was examined below by looking at how he reacted to student reticence and how he followed individual responses.

(Max—Session 5) In the previous session, the question "What is acronym? What is abbreviation? And what is initialism?" was responded to with student silence and was therefore assigned as homework. At the beginning of this session, Max asked the students to share the information they had acquired in groups, and then invited representatives to report to the whole class.

Max, in responding to student silence, did not push the students to come up with an answer, nor did he offer any "teacher expertise." Indeed, he asked students to seek relevant information after class and share it in class. In other words, the un-answered question was not dismissed, but utilized as students' self-exploring and collaborative task.

When responding to individual answers, Max generally acknowledged students' contribution and meanwhile pointed out the weakness in answers. The following classroom data demonstrates this tendency.

(Max—Session 3) Max invited students to share group opinions on the topic "What do you think of the statement that we are deprived of the right to take drugs?"
Max: What's the idea in your group?
S1: We don't think taking drugs is a right because… (Max wrote S1's idea on the PowerPoint)
Max: Any other ideas from your group? How do other people think of it?
S2: Taking drugs is not necessarily a right…. (Max wrote on the PowerPoint)
Max: How about in your group?
S3: I found something interesting from the internet…So whether to take drugs or not is our own rights (Max wrote on the PowerPoint).
Max: All right, it is controversial…… S4, do you have different ideas?
S4: Taking drugs is not your own choice since…
Max: ….Do you have different ideas, S5? In your group?
S5: We are all against this statement……
Max: … OK. So far most students come up with something similar. I'd like to hear different ideas.

TABLE 7 | Factors influencing teacher assessment literacy in questioning practice.

	Influencing factors
At the personal level	Teacher expertise in assessment
	Teacher beliefs
At the institutional level	Student competence and motivation
	Institutional value
At the socio-cultural level	Confucian influenced culture

S6 (voluntarily offered an answer): I think we can peacefully use the drug.
Max: Tell us more about peaceful drug use.
S6: In western countries……but according to Chinese government policies……
Max: S6 just challenged the policy. Do you think all the policies defined by the government are acceptable?…… Have you got more different ideas to support the statement? Try to think critically, try to think differently. Do not just follow the majority.

In this excerpt, Max seemed to take three actions: first, he accepted student ideas by writing the key words on the PowerPoint; second, he perceived the weakness in the responses (answers were similar and lacked critical thinking) and thus kept inviting different ideas; and third, he guided students to explore the question in depth by thinking critically. In this sense, the teacher evaluation was likely to promote deep rather than surface learning and to cultivate learner's critical thinking skills.

Factors Influencing Teacher Assessment Literacy in Questioning Practice

Table 7 summarizes the specific factors found to influence teacher assessment literacy in questioning practice, which are placed at three different levels. Factors at the personal level include teacher expertise in assessment and teacher beliefs; those at the institutional level focus on learner competence and motivation, and institutional value; and at the social-cultural level, Confucian-influenced culture is discussed. In reality, various factors might function simultaneously to exert impact on questioning practice; for the sake of clarity, each factor and its impact is discussed separately, although the interplay among factors is touched upon when necessary.

At the Personal Level

At the personal level, different expertise in assessment influenced teacher questioning practice. The interview data revealed that the four participant teachers had very limited theoretical knowledge of classroom assessment, such as the knowledge of assessment purpose and methods, and knowledge of feedback, though variances in assessment competences could also be noted. Fiona and Rick shared that they had never received any training related to testing or assessment, and their classroom practices were guided by their accumulated knowledge as language learners and teachers. Lily, who had learned a course of assessment in her MA program, commented that the course focused primarily on theories of testing, and in practice, she was not sure about how

to conduct classroom assessment effectively. Max believed the assessment knowledge and skills he had were borrowed from a British teacher expert in language teaching and assessment when he was an undergraduate student. Max was also in charge of an MA program involving a course named *Language Assessment* when the study was conducted. His learning and working experiences, as well as regular reflections in teaching facilitated his development in assessment capacities.

The differences in teachers' assessment competence seemed to result in variability in questioning patterns. At initiation stage, Fiona, Lily, and Rick chose text-oriented convergent questions as the focus of their lessons. One common reason was that they all believed recall questions helped assess students' knowledge of the text and laid the foundation for further learning. Apart from that, Fiona believed that convergent questions facilitated teachers' control of classroom management; Lily thought these questions were less demanding and thus suitable for her students of limited language competence; and Rick maintained he had to deal with textbook and complete the teaching task specified by school administration. It seemed that the emphasis of convergent questions was a result of the interplay among teacher beliefs, student competence, and institutional regulation. Nevertheless, it should be also noted that these teachers did not seem aware of different question types in relation to student learning or skillful in designing both low- and high-order thinking questions to achieve different educational goals, which could be another reason for the dominance of text-oriented recall questions.

