

Project Title : An exploratory study of language assessment training in Hong Kong: Trends, quality and development

Grantee : Hong Kong Baptist University

Principal Investigator : LAM Che-keung, Ricky
Department of Education Studies
Hong Kong Baptist University

Co-investigator : LI Siu-cheung, Sandy
Department of Education Studies
Hong Kong Baptist University

Final Report

by

Principal Investigator

An Exploratory Study of Language Assessment Training in Hong Kong: Trends, Quality and Development

Abstract

Assessment literacy (AL) is about teachers' assessment knowledge, skills and principles when conducting internal and external assessments. There has been no shortage of empirical studies investigating how serving teachers perceive, understand, and utilize AL in various educational contexts. Despite this, not much has been done to understand the trends, quality and development of language assessment training catered for preservice English language teachers in non-Anglophone contexts such as Hong Kong. The aims of this project are threefold. First, the study investigated the overall language assessment training landscape in Hong Kong. Second, it examined facilitating and constraining factors that affected the quality of university-based language assessment training in five Hong Kong teacher education institutions. Third, the study evaluated the extent to which the current language assessment training could equip preservice teachers with knowledge and skills when they transitioned into first-year full-time teaching. Data collection included a questionnaire, preservice teachers' and university instructors' interviews, classroom observations across two years (final-year practicum and first-year teaching), and documentary analysis of the five local teacher education programmes. An exploratory case study approach was adopted to generate in-depth and illuminating data, which significantly contributed to a comprehensive makeover of content and delivery in language assessment training. Findings reflect that the

university-based assessment training remained relevant, but preservice teachers felt that they found peer learning through mentoring, observations, professional dialogues, and standardization procedures more constructive. Recommendations suggest provision of further hands-on opportunities for preservice teachers to create and critique authentic assessment papers both in coursework and in practicum.

Keywords: assessment literacy, language assessment training, English language education, preservice teachers, assessment literacy trends

Introduction

To start with, the relevance of the project to language education development in Hong Kong is described below. First, despite extensive research on pedagogical innovations, not much has been done to investigate how teachers master the state-of-the-art knowledge and skills in using assessment to enhance teaching and learning in their work contexts. Given the prominent role of assessment in language education, this project surveys the language assessment training terrain in the five teacher education institutions, which provide teacher trainers, university administrators, policymakers, and assessment scholars with unique insights into the benefits and issues of language assessment education in Hong Kong. Second, from educational assessment literature, there has been a constant theme arguing what constitutes succinct assessment competencies and well-defined core standards in assessment

literacy to which teacher educators, policymakers and Education Bureau personnel should conform. On this note, the project reviews the content and delivery of language assessment training in the Hong Kong teacher education institutions and substantiates whether this content converge with the core assessment literacy standards advocated by international scholars. Third, much has been done to understand inservice teachers' perceptions towards assessment literacy, yet scholars know little about how preservice teachers perceive the effectiveness of course-based assessment training and in what ways this mode of training can sustain after they become full-time English teachers. This project fills this gap by exploring the extent to which university-based assessment training could equip preservice teachers with contextualized knowledge and skills in conducting classroom-based assessment and utilizing test results for improving teaching and learning. Drawing upon both quantitative and qualitative data, the study identifies the assessment needs of preservice teachers in Hong Kong, which enable the PI to compile timely project findings for wider dissemination.

Review of literature of the project

This literature review discusses three aspects of assessment literacy scholarship including language assessment training landscape, factors influencing the quality of language assessment training, and development of teacher assessment literacy. Before we review the literature, we define two key terms - assessment literacy and language assessment training.

Assessment literacy typically refers to a continuum of differential levels of assessment knowledge, skills and principles in specific subject disciplines (Davies, 2008). In English language education, scholars commonly use a term “language assessment literacy” to imply teacher understanding and acquisition of measurement theory in language testing, practical skills to assess pupils, and social perspectives of assessment (Fulcher, 2012). In this study, we refer assessment literacy to preservice teachers’ mastery of knowledge, skills and principles in the Hong Kong language teacher education context. For language assessment training, it takes an array of modes to equip teachers with assessment knowledge and skills, namely university-based lectures, teaching practicum, text-based resources, professional development events, and online courses.

Language assessment training landscape

In the past decades, language teacher education has primarily focused on training preservice teachers to manipulate up-to-date instructional theories rather than to acquire sound assessment knowledge and skills to support student learning (Weigle, 2007; White, 2011). In many educational jurisdictions, teacher education programmes include child psychology, subject teaching, micro-teaching, teaching practicum, immersion programmes, but rarely cover any content of language assessment except in a course on assessment in general education (Hamp-Lyons, 2017a). As argued by DeLuca and Johnson (2017), assessment or

language assessment is always a neglected area for study. Because of that, it is not surprising to see that language assessment could be the least taught area in preservice teacher education programmes.

Since the early 1990s, Stiggins (1991) identified the issues and challenges when teachers were not assessment capable in the US school settings. One of the key issues is that without sound assessment knowledge and skills, student learning may sacrifice and the overall quality of education could cripple, which affect a country's long-term social and economic development. Despite the rise of global assessment reforms like using assessment to support learning, teacher expertise in this formative aspect has progressed somewhat slowly because of individual, professional, and societal barriers (Hamp-Lyons, 2016; Stiggins, 1995, 2014). First, frontline teachers tend to be afraid of measurement theories and do not want to be involved in any external assessments. Second, there are insufficient instructional hours to assess well and effectively. Third, the public still trusts that teachers, albeit lacking assessment training, are skillful in differentiating various levels of student performance. Further, teacher's impoverished assessment skills in conducting classroom-based assessment for promoting learning remain a cause for concern (Coombe, Troudi, and Al-Hamly, 2012; Qian, 2014).

Globally, there are studies reporting a lack of adequate assessment training in preservice teacher education. In the US, Schafer (1993) found that only half of teacher education programmes included a course on assessment skills. Thus, he advocated an inclusion of sufficient coursework components on assessment into the domain of teacher training. In China, Jin (2010) indicated that while the nationwide reform promoted formative assessment, teachers lacked basic skills and understanding of alternative assessments owing to the absence of language assessment training in their undergraduate studies. In surveying seven European countries, Vogt and Tsagari (2014) revealed that the participating foreign language teachers displayed certain levels of competency in language assessment. Yet, they discovered that because of insufficient training, most of them acquired the assessment skills on the job or used ready-made teaching materials as assessments. In Singapore, Koh (2011) found that the experimental group of teachers who engaged in designing classroom-based assessment and rubrics had enhanced levels of assessment literacy towards the end of the project. Nonetheless, the participants lamented that they did not receive assessment training in their preservice teacher education programmes. In the United Arab Emirates, Troudi, Coombe, and Al-Hamly (2009) reported that 21 EFL teachers were not involved in the assessment decision-making process and excluded from test design. The authors suggested that more training in the rationale and practices of classroom-based assessment be provided. In view of these studies, there is a pressing need to examine the language assessment training landscape

in Hong Kong, given the government has advocated formative assessment over a decade and its actual implementation remains questionable. The next sub-section discusses related factors that affect the quality of language assessment training.

Factors affecting quality of language assessment training

Concerning the factors that influence the quality of language assessment training, we divide the subsequent discussion into three categories, i.e. micro (teacher factors), meso (resource-related factors) and macro (system/policy factors) levels. The micro level refers to teacher conceptions of assessment capability; the meso level is about multiple forms of resources used in language assessment training; and the macro level concerns how a larger cultural context affects the mastery of knowledge, skills and principles in language assessment.

In this part, we look into the interplay between teacher conceptions of assessment (the micro-level) and language assessment training. Volante and Fazio (2007) investigated preservice teachers' conceptions of assessment literacy development in one teacher education programme, finding that low levels of self-efficacy beliefs were recorded in each year of the programme. Based upon another survey study, DeLuca and Klinger (2010) revealed that the participants were generally more confident in performing summative assessment than

formative assessment. More recently, Deneen and Brown (2016) reported that while 32 preservice and inservice teachers had improvement in assessment knowledge and skills, their beliefs in using assessment to serve the accountability purpose remained predominant. Hence, they tended to stick to conventional modes of assessment instead of trying out alternative assessments. From these studies, preservice teachers' conceptions are heavily shaped by their former perceptions and experiences of assessment rather than language assessment training. Moreover, some preservice teachers lack confidence when it comes to evaluating student learning formatively.

The meso-level factors consist of text-based resources, course-based resources and professional development events. In their study, Brown and Bailey (2008) found that between 1996 and 2007, the contents of language assessment courses are almost identical, implying that the field of language testing was expanding steadily but not radically. For the textbook trends, Davies (2008) observed that many current commercial textbooks included assessment principles, such as ethics, fairness and test impacts, and practical materials on implementing classroom-based assessment. Notwithstanding these well-informed text-based resources, Davies claimed that teachers should have adequate opportunities to try out testing theories with language learners. In Malone's (2013) study on an online assessment tutorial, language testers and language teachers had diverse views towards the usefulness of the course, with the

former emphasizing the fidelity of definitions and appropriate test use and the latter expecting user-friendliness of common assessment ideas. Koh et al. (2018) revealed that in-house professional development activities would enhance teacher assessment literacy, provided that teachers need specific learning in the design and use of authentic classroom assessment tasks, which aligned with nationwide educational reforms.

