Measuring Pre-Service English Teachers' Assessment Competence in The Competence-Based English Curriculum in Vietnam

Duong Thu Mai[1], Pham Thi Thu Ha[2]

[1] maidk2018@gmail.com
University of Languages and International Studies,
Vietnam National University
(Vietnam)

[2] haptt1983@vnu.edu.vn
University of Languages and International Studies,
Vietnam National University
(Vietnam)

✉ Corresponding author

ABSTRACT: *Conducting assessment is well-known for posing serious challenges to teachers, particularly with the pluralism in assessment approaches. In Vietnam, English language teachers recently are required to master specified pedagogical competencies, including assessment of students' newly defined language competence, and the national English curriculum has been continuously updated, while most teachers are not prepared for conducting classroom assessment in a competence-based approach. Local English teacher training universities have responded to this need by implementing language assessment courses. The conceptualization of pre-service teachers' assessment competence and the effects of these courses, however, remains fairly unknown. This study examines the construct of pre-service teachers' assessment competence and the results of one introductory language assessment course at a leading language teacher education institution in Vietnam, focusing on measuring the competence of teachers-to-be in language assessment. Both Rasch analysis and descriptive statistics were used to analyze the functioning of the items in a localized construct of assessment literacy for pre-service English teachers and these teachers' self-assessment results, which revealed that a) many practical components of the construct of language assessment competence can function well, but some requirements need to be adapted and instructed with more attention for students in the future courses, and b) while many student teachers could sufficiently perform the required language assessment tasks and were fundamentally ready for the job, others faced serious difficulties in applying the more abstract principles. The findings shed light on the necessary adjustments to be made in defining assessment literacy for the target teachers and offer suggestions for the future professional development of both these teachers and in-service ones.*

1. Introduction

In modern education, contrasting paradigms of assessment co-exist, namely traditional and alternative assessment (McNamara, 2003). Under their influence, educational assessment has gained a new status less peripheral to teaching. In this context, the role of teachers as assessors changes significantly. Conducting their familiar tasks of classroom assessment properly becomes a challenge to most teachers, while their assessment competence, also popularly named as assessment literacy, remains limited.

It is therefore not surprising to hear teachers' complaints about assessment burden and their low levels of confidence in assessment (Mertler, 2009). Many teachers feel inadequate or just partially prepared when they have to implement classroom assessments and make assessment-related decisions (Mertler, 1999; Mertler & Campbell, 2005, cited in Siegel &Wissehr, 2011; Berry et al., 2017; Xu and Brown, 2017, cited in Weng & Shen, 2022; Sultana, 2019)). Despite all this, assessment literacy training for teachers is often scarce or inappropriate (Popham, 2006).

Asking the teachers to do a job for which they have not been well trained is unfair and could lead to serious assessment flaws such as poor reliability, questionable validity, inappropriate use, or poor connection to learning (Malone, 2009). Having been reflected in many articles, these issues have attracted widespread attention in language teacher education

Locally, in the last decade, due to the explicit adoption of a competence-based education approach (Educational Law, 2005), Vietnamese teacher training institutions have been required to transform their curricula into those targeting specific professional competencies. In foreign language education, the National Foreign Language Project 2020 has issued and updated the English Language Teacher Competence Framework (ETCF), which serves as the basis for Vietnamese curriculum development at language teacher training institutions. One of the components of this ETCF is language assessment competence. In such contexts, the University of Languages and International Studies - Vietnam National University (ULIS-VNU) has introduced a compulsory Language Assessment (LA) course for final-year student teachers majoring in English language teaching methodology (TEFL students). Each academic year since 2015, about 200 prospective English teachers undertook the course.

LA courses have been included in TESOL curricula around the world since the 1990s (Brown & Bailey, 2008) with many differences. The established course of LA in this study is described to resemble Inbar-Lourie's (2008) description of competence-based courses, with the focus on "learning, negotiating, discussing, experiencing and researching the core LA framework" (p.396). As well, the course is highly localized as students are required to perform authentic assessment activities required for English teachers in Vietnam (described more specifically in the next section). So far, there has been no official and systematic investigation into the appropriateness of the established LA competence and the prospective teachers' ability to perform the targeted assessment task. After eight years of implementation, it is overdue that this investigation was conducted to better inform future teaching and course revision and to provide practical experience to other FL teacher training institutions in teaching LA to their prospective teachers.

2. Background to the study

2.1. Language classroom assessment competence

Assessment often means the process of giving meanings to judgments to make decisions about learners' competence (Griffin & Nix, 1991). While the term "assessment competence" is infrequently used in language education, its synonym, "assessment literacy" (AL), is a more common term defined with varied foci and scope. On the one hand, there are more condense definitions, such as AL conveys an understanding of the principles of sound assessment to appropriately integrate assessment with instruction and to utilize appropriate forms of teaching (Stiggins, 2002; McMillan, 2000). On the other hand, more specific definitions can also be found. Some of these emphasize the key activities teachers should be able to do well in the assessment process. Mertler (2004), for instance, depicts an assessment-literate teacher as someone who can recognize good assessment practices, understand assessment methods, report assessment results, and integrate assessment and learning. According to Boyles (2005), foreign language AL is quite close to testing literacy, including familiarity with testing practices, the use of assessment methods, the explanation and analysis of collected results, the decision-making process and the use of assessment results for teaching.

Oriented towards the theoretical side, the North Central Regional Educational Laboratory in the United States (2007) adds that teachers should understand terms, standards, and alternatives in assessment. Assessment literates, as articulated in many other articles, are able to understand more than the practical knowledge to conduct assessment. Davies (2008) and Taylor (2009) articulate that language AL should include the knowledge, skills, and principles of conducting assessment appropriately. Sharing the same view but attempting to model these

components in a framework, Fulcher (2012) presents three dimensions to classify the types of knowledge in AL: the Practice dimension (including the knowledge, skills and abilities of language testing), the Principle dimension (which indicates an understanding the guidance for conducting language assessment practices, such as the process, the principles and the concepts) and the Contexts dimension (which shows a grasp of the deeper and more influential frameworks in LA and a clear vision of the two lower dimensions, such as the history and philosophies underlying the field). Obviously, this definition is comprehensive and can be used as a framework for extracting assessment literacy definitions for different types of educationalists, such as teachers, testers, or researchers.