At response stage, choral answers became a shared pattern in the three teachers' practices. While Fiona and Lily intended to engage their less motivational or autonomous learners in classroom interactions by inviting whole-class responses; Rick perceived choral answers as time saving, which helped him stick to the university teaching pace. It seemed that the knowledge and skills teacher obtained through their working experiences enabled them to implement assessment activities with the consideration of students' background and institutional regulations. However, the repeated occurrence of choral responses also revealed the teachers' assumption that all the students could proceed at the same pace, and this thinking risked ignoring learner differences and were inconsistent with assessment for learning principles. Additionally, teachers did not seem to realize the strengths and limitations of different response types, and were less competent in collecting rich and accurate information about student learning. As a result, the choral responses they elicited were usually short utterances associated with basic facts in the text, representing limited information on learning and making it difficult for them to make adjustment to serve learner needs.

At evaluation stage, the three teachers preferred positive comments. Given students' limited language competence and motivation, especially in Fiona's class, encouragement-based feedback seemed important to build up learners' confidence. Nonetheless, positive feedback itself was insufficient since learners, especially those low achievers, needed specific comments to tell them where they were in their learning progressions, as well as directive suggestions to guide them for improvement. In other words, the praise lacked diagnostic

capacity, and Fiona appeared less capable of assisting her students in using feedback to feed forward. Lily and Rick's comments, compared to that of Fiona's, were more specific. The evaluation they made helped identify strengths in students' answers, on the one hand, but lack directive value of pointing out weaknesses and guiding students to make improvement, on the other hand.

Max demonstrated more expertise in assessment, and his questioning style was distinct from that of the other teachers. At initiation stage, world-knowledge convergent questions were deliberately posed to foreground students' deficiencies, guide after-class inquiry, and encourage in-class collaboration. Put it well, Max used assessment to diagnose weaknesses in learning, to build interest to learn, and to cultivate student responsibility in learning. Also, divergent questions were more frequently integrated into his instruction, which aimed to provoke thinking and create peer sharing opportunities. At response stage, individual answers were more often invited, especially so after group discussion, which helped to develop an environment of trust and elicit a wealth of learner information conducive to a better teacher decision making. At evaluation stage, learners' contribution was acknowledged, weaknesses identified, and necessary guidance provided to facilitate the closure of the gap between what students had achieved so far and what they were expected to achieve. It seemed that although Max did not possess theoretical knowledge of assessment, the classroom activities he conducted echoed assessment for learning principles and helped the students to advance their learning.

Teacher beliefs turned out to be another factor impacting on teacher assessment literacy in questioning practice. Fiona believed that the teacher should hold classroom control, whereas "divergent questions implied student-centered teaching" and "resulted in less control in teachers' hand." Consequently, she rarely raised divergent questions and instead placed much more emphasis on convergent ones. Nonetheless, under such circumstances, students had fewer opportunities to exhibit reflective thinking or articulate personal opinions.

Lily and Rick shared the view that teachers were resource persons, and covering the curriculum was their duty, as Lily remarked that "I don't have time for this (requesting individual answers) because I have to finish the teaching tasks in the first place," and Rick commented "if students answered the questions one by one, it would drag my progress." Underpinned by this view, both teachers utilized whole-class responses repeatedly to ensure that teaching tasks were accomplished and the teaching pace was kept, whereas individual learning needs were hardly accommodated.

Max expected his students to be independent and maintained that "do not expect to get all the knowledge from the teacher, the biggest contribution a teacher made is to raise your awareness." In his practice, world-knowledge convergent questions were deliberately raised to foreground student deficiency and to cultivate independent and collaborative learning. What is more, divergent questions were employed to activate thinking, to hear students' voices, and collect useful information for the subsequent decision making. As a result, learning needs were better catered for, and students were assisted in becoming self-reliant and autonomous learners.

At the Institutional Level

At the institutional level, the stakeholders' needs to a certain extent influenced TALiP, and this was, in particular, the case in UA where students were generally low in English language proficiency.

For example, the choice of intensive convergent questions was made based on learners' limited language ability.

> Fiona: My students' English is generally poor. To engage them in classroom interactions, I have to take this into consideration and make sure students don't have much difficulty in coming up with an answer to my question.

Similarly, choral answers were somewhat an outcome of the teachers' intention of engaging low achievers in classroom interactions (Lily) and the teacher's expectation that "everyone could come up with an answer" (Fiona).

In addition, positive feedback was preferred for its perceived value of boosting low achievers' self-esteem.

