The Journal of Asia TEFL

Preparing Primary English Teachers in Hong Kong: Focusing on Language Assessment Literacy

Qin Xie
Education University of Hong Kong

Siyan Tan
North Point Methodist Primary School, Hong Kong

Introduction

Language assessment literacy has become a central issue in the area of language assessment and testing in the recent decade. Research investigating teachers' language assessment literacy is accumulating in the literature. Much research was conducted with in-service language teachers, especially those working at the tertiary level. However, little has been done at the primary level. Pre-service language teachers and teachers working at the primary level were the least studied. The present research investigated primary English language teachers' needs of language assessment literacy, taking the perspectives of both pre- and in-service teachers. The research was situated in the unique socio-educational context of Hong Kong.

Literature Review

Writing in 1991, Stiggins defined assessment literacy (AL) as the knowledge and skills needed for measuring student achievement, stressing the knowledge base that teachers need in order to conduct sound assessment. Drawn from this general concept, language assessment literacy (LAL) is a relatively new concept in language assessment and testing (Fulcher, 2012; Inbar-Lourie, 2008, 2013, 2017; Taylor, 2013). Boyles (2005) noted that foreign language teachers need to develop "an understanding of the principles and practices of testing and assessment" and "necessary tools for analyzing and reflecting upon test data in order to make informed decisions about instructional practice and program design" (p. 18). This conception, though useful to refer to, does not set LAL apart from the general concept of AL. Indeed, despite the accumulating writing on LAL, the distinction between AL and LAL has not always been clear. Upon this point, Inbar-Lourie (2008) noted the importance of having 'a language component' in the LAL framework and that language teachers need to become familiar with "contemporary theories in language-related areas". The language component and 'the unique complexities that are entailed in the testing and assessment of linguistic skills, knowledge and communicative competence' justify a recognition of LAL as 'a unique entity' (Harding & Kremmel, 2016, p. 414).

Nonetheless, defining the scope of LAL and identifying an agreed cannon of its core knowledge are still work in progress and remains inconclusive (Inbar-Lourie, 2017). Within the context of a TESOL course, Brindley (2001) offered two core and three optional components of LAL. The two cores are 1) understanding of the social, educational and political context of assessment, including issues of ethics, accountability and standardization and 2) theoretical knowledge about and familiarity with language and language proficiency, as well as key concepts of measurement (such as validity and reliability). The second core component overlaps with the language component in Inbar-Lourie (2008). The three optional components are 1) skills in developing and evaluating language tests; 2) understanding the role of assessment in a language curriculum and the related concepts of criterion-referenced assessment and alternative assessment methods; 3) the ability to put LAL knowledge into practice or to apply it in research. The three optional components stress teachers' ability to apply LAL skills, knowledge and principles (Davies, 2008) to professional practices.

Similarly, Inbar-Lourie (2008) proposed two layers of a knowledge base for LAL: the bottom layer comprises knowing the reasons for assessment (the 'why' category), knowing the traits to be assessed (the 'what' category), and knowing the procedure of assessment (the 'how' category). On top of this foundation, she adds, there should be an understanding of the social role of assessment, of the nature of language (i.e. the language component), and of assessment practices in the real world. Inbar-Lourie put a particular stress on social contexts within which teachers practice assessment.

In a similar vein, Fulcher (2012) situated LAL within the larger historical, social, political and educational context wherein assessment is conducted. His tripartite model comprises 1) the knowledge, skill and practices concerning language assessment practices in classroom and external language testing; 2) the knowledge of the process, principles and concepts of language assessment that guide assessment practices, and 3) an understanding of the larger context. Moreover, he added, assessment knowledge, skills and principles should not be taught in abstract without context, but be embedded within concrete assessment procedures or test developing cycles so that knowledge and skills can be transferred to practice.

Despite the differences in terminology, LAL frameworks proposed thus far overlap, though there is still a lack of consensus regarding a core cannon (Inbar-Lourie, 2013). Harding and Kremmel (2016) noted that context-specific LAL frameworks could be established through negotiation among teachers and other stakeholders at the local level. A 'negotiated syllabus' (p. 424) could comprise elements selected from a core inventory and contextually specific materials. Taylor (2013) went further to stipulate unique LAL profiles corresponding to different stakeholders' roles in assessment process and use. Adopting Pill and Harding's rating scale (2013), she noted that classroom teachers should achieve Level 4 (the highest: multi-dimensional competence) on language pedagogy, Level 3 (procedural and conceptual proficiency) on technical skills, local practices, personal beliefs and attitudes towards assessment, and Level 2 (functional literacy) on knowledge of theory, principles and concepts, and understanding of scores and decision making. In contrast, university administrators who make admission decisions should possess Level 4 competence in terms of their understanding of test scores and competence in making score-based decisions, and Level 3 competence in terms of their awareness and understanding of local sociocultural values regarding assessment and education, but Level 2 and 1 (nominal literacy) for the other dimensions.

Taylor's differentiation of the necessary LAL for different stakeholder groups is useful, though these profiles have not yet been verified empirically within specific educational contexts. At this point, we turn to a review of existing studies.

Existing LAL Studies

Despite variations in research contexts, existing LAL studies have arrived at some consensus findings. Firstly, they found that teachers generally lacked LAL due to inadequate or non-existent training. Secondly, LAL training tended to be dominated by theories about language testing and measurement with inadequate coverage of classroom assessment practices and formative assessment. LAL instructors and

classroom teachers tended to have different perceptions towards the core content (Jeong, 2013). Finally, teachers' career status was identified to have affected teachers' perceptions of prioritized LAL components.

A number of existing studies found teachers' LAL remains underdeveloped (Popham, 2011; Scarino, 2013). Hasselgreen et al. (2004) observed that many assessment activities were carried out by teachers who had received no training in assessment and there was "a need for training across the board" (p. 7). Similarly, Vogt and Tsagari (2014) found foreign language teachers in Europe had low levels of LAL and the majority had received either "a little" or "no" assessment training (p. 391). Moreover, teachers reported having little confidence in any of the assessment areas investigated, especially the more innovative methods of formative assessment such as using student portfolios, self- and peer-assessment (Tsagari & Vogt, 2017; Vogt, 2018). Similarly, Hatipoğlu (2015) and Kömür (2018) noted that pre-service teachers had difficulty in applying assessment knowledge to real classroom practices.

Moreover, needs analysis studies found teachers' assessment needs were closely associated with their contexts. For instance, Tsagari and Vogt (2017) surveyed a large number of English teachers in schools in Cyprus, Germany, Greece, and Hungary regarding their existing assessment practices and needs. Associated with the high-stakes exam culture in Greece, Greek teachers reported a higher demand for advanced training than their German and Cypriot counterparts, most of whom only asked for basic training. Similarly, pre-service teachers in Turkey (Hatipoğlu, 2015) requested a training course that prepared them for assessing learners of all age groups, because teacher education in Turkey is not streamed, meaning that teacher candidates might be assigned to teach in either primary or secondary schools. In Hong Kong, where there is an increasing emphasis on teachers adopting classroom-based formative assessment within a highly competitive examination-driven context, pre-service teachers perceived their training insufficient for both formative and summative assessment (Lam, 2015).

