

Classroom-based language assessment literacy and professional development need between novice and experienced EFL teachers

Ima Fitriyah^{1*}, Fitriatul Masitoh², Utami Widiati³

^{1,2,3} Graduate Program in English Language Education, Universitas Negeri Malang,
Malang 65145, East Java, Indonesia

^{1,2} Department of English Education, Faculty of Tarbiyah, Institut Agama Islam Negeri Kediri, Kediri 64127,
East Java, Indonesia

ABSTRACT

Classroom-based language assessment (CBLA) in the Indonesian context has become more significant since the shift of the assessment model from summative based assessment (Ujian Nasional or the final exam) to formative-based assessment (Asesmen Kompetensi Minimum (AKM)); therefore, teachers need to develop their classroom-based language assessment literacy (CBLAL). The model places a greater emphasis on students' learning outcomes in class rather than final test scores. This study aims to examine the CBLAL level of experienced and novice EFL teachers, their perceptions of CBLAL, and their needs for CBLAL training. Quantitative and qualitative data were used in conjunction with a case study research design. The CBLAL questionnaire, adopted from Lan and Fan's work (2019), collected quantitative data from 55 EFL teachers. Meanwhile, the qualitative data were collected from four of them (novice and experienced) in the interview sessions. The CBLAL levels of 55 teachers are between functional and procedural-conceptual literacy levels. Both experienced (ETs) and novice teachers (NTs) could grasp fundamental principles in language assessment and can use them in the classroom. Although the ETs have higher levels, there is no significant difference in CBLAL levels between ETs and NTs. The study also revealed that the teachers' understandings of technical skills and language pedagogy were among the highest compared to their knowledge of theories and principles on language assessments. However, they indicated that they still require professional development (PD) in CBLAL, despite their functional and procedural-conceptual literacy. The current study has pedagogical implications for both ETs and NTs. They should actively participate in various professional development activities, focusing on classroom evaluation.

Keywords: Classroom-based language assessment; literacy level; professional development

First Received:

29 Juni 2021

Revised:

21 April 2022

Accepted:

11 May 2022

Final Proof Received:

19 May 2022

Published:

31 May 2022

How to cite (in APA style):

Fitriyah, I., Masitoh, F., & Widiati, U. (2022). Classroom-based language assessment literacy and professional development need between novice and experienced EFL teachers. *Indonesian Journal of Applied Linguistics*, 12(1), 124-134.
<https://doi.org/10.17509/ijal.v12i1.46539>

INTRODUCTION

Assessment, as an inherent element of teaching, can be viewed as a vital part of the teaching and learning process since the information gathered through assessments is crucial for many teachings and learning decisions. Teachers spend 30% to 50% of their time assessing their students (Vogt et al., 2020). Similarly, stakeholders frequently utilize assessment data to make decisions on a variety of educational policies. As a result, careful planning, administration, scoring,

and interpretation of assessment results are required. Teachers must be well-versed in assessment concepts and procedures (Giraldo, 2018; Nurdiana, 2021). Ultimately, it is critical that teachers have sufficient literacy of language assessment.

Stiggins (1991) has written a seminal article on assessment literacy and was followed by the emergence of the importance of language assessment literacy (henceforth LAL) in the early 2000s by Brindley (2001). Assessment-literate teachers do not

* Corresponding Author

Email: imafitria@iainkediri.ac.id

lack the tools to be critical in using assessment data (Stiggins, 1991). Inbar-Lourie (2008) argues that when someone is literate in language assessment, they can ask and answer important questions regarding the assessment's goal, the tool's suitability, testing circumstances, and the use of the assessment's result. There has been large body of knowledge concerning LAL studies which are centred on some aspects i.e. contribution of teachers' LAL to students' performance and achievement (Elshawa et al., 2016; Vogt et al., 2020) and teachers' reflection of their assessment literacy which has interrelationships with various components (Ashraf & Zolfaghari, 2018; Coombe et al., 2020; Fitriyah & Jannah, 2021). LAL also brings students into better learning and promotes motivation (Alderson et al., 2017; Fulcher, 2021; Gan et al., 2019).