> Fiona: It is important to involve encouragement in teacher feedback because it builds up students' confidence.
>
> Lily: For those weak students, encouragement is a routine because I don't want to give them up.

The teachers had good intentions, though the pattern of convergent question—choral response—positive feedback did not seem to maximize learning benefits. The dominance of low cognitive questions emphasized knowledge memorization and risked engaging learners in rote learning; choral responses contained very limited information about individual learners and might disguise learning difficulties; and encouragement itself did not foreground weaknesses in current level of learning nor provided guidance for further improvement.

Learner motivation was another factor impacting on TALiP. In UA, students were generally less motivated, as shown in these quotes, "some do not go over the lesson after class" (Fiona) and "students in general are not self-disciplined or autonomous" (Lily). Consequently, both teachers inclined to conduct a review session at the beginning of each lesson, with intensive text-bound questions to check students' mastery and to impel students to adjust their behavior to fall in line with the teachers' expectations.

In UB where learner motivation was notably higher, review sessions were scarcely observed, and the very few instances captured were brief.

> C4-A1: We don't have review sessions, normally. Rick knows that we are diligent, and we'd do that ourselves. As a good learner, you need to go over the material you learned and the notes you took.
>
> C4-L2: The teacher probably thinks it is less of his work but more of student's responsibility. Reviewing should be self-guided, rather than teacher-led.

Embedded in these quotes were a high level of motivation and a perception of learning as students' own responsibility.

Apart from student competence and motivation, the immediate institutional culture could be a facilitating or inhibiting factor to the effectiveness of questioning practice.

In UA where teachers had to follow a required teaching schedule from the institution, time became a repeatedly mentioned concern.

> Fiona: Divergent questions implied student-centered teaching, which requires me to follow students' ideas and spend more time. This will in turn result in less control in teachers' hand and a delayed progress in course syllabus. The administration has rigid requirements for the teaching pace, though I personally do not like being restricted by it.

> Lily: I sometimes communicate with my students about the teaching pace. I let them know that teachers are in a dilemma, it was not because I am unwilling to cater for their needs, but because the college has a stipulated teaching pace for me to follow.

In a university culture where great emphasis was placed on content coverage, teachers may have limited time or freedom to conduct meaningful classroom activities congruent with the development of student dispositions, such as posing divergent questions to provoke thinking and inviting individual opinions revealing students' thought processes. This may partly explain why convergent questions and choral responses constituted the major parts in Fiona and Lily's lesson.

At the Social–Cultural Level

Situated in Chinese educational context, teaching and assessment were also affected by Confucian-influenced culture. An emphasis on memorization and text knowledge, and a whole-class response were culturally related phenomenon identified in teachers' questioning practice.

First, the choice of intensive text-bound questions in Fiona, Lily, and Rick's teaching reflects a cultural value in Chinese education where great importance is attached to repetitive practice, memorization, and textual knowledge. As the classroom data showed, convergent questions accounted for 88, 79, and 90% in these teachers' instruction. Further examination revealed that, on average, 97% of these questions were closely related to text knowledge. Though this type of questions does have their educational merits, such as testing factual knowledge and laying the foundation for future learning, they demand limited cognition and link to fixed answers. Worse, the dominance of these questions in the instruction could discourage students to think deeply and reflectively.

Second, choral answers were an acceptable and a shared pattern in the three teachers' classes, which is in accordance with collectivism in Chinese tradition under which whole-class teaching is appreciated. Both Fiona and Lily favored whole-class responses and they thought "the question was easy enough for everyone to answer," and Rick preferred choral answers because they "saved time." However, students responding in chorus were less helpful for themselves to disclose their learning difficulties or for teachers to make informed decisions.

In brief, factors at the teacher personal level, institutional level, and wider social–cultural level interacted with each other to have impacted on how teachers engaged in questioning practice.

DISCUSSION

This study investigated classroom questioning practice and factors influencing teacher assessment literacy in questioning practices. With respect to RQ1, it was found that overall teachers raised significantly more convergent than divergent questions, the largest proportion of these questions led to students' choral and individual answers, and teacher feedback was largely on tasks. This questioning pattern did not seem to maximize the learning benefit in each stage of initiation, response, and evaluation.