This part entails the macro-level factors, which illustrate how a wider testing regime would affect language assessment training. In Xu and Liu's (2009) study, they identified that an exam-driven culture was likely to affect teachers' assessment practices. In their study, the instructor adopted a "play safe" mentality to assess her students, as she was afraid not to conform to the common testing practices in her university. Likewise, Gu (2014) pointed out that high-stakes exams in China, namely *Gaokao* defined the instructional and assessment practices in the high school context. He argued that in order to facilitate curricular reforms, the importance of incorporating pedagogical content knowledge and assessment literacy as essentials of assessment training was indispensable. In Hong Kong, Qian (2014) investigated how English language teachers perceived and implemented the School-Based Assessment in the senior secondary-level classrooms. The participants felt that they likened the School-Based Assessment to high-stakes testing as students and parents were preoccupied with marks. Taken together, the micro, meso and macro-level factors would play a significant

role in determining *what* constitutes the best language assessment training content for preservice teachers and *how* this content is most properly delivered to them. The following sub-section delineates the development of teacher assessment literacy in terms of its definitions, standards and measures.

Development of teacher assessment literacy

While there have been studies investigating teacher assessment literacy, its definition remains conceptually complex. Fulcher (2012) challenged DeLuca and Klinger's generic definition of assessment literacy, arguing that they neglected the social, historical, and cultural perspectives of language assessment when it was applied in authentic classrooms and large-scale testing environments. Additionally, Inbar-Lourie (2013) contended that scholars still had not reached an agreed-upon definition of assessment literacy, since it involved specific testing/assessment knowledge, skills and principles; language system and use relating to classroom-based assessment; communications between language testing professionals and practitioners; and differential levels of assessment expertise. Taylor (2013) concurred that assessment literacy was a multi-layered construct, which referred to both knowledge-based and skill-based competence. However, the profile and developmental stages of assessment literacy varied, depending on the wants and expectations of different stakeholders.

Besides definitions, the trends of assessment literacy standards have become questionable. Based upon a review of 15 assessment literacy standards, DeLuca et al. (2016) showed that there had been a gradual shift in these standards, namely from the psychometric (standardized testing) to socio-constructivist paradigms (classroom-based assessment). Despite this evidence, they challenged an assumption that most of these standards paid little attention to the learning-oriented aspect of assessment, which would restrict the development of teacher competence in language teaching and assessment education. Xu and Brown (2016) proposed a “teacher assessment literacy in practice” framework, which embraced essential components of assessment literacy standards, one of which was teacher learning as assessors to get involved in decision-making processes for formative assessment. In the US context, Brookhart (2011) stated that the assessment literacy standards were outmoded in (1) conceptions of formative assessment practices, and (2) knowledge and skills about using assessment to advance learning in accountability testing settings.

Thus far, scholars have adopted a number of assessment literacy inventories, e.g. Teacher Assessment Literacy Questionnaire to measure the level of teacher competence in performing classroom-based assessment. Despite these inventories, there has been no rigorous measure that precisely evaluates various levels of assessment beliefs, knowledge, skills, and principles,

except Taylor's (2013) framework of assessment literacy profiles for four constituencies, namely test writers, classroom teachers, university administrators, and language testers. Taylor's (2013) framework provided a starting point for measuring assessment literacy, yet Hamp-Lyons (2017b) argued that almost all stakeholders needed to be assessment literate in order to make sense of internal/external assessment results including parents. Not until administrators, policymakers, and scholars understood what assessment literacy standards entailed, these standards could not properly applied to benchmark preservice teachers' assessment competence as parts of their licensure (Lam, 2015, 2019). Ultimately, the Education Bureau may formulate a set of mandated assessment literacy measures to appraise preservice teachers' levels of proficiency in assessment.

Conceptual framework of the project

The conceptual framework of the project is built upon (a) the “why-what-how” language assessment literacy knowledge base by Inbar-Lourie (2008) and (b) the “skills-knowledge-principles” model by Davies (2008). Inbar-Lourie (2008) model emphasizes the mastery of the “why” (i.e. the reasoning or rationale behind assessment in education), the “what” (i.e. the theoretical basis of language teaching, learning and assessment, the concept of validity and reliability, and the knowledge of language teaching pedagogy), and the “how” (i.e. the assessment skills and methods, applications of different

assessment procedures for both large-scale exams and classroom assessments). Her model was used to guide the revision of questionnaire content. Informed by a review of changes in language testing textbooks, Davies (2008) proposed the trend of skills, knowledge and principles in assessment textbook content. Skills refer to educational methodology and tools used in testing procedures. Knowledge refers to understanding of language measurement theories. Principles refer to test fairness, test ethics, and test impact. Davies's (2008) model served as a unit of analysis for classroom observation data, as well as the major components of the key construct - assessment literacy.

Methodology

Based upon the above discussion, the following questions are used to guide this study.

Research questions

1. What is the overall language assessment training landscape in five Hong Kong teacher education institutions? What are the assessment training needs of preservice teachers in these institutions?
2. What are contributing and constraining factors that influence the quality of language assessment training provided in the Hong Kong teacher education institution context?
3. To what extent does the reported language assessment training equip final-year preservice teachers with relevant knowledge, skills, and principles when they transition into their

first-year teaching?

Case study approach

The project adopted an exploratory case study approach to develop an in-depth and insider's perspective of a problematic issue under investigation, i.e. language assessment training for preservice teachers (Yin, 2014). Case study research facilitated the PI to enrich his conceptual understanding and lived experience of the overall language assessment training landscape and development of teacher assessment literacy in Hong Kong. The selection of a case study approach was methodologically suitable for the objectives and design of the project, because the PI identified the why (rationale), what (content), how (mode), and when (timing) the five teacher education institutions provided preservice teachers with language assessment training (Tight, 2017). The PI also explored the extent to which university-based language assessment training has sustainable impacts on the preservice teachers' pedagogical and assessment practices when they proceed to the first-year full-time teaching. The other reason for utilizing case study was that besides handling quantitative survey data, the PI handled a large amount of qualitative data, including semi-structured interviews, observations and documentary analysis of multimodal sources which unquestionably required a naturalistic and non-experimental approach to addressing the three research questions (Merriam, 1998). The next section describes the procedures of data collection and data analysis.

Data collection and analysis

The data collection took place in four stages, including a questionnaire, semi-structured interviews, classroom observations, and documentary analysis. The questionnaire was designed to examine final-year preservice teachers' (n=87) perceptions towards the content, delivery, timing, quality and quantity of language assessment training provided by respective language teacher education programmes in their universities. Convenience sampling was used as the PI invited these preservice teachers to be the respondents via their programme directors in each teacher education institution. Hard copies of the survey were circulated to maximize the completion and return rate. Then, the PI arranged interviews with eight preservice teachers for six times across two years (n=48) and with three instructors once (n=3) from one teacher education institution. The interviewees expressed interests to be interviewed and observed in their questionnaires. The semi-structured interview protocols included demographic information, conceptions of language assessment, evaluation of assessment knowledge and skills, sources of assessment training, understanding of assessment principles and possible impacts of assessment training.

The PI further observed these eight informants' classroom assessment practices (n=38). The key informants were observed twice during their second practicum in the final year and two

more times when they taught in primary-level and/or secondary-level schools as full-timers. Classroom observations focused on language assessment knowledge and skills, knowledge of integrating pedagogy and assessment, ability to give feedback and ability to use test results for improving learning. Meanwhile, the research team collected multiple sources of evidence, including programme documents, outlines, rubrics, handbooks, programme websites, quasi-government agency websites, preservice teachers' assignments, instructors' PowerPoint notes and professional development training schedules to address the extent to which language assessment training was provided in the Hong Kong teacher education context. In summary, the questionnaire, interview and document data were used to address Research Questions 1 and 2, whereas the observational and interview data were used to address Research Question 3.

Quantitative and qualitative approaches to data analysis were utilized. The questionnaire was adapted from Vogt and Tsagari's (2014) study with slight modifications to suit the Hong Kong context. The questionnaire sections were structured according to Inbar-Lourie's (2008) "why-what-how" language assessment literacy framework. The quantitative part involved the survey data, which were analyzed using (1) descriptive statistics, e.g. mean, mode, and standard deviation to look for the central distribution of preservice teachers' opinions towards language assessment training. To enhance the trustworthiness of the survey data, a

pilot study was administered by inviting sixteen Year 3 education majors to complete the survey. Three respondents were then interviewed to discuss the problems and errors of the draft survey. Difficult and unclear items were reworded. A preliminary analysis was performed to see whether there was any discrepancy in various items across different sections. To facilitate quantitative data analysis, the PI and his two research assistants input, trialed, and analyzed the survey data using SPSS (v. 21).