In the large space of AL created by Fulcher, contents can generally be ranked along a continuum from theoretical/philosophical to practical. Many books on language assessment fundamentals can specify the key components of assessment literacy (such as (Bachman, 1990; Bachman & Palmer, 1996; McNamara, 1996), some of which are also more practical than others. Notably, more recent assessment literacy definitions become more focused on the transformation of the acquired knowledge into the performance of assessment tasks. Malone (2013) defines and shares several authors' definitions that language AL is the familiarity with definitions in language testing or measurement and the application of this knowledge. To obtain a more comprehensible discussion over language AL conceptual frameworks, Giraldo (2021) and Weng & Shen (2022) provide a thorough review of language AL literature.

To emphasize the language teachers' ability to realize authentic assessment tasks in the classroom, the term "assessment competence" is adopted instead of assessment literacy in this study. This is because the core nature of 'competence,' reflected in its various definitions, is the combination of a number of attributes, such as knowledge, skills, attitudes, ability, etc. in order to complete specific major tasks. Pill and Harding (2013) and Taylor (2013) are among the authors advocating this defining approach,

claiming that assessment literacy refers to a bank of competencies that permit the owners to not just understand but also judge, create and analyze tests. Language assessment competence is used to entail the integration of knowledge about the language, theoretical and practical knowledge of assessment, skills and dispositions in conducting language assessment. Language teachers' assessment competence is usually clarified into smaller tasks in the LA practices.

The most relevant rationale for the current study is provided by Inbar-Lourie (2008), who confirms the need to establish standards and proficiency levels in assessment literacy/ competence, which can be of critical use for teacher training and professional development. This call for action is highly justifiable by the comprehensive body of knowledge, skills and tasks, etc., that an assessment literate needs to possess, as in Fulcher (2012). Malone (2013) also emphasizes that LA standards do not yet exist in teacher training and calls for studies to investigate the issue. In more specific terms, there is hardly any attempt to design and study the assessment competence with a specific level of performance for concerned LA protagonists such as pre-service teachers. Stiggin (2004) has also pinpointed one reason for teachers' limited assessment training is the unclear visualization of what makes up good assessment training. This study can be considered an attempt to respond to their calls, reflecting the establishment of assessment competence standards at an English teacher training institution and the achievement level of the pre-service teachers after training.

2.2. Relevant studies on the training of assessment literacy and competence for teachers

To date, a significant number of empirical studies have been disseminated on teachers' assessment literacy/competence for teachers. This study will focus on the literature of AL created by and applied to language teachers, rather than for language testers. Many studies focused on defining the language AL construct. Malone (2013) reported the intensive construction of an online tutorial on AL for US language teachers with the participation of 150 language instruction

experts and language testing experts. The findings of the study highlighted the differences between the two groups in their definitions of language testing basics, with the testing experts favouring more technical aspects and the instruction experts advocating more practical knowledge. A similar study was conducted by Jeong (2013) to compare language-tester and non-language-tester experts in constructing LA course contents. The language instructors were found to be less interested in test theories but more supportive of classroom assessment and test accommodation. In two studies conducted in 1996 and 2007, Brown and Bailey (2008) analyzed teaching experts' judgments on key LA course contents for ESL teacher candidates. They found a stable knowledge basis of five topic areas. Some of the most popular include the traditional types of validity and reliability, the sources of inconsistency, item writing techniques for different language skills and item analysis, language skill measurement, score interpretation frameworks, etc. In their literature review of empirical studies on language teachers' language AL, Weng, and Shen (2022) categorise numerous studies according to language AL levels and factors influencing language AL.

Recent studies also explore the process of teachers' assessment literacy training. Mertler (2003) developed a Classroom Assessment Literacy Inventory with assessment scenarios accompanied by selected-response questions to investigate the differences between in-service and pre-service teachers. He reported that the latter group outperformed the former in five out of seven competence areas, and the most difficult ones for both groups were developing valid scoring procedures and communicating to students. Volant and Fazio (2007) based on a questionnaire with both open and close items to study the development of assessment literacy among pre-service teachers throughout their training and found quite a disappointing result: the teacher candidates did not show strong confidence, especially in formative assessment, and desired for more practical training in this field.

Xu and Liu (2009) reported a case study of a college EFL teachers' assessment knowledge and practice from three dimensions: temporality, sociality, and place. The studied teacher told three stories reflecting her knowledge, practice and development in language assessment. The study brought to light that in-service teachers might be affected by previous education, colleagues and working contexts in conducting assessments, and their assessment literacy was a highly dynamic concept. Kiomrs, Abdolmehdi and Naser (2011) conducted another study on the effect of Iranian EFL teachers' assessment literacy on their teaching practices. The research instrument was an objective test. It was found that the teachers had low levels of literacy and tended to suffer from negative washback, i.e. to manipulate their teaching according to the requirements of major standardized tests.

Deluca and Linger's (2010) survey study of 228 pre-service teachers ' teacher's confidence levels and needs in assessment literacy is highly relevant to our study. The research instrument consists of 45 Likert-scale items about the practice, theory and philosophy in two major domains namely Assessment for Learning and Assessment of Learning, and some other items on the participants' desired assessment areas. The data collected were very rich, and the researchers could generate a vivid picture of the differences between groups of teacher candidates in the study. The study found most teachers to be highly confident in the areas of AL surveyed, particularly in the assessment of learning, and some differences existed between those with official training and those without. This finding does not simply reflect the high ability of the candidate teachers, but claims that such a positive result was caused by their overconfidence.

A thorough review of language teacher training on language AL in Weng and Shen (2022) also identifies that "the vast majority of language assessment programs were designed for in-service teachers…with fewer programs for pre-service teachers and student teachers" (p.5). It is agreed that these target groups should be provided with more training since they would carry out assessments as long as they start their career in the early stages (Weng & Shen, 2022).