The LAL level of a language instructor can be used to assess how effectively an EFL teacher accurately assesses students' skills. Trisanti (2019) discovered that the higher teacher's LAL level, the greater his capacity to create inquiries. She found that the level of teacher's knowledge in making questions in her study was moderate. However, this finding revealed a dearth of assessment knowledge among language teachers and test creators in terms of how to create reliable and valid tests based on language assessment theories. In addition, Yan and Fan (2021) found that language testing researchers and graduate students had a better understanding of assessment ideas than EFL instructors. Nevertheless, EFL teachers were able to relate such concepts to assessment practices. Moreover, Luthfiyah et al. (2020) examined the EFL secondary teachers' LAL, and the finding denoted that the teachers were in moderate level of LAL.

Concomitantly, LAL must be implemented correctly in the classroom (Fulcher, 2012; Lan & Fan, 2019; Noori et al., 2017). Knowing how literate teachers on LAL also indicates how high the need for teachers professional training on this topic (Anam & Putri, 2021; Latif, 2021; Ogan-Bekiroglu & Suzuk, 2014; Prasetyo, 2018; Widiastuti et al., 2020; Yan & Fan, 2021). Widiastuti et al. (2020) revealed teachers with high professional development (PD) participation levels have stronger formative assessment beliefs than those with lower PD involvement. As a result, with the focus on classroom-based language assessment literacy (henceforth CBLAL), this study was more detailed in assessing LAL in the context of classroom practice.

In consideration to the ongoing debate about the effectiveness of the prevailing national examination at that time, the minister of education and culture of Indonesia abolished the exam (*Permendikbud* No. 22, 2020) starting in 2021 and started to put emphasis on students' learning outcomes as a benchmark for graduating middle school students. This new phenomenon becomes a challenge for teachers at that level to recalibrate their ability to assess students.

Moreover, coupled with the existence of a new assessment model, there is also a new model in Indonesian curriculum. Therefore, it is necessary to assess teachers' levels of CBLAL and their needs to master the expected competency development.

As noted previously, the CBLAL extensive studies have thus far centred on the issues among teachers in general. However, little attention is given to the CBLAL study, which examines the difference between EFL experienced teachers (ETs) and novice teachers (NTs) in carrying out the CBLAL. For both experienced and novice teachers, mapping CBLAL is essential. The capacity to appraise students between ETs and NTs should not be significantly different. Interestingly, Lan and Fan (2019) have developed a CBLAL inventory. The CBLAL primarily has four dimensions: (1) technical skill, (2) scoring and decision making, (3) language pedagogy, and (4) principles in language assessment. Lan and Fan (2019) then delve further into the dimensions used in CBLAL. They define technical skill in CBLAL as skill of language assessment design, developing a language test, and using any alternative assessment. Instructional skills, scoring, making decision based on the scores are examples of scoring and decision making. Language pedagogy includes the ability to assess linguistics skills and language components appropriately. Finally, the CBLAL principles include awareness of the validity and reliability of language assessment.

Lan and Fan (2019) suggest that further research is needed to undertake similar research strategies to provide additional valid evidence for the questionnaire. Their study used an internet survey to distribute the questionnaire, and thus restricting the amount of interview data that could be collected. They believed that further explanatory study may be beneficial in gaining a deeper understanding of the different CBLAL's nature and purpose between ETs and NTs. As most of the previous studies investigated pre-service EFL teachers' LAL (Anam & Putri, 2021; Prasetyo, 2018; Viengsang, 2016), knowing the different CBLAL levels between NTs and the ETs could provide more information about the need to have professional development (PD) in this regard. To bridge that gap, we investigated the CBLAL level of two stakeholder groups; ETs and NTs, and their needs of a PD based on the level of CBLAL. This study employed three research questions:

1. What is the CBLAL level of experienced and novice EFL teachers?
2. In what ways do experienced and novice EFL teachers perceive themselves in CBLAL dimensions?
3. To what extent do both experienced and novice EFL teachers perceive a need for assessment training in CBLAL?