The qualitative part included interview, observational, and documentary analysis. For Research Questions 1 and 2, an inductive approach to handling the qualitative data, e.g. interviews and documents was adopted. Transcripts of informants' interviews, field notes, programme and course documents, and relevant source texts were read and reread by the PI and two research assistants to identify related codes and then emerging themes. Systematic categorization of themes was performed until data saturation. In case of disagreement, the PI and the research assistants would analyze the problematic data sets again until consensus was reached. Around 20 per cent of interview transcripts and 10 per cent of classroom transcripts sent back to the informants for member checking. For Research Question 3, transcripts of classroom data were analyzed against Davies's (2008) model as a measure of the development of assessment literacy. The PI coded all eight preservice teachers' classroom episodes with the "skills-knowledge-principles" model for three times and then the research

assistant verified these codes to enhance consistency. Throughout the process of qualitative data analysis, the PI's objective interpretation and his research insights remained separate in order not to have any bias against a particular emerging theme, individual participants and/or a particular classroom event. This strategy of handling qualitative data objectively could uphold the rigour and trustworthiness of the outcomes of this project.

Results and Discussion

Research Question 1: Assessment training landscape and preservice teachers' assessment needs

A. Document data

We surveyed the five preservice teacher training programmes offered by their respective teacher education institutions (i.e. TE1, TE2, TE3, TE4, and TE5), examining their programme documents, department/faculty websites, course outlines, selected instructional materials, and students' coursework assignments (e.g. term papers). Based upon our document search, each teacher education institution offered one double-degree programme, which trained the local preservice English teachers. In each institution, we found that they provided one assessment-related, credit-bearing course except TE5. TE5 did not offer a stand-alone assessment course. Instead, the assessment components were parts of the four pedagogical courses on teaching of literacy and oracy in primary and secondary schools. We

summarized the content, instructional approaches and assessment methods adopted in these four assessment-related courses, namely C1, C2, C3, and C4 (the course numbers corresponding to their respective university numbers).

Content

C1 was a core course, designed for preservice English teachers in TE1. Its content covered key assessment concepts, test construction skills, assessment of language skills, and feedback for effective student learning. Judging from the teaching topics, we agreed that C1 had a healthy balance of theory and practice in language testing and assessment. C2 was a core course, delivering contents on curriculum and assessment theories. It had 6 major topics, but only two were about assessment, namely (a) Assessment: concepts and functions and (b) Principles of assessment design. Apparently, C2 focused more on curriculum theories than on assessment theories. Further, the assessment component was designed for general education rather than for language education.

C3 was also a core course, designed for education and non-education undergraduates. It had three major elements, including the construction of school curriculum in Hong Kong, curriculum development and implementation/assessment, and the roles of teachers in the school community. As shown, C3 only had limited coverage of assessment implementation

although its assessment component remained practical, e.g. selection and design of effective assessment strategies to support student learning. C4 was a core course about the study of curriculum, assessment, and learning. Unlike C3, two-thirds of C4 topics were about assessment and learning, assessment issues, and assessment design/strategies. The assessment strand of C4 was highly contextualized, discussing lots of controversial assessment issues such as School-based Assessment, Territory-wide System Assessment, and Hong Kong Diploma of Secondary Education Exam.

Instructional approach

As to the instructional approach, all four courses adopted a range of teaching and learning activities, including lectures, teacher-led discussions, oral presentations, tutorials, case analysis, etc. In C1, one major attribute of the course was that lecturing was reduced to around only 20%, since interactive tasks, teacher-led discussions, analysis and critique of language tests occupied 80% of the class time. Such a large ratio of student-centred instruction was not identified in other three assessment courses, since lecturing took up close to 60% to 70% of instructional hours. To examine the course intended learning outcomes, we observed that all four courses required preservice teachers to master several higher-order thinking skills, for instance, “critically examine the roles of teachers” (C3), “analyze the process of curriculum development” (C4), and “evaluate assessment tools applicable to

various key stages” (C4).

When reviewing instructional materials (e.g. PowerPoint notes, outlines, and learning tasks), we found that the content of delivery remained theoretical rather than practical although from time to time, selected lecturers included lots of interactive tasks for students to think about and analyze the issues under study. In C1, the lecturer invited students to peer assess one another self-designed listening tests before final submission. In C3, the lecturer used flipped learning to promote academic reading on assessment topics and utilized Kahoot games to enhance classroom interactions. However, in C2, the lecturer used selected-response items (i.e. fill-in-the-blanks) to check on students’ memory of certain assessment concepts, which only promoted their lower-order thinking skills, not to mention how to deepen the learning of assessment knowledge and skills.

Assessment methods

Regarding the assessment methods, C1 consisted of (1) Design assessment items for a formative listening task (individual work, 20%), (2) Complete a self-designed, peer reviewed English language test (group work, 40%), and (3) Term paper (individual work, 40%). C2 comprised (1) 50% continuous assessment which entailed one in-class task and one 3-part essay about a critical evaluation of the public exam, and (2) 50% end-of-course exam. C3

included 30% group presentation and 70% essay, and C4 required students to do a group exercise (an oral presentation followed by a written report, 40%) and compose an individual term paper with a weighting of 60%. Although in each course, the assessment methods were varied by engaging preservice teachers in numerous application tasks. It can be said that term papers remained heavily weighted ranging from 40% up to 70% except for C1, which had a good balance between formative and summative assessment.

From the course documents, we had no idea of how individual instructors gave feedback to students concerning the results of their assessment tasks. Two instructors told our research team that they simply returned the graded assignments to students via General Office. When asked whether they knew how many students collected the marked scripts, the two instructors said they did not check. Interestingly, in C4, we found that the lecturer deliberately included a feedback session (i.e. two weeks after the last lecture) by handling course evaluation and reporting assessment outcomes to students formatively and summatively. Our research team believed that it was a sound pedagogical practice to consolidate assessment knowledge and skills via verbal feedback.

B. Questionnaire data (selected-response items)

The questionnaire administered in this project was adapted from Giraldo (2018) and Vogt and

Tsagari (2014; see [Appendix 1](#)). There was a total of five sections, including Section 1 Demographics, Section 2 “How” of language assessment, Section 3 “Why” of language assessment, Section 4 “What” of language assessment, and Section 5 Reflections. We first reported Sections 2-4 separately, and then summarized the overall findings of the questionnaire. Section 5 recorded respondents’ qualitative comments, which would be reported in the subsequent section - Research Question 2, part A. In the following, all reported quantitative data were compiled in [Appendix 2](#).

In Section 2, the respondents were asked to rate to what extent they were trained and to what extent they needed more training in the “how” of language assessment. As to the past training opportunities, nearly 50% of the preservice teachers thought that they only received minimal training in using formative assessment (47.13%) and using summative assessment for diagnostic purposes (47.67%). Further, 47.13% of the respondents believed that they did not have sufficient training in incorporating technologies in assessment. They equally felt that they were not sufficiently prepared to improve assessment items after item analysis (e.g. eliminating questions that are too easy or too difficult; 45.53%). Despite this inadequacy, more than half of the respondents considered that they had adequately trained in aligning curriculum objectives, teaching and assessment (64.37%), using multiple assessment methods to inform pedagogical decisions (49.43%), and improving instructions based upon assessment

results and feedback (49.43%). When asked about which aspects they needed more training, more than half of the respondents felt that they needed advanced training in producing original assessment materials (57.47%), giving feedback to students based upon assessment information (48.28%), and designing assessments that are not only in line with course contents but also course tasks (53.49%). Having said that, nearly half of the respondents thought that they still needed basic training in multiple areas, for instance, a, c – i, k, l, and n. In other words, the preservice teachers perceived that despite fundamental training, university-based assessment training might not necessarily meet their needs to handle classroom-based assessment procedures.

In Section 3, the respondents were asked to rate to what extent they were trained and to what extent they needed more training in the “why” of language assessment. Speaking of the past training exposure, slightly more than half of the respondents said that they only received minimal training in 10 out of 13 items. Of these, nearly 60% of them thought that they were not proficient in elaborating on assessment criteria and delivering assessment of learning to evaluate students’ learning. Yet, the other half of the respondents felt that they had adequate training in identifying learning and teaching needs (50.57%) and identifying the purposes of assessment (57.47%). This could be owing to the repeated training in writing lesson plans as well as the sensitivity they had cultivated through micro-teaching and teaching practicum.

Slightly more than half of the respondents felt that they still needed *basic* training in multiple “why” aspects (items a, b, c, e – i, l, m). This could be explained by the rarity of preservice teachers to be involved in running or in preparing assessments during their teaching practicum. Without sufficient hands-on experience, the respondents might feel unfamiliar with the “why” aspect of language assessment. There were also a significant portion of preservice teachers, who aspired to seek more advanced training in assessment of (43.02%), for (53.49%), and as (51.16%) learning. Although the respondents could have some fundamental understanding of these principles, they would find it challenging to put these assessment ideas into practice.