In the context of Vietnam, few studies have

been carried out on language teachers' language AL. In her qualitative study, Bui (2021) explored five English high-school teachers' perceptions of language AL. The results showed that teacher's language AL was not adequate and their language assessment practices were based mainly on traditional testing methods rather than alternative assessment.

In general, the results of the majority of studies suggest a low level of language AL among the participants. Notably, in terms of research methods, most studies used questionnaires with selected response items, Likert-style responses or surveys to investigate the teachers' AL from their own perspective. Few have investigated the teachers' demonstration of the assessment literacy in an assessment course and using a logistic model and statistical analysis to analyze the teacher participants' performance in competence-based assessment methods. It is in this aspect that our study can supplement the current relevant literature, as presented in the following sections.

2.3. The LA competence for pre-service English language teachers at ULIS-VNU

The construct of "LA competence" for the target group was established and validated over one year by the course designers. In terms of contents, the construct consists of five sub-competences (Table 1) corresponding to steps in the language assessment process, each of which can be further specified into smaller and more observable performance-based learning targets for students. Besides, the performance levels of the tasks correspond to *remembering* and *primary processing of knowledge,* the lowest levels in Singer's cognitive competence taxonomy (2006). This selection of low-performance demands for the tasks is justified by the challenging nature of the assessment field and the novelty of assessment training in teacher education in Vietnam. The tasks also align with the basic level of assessment literacy indicated by Inbar-Louri (2013, p307), i.e. an *understanding* of the situated (or local) approach to language assessment and *an awareness* of the impact of assessment decisions. This LA competence construct was verified with and endorsed by six local LA experts in one independent study (Duong et al., 2017). The experts had to comment on different validity aspects of this core LA construct and also its feasibility before it was revised. The final version of the LA competence framework is recorded in Table 1.

Besides three mini-quizzes developed by the lecturers, students are required to write two essay assignments. In essay 1, students need to work in pairs to discuss and co-author an evaluative essay to evaluate a provided authentic final test of English for 10th graders in Vietnam. The assignment 1 essays are

Table 1. The pre-service teachers' targeted LA competence

1. **Evaluating (the strengths and weaknesses of) an English language assessment plan for a specific group of students**
a) *Understand* the process of conducting assessment in English teaching. b) *Identify* various assessment purposes of language assessment. c) *Apply and evaluate* English learning and assessment targets and contents (such as models of communicative language competencies)
2. **Preparing an English language assessment plan to foster students' learning in the classroom**
a) *Comprehend* the reciprocal relationship between assessment and instruction, learning results and learning process. b) *Comprehend* the bases of designing a classroom-based assessment plan for English learning. c) *Select* learner-centered assessment methods in language assessment. d) *Understand* methods to prevent language assessment errors. e) *Make inferences* on students' language competence based on their assessment results.
3. **Designing/selecting language assessment instruments**

a) *Distinguish* the importance of various language assessment instruments for different assessment purposes. b) *Use* some cognitive taxonomies and language competence models to design assessment targets. c) *Evaluate* language assessment instruments. d) *Develop* or select language assessment instruments for learning.	
4. Performing analyses of the assessment results	
Perform basic descriptive statistical analysis on language assessment results to make interpretations on students' competence, instrument quality and other components of assessment process.	
5. Providing feedback to students	
a) *Provide* feedbacks to students according to personalized assessment reference frameworks. b) *Design* intervention strategies based on acquired assessment results.	

scored against a 7-criteria scale with a focus on components 1 to 3 in Table 1. In essay 2, students have to demonstrate a broader LA competence by evaluating an authentic English assessment plan for 10th graders. Students individually write__1500-word essays. The plan has some components such as targets, methods, instruments, etc. The scoring scale for this assignment represents almost all the LA competence targets of Table 1 with 9 criteria. A number of guiding questions and supplementary materials for both essays were provided (such as the national English language targets for 10th graders, local language testing policies, CEFR description of the students' B2 targets). Both the scales for essays 1 and 2 are four-point rating scales, in which the performance levels are ordered according to the accuracy and persuasiveness of the students' arguments. The items in the scales are presented in Table 2.

2. Methods

This research aims primarily to investigate the quality of the items in the construct of pre-service teachers' LA competence, which is revealed in the students' performance of two assessment tasks. It is also expected to discover the student teachers' level of LA competence from their performance in the tasks and their self-assessment. The following questions are addressed:

1. To what extent do the items in the scales succeed in measuring the targeted pre-service teachers' LAL?

2. What is the pre-service teachers' level of language assessment competence?

In the autumn-winter semester of 2022, the data were collected. The main participants comprised of 190 senior pre-service teachers, majoring in teaching English as a foreign language at ULIS - VNU. Almost all of the students are young females with an English proficiency level equal to C1 (Common European Framework of

Table 2. Items in the scoring scales for two assignments

Assignment 1	Assignment 2
1. Evaluate general quality of assessment targets 2. Evaluate content validity of the test 3. Evaluate the quality of selected response questions 4. Evaluate the quality of constructed response questions 5. Evaluate test instruction 6. Evaluate the test administration (logistic and time) 7. Evaluate the conformity to local language testing policies	1. Evaluate the assessment steps/procedure 2. Identify the purposes of assessment and the focus on assessment for learning 3. Evaluate the LA targets 4. Evaluate the use of assessment methods 5. Evaluate the quality of instruments 6. Evaluate the focus on validity in the plan 7. Evaluate the reliability 8. Evaluate the authenticity 9. Use language assessment terms

Reference). Three lecturers with at least ten years of ELT teaching experience also participated in scoring the students' assignments. Two of them are PhDs in applied linguistics and language assessment, one has MA degree majoring in educational assessment.