First, a high proportion of teacher questions was at lower intellectual levels directing at text knowledge, and this corroborates the observation made in the literature that teachers tended to ask lower cognitive questions in both content (Brown and Ngan, 2010) and ESL/EFL classrooms (David, 2007). It is particularly consistent with the results in Tan's (2007) study that low intellectual questions were largely text oriented and risked engaging students in rote learning in Chinese university classrooms. Second, student choral answer was a shared pattern in the current study, which echoes Chick's (1996) proposition that choral responses served more of a social function to engage students in classroom interaction without fearing the loss of face in a public situation with a possible wrong answer, rather than an academic function to diagnose learning difficulties. Third, the participant teachers' feedback was largely task related and encouragement based, which is congruent with the results in the Hattie and Timperley (2007) study that much feedback in current practices involves comments on attitude or praise, whereas opportunities to improve student work based on it seem rare. Although praise or encouragement is necessary, it, itself, is not constructive in guiding learning. In order to make students know "where they are," "where they need to go," and "how to get there," it is essential that teachers point out the learning strengths and weaknesses and provide guidance on improvement (Assessment Reform Group, 2002).

Regarding RQ2, the study supports that factors at different levels exert an influence on teacher assessment literacy in practice (Xu and Brown, 2016). At the personal level, the expertise teachers have in assessment determines the extent of the effectiveness of their classroom questioning. In the current study, none of the participant teachers received relevant training or had sufficient theoretical knowledge of classroom assessment, and their practices were largely guided by their accumulated knowledge from learning and teaching experiences. This supports the concerns expressed in the literature that teachers have restricted repertoire of assessment knowledge and skills (Volante and Fazio, 2007). Although all the participant teachers showed appropriate understanding of basic concepts in assessment and were able to conduct assessment activities by taking student interests into consideration, Max seemed more competent in designing quality questions, eliciting rich learning evidence, and assisting students in using feedback to feed forward.

Teacher beliefs, another influencing factor at the personal level, could bring a good deal of variability in assessment (Leung, 2004). Two broad beliefs were identified in the current study.

Fiona, Lily, and Rick seem to hold a behaviorism orientation that emphasizes the transmission of content and places the teacher at the center of teaching (Carless, 2011). Consequently, intensive text-bound questions and repeated choral response became shared pattern in their questioning. In contrast, Max's practice reflects a view of learning in which assessment helps students learn better (Black et al., 2003). Underpinned by this philosophy, Max thought of his teaching as meeting learning needs rather than covering the curriculum at all cost. He acted as a facilitator to identify learning weaknesses and provide necessary support to assist his students becoming self-reliant and collaborative learners.

At the institutional level, the university culture and stakeholders' need shape teacher assessment literacy in questioning practices. In UA, teachers had to follow an institutional teaching schedule, which partly resulted in a focus of convergent questions in their instruction, even though they would like to try more divergent questions. The culture of emphasizing transmission of content seems to run counter to the notion of assessment for learning (Cizek, 2010). If formative assessment is to be integrated in teaching, time that is currently allocated to the major purposes of covering the curriculum should be reallocated to support instructional planning and modified instructional practice, which means that an institutional culture is expected to change so as to align with the principle of making learning the priority in practice (Cizek, 2010).

Likewise, the stakeholders' needs affect teacher practice. In the current study, learners' limited language competence in UA influenced each stage of teacher questioning practice: less-demanding convergent questions were resorted to, non-threatening whole-class responses were often invited, and praise, rather than critical comments, on the quality of student work was given. In UB, however, a high level of learner motivation made teachers' revision sessions redundant, suggesting that the level of student motivation defines classroom assessment practice (Brown et al., 2009) and determines which variation of formative assessment, the "restricted" or "extended," will be implemented in specific teaching setting (Carless, 2011).

At the wider sociocultural level, traditional Chinese culture influences teacher assessment literacy in practice. The participant teachers' emphasis on covering textbook contents and text-based learning reflect the enduring influences textual knowledge exerted in Chinese education (Han and Yang, 2001). Similarly, the recurring pattern of choral answers is in accordance with the value of collectivism in Chinese tradition, which favors a whole-class instruction and assumes learners can be taught in the same pace toward a common goal (Cheng, 1997). In short, teaching and learning are situated activities, and assessment practice is better understood in specific social context. Teachers, who work in complex contexts, have to make professional decisions in response to various factors facilitating or inhibiting their development in assessment capabilities (Xu and Brown, 2016).

The current study is grounded in Chinese educational context and based on four case teachers' classroom questioning practices. Some suggestions are made for researchers, school administrators, and teacher educators in that specific setting. The study aims to achieve "reader generalizability," meaning "leaving the extent to which a study's findings apply to other situations up to the people in those situations" (Merriam, 1998, p. 211).