In Section 4, the respondents were asked to rate to what extent they were trained and to what extent they needed more training in the “what” of language assessment. Half of the respondents said that they did not receive adequate training in using statistics to evaluate the quality of assessment items, which could be explained that such a technical aspect of language testing was not included in the course-based training. Rather surprisingly, more than half of the respondents considered that they only had minimal training in assessing vocabulary (49.93%), assessing integrated language skills (49.93%), understanding essential assessment principles, i.e. reliability (56.32%), validity (58.62%), authenticity (44.83%), practicality (57.47%), and washback (56.32%). Apparently, the existing assessment-related

courses might lack the aforementioned domains. Yet, the other half of the respondents claimed that they had adequate training in assessing grammar, receptive and productive skills. It could be explained that the respondents might have already learnt how to assess productive and receptive skills in their pedagogy courses. When asked what training needs they wanted more in the “what” aspect, around 55% of the respondents chose to have advanced training in assessing specific language sub-skills, such as receptive, productive, grammar, vocabulary, and integrated language skills. Because nearly half of the respondents lacked ample training in understanding assessment principles (items h, i, l, and m), they preferred having some basic training in these specific domains.

To summarize, nearly half of the respondents felt that they were only minimally trained in terms of the “how” (42.43%), “why” (47.74%), and “what” (48.18%) dimensions of language assessment. There was only a significant portion of students thought that they were adequately trained in terms of the “how” (41.04%) aspect of language assessment. Concerning assessment training, a majority of the respondents considered that they needed more **basic training** in terms of the “how” (49.22%) and “why” (48.00%) of language assessment. While 45.50% of the respondents expressed interest in receiving more **advanced training** in terms of the “what” aspect, close to 43% of them also thought that they still needed more basic training in terms of the “what” aspect of language assessment. This

significant portion of preservice teachers (43%) remained unconfident and uncertain about their acquisition of assessment knowledge, skills, and principles in the teacher education programmes.

Research Question 2: Factors that influence the quality of language assessment training

A. Questionnaire data (constructed-response items)

The questionnaire included two open-ended questions in Section 5. Respondents were asked about their strengths and challenges in the design and delivery of language assessment and the reasons. There were 74 counts of responses from Question 7, Section 5. Our team thematically analyzed all responses to generate a total of 12 categories of strengths. These categories were derived from an in-depth analysis of the keywords and the gist of each response. These categories included: alternative assessment, authenticity of assessment, delivery (general), formative and summative assessment, integration of theories in assessment, interpersonal aspects in assessment, material development, practicality of assessment, subject knowledge, understanding of assessment, implementation of ICT, and validity and reliability of assessment.

The three most popular categories that translated into the facilitative factors in the quality of assessment training consisted of (1) authenticity of assessment (n = 16), material

development (n = 15), and alternative assessment (n = 8). Our respondents thought that they had learnt about how to contextualize assessment practices through coursework and practicum. They believed that setting achievable assessment goals and aligning these goals with the content of test items could warrant a fair and ethical assessment task. Selected respondents stated that they received professional training in designing teaching and assessment materials for their target learners in microteaching, practicum, and immersion programme, especially when the respondents were given autonomy to try out various modes of formative assessment. The respondents also told us that they gained hands-on experience in performing alternative assessments, namely gamified assessments, self- and peer-assessment to enhance the “creativity” and “diversity” sides of language assessment training. Thus, course-based and practicum-based assessment training has a pivotal role to play in enriching preservice teachers’ assessment knowledge and skills.

As to the inhibiting factors, our team examined the greatest challenges when the respondents designed and delivered language assessment. There were 71 counts of responses from Question 8, Section 5. We thematically analyzed all responses to generate a total of 13 categories of challenges. These 13 categories were derived from the same in-depth analysis of the keywords and the gist of each response. They overlapped with the previous 12 categories reported in Question 7, except for 3 new categories including *item-analysis*, *learner diversity*,

and *rendering feedback*.

The three emerging categories that contributed to the inhibiting factors in the quality of assessment training entailed (1) material development (n = 15), (2) learner diversity (n = 11), and (3) emotional aspects of assessment (n = 9). While the respondents felt that developing school-based materials was one of their strengths, quite a number of them considered that this was also one of the challenges. Not surprisingly, they found it taxing to design materials that were not only intriguing and engaging, but were also in keeping with the curriculum and syllabus by large. None of the preservice teachers stated that tending to learner diversity was their strength. Instead, these teachers found it exacting to design assessment tasks that took both the mixed abilities and diverse needs of all students into consideration. Lastly, the respondents expressed concerns that it was not easy to motivate students in completing formative assessment tasks, given that most students were pragmatic and exam-oriented. It was probable that university-based assessment training might include differentiated assessment as one major component, and strengthen preservice teacher capacity in dealing with various affective aspects of language assessment, namely motivation, self-efficacy, and test anxiety.

B. Interview data

To understand both contributing and constraining factors in the quality of assessment training, we interviewed eight key informants for six times across two years. In this sub-section, we reported coded interview data from the first two interviews conducted with six out of eight preservice teachers, namely Teacher A, Teachers B – F. Thus far, we have identified two contributing factors and three constraining factors.

Contributing factors: (a) Sufficient peer support / (b) Positive experience in teaching practicum

All six teachers except one told us that they had positive experience in their teaching practicum, because their mentors were helpful and supportive despite their hectic schedules. Teacher A said that since her mentor had a project on e-Learning with the Education Bureau, she was involved in learning how to use updated software tools to evaluate student learning formatively. Her mentor also gave feedback on Teacher A's lesson plans daily. Similarly, Teacher C had opportunities to design the summative assessment task in her practicum school (a Band 1 secondary school), yet the content should stick to the public exam format. Teacher C felt that she had great autonomy to design all formative and summative assessments as long as she obtained prior endorsements from her mentor.

While Teachers D and E thought that they had not received sufficient assessment training

from their universities, they learnt a lot from their mentors' advice. For instance, Teacher D's mentor would brief Teacher D about the abilities of her students, and coached her how to set homework assignments and pop quizzes which were too neither difficult nor too easy. Teacher E also got ample peer support from his mentor, as she encouraged Teacher E to attempt alternative instructional approaches, e.g. process writing to support student learning. His mentor sat in Teacher E's classes regularly, and provided him with constructive feedback for improvement. Teacher F was the only one who did not get enough support from her mentor, because her mentor was the Panel Chair of English Department. Despite this, she consulted other senior English teachers on how to set appropriate assessment tasks, e.g. grammar quizzes for summative evaluation.

Constraining factors: (a) Lack of opportunities to construct assessments / (b) Course-based assessment training being too theoretical / (c) Lack of confidence

Unanimously, the six teachers thought that they lacked exposure and hands-on experience in constructing uniform tests or parts of the exam papers. Although they knew that they might not be qualified to do so, they expected to be involved in the test construction processes so that they could gain some practical assessment skills. Teacher C once said, "*We are taught how to teach listening, reading, writing and all kinds of stuffs. But then, we are not taught how to set assessments.*" She told us that she actually wanted to learn how to construct

different question types for a summative test and its model answers. Teachers E and F also echoed the same view that they were only involved in designing worksheets not in constructing any formal tests/quizzes throughout the practicum.

Concerning the quality of course-based assessment training, Teacher B pinpointed that the assessment training in her university was too theoretical, rendering little room for practical applications. She emphasized that she needed to learn contingency plans in assessment, suggested adding a stand-alone language assessment course with up to 70% practice components in her teacher education programme. Teacher F proposed that her university should negotiate with placement schools to allow preservice teachers to set parts of the summative test items. Likewise, Teacher E felt that he did not think teaching practicum provided preservice teachers with any opportunities to deliver assessment, which he considered an issue of cultivating teacher assessment literacy. Because of an obvious lack of assessment training and hands-on experience, Teachers A, C, and D thought that they lacked confidence in evaluating their students competently during the practicum. Further, the duration of teaching practicum was transient (i.e. 6 weeks at most). The teachers could not have enough time to witness students' learning progress even after administering a formative or summative assessment task.

Research Question 3: Transition from final-year preservice teachers to first-year inservice teachers

A. Interview data

To understand preservice teachers' transition from trainees to full-time professionals, we reported coded interview data conducted near the end of their teacher education programmes (the third and fourth interviews) and one year after they became full-timers (the fifth and sixth interviews). Then, we compared and contrasted these four rounds of interviews, and identified the progression of the four informants' assessment literacy levels and needs in terms of five emerging themes.

To begin with, the six preservice teachers became more assertive and confident after one-year full-time teaching in school. Four of them (i.e. Teachers A, B, D, and E) unanimously admitted that they had substantial improvement in writing assessment more than speaking assessment, because they were not directly involved in teaching and assessing students' oral English. Nonetheless, Teachers C and F told us that they had improvement in *speaking assessment* more than writing assessment due to their instructional approach and teaching assignments. For instance, since Teacher C taught three classes in a sub-urban primary school and had ample opportunities to polish her questioning techniques. She found that she was able to give verbal formative feedback to students more effectively than before. Similarly,

Teacher F was assigned to mark Grade 6 writing exam scripts (i.e. short compositions) for the whole form level. Because of that, she did not have much chance to practice writing assessment except for exams. On the contrary, she was able to teach and assess her Grades 3 and 6 students' speaking skills almost every day.