Classroom-Based Language Assessment

Assessment is defined as a continuous process of monitoring and tracking learners' development that includes obtaining, analysing, recording, and applying data regarding students' performance on educational activities (Black & Wiliam, 1998). In a larger sense, assessment refers to a variety of ways for gathering evidence about a learner's knowledge, aptitude, understanding, attitude and motivation, and it may be done through formal and informal instruments, such as a portfolio, self-evaluation, and so on (Dysthe et al., 2007). Assessment is intertwined with the teaching and learning process and is a part of both instructors' and students' daily lives in the classroom. Classroom-based assessment, according to Fulcher (2012), focuses on needs, strengths and weaknesses, which is the process of collecting, synthesizing, and analysing data in order to enhance teachers' decision-making with the data acquired in connection to classroom setting. Teachers utilize the data to monitor, measure, and record students' progress in the classroom, as well as to adapt their own teaching methods depending on their own skills, knowledge, and perspectives. Because of this, classroom-based assessment is an ongoing and integrated process that is crucial for gathering information, modifying instruction, and awarding grades to students in the classroom.

One aspect of the interface between assessment methods and second language acquisition is CBLA (Green, 2018), in which instructors are responsible for promoting learning and gathering information about students' progress and accomplishment (Alderson et al., 2017). They are integrated into the classroom's teaching and learning cycle, as opposed to large-scale language exams that attempt to assess general competency and are frequently conducted in highly controlled test environments (Qian & Cumming, 2017). Teachers utilize CBLA not only to get information about their students' language attainment and development, but also to help them learning the language. It is critical for teachers to acquire CBLA skills and methods in order to provide positive effect in classrooms and accomplish the curriculum's ultimate goals (Rea-Dickins, 2004).

Classroom-Based Language Assessment Literacy

In 1991, Stiggins was the first to describe assessment literacy as the capacity to use one's understanding of educational assessment to assess a variety of student performance indicators (Stiggins, 1991; Xu & Brown, 2016). His notion of assessment literacy has been generally recognized. The word is increasingly used to describe the variety of skills and knowledge that many stakeholders need to deal with the new problems in the testing and assessment duties that teachers are now expected to handle (Fulcher, 2012). Language assessment literacy is a combination of assessment literacy abilities and language-specific competences that constitute a unique entity (Inbar-Lourie, 2008). According to Giraldo, (2018), language assessment literacy is a competency that involves a variety of

stakeholders and consists of knowledge, skills, and principles related to assessing language ability. He has offered this framework for teachers and even for future research to measure teachers' CBLAL.

In order to enhance students' language learning outcomes, language teachers, according to Rea-Dickins (2004), are "agents" of classroom assessment by continuously monitoring, assessing, and interpreting learners' performance. Therefore, teachers must undertake the duties of both instructor and assessor as a result of the CBLA (Poehner, 2009). Taylor (2013) defined CBLAL as the knowledge and skills required by language teachers to conduct classroom-based language assessment activities. CBLAL primarily includes the abilities to design, develop, and critically evaluate tests and other assessment procedures, as well as the abilities to monitor, evaluate, grade, and score assessments based on theoretical knowledge, and the abilities to interpret and communicate assessment results (Lan & Fan, 2019).

The combination of technical skills literacy and theoretical knowledge constitutes CBLAL's primary focus, which includes the abilities to design, implement, and critically analyse tests and other assessment methods as well as the abilities to monitor, evaluate, grade, and score assessments based on theoretical knowledge (scoring knowledge), the abilities to assess the language skill and linguistics component (language pedagogy), and the abilities to interpret and communicate assessment results (theories and principles) (Lan & Fan, 2019). This group of scholars (Giraldo, 2018; Lan & Fan, 2019; Pill & Harding, 2013; Taylor, 2013; Vogt & Tsagari, 2014) has provided various distinct aspects of the CBLAL, but Phil and Harding's (2013) classification of literacy levels is the most thorough.