Teachers' assessment needs also vary across career stages; pre- and in-service teachers tend to have different needs. Pre-service teachers tend to ask for a comprehensive assessment training course (DeLuca & Klinger, 2010) with a focus on practical techniques, such as "reporting achievement, modifying assessments, developing constructed-response items, and determining item reliability and validity" (p. 419). In-service teachers' concerns, on the other hand, are reflective of their professional needs and priorities, such as the grade levels and subjects they taught (Zhang & Burry-Stock, 2003). The practicing teachers in Coombs et al (2018) were found adjusting their priorities in assessment practices across different career stages. Despite received pre-service training stressing formative assessment, novice teachers in their first five years of practice were found shifting towards more standardized and summative assessment approaches, probably in response to the high accountability school climate. However, once they had established themselves as professionals and learnt to mediate external and internal accountability pressure, they began to re-prioritize formative and differentiated approaches to assessment.

The Present Study

Existing research offers valuable insights into the conceptualization of LAL and the ways to investigate it. We found most existing studies involved either in-service (e.g., Fulcher, 2012; Hasselgreen et al., 2004; Vogt & Tsagari, 2014) or pre-service teachers (e.g., Graham, 2005; Jeong, 2013). Few studies incorporated the perspectives of both. While the majority of studies seemed to agree that training courses for pre-service teachers should differ from those for in-service teachers, exactly what the differences should be remains unclear. Moreover, empirical studies from contexts in Asia, Africa and South America were in paucity (Harding & Kremmel, 2016). In Asia, we found one study looks into the LAL training of tertiary English instructors in China (Jin, 2010) and another one examined pre-service teacher training in Hong Kong (Lam, 2015). In the latter study, the researcher examined course outlines and interviewed both tertiary lecturers and teacher candidates, providing important insights about the marginalized status of LAL training in teacher education programs. However, the study did not involve practicing teachers

into the 'negotiation' of LAL syllabus. Notwithstanding the importance of their perspectives, tertiary lecturers and pre-service teachers may not have adequate or updated knowledge about existing assessment practices in schools or novice teachers' assessment responsibilities. Such knowledge could be sought from practicing teachers. Capitalizing on in-service teachers' assessment experience (Fulcher, 2012) can help develop a LAL syllabus that is both context sensitive and forward-looking.

Aiming at identifying an appropriate set of LAL components and a suitable LAL profile for pre-service primary English teachers in Hong Kong, the present research investigated both pre- and in-service teachers with the guidance of the following research questions:

1. From the in-service teachers' perspective, what is the necessary knowledge base for novice teachers to fulfill their assessment-related duties?
2. To what extent, if at all, do teacher candidates perceive themselves to be ready for the upcoming assessment responsibilities in schools?

Methodology

The study adopted a mixed-method approach that began with a qualitative phase of interviews with in-service teachers to address RQ1, followed by a quantitative phase using a questionnaire with pre-service teachers to address RQ2. As findings from the qualitative phase informed the design of the questionnaire used in the quantitative phase, the research can be characterized as an exploratory sequential design (Creswell, 2015).

Participants

Eleven teachers from eleven primary schools participated in phase one (Table 1). They were recruited based on a combination of convenience sampling and maximum variation sampling (Merriam & Tisdell, 2015) in order to involve teachers at different career stages and from different types of schools. There are four types of primary school in Hong Kong: government, aided, private and Direct Subsidy Scheme (DSS) schools. Government and aided schools are funded by public money and required to closely follow the central curriculum specified by the government. DSS schools are funded jointly by the public and the private sectors and enjoy greater flexibility in "curriculum design and student admission" (EDB, 2018). Private schools (including international schools) are self-financed and have the greatest flexibility in curriculum and student admission.

TABLE 1
Profiles of the In-service Teacher Participants

#	Gender	Teaching experience	School type
T1	F	20+ years	Government
T2	F	3 years	
T3	F	13 years	Aided
T4	F	4 years	
T5	F	2 years	
T6	M	9 months	
T7	F	6 years	DSS
T8	F	9 months	
T9	M	6 years	Private
T10	F	2 years	
T11	F	9 months	

Care was taken to involve teachers at different career stages, including novice (with less than 3 years of teaching experience), junior (having 3 to 6 years of teaching experience) and seasoned teachers (with more than 10 years of teaching experience). Of the eleven participants, two were senior teacher (T1 and T3), four junior (T2, 4, 7 and 9), and five novice (T5, 6, 8, 10 and 11). As much as possible, for each school type, we involved teachers at three career stages. For instance, the four teachers from aided schools covered from novice, junior to seasoned teachers.

The participants of Phase 2 were pre-service student teachers studying for a Bachelor of Education degree in English Language Education (the primary stream) in a university in Hong Kong. At the time of data collection, they were expected to graduate and enter the profession in a month's time. A total of 55 student teachers responded to our questionnaire, representing 51.9% of the program's year cohort; 52 questionnaires were valid. All student teachers had completed their teaching practicum of 20 weeks. The majority had received assessment-related training albeit of a limited nature. All had attended one mandatory general course titled *Curriculum and Assessment,* but only nine classroom hours (out of 39) were spent on introducing basic knowledge about assessment. Considerably fewer participants took the other assessment-related but elective courses, namely, *Language Assessment* (36.5%), *Assessment for Productive Learning* (1.9%) and *School-based Assessment: Theories and Perspectives* (1.9%). All courses, except for *Language Assessment*, were non-language specific and theory-oriented. The majority of students (82.7%) claimed to have experience of designing assessment tasks or test papers through an assignment for a university course (53.8%), during their teaching practicum (53.8%) or as part of their private tutoring experience (38.5%).

Instruments and Data Collection

Qualitative data were collected through semi-structured, face-to-face, individual interviews. Each interview lasted between 45 and 60 minutes. Interviews were conducted in the participants' first language (Cantonese) to ensure as far as possible that their responses accurately and adequately reflected their views. All interviews were audio-taped and fully transcribed.

Interview questions were structured with reference to the key LAL components as reviewed earlier, which could be grouped according to Fulcher's tripartite model (2012) of assessment practices and procedures (see Q4 to 10 in Appendix A), assessment principles and concepts (Q11 and 13), and issues relating to context (Q2, 3). Two main types of question were asked in the interviews: (i) about existing assessment practices in schools and those required of novice teachers, and (ii) about the difficulties encountered by novice teachers in conducting assessment-related duties. For the latter, novice and junior teachers reported their own experiences, while seasoned teachers recalled their own experience when they were novice as well as to make comments based on their working experience with their junior colleagues.

Based on the interview findings as well as references to existing AL measures (e.g., DeLuca et al., 2016), we developed a questionnaire to survey pre-service teachers. The initial version of the questionnaire was piloted with five student teachers and subsequently modified of its wordings and formats. The first two questions on the questionnaire (see Appendix B) elicit information about respondents' training experience. The other questions cover four areas of language assessment practices deemed important by the teachers interviewed. Each area is further divided into sub-questions relating to specific knowledge and language skills. A 4-point Likert scale, from 0 (Level 0: I don't know about it) to 3 (Level 3: I am very good at it) corresponding to the first four levels of LAL in Pill and Harding (2013) and Taylor (2013) as reviewed earlier was adopted for participants to estimate their competence. Table 2 (to appear in Section 6) presents the items in categories. It is worth clarifying that the slightly uneven distribution of items stems from some assessment components which, though deemed important, are associated with only a small number of skills that can be investigated by a questionnaire.