By comparing and contrasting the interviews conducted before graduation and the interviews conducted after one-year teaching, our team coded and categorized these longitudinal, qualitative data into five themes, namely (1) reality shock, (2) innovative practices, (3) compliance with school assessment policies, (4) understanding of students' academic levels, and (5) assessment literacy levels and needs. In the first theme, all six teachers except one felt that when they worked as full-time English teachers, the teaching workload was so much heavier than they could imagine and handle. Teachers D and E emphasized that they struggled a lot amid doing three things concurrently, including teaching of three full classes, assessing students' compositions and assignments, and performing various administrative duties. They particularly found doing administrative duties (i.e. joining committees and handling class affairs) time-consuming, demanding, and distracting as they were very new to their fellow colleagues and respective school culture. Teacher D further added that because of those administrative duties, she really had no time and energy to enhance her assessment capacity, not to mention to allocate time for upgrading her instructional approach.

Despite the reality shock, under the second theme, four teacher informants told us that they tried out different innovative assessment practices. For example, Teacher A invited students to correct peers' grammatical errors on Google Form with positive feedback. Teacher B attempted self- and peer-assessment with reference to a scoring rubric although some students did not buy peer feedback owing to its accuracy. Teacher C utilized interactive questioning techniques to encourage primary students with special needs to speak up. Teacher E implemented process writing in a Grade 10 class for one year with a satisfactory outcome. Although it was premature to judge whether these four informants had built stronger assessment literacy capacity, it was unquestionable that they were motivated to try out innovative assessment practices against all odds. In the third theme, all six informants expressed concerns that although they were given autonomy to teach and assess students, they still had to comply with the school-based assessment policies. Teacher A described that she found setting the uniform test very frustrating, since both in-house style and test syllabus were somewhat restrictive. Teacher A's assessment construction experience ended up imitating the past papers' format and contents. The same view and experience were echoed by Teachers B and F.

The fourth theme was about understanding of students' academic abilities. The six informants

considered regardless of how much assessment training they received, they were still uncertain about their students' abilities even after one-year of teaching. They also had difficulties in assessing mixed ability classes, especially those with SEN students. To this end, they needed more time and experience in comprehending students' levels in order to construct effective assessment tasks. In the fifth theme, we categorized the six informants' assessment literacy levels and needs when they transitioned from trainees to full-timers. Among a four-level scale of "emerging", "developing", "accomplished", and "exemplary", the informants self-evaluated themselves as "developing" in acquiring assessment knowledge and skills. In fact, while Teacher B was assertive, she only rated herself in-between the two levels of "emergent" and "developing". For Teacher D, after receiving assessment training and having hands-on experience, she rated herself close to "accomplished" albeit partially. Overall speaking, the informants' assessment literacy levels were somewhat *stable* and *promising* throughout the transition. As to assessment needs, the informants expected to attend workshops on classroom-based assessment, learn about online assessment methods, and work closely with senior colleagues on marking and standardization procedures.

B. Classroom observation data (practicum teaching and first-year full-time teaching)

Because of designated word limit, we only reported the analysis of four key informants' classroom episodes across two years, namely Teachers C, D, E, and F. By using Davies's

(2008) “skills-knowledge-principles” model, we outlined whether the four informants had smooth professional transition and the extent to which they became assessment literate. Concerning the transition, Teachers C, E and F had a more pleasant transition experience than Teacher D, who struggled among teaching, assessing, and administrative work and had an obvious lack of confidence in evaluating students. For each participant, we observed their classes three to four times, and then identified two major reasons for contributing to their transition. The first one was about their abilities to understand students’ academic levels (i.e. skills), and the second one their capacity to align teaching and assessment within a very stringent exam-oriented system (i.e. knowledge and skills). By adopting the formative assessment approach, Teachers C and E did particularly well in analyzing students’ assessment data to fine-tune their pedagogies. Although Teacher F remained textbook-bound and adopted intensive drilling, she could distinguish the students’ diverse educational needs and perform summative assessment by taking care of their mental well-beings (i.e. use of positive reinforcement).

As to the assessment literate levels, it appears that Teachers C and E had noticeable enhancement in all three aspects of knowledge, skills, and principles when evaluating their students. For instance, Teacher C had a much better understanding of formative assessment and ways of giving feedback to young learners (i.e. Grade 2 students). She became more

interactive in class and improved quite a lot in her questioning techniques. More importantly, she encouraged students to do peer assessment tasks and to have post-assessment reflection without threatening student self-esteem and creating test anxiety. Likewise, Teacher E was competent in applying differentiated instruction and assessment in both primary and senior secondary classrooms. Besides, he could provide his Grade 10 students' with scaffolding and instant feedback when teaching them how to write a restaurant review using a genre-based approach.

On the other hand, Teacher F had improvement in assessment skills and principles, but not assessment knowledge. Teacher F's confidence and competence was the best among the four participants. Her instructional approach was effective and her classroom management skills were indisputable albeit exam-oriented and teacher-led. After four rounds of observation, Teacher F could not align teaching and assessment constructively, and believed that formative and summative assessment should separate. Her conceptual understanding of classroom-based formative assessment practices remained weak. Among the four informants, Teacher D was the most vulnerable, provided that she did not have improvement in her assessment literacy capacity, e.g. lack of understanding of students' levels, ineffective teacher-student interaction, reliant on publishers' instructional materials, and inability to implement differentiated assessment practices. It was likely that teacher D was overwhelmed

with the “reality shock” so that she had no time to get involved in assessment training upon graduation.

Conclusions and Recommendations

In brief, the five teacher education institutions did provide preservice teachers with generic assessment-related training. The objectives, contents, and assessments of these assessment courses were well-designed and well-delivered albeit theoretical. The questionnaire data indicated that the majority of respondents required **basic** training in the “how” (assessment skills) and “why” (theoretical rationale) aspects of language assessment, and aspired to have **advanced** training in “what” aspects (e.g. reliability, fairness, ethics). Regardless of this programme-wide training, the preservice teachers felt that they lacked hands-on experience in constructing summative assessment tasks and ability to understand students’ academic levels owing to their limited exposure. Further, they expected to receive practically-oriented, contextualized, and mentor-supervised assessment training. Three out of four key informants had smooth professional transition from teacher trainees to full-time teachers although they had diverse assessment literacy capacities, mainly evolving from “developing” to slightly close to being “accomplished”.

Based upon the findings, we would like to make the following recommendations which can

broadly enhance Hong Kong preservice teachers' assessment literacy. First, course-based language assessment training should be included and made compulsory, like a core, credit-bearing course in the 5-year programme. Second, preservice teachers are allowed to construct parts of the summative assessment during their teaching practicum or immersion programmes so that they can experience how to generate an “authentic” test/exam paper. Third, assessment-related workshops or webinars (e.g. how to conduct classroom-based formative assessment practices) could be organized to facilitate preservice teachers' assessment literacy by local universities. Fourth, teacher educators' assessment literacy needs to be upgraded and warranted via systematic professional development programmes, because they are preservice teachers' role models and academic facilitators in the aforementioned course-based training.

Bibliography

Brookhart, S.M. (2011). Educational assessment knowledge and skills for teachers.

Educational Measurement: Issues and Practices, 30(1), 3-12.

Brown, J.D., & Bailey, K.M. (2008). Language testing courses: What are they in 2007?

Language Testing, 25(3), 349-383.

Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher

assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P. Davidson,

B. O'Sullivan, & S. Stoyloff (eds.), *The Cambridge Guide to Second Language*

Assessment (pp. 20-29). Cambridge: Cambridge University Press.

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

327-347.

DeLuca, C., & Johnson, S. (2017). Developing assessment capable teachers in this age of

accountability. *Assessment in Education: Principles, Policy & Practice*, 24(2), 121-126.

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.

- Deneen, C.C., & Brown, G.T.L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education programme. *Cogent Education*, 3, <https://doi.org/10.1080/2331186X.2016.1225380>.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.
- 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? *Assessment in Education: Principles, Policy & Practice*, 21(3), 286-305.
- Hamp-Lyons, L. (2016). Purposes of assessment. In D. Tsagari & J. Banerjee (ed.), *Handbook of Second Language Assessment* (pp. 13-27). Germany: De Gruyter.
- Hamp-Lyons, L. (2017a, February). *How assessments can support learning: teachers' roles, teachers' knowledge*. Seminar presented at Distinguished Professor Seminar Series by the Research Institute for Bilingual Learning and Teaching, the Open University of Hong Kong, Hong Kong.
- Hamp-Lyons, L. (2017b, February). *Why is assessment literacy important? And to whom?* Seminar presented at Distinguished Professor Seminar Series by the Research Institute for Bilingual Learning and Teaching, the Open University of Hong Kong, Hong Kong.
- Inbar-Lourie, O. (2008). Constructing an 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.
- 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.
- Koh, K., Burke, L.E.C., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teacher Research*, 22(3), 264-288.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197.
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81, 78-89.
- Merriam, S.B. (1998). *Qualitative Research and Case Study Applications in Education: Revised and Expanded from Case Study Research in Education*. San Francisco, CA: Jossey-Bass Publishers.
- 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 & Practice*, 21(3), 251-270.