First, dialogues disseminating information about formative assessment could be conducted between researchers and teachers on the basis of respect for teacher's perspectives and beliefs. Teachers in the current study have limited knowledge of formative assessment, and they are doing what they perceived beneficial for their students. The assessment activities they carried out in some cases do support student learning, though they, themselves, are not aware these activities could be called formative assessment. Thus, researchers could share the key formative assessment principles and strategies with teachers, and encourage them to reflect on their approach to pedagogy and in particular their classroom questioning. Also, based on the research evidence on educational innovations that teachers are likely to change and enhance their repertoire of teaching, learning, and assessment strategies when they experience dissatisfaction from student response (Carless, 2011), researchers could advise teachers to hold regular discussions with students, checking whether students are satisfied with their patterns of teaching and whether there is a need to improve their existing practice. It is hoped the professional dialogue between researchers and teachers, as well as the constant discussions between teachers and students, contribute to a better understanding of teachers' own day-to-day practices and further to the improvement in formative assessment practices.

Second, school administrators need to provide a supportive immediate educational environment for teacher professional development. For example, workshops on assessment could be held to enhance teacher assessment literacy. The teachers in the current study have received no relevant training on formative assessment before being involved in classroom assessment activities, and there is little preservice or in-service training in China in relation to formative assessment (Xu and Liu, 2009). However, if formative assessment is to be implemented effectively, teachers need up-to-date theoretical knowledge to help them reexamine their classroom practice and an array of skills to modify that practice. For another instance, conditions for facilitating collective teacher learning of assessment need to be created. Community of practice, the concept of professionals working in teams to share knowledge and enhance practice, is very effective in embedding formative assessment in schools (Leahy and Wiliam, 2009). Participation in community activities enables teachers to have professional conversations about their assessment practices, provides them opportunities to understand alternative thinking and practice of assessment, and allows them to negotiate their ideas with colleagues (Leahy and Wiliam, 2012). Additionally, and perhaps more importantly, the school assessment policies and institutional approaches to pedagogy could be adjusted. New assessment policies can be carried out to encourage teachers to try assessment for learning innovations. More learning-oriented approaches to pedagogy and a less packed curriculum may be adopted to

empower teachers with greater autonomy and assist them in claiming more ownership over what they could do in their classrooms.

Third, teacher educators need to diagnose problems in pre- and in-service assessment education in Chinese educational context and design specific courses to promote teacher assessment literacy. For example, training sessions on effective questioning could be conducted, which include how to formulate quality questions, how to collect responses representing student learning, and based upon the information gathered, what actions to take to better facilitate learning. Drawing on authentic classroom examples, the current study proposes the assessment literacy teachers may need to conduct effective questioning as follows:

- Teachers should understand central concepts and key principles of formative assessment so as to better guide their questioning practice.
- Teachers should be clear about different taxonomies of classroom questions and be able to construct quality question of different types to achieve various teaching/assessment purposes.
- Teachers need to know the strengths and limitations of different response options and be able to elicit rich and accurate evidence of student learning.
- Teachers need to understand different feedback types in relation to student learning. Based on student response elicited, they should be able to identify the gap between student current level and the target level of knowledge and skills, and conduct meaningful intervention to close that gap.

Nevertheless, improving teacher assessment literacy in practice is not merely providing courses assisting teachers with the mastery of a body of knowledge, such as the knowledge of assessment, disciplinary knowledge, and pedagogical content knowledge. It involves different ways and individualized suggestions for teachers to carry out culturally and contextually grounded formative assessment based on their existing practices, given that different teachers may hold different assessment beliefs and expertise that are not easily changed through professional development. It is suggested that in the context where teachers are less likely to have well-prepared capacities for implementing formative assessment, or the institutional and social values are less directly supportive of formative assessment, teachers could start with "restricted" form and gradually introduce "extended" form of formative assessment (Carless, 2011). However, if the "extended" version is to be achieved, joint efforts from different stakeholders, including teacher educators, school administrators, and students, are needed to establish a social and institutional culture aligning with assessment for learning principles.

CONCLUSION

This study examines classroom questioning practice through a formative assessment lens and investigates potential factors influencing teacher assessment literacy in questioning practice. The finding showed that different teacher assessment competences resulted in variability in teacher questioning practices, which further impacted on student learning. The results also indicated that teacher assessment literacy was dynamics situated in context. It interrelated with other mediating factors at the personal, institutional, and socio-cultural levels.

Based on an in-depth analysis of the empirical evidence, the study discusses the assessment literacy teachers may need to conduct effective questioning and highlights the legitimacy of teacher questioning as a sub-construct of teacher assessment literacy. Findings in this study also carry practical implications for the establishment of a social and institutional culture aligning with assessment for learning principles.