Data Analysis

All interviews were audiotaped and fully transcribed verbatim in Chinese, which generated 94 pages of transcripts for analysis. Only the extracts presented in the result section were translated into English. Analysis of the qualitative data was both deductive, that is, using interview questions and the LAL components as categories to pool data, and inductive. Although the categories were predetermined, specific codes were assigned to emerging themes and patterns. Specifically, the analyst first read through the transcripts and allocated meaningful responses to relevant LAL components. The data within each category were then read multiple times to identify recurring themes across participants that were responsive to each research question (Merriam & Tisdell, 2015). Coding reliability was ensured by having two other researchers read through the codes and the coding results to identify any flaws in the coding. Differences in opinions were resolved via discussions. Analysis of questionnaire data aimed at describing the general views. Therefore, only descriptive statistics were generated, which included means, standard deviations at item level, and their aggregations at the category level.

Phase 1 - Qualitative Findings

Perceived Difficulty and Necessary Assessment Skills of Novice Teachers

The teachers brought up a range of necessary knowledge and skills, from assessment design to communication of assessment results to students and parents. In terms of assessment design, the majority indicated an increasing need to adopt a variety of formal and informal **assessment methods** in their teaching and assessment, such as *tests, dictation exercises, classroom observation, homework, and even drama performance*. As formal testing remains the main evaluation method in schools, all teachers emphasized the importance of the skills to design and develop test papers, with Teacher 1 and 8 explicitly referring to them as "*the most important*", and Teacher 8 stating that '*familiarizing with various question types is necessary*'. They noted a general need for novice teachers to be familiar with a repertoire of test/exam items, including but not limited to *multiple choice, fill in the blank, error correction, mind-maps and jigsaw puzzle questions*. Compared with the teachers from public schools, those from private and DSS schools put even more emphasis on having a variety of item types on test papers. Finally, participants noted that novice teachers should also learn to apply formative assessment strategies in teaching and assessment because they were required by the central curriculum and gained acceptance and popularity in schools as an effective means to enhance teaching and learning.

Another area of knowledge frequently mentioned by the participants is **content/subject knowledge**, which corresponds to the 'language component' in the literature (Inbar-Lourie, 2008). Teacher 4 gave an example of conducting assessment of speaking performance, where teachers needed to master "*accurate pronunciation and use of tones so that they can effectively assess these aspects of students' performance*". Quite a few teachers (T1, 4, 5, 8 and 10) also stressed the ability to write "*simple, concise and grammatically correct*" instructions, which pertains to teachers' English proficiency. Familiarity with different target text types was also mentioned as important, which is associated with the requirement of the local primary English curriculum for pupils to read and write a range of text types, such as emails, letters, invitation cards, book covers and picture-cued stories.

In terms of **marking**, all teachers reported that they were expected to follow marking schemes designed by the school's subject panel. In some schools, teachers "*determine marking criteria together*" (T4); in others, they need to fine-tune the given rubrics (Extract 1) as well as analyze and interpret test results.

Extract 1

就咽個 paper 嘅老師擬題老師去定嘅，（即係無話 6 個年級固定晒噶？）無年級去定晒
噶。咁例如話 content 系 12 分，language 系 12 分，跟住 organization，有時有 d 有 format
噶嘛，譬如 letter 有 format 咁你咪唔同咗咯。譬如你譬如 article 無 format 噶嘛，咁所以你
咪可以 skip 咽個 part 嘅 format 嘅分數咁咯。 [There is an existing marking scheme for writing
but) the teacher who designs the test paper also needs to decide how to assign weight to different
aspects of writing. It is not set by the subject panel. For example, usually content and language
each takes up 12 points. When the task is letter writing, we also need to assign marks to the
target textual features *(i.e. the format features of a letter)*. When assessing a short essay, we can
skip the format part] (T1, italicized added by the researcher).

Two issues concerned both novice and experienced teachers. As Extract 3 indicates, at least some
schools did not provide necessary training or standardization sessions to their teachers, not even to
beginner teachers (T10 and T11) but assumed they could follow the given marking schemes by
themselves.

Extract 2

我們有張 marking scheme，但無人同我講過點比分，開會亦無講…通常都無 share 個情況
嘅…憑感覺咯. 其實我系呢睇翻佢地以往嘅分系大概邊個 range 咁樣，咁樣俾咯。[We do
have marking schemes, but no one told us what is expected for each banding, nor do we discuss
it in the department meetings. Honestly, I had no idea how to mark students' performance in the
beginning. So, I reviewed the previous marking results, got to know the usual range of scores
and marked my students' performance according to it] (T10).

Another issue is related to **communicating assessment results to parents and students.** Most
teachers believed being skilled in this area was essential yet challenging even for seasoned teachers.
Teachers need to explain not only the assessment results but also the decisions made based on the results.
Given the generally competitive school culture in Hong Kong, parents are keen to know their children's
academic performance compared to their peers. Teachers, however, were reluctant to reveal student
grades, fearing that this would intensify competition and bring about negative consequences. Protecting
student assessment results was the most frequently mentioned issue when teachers were asked about
ethical practices in assessment. They were vigilant towards the stress that assessment can bring to pupils
and reported that they discouraged unnecessary competition among students. For instance, teachers "*told
students not to look at or talk about others' results*" (T1, 2, 7). They also exercised caution when talking
to parents, such as "*refrain from disclosing confidential information such as the highest or lowest scores
to parents*" (T9). When information of this kind was sought by parents, teachers reported that they would
steer the focus of communication towards the follow-up actions that parents could adopt to improve
students' performance.

Consistent with Harding and Kremmel's prediction (2016), teachers, as the primary contacts in dealing
with key stakeholders, possessed adequate knowledge about the socio-educational contexts and local
values and keen awareness of the impact of assessment. The majority, if not all, of teacher participants
possessed at least Level 3 (i.e. procedural and conceptual literacy) knowledge on sociocultural values and
local practices (Taylor 2013). In comparison, their knowledge about assessment theory and technical
skills was not as good.

Perceptions of Assessment Theory and Concepts

Comparing to the practical know-hows of conducting assessment, teachers' attitudes towards
assessment theory and concepts were much less enthusiastic. When asked, they endorsed the importance

of having a good understanding of different **assessment purposes**, and this understanding should guide the design of assessment or the selection of assessment methods. Some teachers noted that novice teachers seemed to be unaware of the purposes of assessment, let alone capable of matching assessment strategies with specific purposes.

Extract 3

> 我覺得卷就一定出到噶，但係你話叫佢地諗背後嘅目的就未必會諗到咯，佢可能即係為測而測啊，背後嗰個意思可能都唔系好清楚咯。 [They - *novice teachers* - can design test papers, but they may not understand the purposes behind a particular assessment task. They design an assessment task only because they have to; they may not know very clearly why they assess this but not that] (T4).