As students' assignments were being collected, students and lecturers were informed of the use of assignments for research purposes. The meaning of the rubrics was agreed among the lecturer raters before scoring the assignments. 190 students' assignments were scored. After being collated, students' scores in two essay assignments were entered into data files. Students' competence in the targeted LA competence is estimated by analyzing the scores of two essay assignments. The data analysis was conducted in two steps. Firstly, to obtain evidence for the validity of the two scoring scales, an analysis of the scores of both assignments was implemented with a Rasch logistic model. Item response theory models, including Rasch models, were introduced into second language testing in the middle 1980s and received widespread support. Rasch (1960, 1980) is one of the authors proposing the concept of underlying ability continua, which he named "latent traits", and empirical mathematical models (or statistical patterns) to predict students' probability of success on one item based on only their observed "ability" and the item "difficulty", which are later known as Rasch models. By analysing the data with Rasch models, the study can discover the relation between the demands of each criterion in the scales and the students' capacity to perform it. The models have been used extensively in language test validation (McNamara, 1996) thanks to many advantages. One remarkable strength of the Rasch model application in this study is that besides the sound estimation of students' ability and criteria difficulty, the results can provide the evidence for the validity of these criteria/items in measuring the pre-service teachers' language assessment competence - the core construct or latent trait. In this study, the Rasch model to be used is the unidimensional rating-scale model to match the numbers of dimensions in the scoring scales of two assignments (both of which measure one-dimension constructs) and the ordinal data (the performance levels). Item reliability and fit statistics are obtained from the application of Rasch model to confirm the functioning of the criteria in the two scoring scales. The data analysis was based on the software Conquest (Wu, 2003). To provide more evidence for the functioning of the items, evidence from item-person maps was also analyzed.

To answer research question 2, firstly, descriptive statistics of students' assignment scores were interpreted. Also, the students were asked to respond to a self-evaluation questionnaire with 16 five-point Likert-style questions paraphrasing and clarifying the LA competence targets in Table 1. The use of self-assessment questionnaires could shed more light on the pre-service teachers' abilities with the targets unassessed in the course assignments. For instance, in both essay assignments, the pre-service teachers were required only to evaluate, rather than create an assessment instrument. In the course, they actually had more demanding practices, such as writing test questions or giving feedback. The self-assessment questionnaire could provide evidence in these otherwise unassessed aspects. The students received the link to the survey questionnaire in their emails after the course finished. However, only 87 students responded to this voluntary survey, so data collection had to be further conducted with a hard copy version in the second semester for those who did not respond to the online survey. Due to the absence of students who had to participate in the practicum, the final number of respondents was 108. The survey results were analyzed for descriptive statistics and trends.

4. Findings

4.1. Research question 1 - The construct of Students' success with the targeted language competence

Rasch model analysis

Assignment 1: Students' competence in test evaluation

The results reveal that the assignment scale has good reliability, with separation reliability in Rasch model analysis being 0.988. This means that the criteria in the scale are consistent

Table 4. Item analysis results of assignment 1 scale

Item	Name	ESTIMATE	ERROR	MNSQ	CI	t
1	Evaluate targets	0.534	0.049	0.81	(0.88, 1.12)	-3.4
2	Evaluate content validity	-0.615	0.059	1.08	(0.83, 1.17)	0.9
3	Evaluate selected-response questions	-0.214	0.054	0.8	(0.86, 1.14)	-2.9
4	Evaluate constructed-response questions	0.341	0.05	0.75	(0.88, 1.12)	-4.5
5		0.023	0.052	1.1	(0.87, 1.13)	1.5
6	Evaluate test administration	0.675	0.049	1.33	(0.89, 1.11)	5.1
7	Evaluate conformity to local policies	-0.744*	0.128	1.19	(0.82, 1.18)	1.9

with each other and range significantly in their difficulty, and they can discriminate students into some levels of ability in language test evaluation. However, the students' separation reliability is not very high, at 0.6, which can be explained by the small number of items. The scale does not seem to separate students well into many levels of LA competence.

Table 4 reveals how the criteria function together in measuring the pre-service teachers' test evaluation competence and how difficult the criteria are to the teachers. When the 0.7-1.3 range of acceptable fit statistics in a performance task by Adam and Khoo (1997) are referred to, only item 6 (evaluating test administration) slightly functions unpredictably in this assignment scoring. However, considering the weighted mean-square fit statistics (MNSQ), five out of seven are marginally outside the confidence intervals (CI), which shows that they are significantly different from 1 given this sample of student teachers, and that the items do not appear to work well with one another to represent this underlying construct. However,

As regards the item difficulty estimates, the most difficult item is 6 (evaluating the test administration) while items 2 (assessing the content validity) and 7 (evaluating the conformity of the test to the local polities of language assessment) are the least challenging to the pre-service teachers. Also, figure 1 visually demonstrates whether the criteria (items) match the students' ability. It represents the distribution of students and items in the assignment. Each X in the left represents 7.4 students. Items vary significantly in the difficulty, with item 6 being the hardest and item 2 and 7 being the easiest.

Logit	Students	Items
2	XXXXXX	
	XXXXXX	
1	X	
	XXXXX	
	XXXX	
		6
	X	
	X	
	XXXXX	1
	XXXX	4
	X	
	XX	
0		5
	X	
	X	3
	X	2 7

Figure 1. Assignment 1 Item-person map

Table 5. Item analysis results of assignment 2 scale

	Item	ESTIMATE	ERROR^	MNSQ	CI
1	Assessment Procedure	0.176	0.083	0.9	(0.81, 1.19)
2	Assessment purposes	0.073	0.084	1.04	(0.80, 1.20)
3	Assessment targets	-0.544	0.09	1.13	(0.77, 1.23)
4	Assessment methods	-1.113	0.097	1.14	(0.72, 1.28)
5	Assessment instruments	0.005	0.084	0.82	(0.80, 1.20)
6	Validity	0.55	0.081	0.74	(0.81, 1.19)
7	Reliability	0.64	0.081	1.02	(0.81, 1.19)
8	Authenticity	0.775	0.081	1.52	(0.81, 1.19)
9	Assessment Terminology	-0.560*	0.241	0.67	(0.77, 1.23)

The item difficulty range is located quite lower than the student ability range, so it can be concluded that some students performed this essay assignment with high probability of success while others students had a lower probability and some even struggled.