- Schafer, W.D. (1993). Assessment literacy for teachers. *Theory into Practice*, 32(2), 118-126.
- Stiggins, R.J. (1991). Assessment literacy. *The Phi Delta Kappan*, 72(7), 534-539.
- Stiggins, R.J. (1995). Assessment literacy for the 21st century. *The Phi Delta Kappan*, 77(3), 238-245.
- Stiggins, R. (2014). Improve assessment literacy outside of schools too. *The Phi Delta Kappan*, 96(2), 67-72.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412
- Tight, M. (2017). *Understanding Case Study Research: Small-scale Research with Meaning*. London, UK: Sage.
- Troudi, S., Coombe, C., & Al-Hamly, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546-555.
- Vogt, K., & Tsigari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402.
- 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.
- Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second*

Language Writing, 16(3), 194-209.

White, E. (2011). Are you assessment literate? Some fundamental questions regarding effective classroom-based assessment. *On CUE Journal*, 3(1), 3-25.

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

Xu, Y.T., & Liu, Y.C. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 493-513.

Yin, R. (2014). *Case Study Research: Design and Methods* (5th ed.). Thousand Oaks, CA: Sage.

Appendix 1

An Exploratory Study of Language Assessment Training in Hong Kong

This questionnaire aims to find out how **pre-service** teachers perceive their language assessment training with regards to the 'how', 'why', and 'what' aspects. It sets out to see the extent to which they feel that they have adequate training as **frontliners**.

No names of participants and institution information will be disclosed in any official document including interim and final project reports to SCOLAR and research output (e.g., scholarly publications). All research data will be kept confidential.

Section 1: Basic information

Name

Email

Type of school you are working in ☐ () Government school ☐ () Aided school ☐ () Direct Subsidy Scheme (DSS) School ☐ () International School

Level of class(es) you are responsible for (e.g., P.1 or S.1)

Section 2: 'How' of language assessment

1. Please rate to what extent you think you WERE trained in the following domains.

**e.g., 1 – 2 courses.*

Statements	Not adequate	*Minimal	Adequate
a. Aligning curriculum objectives, teaching, and assessments.			
b. Producing original assessment materials.			
c. Using ready-made tests from textbook packages or from other sources.			
d. Using self- or peer-assessment.			
e. Using formative assessment (i.e. informal, continuous, non-test type assessment).			
f. Using summative assessment for diagnostic purposes.			
g. Using portfolio assessment.			
h. Using multiple assessment methods to inform pedagogical decisions.			
i. Incorporating technologies in assessment.			
j. Giving feedback to students based on assessment information.			

k. Improving instructions or quality of input based on assessment results and feedback.	1	2	3	4
l. Improving assessment items after item analysis (e.g., eliminating questions that are too easy or too difficult).	1	2	3	4
m. Designing assessments that are not only in line with course contents but also course tasks.	1	2	3	4
n. Using past papers as teaching and/or assessment materials.	1	2	3	4

2. Please rate to what extent you think you NEED more training in the following domains.

e.g., 1 – 2 seminars/workshops; ^ e.g., 1 – 2 full courses.

Statements	None	# Yes, basic training	^ Yes, advanced training
a. Aligning curriculum objectives, teaching, and assessments.	1	2	3
b. Producing original assessment materials.	1	2	3
c. Using ready-made tests from textbook packages or from other sources.	1	2	3
d. Using self- or peer-assessment.	1	2	3
e. Using formative assessment (i.e. informal, continuous, non-test type assessment).	1	2	3
f. Using summative assessment for diagnostic purposes.	1	2	3
g. Using portfolio assessment.	1	2	3
h. Using multiple assessment methods to inform pedagogical decisions.	1	2	3
i. Incorporating technologies in assessment.	1	2	3
j. Giving feedback to students based on assessment information.	1	2	3
k. Improving instructions or quality of input based on assessment results and feedback.	1	2	3
l. Improving assessment items after item analysis (e.g., eliminating questions that are too easy or too difficult).	1	2	3
m. Designing assessments that are not only in line with course contents but also course tasks.	1	2	3
n. Using past papers as teaching and/or assessment materials.	1	2	3

Section 3: 'Why' of language assessment^{4J}

3. Please rate to what extent you think you WERE trained in the following domains.^{4J}

^{*}e.g., 1 – 2 courses.

^{4J}

Statements	Not adequate	Minimal [*]	Adequate
a. Giving grades.	1	2	3
b. Identifying learning and teaching needs.	1	2	3
c. Identifying the purpose(s) of assessments.	1	2	3
d. Identifying students' strengths and weaknesses based on assessment results.	1	2	3
e. Placing students onto groups according to their abilities.	1	2	3
f. Being aware of the impact and power of public examinations.	1	2	3
g. Elaborating on assessment criteria (i.e. transparent assessment practices).	1	2	3
h. Giving students opportunities to share the voice about assessment (i.e. democratic assessment practices).	1	2	3
i. Delivering assessment of learning to evaluate students' learning.	1	2	3
j. Delivering assessment for learning to support students' learning.	1	2	3
k. Delivering assessment as learning to promote students' learning.	1	2	3
l. Delivering norm-referenced assessments to compare students' learning among one another.	1	2	3
m. Delivering criterion-referenced assessments to measure students' learning in relation to assessment rubrics.	1	2	3

^{4J}

4. Please rate to what extent you think you NEED more training in the following domains.^{4J}

[#] e.g., 1 – 2 seminars/workshops; [^] e.g., 1 – 2 full courses.

^{4J}

Statements	None	[#] Yes, basic training	[^] Yes, advanced training
a. Giving grades.	1	2	3
b. Identifying learning and teaching needs.	1	2	3
c. Identifying the purpose(s) of assessments.	1	2	3
d. Identifying students' strengths and weaknesses based on assessment results.	1	2	3
e. Placing students onto groups according to their abilities.	1	2	3
f. Being aware of the impact and power of public examinations.	1	2	3

g. Elaborating on assessment criteria (i.e. transparent assessment practices).
h. Giving students opportunities to share the voice about assessment (i.e. democratic assessment practices).
i. Delivering assessment of learning to evaluate students' learning.
j. Delivering assessment for learning to support students' learning.
k. Delivering assessment as learning to promote students' learning.
l. Delivering norm-referenced assessments to compare students' learning among one another.
m. Delivering criterion-referenced assessments to measure students' learning in relation to assessment rubrics.

Section 4: 'What' of language assessment

5. Please rate to what extent you think you WERE trained in the following domains.

*e.g., 1 – 2 courses.

Statements	Not adequate	Minimal*	Adequate
a. Assessing receptive skills (i.e. reading and listening).
b. Assessing productive skills (i.e. speaking and writing).
c. Assessing grammar.
d. Assessing vocabulary.
e. Assessing integrated language skills.
f. Integrating language assessment theories to assessments.
g. Using statistics to evaluate the quality of assessment items.
h. Establishing reliability of assessments.
i. Establishing validity of assessments (i.e. to only assess what has been taught).
j. Establishing authenticity of assessments (i.e. assessments related to real-life situations).
k. Establishing practicality of assessments (i.e. assessment manageable for both teachers and students).
l. Evaluating washback effect on learning, teaching, curricula, and institutions (e.g. impacts of following a test-driven syllabus).
m. Evaluating innovations and trends in

assessment.

6. Please rate to what extent you think you NEED more training in the following domains.
e.g., 1 – 2 seminars/workshops; ^ e.g., 1 – 2 full courses.

Statements	None	# Yes, basic training	^ Yes, advanced training
a. Assessing receptive skills (i.e. reading and listening).			
b. Assessing productive skills (i.e. speaking and writing).			
c. Assessing grammar.			
d. Assessing vocabulary.			
e. Assessing integrated language skills.			
f. Integrating language assessment theories to assessments.			
g. Using statistics to evaluate the quality of assessment items.			
h. Establishing reliability of assessments.			
i. Establishing validity of assessments (i.e. to only assess what has been taught).			
j. Establishing authenticity of assessments (i.e. assessments related to real-life situations).			
k. Establishing practicality of assessments (i.e. assessment manageable for both teachers and students).			
l. Evaluating washback effect on learning, teaching, curricula, and institutions (e.g. impacts of following a test-driven syllabus).			
m. Evaluating innovations and trends in assessment.			

Section 5: Reflections

7. What is your strength in the design and delivery of language assessment? Why?

8. What is your greatest challenge in the design and delivery of language assessment? Why?

Appendix 2

LAT Project: Survey Data Project AY2019-2020

This report covers survey data of **pre-service** teachers (n = 87) only.