Nonetheless, these teachers considered the key **assessment concepts** such as validity and reliability non-essential, though some felt having a good understanding could be helpful. Teacher 7 noted that the terms of "reliability" and "validity" were not used in schools though teachers were expected to adhere to these principles when conducting assessment. For instance, he added, the content of assessment should fall within the scope defined by the syllabus as well as be representative of the central curriculum. Teachers were also acquainted with the importance of rater consistency in marking (pertaining to reliability). Another teacher remarked that having 'common sense' about assessment would be sufficient. Their views suggest that the two assessment concepts, being closely tied to the long-established exam tradition in Hong Kong schools, had become tacit and shared knowledge of schoolteachers probably as part of their growing up in this system. Hence, they referred to them as 'common sense' and were ready to provide pertinent practices and procedures as evidence of their knowledge and skills in association with the two concepts.

In terms of **fairness in assessment**, the aspect most often mentioned by the teachers is associated with Special Educational Needs (SEN) students. While maintaining fairness in school assessment was upheld with little dispute about its importance, teachers' conceptualization of fairness was essentially equal treatment to all (Davison, 2010). Fair assessment is associated with standardization, such as using the same test papers and the same marking schemes with all students by the same teacher. Although some accommodation measures were adopted for SEN students, these were reminiscent of those implemented in standardized testing situations. Allowances were made, such as allowing extra time for testing, providing a separate test room, printing test papers with enlarged fonts and having teachers read out questions for the students.

Teachers were familiar with the concept of differentiated teaching (Yin Cai Shi Jiao, 因材施教), a Confucius educational notion that students of different learning aptitudes should be taught differently. However, the concept of differentiated assessment, e.g. adjusting assessment methods and difficulty in order to elicit a student's best performance on assessment, was still considered at odds with the notion of fairness. Although it has been advocated via the large-scale school-based assessment innovation introduced to Hong Kong secondary schools in 2006 (Qian, 2014), differentiated assessment had not found its way into primary schools.

Perception of Context-specific Challenges

Among the wide range of LAL components, teachers were most aware of their professional contexts and were vocal about constraints that they faced in conducting assessment. According to them, their practices were influenced by existing school assessment practices, stakeholders' concerns, as well as educational and assessment policies. All teacher participants reported that they were obliged to implement school policies in assessing students and that novice teachers should be proactive to learn and quickly adapt themselves to existing practices (T3, 5 and 7).

Extract 4

間間學校你入到去文化唔同啦，有 D 學校鍾意做 formative assessment 唔考試㗎，有 D 學校系要考紙筆，有 D 學校鍾意 perform 比家長睇，各種類型㗎都唔同咯。因為間間學校都唔同，我們轉一間學校都要重新學過另外一套 assessment 嘅嘢，間間學校嘅唔同。[School cultures vary considerably. Some schools prefer formative assessment and do not use exams, some require paper-and-pencil written tests, others like assessing students through performing in front of parents. Because assessment practices are widely different from school to school, we need to learn a new set of assessment practices when transferring to a new school] (T3).

Besides being aware of their accountability to existing school practices, the teacher participants also reported their accountability to students. In terms of designing test papers for pupils, all teachers stressed the importance of adjusting the difficulty levels of test items to cater for the ability of students. They reported adopting various strategies to achieve this, including selecting suitable source materials (T10), replacing difficult words with simple ones (T2), and giving clues and hints as part of test input (T2 and T6). Setting test papers at the right levels of difficulty was considered challenging as it required sufficient tacit knowledge about students' ability as well as a solid grasp of textbook and curriculum targets, which novice teachers and teachers who were new to a school tended not to have.

Extract 5

睇佢程度咯，難度咯… …主要系驚出得太深咯，因為你未摸熟晒呢間學校嘅學生嘅底。[I'll look at students' ability to adjust test difficulty. My major concern is that the test should not be too difficult because I don't know very well about the ability of the students in this school] (T4).

Extract 6

有時有 d 真係會覺得佢無系學生嘅角度出發黎諗咯，因為有時系可能根本老師望落份卷都會覺得，哈？點解呢一 part 突然間放咗系呢度㗎咁樣？或者呢一個，即係佢比嘅提示唔足以比學生估到呢度寫呢一個答案咁樣咯。[I feel new teachers are not thinking from the students' perspective. When I read the test papers they designed, I would wonder why this part or this item was put here, or sometimes there weren't enough hints for students to figure out the answers] (T1).

Also noted is that novice teachers tended to design assessment in isolation without due consideration of the curriculum as a whole. This is related to the core component of knowledge about the relationship between assessment and curriculum in Brindley (2001). Here the teachers stressed that the novice were not aware of the vertical alignment of the curriculum across different grades and the corresponding alignment of school tests across different grades. Teacher 9 said:

Extract 7

但係你需要有一個全盤嘅計劃。 譬如你呢條考左，咁三年班嘅老師會考 d 咩呢？因為經歷過一輪，之後你就會知道哦，原來你出呢條題目係會影響到三年班嘅老師… 譬如話你出 vocabulary 你出 comprehension 你難到好誇張㗎。仲難過三年班嗰 d 㗎。所以你要出到，即係出到咁上下，你就會抓到嗰個難處系邊㗎，嗰個程度拿捏到邊個位，或者啱你嗰個 level。[You need to think about the teaching and assessment holistically when designing assessment. For instance, if you assess this aspect (in Grade 2), what the teacher of Grade 3 will assess? … You will realize later that your choice has an impact on teachers of the other grades… For example, the vocabulary and comprehension tasks (in your tests for the second-graders)

can't be more difficult than those for the third-grade. So, you need to get the right levels of difficulty] (T9).

Parents were another factor which could affect teacher assessment practices. Teachers, however, appeared to be rather critical of parental influence. Teacher 3 stated, communicating with parents was one of her major challenges when she started teaching 13 years ago, and she still spent much of her work time dealing with parents, who, according to her, made far too many requests and complaints over trivial issues. A challenge she perceived was associated with inadequate support for students.

Extract 8

> 點講呢，家長會有要求啦，當然系啦，咁啊，可能同埋確實系，宜家嘅家長系比較溺愛小朋友嘅，我覺得，咁所以呢，他們可能會有小小野都會去投訴，系喇，小小野都投訴。咁同埋宜家小朋友系真係抗逆能力系比較差嘅，咁佢地好多時接受唔到一 d 失敗，系喇。咁同埋呢，即如果系教育上面呢，其實呢，宜家個別差異系多嘅，我覺得，個別差異系比以往系更加大嘅，因為系確實系多咗一 d 雙非嘅小朋友，咁可能呢，佢地系入一年班嘅時候呢，佢地系完全未學過英文嘅，咁但係就會變咗同香港小朋友會有一 d 差異咯。咁同埋因為宜家系多咗有特殊學習需要嘅小朋友嘅我都覺得，系啦，咁系支援呢 d 小朋友上面呢系吃力咯，我覺得。[Parents made requests, of course (it's their right), but probably because todays' parents spoil their children more than before, I think, they would launch complaints even for very trivial things, yeah, very trivial matters. Moreover, children's ability to resist frustrations and set-backs is worse than before. They often couldn't handle failures. Also, there is far more diversity among students than before. There are children coming from the mainland who only start learning English upon entering into Grade 1, hence have considerable gaps with local children (*who start learning English at K1 in kindergarten*). Also, we have more SEN children… it is much harder to support (such diverse) learning needs] (T3).