Assignment 2: Students' competence in assessment plan evaluation

The findings reveal that the assignment scale has good reliability, with item separation reliability in Rasch model analysis being 0.982. This means that the criteria in the scale are consistent with each other in measuring the construct but significantly different in their difficulty. The students' separation reliability, however, is not very high, at 0.58, which can be explained by the small number of items. In other words, the items are quite distinct in their difficulty, but the students' ability range cannot be divided into many levels.

Table 5 shows the evidence for the construct validity of the LA competence in assignment 2. As regards the weighted MNSQ, seven out of nine are within the range of acceptable fit statistics (0.7-1.3) suggested by Adam and Khoo (1997), and most of them are within the range of confidence intervals (CI), indicating that they are not significantly different from 1. These findings are considered good evidence for 1) the existence of one underlying construct of LA competence in the assignment and 2) the co-functioning of items with one another to represent this underlying construct. In short, only the fit statistics of item 8 (authenticity) need to be seriously examined. This item does not seem to measure the pre-service teachers' LA competence as well as others. Besides, the fit statistics also indicate good discrimination of most items. Item 8 is a poorly discriminating item for students' LA competence, and item 9 is less discriminating than the model predicts. In other words, the distinction of students' ability in LA competence is not predictable by item 8 (their evaluation of the authenticity of the assessment plan), which implies that the student's ability to evaluate test task authenticity needs to be attended to in future courses. When students' assignments are revisited, it is quite clear that many students misunderstood the term "authenticity" to be just a characteristic of the topics in a language task and not a quality of the tasks. Item 9 on the use of assessment terms is quite easy and cannot discriminate against students because almost all students received high raw scores of 3 or 4 for this criterion. This is explicable by the students' advanced level of English proficiency and possibly the lecturers' attention to explaining the terms in this introductory course. This item may be removed from the scale for future students.

Regarding item difficulty estimate, the most difficult item is item 8 (evaluating the authenticity), followed by item 7 (assessing the reliability) and item 6 (assessing the content and scoring validity). Students showed a good understanding of content validity in assignment 1, but it can be inferred from assignment 2 analysis that they found

scoring validity a hard topic. It is found from their assignment papers that scoring validity is often mistaken for reliability. Only a small number of students could analyze the reliability of the rubrics in the assessment plan, while many of them only consider the assessment plan's reliability from the lack of strict requirements for double scoring and administration. The lectures on the qualities of good assessment, whose contents were based on Bachman (1990), were in the middle of the course and the contents are deliberately reviewed through the later half of the course, but it seems assessment principles are still the most challenging lesson to the pre-service teachers, who have scant experience in LA practices. The easiest item is item 4 (assessing the methods of assessment) and item 3 (assessing the targets in the assessment plan). Students seem to master these fundamentals of language assessment quite well.

Logit	Students	Items/criteria
3		
	XXX	
	XXXXX	
2	XXXX	
	XXXXXXXXXX	
	XXXXXXXXX	
	XXXXXXX	
	XXXXX	
1		
	XXXXX	
		8
	XXXXX	
		7

	XXXX	6
	X	
	XXX	
	X	1
		2
0		5
		3 9
	X	
		4

Figure 2. Assignment 2 item-person map

Figure 2 shows whether the criteria (items) match the students' ability in assignment 2. It represents the distribution of students and items in the assignment. Each X on the left represents 2.4 students. Obviously, the range of item difficulty is quite large, with item 4 being the easiest and item 8 being the hardest. As the item difficulty range is located quite lower than the student ability range, it can be concluded that many students perform this essay assignment with a high probability of success, while a significant group of students may still struggle with all the requirements of the assignment.

The results from Rasch analyses of two assignment scores yield fairly inconclusive results on the functioning of items in the rating scales. There is more supporting evidence for the existence of the underlying constructs of LA competence (assignment 2) than for test evaluation (assignment 1). Within these constructs, the items relating to more abstract principles of LA are more challenging and less predictable by the model, while those relating to LA practices such as evaluating assessment methods and instruments, evaluating contents validity of assessment, using LA terms, understanding local policies in LA are more suitable for the student teachers' LA competence.

4.2. Pre-service teachers' language assessment competence.

Rasch analysis of student teachers' scores

Figures 1, 2 and 3 represent the distribution of students' ability measures in the two assignments against the item difficulty measures, the results from Rasch rating scale model analysis. In both
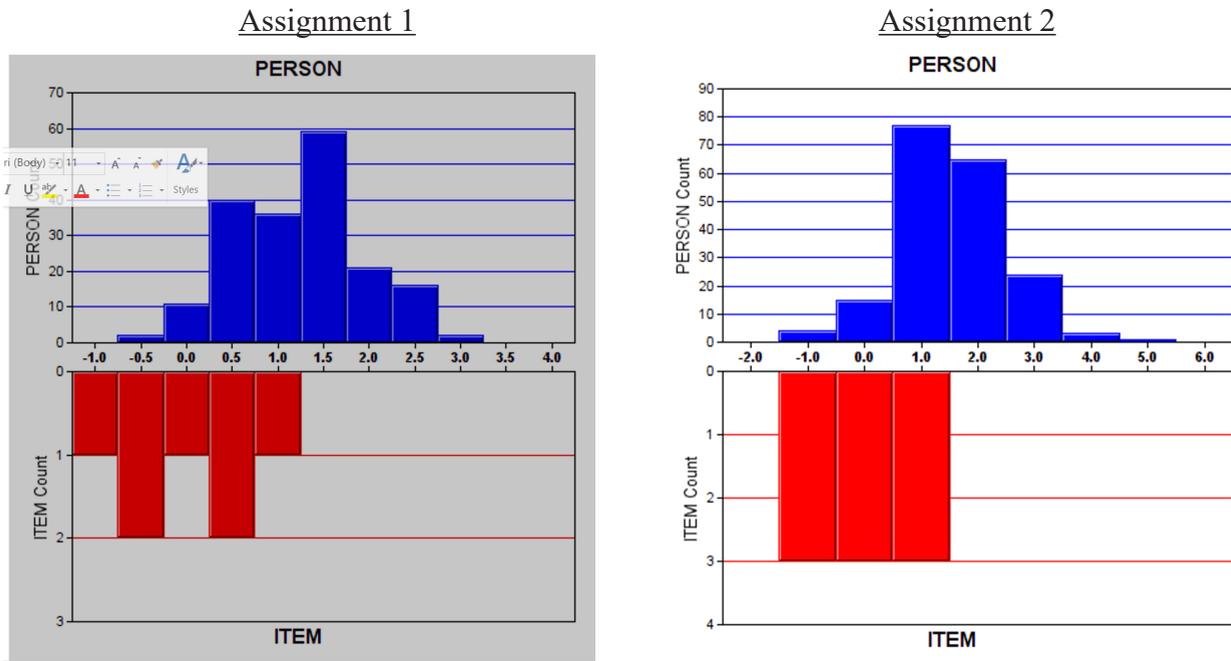
Figure 3. Student-item histogram of Rasch model estimates in the two assignments