Section 0: Overview

Statements		Not adequate	Minimal	Adequate	Total no. of response
Q1: Rate to what extent you think you WERE trained	Section 2: 'How' of language assessment	201	516	499	1216
Q3: Rate to what extent you think you WERE trained	Section 3: 'Why' of language assessment	16.53%	42.43%	41.04%	1127
Q5: Rate to what extent you think you WERE trained	Section 4: 'What' of language assessment	228	538	361	1127
		20.23%	47.74%	32.03%	
		232	543	352	1127
		20.59%	48.18%	31.23%	

Statements		None	Basic training	Advanced training	Total no. of response
Q2: Rate to what extent you think you NEED more training	Section 2: 'How' of language assessment	134	596	481	1211
Q4: Rate to what extent you think you NEED more training	Section 3: 'Why' of language assessment	150	541	436	1127
Q6: Rate to what extent you think you NEED more training	Section 4: 'What' of language assessment	128	478	506	1112
		11.51%	42.99%	45.50%	

Section 2: 'How' of language assessment

1. Please rate to what extent you think you WERE trained in the following domains. *e.g., 1 – 2 courses

Statements		Not adequate	*Minimal	Adequate	n =
a. Aligning curriculum objectives, teaching, and assessments	No. of response	2	29	56	87
	Weighting	2.30%	33.33%	64.37%	
b. Producing original assessment materials	No. of response	15	39	33	87
	Weighting	17.24%	44.83%	37.93%	
c. Using ready-made tests from textbook packages or from other sources	No. of response	16	31	40	87
	Weighting	18.39%	35.63%	45.98%	
d. Using self- or peer-assessment	No. of response	9	36	42	87
	Weighting	10.34%	41.38%	48.28%	

e. Using formative assessment (i.e. informal, continuous, non-test type assessment)	No. of response	7	41	39	87
	Weighting	8.05%	47.13%	44.83%	
f. Using summative assessment for diagnostic purposes	No. of response	14	41	31	86
	Weighting	16.28%	47.67%	36.05%	
g. Using portfolio assessment	No. of response	33	38	16	87
	Weighting	37.93%	43.68%	18.39%	
h. Using multiple assessment methods to inform pedagogical decisions	No. of response	11	33	43	87
	Weighting	12.64%	37.93%	49.43%	
i. Incorporating technologies in assessment	No. of response	13	41	33	87
	Weighting	14.94%	47.13%	37.93%	
j. Giving feedback to students based on assessment information	No. of response	10	36	41	87
	Weighting	11.49%	41.38%	47.13%	
k. Improving instructions or quality of input based on assessment results and feedback	No. of response	8	36	43	87
	Weighting	9.20%	41.38%	49.43%	
l. Improving assessment items after item analysis (e.g., eliminating questions that are too easy or too difficult)	No. of response	16	39	31	86
	Weighting	18.60%	45.35%	36.05%	
m. Designing assessments that are not only in line with course contents but also course tasks	No. of response	13	39	35	87
	Weighting	14.94%	44.83%	40.23%	
n. Using past papers as teaching and/or assessment materials	No. of response	34	37	16	87
	Weighting	39.08%	42.53%	18.39%	

2. Please rate to what extent you think you NEED more training in the following domains.

e.g., 1 – 2 seminars/ workshops; ^ e.g., 1 – 2 full courses

Statements		None	# Yes, basic training.	^ Yes, advanced training.	n =
a. Aligning curriculum objectives, teaching, and assessments	No. of response	11	50	25	86
	Weighting	12.79%	58.14%	29.07%	
b. Producing original assessment materials	No. of response	5	32	50	87
	Weighting	5.75%	36.78%	57.47%	
c. Using ready-made tests from textbook packages or from other sources	No. of response	17	45	25	87
	Weighting	19.54%	51.72%	28.74%	
d. Using self- or peer-assessment	No. of response	13	49	24	86
	Weighting	15.12%	56.98%	27.91%	
e. Using formative assessment (i.e. informal, continuous, non-test type assessment)	No. of response	9	43	35	87
	Weighting	10.34%	49.43%	40.23%	
f. Using summative assessment for diagnostic purposes	No. of response	10	45	32	87
	Weighting	11.49%	51.72%	36.78%	

g. Using portfolio assessment ^o	No. of response ^o 9 ^o	49 ^o	29 ^o	87 ^o
	Weighting ^o 10.34% ^o	56.32% ^o	33.33% ^o	
h. Using multiple assessment methods to inform pedagogical decisions ^o	No. of response ^o 5 ^o	42 ^o	39 ^o	86 ^o
	Weighting ^o 5.81% ^o	48.84% ^o	45.35% ^o	
i. Incorporating technologies in assessment ^o	No. of response ^o 10 ^o	40 ^o	37 ^o	87 ^o
	Weighting ^o 11.49% ^o	45.98% ^o	42.53% ^o	
j. Giving feedback to students based on assessment information ^o	No. of response ^o 9 ^o	36 ^o	42 ^o	87 ^o
	Weighting ^o 10.34% ^o	41.38% ^o	48.28% ^o	
k. Improving instructions or quality of input based on assessment results and feedback ^o	No. of response ^o 7 ^o	42 ^o	37 ^o	86 ^o
	Weighting ^o 8.14% ^o	48.84% ^o	43.02% ^o	
l. Improving assessment items after item analysis (e.g., eliminating questions that are too easy or too difficult) ^o	No. of response ^o 5 ^o	45 ^o	36 ^o	86 ^o
	Weighting ^o 5.81% ^o	52.33% ^o	41.86% ^o	
m. Designing assessments that are not only in line with course contents but also course tasks ^o	No. of response ^o 6 ^o	34 ^o	46 ^o	86 ^o
	Weighting ^o 6.98% ^o	39.53% ^o	53.49% ^o	
n. Using past papers as teaching and/or assessment materials ^o	No. of response ^o 18 ^o	44 ^o	24 ^o	86 ^o
	Weighting ^o 20.93% ^o	51.16% ^o	27.91% ^o	

Section 3: 'Why' of language assessment^o

3. Please rate to what extent you think you WERE trained in the following domains.^o

* e.g., 1 – 2 courses^o

Statements ^o	Not adequate ^o	Minimal ^o	Adequate ^o	n = ^o
a. Giving grades ^o	No. of response ^o 39 ^o Weighting ^o 45.88% ^o	36 ^o 42.35% ^o	10 ^o 11.76% ^o	85 ^o
b. Identifying learning and teaching needs ^o	No. of response ^o 7 ^o Weighting ^o 8.05% ^o	36 ^o 41.38% ^o	44 ^o 50.57% ^o	87 ^o
c. Identifying the purpose(s) of assessments ^o	No. of response ^o 4 ^o Weighting ^o 4.60% ^o	33 ^o 37.93% ^o	50 ^o 57.47% ^o	87 ^o
d. Identifying students' strengths and weaknesses based on assessment results ^o	No. of response ^o 17 ^o Weighting ^o 19.54% ^o	43 ^o 49.43% ^o	27 ^o 31.03% ^o	87 ^o
e. Placing students onto groups according to their abilities ^o	No. of response ^o 14 ^o Weighting ^o 16.09% ^o	45 ^o 51.72% ^o	28 ^o 32.18% ^o	87 ^o
f. Being aware of the impact and power of public examinations ^o	No. of response ^o 25 ^o Weighting ^o 28.74% ^o	35 ^o 40.23% ^o	27 ^o 31.03% ^o	87 ^o
g. Elaborating on assessment criteria (i.e. transparent assessment practices) ^o	No. of response ^o 17 ^o	50 ^o	19 ^o	86 ^o

3^o

	Weighting ^o 19.77% ^o	58.14% ^o	22.09% ^o	
h. Giving students opportunities to share the voice about assessment (i.e. democratic assessment practices) ^o	No. of response ^o 21 ^o	40 ^o	26 ^o	87 ^o
	Weighting ^o 24.14% ^o	45.98% ^o	29.89% ^o	
i. Delivering assessment of learning to evaluate students' learning ^o	No. of response ^o 8 ^o	47 ^o	32 ^o	87 ^o
	Weighting ^o 9.20% ^o	54.02% ^o	36.78% ^o	
j. Delivering assessment for learning to support students' learning ^o	No. of response ^o 10 ^o	41 ^o	36 ^o	87 ^o
	Weighting ^o 11.49% ^o	47.13% ^o	41.38% ^o	
k. Delivering assessment as learning to promote students' learning ^o	No. of response ^o 17 ^o	45 ^o	25 ^o	87 ^o
	Weighting ^o 19.54% ^o	51.72% ^o	28.74% ^o	
l. Delivering norm-referenced assessments to compare students' learning among one another ^o	No. of response ^o 27 ^o	44 ^o	16 ^o	87 ^o
	Weighting ^o 31.03% ^o	50.57% ^o	18.39% ^o	
m. Delivering criterion-referenced assessments to measure students' learning in relation to assessment rubrics ^o	No. of response ^o 22 ^o	43 ^o	21 ^o	86 ^o
	Weighting ^o 25.58% ^o	50.00% ^o	24.42% ^o	