One limitation is that the study adopted a fairly broad perspective. It examined the three stages of teacher questioning process, investigated teacher assessment literacy, and explored mediating factors influencing the classroom practice. If the study had focused on fewer strands, it could have explored each more deeply. Also, the study is a small-scale exploratory case study conducted with four EFL university teachers in two Chinese institutions, so its findings cannot be generalized to a larger population or other contexts. Further studies could investigate how teacher training programs might promote teacher assessment literacy in classroom questioning. Future efforts could also be made to explore the ways various levels of contexts facilitate or inhibit teacher professional development in assessment literacy so that we could better understand the contextual factors and create a facilitating environment for the implementation of formative assessment.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The University of Hong Kong. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YJ is the solo author who contributed the conception and design of the study, organized the database, performed the statistical analysis, wrote the draft, and revised the manuscript. She also takes the responsibility for communication with the journal and editorial office during the submission process, throughout peer review, during and post publication.

REFERENCES

Adamson, B., and Morris, P. (1998). "Primary schooling in Hong Kong," in *The Primary Curriculum: Learning From International Perspectives,* eds J. Moyles and L. Hargreaves (London: Routledge), 181–204.

American Federation of Teachers, National Council on Measurement in Education, and National Education Association (AFT, NCME, & NEA) (1990). Standards for teacher competence in educational assessment of students. *Educ. Measure. Iss. Pract.* 9, 30–32. doi: 10.1111/j.1745-3992.1990.tb00391.x

Assessment Reform Group (2002). *Assessment for Learning: 10 Principles Research-Based Principles to Guide Classroom Practice.* Cambridge: University of Cambridge.

Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2003). *Assessment for Learning: Putting it Into Practice.* Buckingham: Open University Press.

Black, P., and Wiliam, D. (1998). *Inside the Black Box: Raising Standards Through Classroom Assessment.* London: King's College London School of Education.

Black, P., and Wiliam, D. (2009). Developing the theory of formative assessment. *Educ. Assess. Eval. Account.* 21, 5–31. doi: 10.1007/s11092-008-9068-5

Bloom, B. S. (1956). *Taxonomy of Educational Objectives.* New York, NY: Longmans, Green.

Boyles, P. (2005). "Assessment literacy," in *National Assessment Summit Papers,* ed M. Rosenbusch (Ames, IA: Iowa State University), 1–15. Available online at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.883.1970&rep=rep1&type=pdf#page=18

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educ. Measure. Iss. Pract.* 30, 3–12. doi: 10.1111/j.1745-3992.2010.00195.x

Brown, G. T. L. (2008). *Conceptions of Assessment: Understanding What Assessment Means to Teachers and Students.* New York, NY: Nova Science.

Brown, G. T. L., McInerney, D. M., and Liem, G. A. D. (2009). "Student perspectives of assessment," in *Student Perspectives on Assessment: What Students Can Tell Us About Assessment for Learning,* eds D. M. McInerney, G. T. L. Brown, and G. A. D. Liem (Charlotte, NC: Information Age Publishing), 1–21.

Brown, G. T. L., and Ngan, M. Y. (2010). *Contemporary Educational Assessment: Practices, Principles, and Policies.* Pearson: Hong Kong.

Campbell, C., Murphy, J. A., and Holt, J. K. (2002). Psychometric analysis of an assessment literacy instrument: applicability to preservice teachers. *Paper Presented at the Annual Meeting of the Mid-Western Educational Research Association* (Columbus, OH).

Carless, D. (2011). *From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-Heritage Settings.* New York, NY: Routledge.

Cheng, K. M. (1997). "The education system," in *Schooling in Hong Kong: Organization, Teaching and Social Context,* eds G. Postiglione and W. O. Lee (Hong Kong: Hong Kong University Press), 25–42.

Chick, J. (1996). "Safe-talk: collusion in apartheid education," in *Society and the Language Classroom,* ed H. Coleman (Cambridge: Cambridge University Press), 21–39.

Cizek, G. J. (2010). "An introduction to formative assessment: history, characteristics, and challenges," in *Handbook of Formative Assessment,* eds H. L. Andrade and G. J. Cizek (New York, NY: Routledge), 3–17.

David, O. F. (2007). Teacher's questioning behavior and ESL classroom interaction pattern. *Hum. Soc. Sci. J.* 2, 127–131.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Lang. Assess. Q.* 9, 113–132. doi: 10.1080/15434303.2011.642041

Gall, M. D. (1970). The use of questions in teaching. *Rev. Educ. Res.* 44, 707–721. doi: 10.3102/00346543040005707

Gu, P. Y. (2014). The unbearable lightness of the curriculum: what drives the assessment practices of a teacher of English as a Foreign language in a Chinese secondary school. *Assess. Educ. Principles Policy Pract.* 21, 286–305. doi: 10.1080/0969594X.2013.836076

Han, M., and Yang, X. (2001). Educational assessment in China: lessons from history and future prospects. *Assess. Educ. Principles Policy Pract.* 8, 5–10. doi: 10.1080/09695940120033216

Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487

Hill, K., and McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Lang. Test.* 29, 395–420. doi: 10.1177/0265532211428317

Hodgen, J., and Webb, M. (2008). "Questioning and dialogue," in *Unlocking Assessment: Understanding for Reflection and Application,* ed S. Swaffield (New York, NY: Routledge), 73–89.