Table 3. Descriptive statistics of assignments 1 and 2

Assignment 1					
	N	Min	Max	Mean	SD
Evaluate assessment targets	185	1	4	2.86	.826
Evaluate content validity	185	1	4	3.55	.674
Evaluate the quality of selected response questions	187	2	4	3.35	.696
Evaluate the quality of constructed response questions	187	1	4	3.00	.766
Evaluate the test instruction	187	1	4	3.22	.868
Evaluate the test administration	185	1	4	2.76	1.063
Evaluate the conformity to local LA policies	181	1	4	3.60	.679
Assignment 2					
	N	Min	Max	Mean	SD
Evaluate assessment procedure	189	1	4	3.18	.736
Identify and evaluate assessment purposes	187	1	4	3.19	.759
Evaluate assessment targets	189	2	4	3.48	.726
Evaluate assessment methods	189	1	4	3.62	.612
Evaluate assessment instruments	189	2	4	3.30	.699
Evaluate content and scoring validity	189	1	4	2.96	.725
Evaluate reliability	188	1	4	2.89	.886
Evaluate authenticity	189	1	9	3.03	1.476
Use LA terminology	189	2	4	3.48	.561

assignments, the range of estimates for students' ability is larger than for item difficulty, and the means of students' ability estimates are also higher, so there appear to be more students with a high probability of success in the two tasks than those who have difficulties, but a significant number of students still have less than 50% of success with the items.

Descriptive statistics of the assignment scores

The descriptive statistics of students' scores in two assignments are presented in Table 3.

All the LA competence components in assignment 1 received above average scores, suggesting the competent level of the pre-service teachers in this construct. The criteria receiving the highest scores are evaluating the conformity of the test to local policies (comparing the test targets against the national requirement for assessing 10th graders), and evaluating the content validity of the test (the compatibility of the test and the national course targets for 10th graders). These two tasks require the students to compare the test targets to a set of specific contents, so they pose a few challenges. For instance, most students could compare the number of items/skills/sections in the test and the required number of items/skills/sections by the Ministry of Education and Training. The most challenging criterion is evaluating the test administration such as time allowance, printing mistakes, test format, etc. The low mean of scores in this criterion may be due to students' negligence of evidence for their claims on the test administration aspects. Many of them just provided general comments on the test without giving supporting ideas, so the scores for them could only be 2 or 3. Assessment target evaluation was the area with the second lowest mean score, at 2.86. As stated, the provided test did not have a test specification, and students had to deduce the targets from the questions provided before evaluating them, which may be a hard job for even in-service experienced teachers.

Assignment 2 scores also receive positive descriptive statistics, with means ranging from 2.91 to 3.68. Thus, similar to the findings in Rasch model analysis, in general, the student's achievement in this assignment is significant. The most challenging tasks for students in this assignment are evaluating the reliability of the plan, evaluating the content, and scoring validity, which are two abstract qualities of language assessment and may be hard for them to understand. The easier tasks are evaluating the methods of assessment, evaluating the targets, and using assessment terms. It is interesting to note that "evaluating the targets" has changed the status from one of the hardest items in assignment 1 to one manageable one in assignment 2. Also noteworthy is the outstandingly high standard deviation of the item "evaluating the authenticity" of the assessment tasks. When students' assignments were more carefully studied, they not only revealed students' misconceptions or inadequate understanding of this quality of good language assessment but also the lecturer's differences in leniency. Authenticity in language assessment tasks means that the task aspects (the task and the conditions) resemble real-life communication. Student's weak performance for this criterion is mainly due to their inadequate understanding of the principle. The lecturers also showed a certain level of inconsistency as some only required students to understand material authenticity to give them the highest score, while others demanded a full understanding of the term.

Students' self-assessment of LA competence

It can be seen from Table 6 and Figure 4 that most students showed a positive perception of their competence. They were most confident in understanding the LA purposes, the steps and bases of an assessment plan. They also ranked themselves high on evaluating assessment instruments and selecting/ developing assessment instruments for student's learning. This may be explained by the focus of both assignments on test evaluation. Student's lower confidence was revealed in questions 9, 14 and 16 (making inferences about students' competence, designing instruction after assessment and performing basic descriptive statistical analysis). Interestingly, they assessed themselves obviously better in evaluating LA instruments than in developing/ selecting the instruments. The difficulty to them seem to be 'applying', a higher level of performance than "understanding" - the targeted level in this introductory course, also the level not emphasized by the lecturers. The application of assessment results to teaching is also not yet adequately covered in the course because this is interdisciplinary between LA and other English language teaching courses.

In comparison with the assignment results, the students' self-assessments shares some features in common while presenting new findings.