4. Please rate to what extent you think you NEED more training in the following domains.^o

e.g., 1 – 2 seminars/ workshops; ^ e.g., 1 – 2 full courses^o

Statements ^o	None ^o	# Yes, basic training. ^o	^ Yes, advanced training. ^o	n = ^o
a. Giving grades ^o	No. of response ^o 6 ^o Weighting ^o 6.90% ^o	45 ^o 51.72% ^o	36 ^o 41.38% ^o	87 ^o
b. Identifying learning and teaching needs ^o	No. of response ^o 14 ^o Weighting ^o 16.09% ^o	39 ^o 44.83% ^o	34 ^o 39.08% ^o	87 ^o
c. Identifying the purpose(s) of assessments ^o	No. of response ^o 14 ^o Weighting ^o 16.09% ^o	46 ^o 52.87% ^o	27 ^o 31.03% ^o	87 ^o
d. Identifying students' strengths and weaknesses based on assessment results ^o	No. of response ^o 10 ^o Weighting ^o 11.49% ^o	36 ^o 41.38% ^o	41 ^o 47.13% ^o	87 ^o
e. Placing students onto groups according to their abilities ^o	No. of response ^o 15 ^o Weighting ^o 17.24% ^o	44 ^o 50.57% ^o	28 ^o 32.18% ^o	87 ^o
f. Being aware of the impact and power of public examinations ^o	No. of response ^o 24 ^o Weighting ^o 27.59% ^o	40 ^o 45.98% ^o	23 ^o 26.44% ^o	87 ^o
g. Elaborating on assessment criteria (i.e. transparent assessment practices) ^o	No. of response ^o 10 ^o Weighting ^o 11.49% ^o	46 ^o 52.87% ^o	31 ^o 35.63% ^o	87 ^o
h. Giving students opportunities to share the voice about assessment (i.e. democratic assessment practices) ^o	No. of response ^o 7 ^o Weighting ^o 8.05% ^o	47 ^o 54.02% ^o	33 ^o 37.93% ^o	87 ^o
i. Delivering assessment of learning to evaluate students' learning ^o	No. of response ^o 12 ^o Weighting ^o 13.95% ^o	37 ^o 43.02% ^o	37 ^o 43.02% ^o	86 ^o

4^o

j. Delivering assessment for learning to support students' learning ^o	No. of response ^o	12 ^o	28 ^o	46 ^o	86 ^o
	Weighting ^o	13.95% ^o	32.56% ^o	53.49% ^o	
k. Delivering assessment as learning to promote students' learning ^o	No. of response ^o	10 ^o	32 ^o	44 ^o	86 ^o
	Weighting ^o	11.63% ^o	37.21% ^o	51.16% ^o	
l. Delivering norm-referenced assessments to compare students' learning among one another ^o	No. of response ^o	6 ^o	53 ^o	28 ^o	87 ^o
	Weighting ^o	6.90% ^o	60.92% ^o	32.18% ^o	
m. Delivering criterion-referenced assessments to measure students' learning in relation to assessment rubrics ^o	No. of response ^o	10 ^o	48 ^o	28 ^o	86 ^o
	Weighting ^o	11.63% ^o	55.81% ^o	32.56% ^o	

Section 4: 'What' of language assessment^o

5. Please rate to what extent you think you WERE trained in the following domains.^o

* e.g., 1 – 2 courses^o

Statements ^o		Not adequate ^o	Minimal ^o	Adequate ^o	n = ^o
a. Assessing receptive skills (i.e. reading and listening) ^o	No. of response ^o	11 ^o	34 ^o	42 ^o	87 ^o
	Weighting ^o	12.64% ^o	39.08% ^o	48.28% ^o	
b. Assessing productive skills (i.e. speaking and writing) ^o	No. of response ^o	6 ^o	35 ^o	46 ^o	87 ^o
	Weighting ^o	6.90% ^o	40.23% ^o	52.87% ^o	
c. Assessing grammar ^o	No. of response ^o	10 ^o	33 ^o	44 ^o	87 ^o
	Weighting ^o	11.49% ^o	37.93% ^o	50.57% ^o	
d. Assessing vocabulary ^o	No. of response ^o	12 ^o	43 ^o	31 ^o	86 ^o
	Weighting ^o	13.79% ^o	49.43% ^o	35.63% ^o	
e. Assessing integrated language skills ^o	No. of response ^o	21 ^o	44 ^o	31 ^o	86 ^o
	Weighting ^o	24.14% ^o	49.43% ^o	35.63% ^o	
f. Integrating language assessment theories to assessments ^o	No. of response ^o	16 ^o	44 ^o	27 ^o	87 ^o
	Weighting ^o	18.39% ^o	50.57% ^o	31.03% ^o	
g. Using statistics to evaluate the quality of assessment items ^o	No. of response ^o	44 ^o	31 ^o	12 ^o	87 ^o
	Weighting ^o	50.57% ^o	35.63% ^o	13.79% ^o	
h. Establishing reliability of assessments ^o	No. of response ^o	25 ^o	49 ^o	13 ^o	87 ^o
	Weighting ^o	28.74% ^o	56.32% ^o	14.94% ^o	
i. Establishing validity of assessments (i.e. to only assess what has been taught) ^o	No. of response ^o	19 ^o	51 ^o	17 ^o	87 ^o
	Weighting ^o	21.84% ^o	58.62% ^o	19.54% ^o	
j. Establishing authenticity of assessments (i.e. assessments related to real-life situations) ^o	No. of response ^o	11 ^o	39 ^o	37 ^o	87 ^o
	Weighting ^o	12.64% ^o	44.83% ^o	42.53% ^o	
k. Establishing practicality of assessments (i.e. assessment manageable for	No. of response ^o	12 ^o	50 ^o	25 ^o	87 ^o

5^o

both teachers and students) ^o	Weighting ^o	13.79% ^o	57.47% ^o	28.74% ^o	
l. Evaluating washback effect on learning, teaching, curricula, and institutions (e.g. impacts of following a test-driven syllabus) ^o	No. of response ^o	16 ^o	49 ^o	22 ^o	87 ^o
	Weighting ^o	18.39% ^o	56.32% ^o	25.29% ^o	
m. Evaluating innovations and trends in assessment ^o	No. of response ^o	29 ^o	41 ^o	15 ^o	85 ^o
	Weighting ^o	33.33% ^o	47.13% ^o	17.24% ^o	

6. Please rate to what extent you think you NEED more training in the following domains.^o

e.g., 1 – 2 seminars/ workshops; ^ e.g., 1 – 2 full courses^o

Statements ^o		None ^o	# Yes, basic training ^o	^ Yes, advanced training ^o	n = ^o
a. Assessing receptive skills (i.e. reading and listening) ^o	No. of response ^o	10 ^o	29 ^o	47 ^o	86 ^o
	Weighting ^o	11.63% ^o	33.72% ^o	54.65% ^o	
b. Assessing productive skills (i.e. speaking and writing) ^o	No. of response ^o	10 ^o	28 ^o	48 ^o	86 ^o
	Weighting ^o	11.63% ^o	32.56% ^o	55.81% ^o	
c. Assessing grammar ^o	No. of response ^o	11 ^o	29 ^o	45 ^o	85 ^o
	Weighting ^o	12.94% ^o	34.12% ^o	52.94% ^o	
d. Assessing vocabulary ^o	No. of response ^o	7 ^o	31 ^o	47 ^o	85 ^o
	Weighting ^o	8.24% ^o	36.47% ^o	55.29% ^o	
e. Assessing integrated language skills ^o	No. of response ^o	4 ^o	30 ^o	52 ^o	86 ^o
	Weighting ^o	4.65% ^o	34.88% ^o	60.47% ^o	
f. Integrating language assessment theories to assessments ^o	No. of response ^o	10 ^o	44 ^o	32 ^o	86 ^o
	Weighting ^o	11.63% ^o	51.16% ^o	37.21% ^o	
g. Using statistics to evaluate the quality of assessment items ^o	No. of response ^o	10 ^o	46 ^o	30 ^o	86 ^o
	Weighting ^o	11.63% ^o	53.49% ^o	34.88% ^o	
h. Establishing reliability of assessments ^o	No. of response ^o	9 ^o	43 ^o	34 ^o	86 ^o
	Weighting ^o	10.47% ^o	50.00% ^o	39.53% ^o	
i. Establishing validity of assessments (i.e. to only assess what has been taught) ^o	No. of response ^o	10 ^o	39 ^o	37 ^o	86 ^o
	Weighting ^o	11.63% ^o	45.35% ^o	43.02% ^o	
j. Establishing authenticity of assessments (i.e. assessments related to real-life situations) ^o	No. of response ^o	13 ^o	35 ^o	37 ^o	85 ^o
	Weighting ^o	15.29% ^o	41.18% ^o	43.53% ^o	
k. Establishing practicality of assessments (i.e. assessment manageable for both teachers and students) ^o	No. of response ^o	9 ^o	37 ^o	39 ^o	85 ^o
	Weighting ^o	10.59% ^o	43.53% ^o	45.88% ^o	
l. Evaluating washback effect on learning, teaching, curricula, and institutions (e.g. impacts of following a test-driven syllabus) ^o	No. of response ^o	15 ^o	41 ^o	29 ^o	85 ^o
	Weighting ^o	17.65% ^o	48.24% ^o	34.12% ^o	
m. Evaluating innovations and trends in assessment ^o	No. of response ^o	10 ^o	46 ^o	29 ^o	85 ^o
	Weighting ^o	11.76% ^o	54.12% ^o	34.12% ^o	

6^o