Similarly, Teacher 11 added that teachers marked test papers very carefully to avoid mistakes that might invite parents' criticism and complaints. When test papers were found too difficult for the whole cohort of students, adjustments were made to lower the marking standards so that the test scores would not be too low. This remark suggests that teachers' assessment practices are made accountable to parents primarily via their marking and feedback written on test papers and homework. Through this channel, parents can exert some influence on teachers' assessment practices.

Of all the contextual factors, perhaps those associated with school logistics were the most often reported. Resonant with previous studies (e.g., Qian, 2014), the interviewer heard teacher participants talking about their heavy workload, crammed teaching schedules and lack of administrative support for test paper design. Administrative support for test paper design varied considerably from school to school. In some schools, the task of designing test papers for different language skills was a team effort with different teachers responsible for different papers, while in others, the same task was given to one teacher for completion within a very short period of time. The former scenario provided better support for designing quality assessment tasks or test papers and potential opportunities for novice teachers to learn. In the latter scenario, however, novice teachers were exposed to a rather challenging task with little assistance. Consequently, some chose to play safe, copying previous test papers with minimal editing. Recalling her first test design experience, both Teacher 6 and 10 stated with some bitterness that they received minimal support from their schools; both resorted to adapt past papers.

Extract 9

> 完全無從入手咯。完全唔了解呢間學校想要 D 乜嘢嘅要求咁樣咯… …咁我會有一 D 舊卷睇到做 reference。 [I had no idea what to do (when I first designed test papers). I didn't know

what the school expected or required…so I reviewed the past papers and took them as reference] (T10).

Extract 10

果次就真係好辛苦嘅，因為從來無出過啦，然後我呢間學校系零 support 嘅，出卷呢方面系零 support 嘅，佢淨系會話比你聽你要出由呢一課至呢一課嘅所有野… … 咁所以對於我來講系新嘗試黎嘅，完全唔識嘅。 [My first time of designing test papers was exhausting because I had no experience with preparing test papers before, and there's no support from the school at all. They only told you to cover this lesson to that lesson…I had no knowledge about it at all] (T6).

According to the participants, schools usually had a quality assurance (QA) mechanism to maintain the quality of test papers, though its rigor varied. Teacher 9's school might be one of the most rigorous. After compiling a set of test papers, the teacher in charge was expected to pass the draft for review to her colleagues and the year coordinator, then to senior teachers and the panel chair in turn, and finally to the school principal. This meant that one test paper was often edited and revised 10 to 12 times before it was finalized. Such a QA mechanism placed enormous pressure on individual teachers assigned to design a test. While this certainly helped maintain the standard of school-based test papers, it also promoted conservatism because any innovative departure from established practices was risk taking, which demands considerable courage and confidence on the part of the teacher. Teacher 6 added that, more often than not, the feedback they received from the QA mechanism described above were so general and vague that they felt "*helpless*" and did not know how to respond or follow it up. They ended up becoming more concerned with "*what the panel wants*" rather than how to improve the test paper.

Quantitative Findings

Overall, pre-service teacher participants reported moderate levels of confidence towards the four areas of assessment practices and skills perceived to be important by practicing teachers. On the four-point scale, where Level 0 stands for illiteracy, Level 1 stands for nominal literacy, Level 2 stands for functional literacy, and Level 3 stands for procedural and conceptual literacy, their self-ratings were mostly between the nominal (Level 1) and functional literacy (Level 2). They were most confident with their generic assessment design skills and their competence of using existing assessment instruments and providing feedback to students and parents, but less confident in their skills to design assessment tasks and rubrics for assessing specific language skills and least confident with interpreting test scores. Comparing this profile with Taylor's expectation, especially her view that teachers should reach Level 3 on technical skills of assessment (2013), the student teachers did not reach that level of sophistication.

Competence in Using Existing Assessment Tools

Participants reported moderate levels of confidence about using appropriate assessment methods to serve the four common assessment purposes (Mean = 1.71). In relation to the purposes of 1) evaluating students' achievement, 2) diagnosing problems in students' learning, and 3) streaming students of different proficiency, the majority reported feeling adequate at Level 2 (59.6%, 69.2% and 63.5% respectively). However, up to 61.5% reporting that they were not skilled at all (Mean = 1.31) in conducting admission assessment with school applicants, the fourth main assessment purpose that received the lowest rating of all.

TABLE 2
Item Distribution, Content and Descriptive Statistics

Dimensions	item #	content	mean	SD
competence of using existing assessment tools (12 items)	*8a-d*	*using existing rubrics to assess speaking in English*	*2.11*	*0.61*
	10a-d	*using existing rubrics to assess writing in English*	*1.93*	*0.63*
	3abcd	*using/selecting assessment for specific purposes*	*1.71*	*0.65*
generic assessment design skills and competence (17 items)	4g	write clear instructions for assessment tasks	**2.21**	0.63
	4a	align the content of assessment tasks with learning objectives/syllabus	**1.94**	0.64
	5a-j	familiar with broad assessment types	1.88	0.61
	4b-f	consider different context factors in assessment design	1.81	0.52
skills to design assessment tasks for specific language skills (19 items)	*6a-g*	*designing tasks for assessing reading and listening*	*1.96*	*0.62*
	9a-d	*designing tasks for assessing writing*	*1.81*	*0.72*
	7a-f	*designing tasks for assessing speaking*	*1.77*	*0.74*
	4h-i	*design rubrics for assessing productive skills*	*1.52*	*0.78*
ability to interpret test scores and provide feedback (5 items)	4j-k	interpret raw test scores	1.44	0.75
	4l	interpret band scores in standardized tests	1.35	0.76
	4m-n	provide feedback to students and parents	**1.90**	0.75

Generic Assessment Design Skills and Competence to Follow Key Principles

Student teachers showed higher levels of confidence in ensuring content validity and taking contextual factors into consideration (Mean = 1.94, 1.81). A majority (76.9%) felt themselves adequate (Level 2) for or even very good at aligning the content of assessment with syllabus objectives to ensure content validity (Level 3). Similarly, they considered themselves capable of accommodating various contextual factors (such as students' language abilities, test lengths and available time, and the coherence of assessment tasks across grades), with 69.2% feeling adequate and 5.8% feeling very good at it.

They also reported feeling capable of writing clear instructions for assessment tasks on test papers (Mean = 2.21). The average rating of their competence in using appropriate language in writing assessment tasks was higher than for the other components (63.5% at Level 2 and 28.8% at Level 3), showing strong confidence in their own English language proficiency.

However, they considered their knowledge of assessment methods and item types inadequate. Of the ten assessment types identified from phase 1, they were confident with only four of them, namely, dictations, homework, class interaction and class observation (Mean = 2.31, 2.19, 2.12 and 2.04 respectively). Formative assessment techniques, such as using questionnaires[1], learning journals and portfolios, received much lower ratings (Mean = 1.58, 1.58 and 1.54), with nearly half of them estimating having only nominal literacy (Level 1: 46.1%, 44.2% and 46.2%). This finding indicates that LAL training courses can spend more time on formative assessment techniques.