Hofstede, G. H. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizing Across Nations.* Thousand Oaks, CA: Sage.

Hoyt, K. (2005, April). Assessment: impact on instruction. *New Vision on Action.* National Assessment Summit, Alexandria, VA.

Hu, Q., Nicholson, A., and Chen, W. (2004). A survey on the questioning pattern of college English teachers. *For. Lang. World* 6, 22–27.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: a focus on language assessment courses. *Lang. Test.* 25, 385–402. doi: 10.1177/0265532208090158

Jackson, J. (2002). Reticence in second language case discussions: anxiety and aspirations. *System* 30, 65–84. doi: 10.1016/S0346-251X(01)00051-3

Jiang, Y. (2014). Exploring teacher questioning as a formative assessment strategy. *RELC J.* 45, 287–304. doi: 10.1177/0033688214546962

Leahy, S., and Wiliam, D. (2009). From advocacy to adoption: getting formative assessment into more classrooms. *Symposium Presented at American Educational Research Association (AERA) Conference* (San Diego, CA).

Leahy, S., and Wiliam, D. (2012). "From teachers to schools: scaling up professional development for formative assessment," in *Assessment and Learning,* ed J. Gardner (London: SAGE), 49–72.

Leong, W. S., Ismail, H., Costa, J. S., and Tan, H. B. (2018). Assessment for learning research in East Asian countries. *Stud. Educ. Eval.* 59, 270–277. doi: 10.1016/j.stueduc.2018.09.005

Leung, C. (2004). Developing formative teacher assessment: knowledge, practice, and changes. *Lang. Assess. Q.* 1, 19–41. doi: 10.1207/s15434311laq0101_3

Leung, C. (2014). "Classroom-based assessment issues for language teacher education," in *The Companion to Language Assessment,* Vol. 3, ed A. J. Kunnan (Chichester: Wiley Blackwell), 1510–1519.

Leung, C., and Mohan, B. (2004). Teacher formative assessment and talk in classroom context: assessment as discourse and assessment of discourse. *Lang. Test.* 21, 335–359. doi: 10.1191/0265532204lt287oa

Long, M. H., and Sato, C. J. (1983). "Classroom foreigner talk discourse: forms and functions of teachers' questions," in *Classroom Oriented Research in Second Language Acquisition,* eds H. W. Seliger and M. H. Long (Rowley, MA: Newbury House), 268–285.

McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Prac. Assess. Res. Eval.* 7. Available Online at: https://scholarworks.umass.edu/pare/vol7/iss1/8.

Mehan, H. (1979). *Learning Lessons: Social Organization in the Classroom.* Cambridge, MA: Harvard University Press.

Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education.* San Francisco, CA: Jossery-Bass Publishers.

Mertler, C. A. (2004). Secondary teacher's assessment literacy: does classroom experience make a difference? *Am. Sec. Educ.* 33, 49–64.

Miles, M. B., and Huberman, A. M. (1994). *Qualitative Data Analysis, 2nd Edn.* London: Sage Ltd.

Musumeci, D. (1996). Teacher-learner negotiation in content-based instruction: communication at cross-purposes? *Appl. Linguist.* 17, 286–325. doi: 10.1093/applin/17.3.286

Peng, J. E. (2012). Towards an ecological understanding of willingness to communicate in EFL classrooms in China. *System* 40, 203–213. doi: 10.1016/j.system.2012.02.002

Pierson, H. D. (1996). "Learner culture and learner autonomy in the Hong Kong Chinese context," in *Taking Control: Autonomy in Language Learning,* ed R. Pemberton (Hong Kong: Hong Kong University Press), 48–58.

Pill, J., and Harding, L. (2013). Defining the language assessment literacy gap: evidence from a parliamentary inquiry. *Lang. Test.* 30, 381–402. doi: 10.1177/0265532213480337

Plake, B. S., Impara, J. C., and Fager, J. J. (1993). Assessment competencies of teachers: a national survey. *Educ. Measure. Iss. Pract.* 12, 10–12. doi: 10.1111/j.1745-3992.1993.tb00548.x

Pryor, J., and Crossouard, B. (2008). A socio-cultural theorization of formative assessment. *Oxf. Rev. Educ.* 34, 1–20. doi: 10.1080/030549807014 76386

Richards, J. C., and Lockhart, C. (1994). *Reflective Teaching in Second Language Classrooms*. Cambridge: Cambridge University Press.