Table 6. Descriptive statistics of students' self-assessment survey

| Criteria	N	Mean	SD
1. Understanding the LA process	108	3.87	.597
2. Understanding the bases/steps to conduct a classroom assessment plan to foster students' learning	108	3.81	.686
3. Identifying purposes	108	4.12	.806
4. Evaluating targets	108	3.72	.863
5. Applying assessment targets	106	3.75	.677
6. Comprehending the relationship between assessment and instruction	108	3.77	.882
7. Using learner-centered LA methods	108	3.68	.593
8. Understanding errors and the methods to prevent them	108	3.85	.653
9. Making inferences on students' language competence	108	3.58	.613
10. Distinguish LA instruments	108	3.84	.672
11. Design targets for assessment instruments	108	3.62	.782
12. Evaluating LA instruments	108	3.90	.546
13. Developing/ selecting LA instruments	107	3.68	.560
14. Performing basic descriptive statistical analysis	108	3.46	.942
15. Providing feedback	108	3.77	.650
16. Designing appropriate teaching strategies	108	3.58	.628
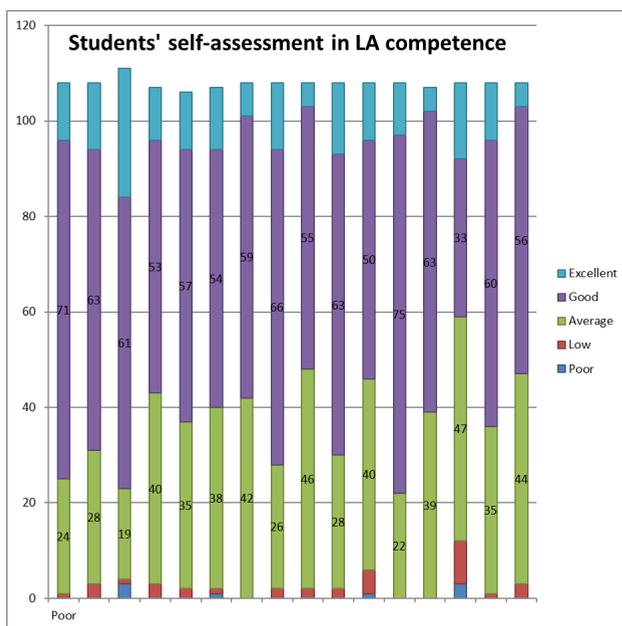
Figure 4. Student's self-assessment on their LA competence targets

The consistent findings of two assignments say that the pre-service teachers are confident and successfully respond to most of the required targets. Besides reflecting that positive trend, the self-assessment also adds that they are not ready to apply several LA tasks and theoretical principles. The pre-service teachers' difficulties with the more theoretical principles of LA have been reflected in previous studies (Mertler, 2003; Volant & Fazio, 2007). Mertler (2003) claimed that some of these difficulties could be alleviated when the teachers start working and are exposed to practical tasks. Moreover, the theoretical principles of LA are arranged to be in the higher level of knowledge in Fulcher (2012), which suggests their complexity. They may be introduced to teachers in their training, but understanding and applying them thoroughly can be a life-long professional development issue.

5. Conclusion and lessons for future implementation

This study has examined a localized construct of assessment literacy for Vietnamese pre-serviced English teachers and their performance in this construct in an introductory language

assessment course. From the results, the scales for measuring the student teachers' assessment literacy have generally good evidence for their quality. The young undergraduate teachers could evaluate an LA plan (the process, purposes, targets of assessment), understand and evaluate LA targets, comprehend the bases of designing an LA plan, select assessment methods to foster students' learning, distinguish and evaluate major types of LA questions/tasks with the focus on students' learning. Moreover, they show their competence to evaluate the other steps in the LA process such as analyzing LA results, reporting results, and connecting assessment to teaching.

Despite the modest scope of this study, some suggestions can be made for future LA course developers. The major lessons to share are related to course design. Firstly, regarding students' attainment of course targets, as reflected above, the tasks in this course can be considered suitable for most students. In our opinion, the success is largely attributed to the appropriate selection of performance levels in the course targets to match the students' limited experience. The course aims mostly at "understanding" rather than 'applying' the key knowledge in LA. We plan that our future courses maintain this level of performance level. However, as the areas of most challenges to the student teachers tend to be the understanding of the quality principles of good LA, and the application of LA results in improving language instruction, future courses should be adjusted in the theory-practice time allocation. Both lecturers and course designers for pre-service teachers, and even those for in-service Vietnamese teachers untrained in LA as well, should make sure to embed the lessons even further into the English teaching contexts and provide more situated and interesting examples in the theoretical lessons (such as the lecture on the assessment qualities) or more practice exercises (such as in developing more English tests or assignments).

The study addresses the calls for evaluating pre-service teachers' assessment literacy, which is a quite new construct compared to in-service teachers' assessment literacy. The course designers had to overcome some hurdles, including the time and effort consumption, iterative adjustment/clarification of the construct of language assessment competence, the need for constant lecturer team discussion to align our teaching and assessment to the targeted competence, and the modeling role of good assessors we have to perform to the students. We believe that many of these are also obstacles for developers and lecturers of LA courses in other contexts. The difficulties, rather than being daunting, are lessons for the future. They also stimulate us to conduct more empirical studies on the course in the future, one of which may be the analysis of pre-service teachers' misconceptions of LA knowledge, and a comparison between pre-service and in-service teachers' LA competence.

References

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2007). Evaluation assessment quality in competence-based education: a qualitative comparison of two frameworks. *Educational Research Review, 2,* 114-129. doi:10.1016/j.edurev.2007.06.001

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

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

Berry, V., Sheehan, S., & Munro, S. (2017). "Exploring teachers' language assessment literacy: a social constructivist approach to understanding effective practices." In ALTE (Ed.), *Learning and Assessment: Making the Connections - Proceedings of the ALTE 6 Th International Conference. Association of Language Testers in Europe*. 3-5 May 2017, 201-207.

Boyle, A. (2005). Sophisticated Tasks I E-Assessment: What are they? And what are their benefits? In: *Proceedings of the 9th CAA Conference, Loughborough: Loughborough University*. Retrieved 3rd September, 2023 from https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/1979/1/BoyleA2.pdf

Brindley, G. (2001a). Language assessment and professional development. In C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara, &K. O'Loughlin (Eds.). *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126-136). Cambridge: Cambridge University Press.