Skills to Design Assessment Tasks for Specific Skills

Consistent with the finding in the previous section, student teachers' confidence regarding the design of various types of test item vary. For example, item types used in reading and listening papers (Mean = 1.96) received slightly higher ratings than those used to assess writing and speaking (Mean = 1.81, 1.77). Also consistent is that traditional assessment types tended to receive higher ratings than more innovative task types. For reading and listening, student teachers considered themselves to have adequate skills for the *fill-in-the-blanks*, *multiple choice questions* and *short answer questions* (Mean = 2.31, 2.17, 2.17),

---

[1] Simple questionnaires can be used to collect student information such as their use of reading strategies and their frequency of reading books at home.

less so with item formats such as *proofreading, mind-maps* and *jigsaw puzzle* (Mean = 1.79, 1.79, 1.50), although the latter set of item types were also widely used in primary schools in Hong Kong. Similarly, when it comes to speaking, pre-service teachers were also more confident in designing conventional assessment tasks, such as *reading aloud, Q&A, individual oral presentation* and *picture description* (Mean = 2.10, 1.92, 1.81 and 1.77), than the more innovative ones, such as *group discussion and drama performance* (Mean = 1.71, 1.31). In terms of assessing writing, they felt competent in designing *guided writing tasks* and *picture description tasks* (Mean = 2.04, 1.96), much less so with *parallel writing*[2] and *creative writing* (Mean = 1.71, 1.54).

Overall, student teachers had a rather low level of confidence in designing rubrics to assess productive skills (Mean = 1.52). About half claimed to have either no knowledge at all (Level 0, 7.7%) or only have nominal literacy (Level 1: 42.3%). Though, most believed they were able to follow existing rubrics to assess speaking and writing (Mean = 2.02), they seemed to have more confidence in assessing speaking than writing (Mean = 2.11 vs. 1.93). Their self-ratings for assessing different aspects of speaking were similar, with intonation rated the lowest (Mean = 1.98). In terms of assessing writing, they were most confident about assessing the language of writing (Mean = 2.12), and least confident about assessing the creativity in writing (Mean = 1.77). For the latter aspect, more than 34.6% indicated that they were only aware of it. The findings are not surprising given the little training most students had received in terms of assessing language skills. Students were relatively confident in using the rubrics probably because rubrics have been widely used, from primary to tertiary, in the Hong Kong education system.

Interpreting Test Scores and Providing Feedback

Despite the low confidence in their competence in interpreting assessment scores, respondents generally considered themselves to be competent in providing suggestions and feedback to parents and students. They were slightly more confident about interpreting raw scores (Mean = 1.44) than interpreting bandings in standardized tests (Mean = 1.35) and appeared to be rather comfortable with communicating assessment results and providing feedback (Mean = 1.90). More than half (55.8%) felt they were able to provide assessment feedback to parents (Level 2) and 19.2% believed they were "very good at it" (Level 3).

Discussion

In discussing the findings, we make references to our research questions as well as three areas where further research is needed (Inbar-Lourie, 2013), namely, the breadth of the LAL framework, the core components, and the language component.

Our first RQ seeks practicing teachers' views on the core LAL necessary for novice teachers. In previous research, practicing teachers' perceptions were compared with those of tertiary instructors and interesting differences were identified (e.g., Jeong, 2013). In the present study, their perceptions were compared to those of the pre-service teachers and two major disagreements were noticed. Firstly, the practicing teachers considered interpreting and following marking schemes for assessing students' speaking and writing skills to be a potential weakness of the novice. However, our would-be teachers generally considered themselves competent in this regard. We think their confidence probably stems from having direct and extensive experience with marking rubrics from growing up in an education system driven by examinations; as they were accustomed to teachers' using rubrics to assess their work, they felt confident that they could do it too.

---

[2] Parallel writing refers to providing model sentences in the writing instruction and requiring students to produce similar (parallel) sentences in their writing.

Student teachers also disagreed with seasoned teachers in terms of their ability to accommodate contextual factors and design context-appropriate assessment tasks. They were rather confident that they could fit in with little difficulty. However, in our view they may have underestimated the complexity of real classroom and school contexts, a phenomenon observed and described by DeLuca and Klinger (2010) as pre-service teachers' unrealistic optimism. Indeed, the junior teachers who we interviewed confirmed having encountered much difficulty with their assessment duties in their first two years.

The disagreements between the two teacher groups indicate the evolving nature of teachers' LAL at different stages of their careers. Typically, being optimistic before entering practice, struggling during the first two to three years, becoming increasingly complying to standardized testing norms in schools in the first five years (Coombs et al., 2008), until establishing themselves as professionals and began to reprioritize formative and differentiated assessment. On one hand, the availability or absence of school support seems to play an important role in facilitating or impeding the development of their practical assessment skills. On the other hand, their pre-service training of innovative assessment theories only exerts influence on them when they have gained solid grasp of context knowledge and able to see suitable application of theories. To easing their pain of transiting from academia to practice, case analyses of school assessment practices and sharing of guest lecturers invited from schools could be provided to pre-service teachers, preparing them for the myriad of contextual factors that they need to be aware of and in developing corresponding skills and strategies.

Relating to the breadth and the core content of a LAL course, we found that the technical knowledge and skills of test design and analysis were still regarded as the most useful skills to be acquired, resonant with previous research (Fulcher, 2012; Lam, 2015; Vogt & Tsagari, 2014). However, student teachers did not consider themselves to be adequately prepared in this regard. Both groups concurred that teachers should be exposed to a diverse range of test items commonly adopted in school-based tests and that novice teachers needed to strengthen their technical skills, especially those related to analyzing and interpreting test scores. They should also achieve better understanding of the relationships between assessment purposes and test design.

Both groups agreed on the importance of designing and conducting formative assessment, a finding that indicates a gradual acceptance of FA by teachers after more than a decade since its introduction to Hong Kong schools, and correspondingly, the need to grant a duel priority to both traditional testing and innovative FA techniques in pre-service training. The demand for an expanded range of LAL skills, however, poses a challenge to teacher training courses, which tend to have very limited contact hours. Expanding one LAL component inevitably necessitates comprise elsewhere. As reported in Lam (2015), the student teachers who completed LAL courses incorporating both priorities did not feel they had reached an adequate level of competence in either area. Lam related this to a lack of policy recognition of LAL in teacher qualification certification, the marginalized status of LAL in teacher training programs, as well as the instructional mode (i.e. teacher-front lectures). He suggested incorporating LAL into the assessment of the teaching practicum for pre-service teachers. Other LAL researchers recommended adopting a process-based and student-centered approach to course delivery (Kleinsasser, 2005) and harnessing the power of web technologies to make LAL resources available online (Malone, 2017).

We have no doubt that these research-based insights would make pre-service teacher training more accessible, effective and efficient, but we have reservations whether the daunting task of equipping teachers with sufficient LAL could be accomplished by pre-service training alone, especially in light of its marginal status. In addition to pre-service training that provides the essential knowledge base, school administrators and subject panels have an important role to play in supporting the continuing development of novice teachers' LAL. Concurring with Koh (2011), we believe that ongoing, sustained assessment training on the job has the potential to promote teachers' AL/LAL along with other aspects of their professional development. Continuous assessment training at different stages of a teacher's professional careers should be in place, synergizing the efforts of both universities and schools.