Ruiz-Primo, M. A., and Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: exploring teachers' practices and student learning. *Educ. Assess.* 11, 237–263. doi: 10.1080/10627197.2006.9652991

Ruiz-Primo, M. A., and Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *J. Res. Sci. Teach.* 44, 57–84. doi: 10.1002/tea.20163

Sanders, N. M. (1966). *Classroom Questions: What Kinds?* New York, NY: Harper and Row.

Shomoossi, N. (2004). The effect of teachers' questioning behavior on EFL classroom interaction: a classroom research study. *Read. Matrix* 4, 96–104.

Smith, F., Hardman, F., Wall, K., and Mroz, M. (2004). Interactive whole class teaching in the national literacy and numeracy strategies. *Br. Educ. Res. J.* 30, 395–411. doi: 10.1080/01411920410001689706

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

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

Stiggins, R. J. (2010). "Essential formative assessment competencies for teachers and school leaders," in *Handbook of Formative Assessment*, eds H. L. Andrade and G. J. Cizek (New York, NY: Routledge), 233–250.

Tan, Z. (2007). Questioning in Chinese university EL classrooms: what lies beyond it? *RELC J.* 38, 87–103. doi: 10.1177/0033688206076161

Taylor, L. (2009). Developing assessment literacy. *Annu. Rev. Appl. Linguist.* 29, 21–39. doi: 10.1017/S0267190509090035

Volante,. L., and Fazio, X. (2007). Exploring teacher candidates' assessment literacy: implications for teacher education reform and professional development. *Can. J. Educ.* 30, 749–770. doi: 10.2307/20466661

Wintergerst, A. C. (1994). Second *Language Classroom Interaction*. Toronto, ON: University of Toronto Press.

Xu, Y., and Brown, G. T. L. (2016). Teacher assessment literacy in practice: a reconceptualization. *Teach. Teacher Educ.* 58, 149–162. doi: 10.1016/j.tate.2016.05.010

Xu, Y., and Liu, Y. (2009). Teacher assessment knowledge and practice: a narrative inquiry of a Chinese college EFL teacher's experiences. *TESOL Q.* 43, 493–513. doi: 10.1002/j.1545-7249.2009.tb00246.x

Yu, W. M. (2015). Teacher leaders' perceptions and practice of student assessment reform in Hong Kong: a case study. *Plann. Chang.* 46, 175–192.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

Teacher Interview Schedule

Date of the interview: Name of the teacher interviewed:

Part One: General Questions About the Teacher

1. Would you please briefly introduce yourself, e.g., your professional title; educational background; years of English teaching?
2. Have you received any assessment-related training before (If "Yes," please describe briefly the focus of the training. If "No," please describe where you think the knowledge and skills that you use to assess your students come from)?

Part Two: General Questions About the Students

1. How long have you been teaching this class?
2. Would you introduce your students in general, e.g., their English proficiency, their involvement in the classroom activities, their motivation in learning, the extent they take responsibility for their own learning?

Part Three: Formative Assessment Practice and Perception

1. What do you normally do to check or monitor student learning?
2. Based on the characteristics of your students, what assessment activities do you employ to help with their learning? Would you list the top three that you think are most helpful for students' learning?

3. What do you understand by effective assessment?

Part Four: Teacher Questioning

1. I noticed that at the beginning of each lesson, you normally raised a lot of questions to help students review what they learned in the previous lesson. Do you agree upon my observation? What are the purposes of asking those questions? (For Fiona and Lily)
2. I noticed that at the beginning of your lesson, there is no obvious revision part. Why didn't you review what had been taught before? (For Max and Rick)
3. Closed or open questions, which type of questions did you utilize more in your instruction? Why?
4. Do you think, in general, you provide enough time for students to think and formulate answers to your questions? Why?
5. I noticed that you sometimes invited students to answer your question as a whole class, and sometimes you invited individual student to answer the question. Is that so? Could you explain why? (For Fiona, Lily, and Rick)
6. I noticed that before asking individual students to answer your questions, you usually encouraged students to exchange ideas in pairs. What is your intention of doing this? (For Max)
7. Could you elaborate on the elements that are usually involved in your oral feedback?
8. If the student's answer is not correct or different from your expected answers, how will you react to it? Why?
9. What do you understand by constructive feedback?
10. How does your classroom questioning impact on students' learning (positively and/or negatively)?