Brown, J.D., & Bailey, K.M. (2008). Language Testing Course: What are they in 2007? *Language Testing, 25*(3), 349-383. doi: 10.1177/0265532208090157

Brown. JD, & Hudson, T. (1998). *The alternatives in language assessment*. University of Hawaii Working

Papers in ESL, 16(2), 79-103.

Bui, T.H. (2021). Teachers' language assessment literacy: a case study at a high school in Halong city, Vietnam. *Viettesol International Convention 2021 Rethinking English Language Education in The Covid Era*, p57-70.

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing, 25*(3), 327-347. doi: 10.1177/0265532208090156

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. doi: 10.1080/0969594x.2010.516643

Duong Thu Mai, Nguyen Thi Chi & Pham Thi Thu Ha (2017). The construction of a language assessment competence framework for pre-service English language teachers at ULIS-HNU: a focus on validity. *Foreign Studies, 1*(1), 74-88.

Fulcher, G. (2012). Assessment Literacy for the Language Classroom. *Language Assessment Quarterly, 9*(2), 113-132. doi:10.1080/15434303.2011.642041.

Giraldo, F. (2021). Language assessment literacy and teachers' professional development: A review of the literature. *Profile: Issues in Teachers' Professional Development, 23*(2), 265-279. https://doi.org/10.15446/profile.v23n2.90533

Griffin, P., & Nix, P. (1991). *Educational assessment and reporting: a new approach*. Sydney: Harcourt Brace Jovanovich.

Hirvela, A., & Pierson, H. (2000). *Portfolios: Vehicles for authentic self-assessment*. In G. Ekbatani & H. Pierson (Eds.). Learner-directed assessment in ESL (pp.105-126). Mahwah, NJ: Lawrence Erlbaum.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment course. *Language testing, 25*(3), 385-402. doi: 10.1177/0265532208090158.

Inbar-Lourie, O. (2013). Guest Editorial to the special issue on language assessment literacy. *Language Testing, 30*(3), 301-307. doi: 10.1177/0265532213480126.

Kiomrs, R., Abdolmehdi, R., & Naser, R. (2011). On the Interaction of Test Washback and Teacher Assessment Literacy: The Case of Iranian EFL Secondary School Teachers. *English Language Teaching, 4*(1). doi:10.5539/elt.v4n1p156.

Kouwenhoven, W. (2010). Competence-based curriculum development in higher education: some African experience. In M. Cantrell, M. Kool & W. Kouwenhoven (Eds.), *Access & Expansion: Challenges for Higher Education Improvement in Developing Countries* (pp.125-146). Amsterdam, The Netherlands: VU University Press.

Malone, M.E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language testing, 30*(3), 329-344. doi: 10.1177/0265532213480129.

McNamara, T. F. (1996). *Measuring second language performance*. Harlow, Essex, UK: Addison Wesley Longman Ltd.

McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation, 7*(8). Retrieved from: http://PAREonline.net/getvn.asp?v=7&n=8.

McNamara, T. (2003). Tearing us apart again. The paradigm war and the search for validity. *EUROSLA Yearbook, 3*, 229-238. doi: 10.1075/eurosla.3.13mcn

Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education, 120*, 285-296.

Mertler, C. A. (2003). *Pre-service versus in-service teachers' assessment literacy: does classroom experience make a difference?* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH (Oct. 15-18, 2003). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.556.4217&rep=rep1&type=pdf.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: does classroom experience make a difference? *American Secondary Education, 33*(1), 49-64.

Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory*. Paper presented at the annual meeting of the American Educational Research Association, April, Montreal, QC, Canada.

Mertler, C.A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools, 12*(2), 101-113. doi:10.1177/1365480209105575

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing, 30*(3), 381-402. doi: 10.1177/0265532213480337.

Rasch, G. (1960). *Probabilistic model for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press (original work published 1960).

Richards, J.C. (2013). Curriculum approaches in language teaching: forward, central backward design. *RELC Journal, 44*(1), 5-33. doi:10.1177/0033688212473293.

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. doi: 10.1177/0265532213480128

Siegel, M., & Wissehr, C. (2011). Preparing for the Plunge: Pre-service Teachers' Assessment Literacy. *Journal of Science Teacher Education, 22*(4), 371-391. doi:10.1007/s10972-011-9231-6.

Singer, M. (2006). A cognitive model for developing a competence-based curriculum in secondary education. In Al. Crisan (Ed.), *Current and Future Challenges*

in Curriculum Development: Policies, Practices and Networking for Change (pp. 121- 141). Bucureşti: Education 2000+ Publishers. Humanitas Educational, ISBN-(13) 978-973-689-104-5.

Sultana, N. Language assessment literacy: an uncharted area for the English language teachers in Bangladesh. *Lang Test Asia 9*, 1 (2019). https://doi.org/10.1186/s40468-019-0077-8

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*(3), 238-245.

Stiggins, R. J. (2002). Assessment Crisis: The Absence of Assessment for Learning. *Phi Delta Kappan, 83*(10), 758-765. Doi: 10.1177/003172170208301010.

Stiggins, R.J. (2004). New Assessment Beliefs for a New School Mission. *Phi Delta Kappan, 86*(1), 22-27. Doi: 10.1177/003172170408600106.

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics, 29*, 21-36. doi: 10.1017/s0267190509090035

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Lang Test. 30*, 403-412. doi: 10.1177/0265532213480338

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. doi 10.2307/20466661

Weigle, S. C. (2002). *Portfolio assessment*. Cambridge: Cambridge University Press.

Weng, F. & Shen, B. (2022). Language assessment literacy of teachers. *Frontiers in Psychology, 13*. Retrieved from https://doi.org/10.3389/fpsyg.2022.864582

Xu, Y., & Brown, G. T. L. (2017). University English teacher assessment literacy: a survey-test report from China. Pap. *LangTest. Assess, 6*, 133-158.

Yu, C. (2010). Business curriculum and assessment reform in Hong Kong schools: a critical review from a competence-based perspective. *Journal of Vocational Education & Training, 62*(1), 27-50. doi: 10.1080/13636821003605387.