Moreover, a more realistic LAL profile should be specified for pre-service teacher training, which should be different from that for experienced teachers as specified in Taylor (2013). Teacher candidates

should be aware of existing assessment practices and have more opportunities to discuss both macro sociocultural, educational, and micro school contexts, as well as practical strategies to deal with issues relating to ethics and fairness in assessment and techniques to ensure consistency in marking. On the other hand, given the considerable variance among schools in their assessment practices, context-specific practices, such as marking against pre-set school rubrics, following schools' requirements for item variety and formatting, may be more effectively learnt on the job.
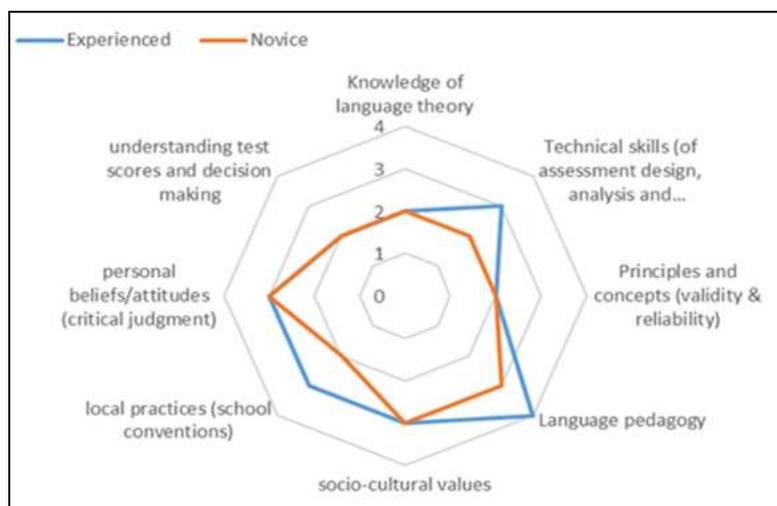

Figure 1. LAL profiles for novice vs. experienced language teachers.

Figure 1 presents a LAL profile stipulated for pre-service and novice teachers, in contrast to the one stipulated by Taylor (2013) for experienced teachers. While the conceptualization of Taylor (2013) was based on her decades of experience as a teacher educator, our stipulation was based on the empirical findings of the present study. It is our view that the areas that require extensive practice and experience could be lowered by one level for pre-service and novice teachers. Specifically, the dimensions of language pedagogy and local practice could be adjusted from Level 4 (expert or multi-dimensional literacy) to 3 (conceptual and procedural literacy) and that of technique skills from Level 3 to 2 (functional literacy). On the other hand, expectations towards the other five dimensions that do not require extensive teaching practice (namely, knowledge of language theory, understanding test scores and decision-making, socio-cultural values, and personal beliefs and attitudes for critical judgment) can remain at the same levels as those for experienced teachers.

Personal beliefs and attitudes for critical judgment is crucial to avoid what Tsagari and Vogt (2017) described as 'uncontested adoption' of existing practices or reverting to 'test as you were tested' (p. 54), which inhibited teacher development as well as the adoption of innovative assessment ideas in schools. Teacher training programs should enable future teachers to acquire the knowledge about contemporary theories of assessment and language (Inbar-Lourie, 2017) and introduce innovative assessment ideas and practices so that student teachers become capable of critical evaluation of existing practices in schools (such as T3 and T6) and to innovate and initiate changes when they are ready.

Thus, although certain aspects of assessment practices are better learnt on the job, knowledge of current assessment theories and innovative assessment practices, personal beliefs, critical and independent judgment about whether existing assessment practices offer the long-term benefits for teaching and learning should be fostered in the pre-service stage. Without such competence, novice teachers may have to rely on school mentors or their own educational experience, which perpetuates existing assessment practices and leaves little room for innovation (Vogt & Tsagari, 2014).

Conclusion

Following the socio-cultural turn, the language assessment community has recognized the importance of in-depth investigation within specific socio-educational contexts to inform both international practices and local needs. The present study is situated in a context where LAL education remains under-investigated and under-documented. It investigated pre-service teacher training needs through capitalizing on the context knowledge of in-service teachers. Useful information regarding existing practices in schools as well as teacher candidates' readiness for their upcoming assessment duties could inform the design of appropriately prioritized pre-service training syllabus.

Like other research in education, this study is not without its limitations. Firstly, it is a small-scale research. We conducted in-depth interviews with teachers from eleven primary schools and surveyed 55 pre-service teachers. As teacher LAL evolves along different stages of their careers, further research can consider conducting continuous and longitudinal monitoring of teachers' LAL development upon graduation and a few years after entering the profession. Another limitation of this study is that it only investigated teachers' self-perceived competence, which, though a useful indicator of their readiness, is constrained by the self-reported nature of the data. Additional data regarding teacher candidates' actual assessment practices, if available, would add strength to the findings. Further research can also consider collecting such data from teaching practicum, through analyzing student teachers' lesson plans, observing their assessment activities, or analyzing test papers that they developed.

Notwithstanding its limitations, we believe this research is a useful addition to the literature on LAL as one of the few investigations into an area under-researched and under-documented.

Acknowledgment

The Authors

Qin Xie (first and corresponding author) is Assistant Professor at the Education University of Hong Kong. She has published on peer-reviewed journals such as *Language Testing*, *Language Assessment Quarterly*, and *Assessing Writing*, focusing on washback, test validation, and diagnostic language assessment. She is also interested in Structural Equation Modelling, Rasch Modeling, and Cognitive Diagnostic Modeling including their applications in language education and assessment.

Department of Linguistics and Modern Language Studies
Education University of Hong Kong
B4-2/F-12
NO. 10 Lo Ping Road, Tai Po Market,
New Territories, Hong Kong
Email: qxie@eduhk.hk

Siyan Tan (second author) is a frontline English language teacher working in a primary school in Hong Kong. She is interested in teacher assessment literacy, language assessment and teaching pedagogy.

North Point Methodist Primary School,
2A Pak Fuk Road, North Point, Hong Kong
Email: tsycynthia@gmail.com

References

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Berry, R. (2011). Assessment trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Principles, Policy & Practice, 18*(2), 199-211.

Brindley, G. (2001). Language assessment and professional development. In C. Elder et al. (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp.126-136). Cambridge, UK: CUP.

Boyles, P. (2005). Assessment literacy. In M. Rosenbusch (Ed.), *National assessment summit papers* (pp. 11-15). Ames, IA: Iowa State University.

Coombs, A., Deluca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, *71*, 134-144.

Curriculum Development Institute. (2004). *Promoting assessment for learning in English language education*. Hong Kong: Hong Kong Government Printer.

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing, 25*(3), 327-347.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419-438.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251-272.

EDB. (2018). *General information on DSS*. Accessed 6 June 2018 http://www.edb.gov.hk/en/edu-system/primary-secondary/applicable-to-primary-secondary/direct-subsidy-scheme/index/info-sch.html

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113-132.

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, *21*(6), 607-621.

Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 413-428). Berlin & New York: Mouton de Gruyter.

Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and assessment needs. Part one: General findings*. Gothenburg, Sweden: European Association for Language Testing and Assessment. Accessed 7 June 2018 *http://www.ealta.eu.org/documents/ resources/ survey-report-pt1.pdf.*

Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: Expectations and needs of pre-service English language teachers. *ELT Research Journal*, *4*(2), 111-128.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, *25*(3), 385-402.

Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing*, *30*(3), 301-307.

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 257-270). Cham: Springer International Publishing.

Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing, 30(*3), 345-362.

Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555-584.

Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, *22*(3), 255-276.

Kömür, S. (2018). Preservice English teachers' assessment awareness: Level of readiness for classroom. *The Journal of Language Teaching and Learning, 8*(1), 109-121.

Kleinsasser, R. C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect, 20*, 77-102.

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing, 32*(2), 169-197.

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing, 30*(3), 329-344.

Malone, M. E. (2017). Training in language assessment. In E. Shohamy et al. (Eds.), *Language testing and assessment*, *encyclopedia of language and education* (pp. 225-239). Cham: Springer International Publishing.

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. San Francisco: John Wiley & Sons.

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, *30*(3), 381-402.

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator, 46*(4), 265-273.

Qian, D. D. (2014). School-based English language assessment as a high-stakes examination component in Hong Kong: Insights of frontline assessors. *Assessment in Education: Principles, Policy and Practice*, *21*(3), 251-270.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(3), 309-327.

Stabler-Havener, M. L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Working Papers in Applied Linguistics and TESOL*, *18*(1), 1-22.

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

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, *30*(3), 403-412.

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, *6*(1), 41-63.

Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education, 30*(3), 749-770.

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374-402.

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149-162.

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, *16*(4), 323-342.

Appendix A

Main Interview Questions

Warm-up	1. When did you first come to work in this school? Had you been working in other schools before joining this one?
Assessment-related duties	2. What are your major duties here? Do you have any challenges in your work now? 3. What duties are usually allocated to new teachers who have just graduated?
Assessment purpose	4. Can you describe the assessment policy of this school for the English subject? What are the usual purposes for giving assessment to students?
Assessment design	5. What is the usual practice in your school in preparing for exam/test papers? Have you ever been involved in preparing exam papers or assessment tasks? Can you recall and describe the first time (and the latest time) you were involved? 6. Apart from those from the school, are there any other assessment types that you have used with your students?
Marking	7. What about the marking, what is the usual practice in your school in marking? Do you think the marking procedures good enough? Why or why not?
Communication of assessment results	8. After marking, what do you usually do? 9. I learnt that you've just completed one major exam, how do your students do this time? Do you have any plan to follow up with the results? 10. What's the usual practice in your school with regards to releasing assessment results to students and parents?
Assessment fairness and ethics	11. Are there any special arrangements for students with special needs (SEN)? How about with students newly emigrated from the mainland, South Asian students?
Assessment knowledge and theory	12. What aspects of your assessment knowledge and skills do you want to improve? 13. Pre-service teacher education courses on language assessment tend to introduce the key concepts such as the validity and reliability of testing and assessment. How do you think about these concepts and theories? Are they helpful for the assessment-related duties in your school? Why or why not?
Novice teachers' assessment competence and training needs	14. Can you recall when you were a freshly graduated English teacher? At that time, were you able to handle your assessment-related duties? 15. Have you ever worked with freshly graduated English teachers? What do you think about their ability to handle assessment? 16. In your view, what assessment skills are necessary for a freshly graduated teacher?

Appendix B

Main Questionnaire Items

Items 1-2 collect students' training background

1.   Have you ever attended any assessment-related courses? If yes, please indicate the course(s) you have taken.
2.   Have you ever designed a set of language assessment papers/tasks? If yes, please indicate the occasion(s) when you designed the assessment tasks

Items 3-10 require students to rater their competence on a 4-point scale. The labels of each point on the scale are phrased separately to be appropriate for each set of items but are equivalent to the following:

- Level 0 illiteracy (I don't know …);
- Level 1 nominal literacy (I know it but I am not skilled at all…);
- Level 2 functional literacy (I have adequate knowledge and skill for…);
- Level 3 procedural/conceptual competence (I am very good at …)

3.   Assessment are used to serve various purposes in schools. Please indicate how well you know each of the following purposes or how skillful you are in using different assessment to fulfill each purpose
    a)   To evaluate students' achievement
    b)   To diagnose problems in students' learning
    c)   To stream students of different English proficiency
    d)   To select students for admission
4.   Please indicate your level of skills regarding to the following aspects of assessment:
    a)   Aligning the content of assessment with learning objectives on the syllabus
    b)   Considering different factors in designing assessment (e.g. authenticity, practicality, validity and reliability)
    c)   Designing assessment or adjusting the level of difficulty of existing assessment according to students' ability
    d)   Designing or adjusting a language test according to available testing time.
    e)   Designing assessment tasks to achieve task coherence across a test paper.
    f)   Designing a test in light of the tests of the lower and upper grades
    g)   Writing simple and precise instructions for assessment tasks
    h)   Designing rubrics for assessing speaking
    i)   Designing rubrics for assessing writing
    j)   Interpreting raw scores of a test
    k)   Interpreting raw scores of standardized tests (e.g. TSA)
    l)   Interpreting bandings of standardized tests
    m)   Providing feedback to parents about students' assessment results
    n)   Providing feedback to students about their assessment results

5.   Please indicate your level of skills regarding to designing the following assessment types
    a)   tests
    b)   dictations
    c)   class observation
    d)   class interactions (e.g. questioning)
    e)   questionnaires
    f)   homework
    g)   portfolios
    h)   project work

   i)   journals (e.g. learning journals; students' diaries)
   j)   performance tasks (e.g. student presentation)

6. Please indicate your level of skills in designing the following tasks for **assessing reading and listening**
   a) Multiple choice questions
   b) Fill-in-the-blanks
   c) Short answer questions
   d) Long answer questions (i.e. giving answers in complete sentences)
   e) Proofreading or editing task (i.e. asking students to correct mistakes)
   f) Mind map
   g) Jigsaw puzzle

Please indicate your level of skills **in assessing speaking and writing**, in terms of

7. Designing the following speaking assessment tasks
   a) Reading aloud
   b) Picture description
   c) Individual oral presentation
   d) Group discussion
   e) Q&A (i.e. students answer the questions asked by a teacher)
   f) Drama

8. Making use of **existing rubrics** to assess different aspects of speaking, such as
   a) Accuracy
   b) Fluency
   c) Pronunciation
   d) Intonation

9. Designing the following writing assessment tasks
   a) Picture description
   b) Parallel writing (e.g. students are required to write parallel to the given model writing)
   c) Guided writing (e.g. students are required to respond to guiding questions)
   d) Creative writing (e.g. students are asked to write an ending to a story)

10. Making use of **existing rubrics** to assess different aspects of writing, such as
    a) Content (sufficient and relevant details)
    b) Language (accuracy, fluency etc.)
    c) Organization
    d) Creativity