It is the contention of Pill and Harding (2013), that a five-level continuum of literacy may be productively used to language assessment literacy. Literacy is divided into five categories: illiteracy, nominal literacy, functional literacy, procedural-conceptual literacy, and multidimensional literacy. Illiteracy is defined as the inability to understand or apply language assessment concepts and methods. Nominal literacy, on the other hand, is defined as knowing that a specific term has some connection to assessment, but it may indicate a misinterpretation. To be functionally literate, one must have a strong grasp of the most basic vocabulary and concepts. It is a combination of procedural and conceptual literacy that enables one to grasp and apply the fundamental principles in the area. Lastly, multidimensional literacy, encompasses philosophical, historical, and social evaluations that go beyond the conventional ideas. Assessment is conducted and facilitated by teachers to provide numerous samples of learner performance, to offer learners with rapid and constructive feedback, and to analyse and adapt the

assessment methods to maximize teaching/learning in the CBLA environment (Stephen, 2012).

METHOD

This study employed a case study design to examine teachers' CBLAL in an Indonesian middle education setting. According to Creswell (2012), an investigation can gather a complete understanding of a case by comparing different perspectives drawn from quantitative and qualitative data. The goal of using both quantitative and qualitative data was to prove the study's trustworthiness by gathering data from different sources, allowing for in-depth viewpoints on CBLAL.

Participants

English teachers participating in this study came from middle schools currently residing in one district in East Java Province, Indonesia. The district was selected as it is famous as an education city. Fifty-five EFL teachers voluntarily did and returned the CBLAL questionnaire and test. Meanwhile, in the interview stage, two female ETs (ET1 and ET3) and two female NTs (NT2 and NT4) of the participants willingly took part. The participants' perception of CBLAL is therefore adequately represented in this number. The categorization of teachers is based on their teaching experience. For this research, NTs are teachers who have been teaching for five years or less (Ingersoll & Smith, 2003), whereas the ETs are those who have more than five years of teaching experience.

Instruments

The instruments consisted of a questionnaire, test, and interview guide. The present study partly replicated the questionnaire used by Lan and Fan's (2019) study in CBLAL. Concerning the focus of the present study and to meet the teaching and assessment context in Indonesia, we adapted the questionnaire of Lan and Fan (2019). First, in part one, we added the information of the teachers' experience in teaching English. Secondly, for the level of CBLAL and PD, we omitted two items about placing students into programs and giving a certificate. Both practices do not exist in the Indonesian language classroom context. Third, we provided an idea for assessing students' attitudes. This part is essential since the Indonesian curriculum emphasizes students' attitudes integrated into the subject of study. The questionnaire of the current study consisted of two main parts. Part one consisted of sixteen questions asking about CBLAL level (QA), and part two was used to identify CBLAL training (QB) (see Appendix 1). Both parts were based on the dimensions of CBLAL; (1) technical skills, (2) scores and decision making, (3) language pedagogy, and (4) theories and principles in CBLA.

Additionally, Latif (2021) asserted that the level of literacy could also be seen from the ability to understand the knowledge of CBLA. Thus, in addition

to the questionnaires, the teacher's knowledge was also measured using the CBLAL test, which consisted of fifteen items. The test was adjusted to the dimension of CBLAL. Moreover, to measure the reliability of the questionnaires and tests, the Cronbach's alpha was computed. With a score of 0.947, the questionnaire and the test had a high level of internal consistency. Finally, an interview guide based on the CBLAL dimension was employed to answer the second research problem. The tool was utilized to unearth qualitative data and get the teachers' perspectives of CBLAL through their narrative.

Data collection

In collecting the data, quantitative and qualitative data were collected in two periods. The first part consisted of gathering quantitative data via a questionnaire. It was a self-reported questionnaire of CBLAL level and PD. In addition, the CBLAL test was utilized to assess teachers' CBLAL levels. In addition, the qualitative data were gathered through interviews to explain the questionnaire results in greater depth. From 55 teachers who returned the questionnaire and test, the data revealed that 56.4 % or 32 teachers were experienced, and 43.6% or 23 teachers had less than five years of teaching. Most of them (83.6%/46 teachers) graduated with a Bachelor (S1) degree, and the rest were Master in ELT.

The CBLAL level is based on a scale that includes (1) illiteracy, (2) nominal literacy, (3) functional literacy, (4) procedural-conceptual literacy, and (5) multidimensional literacy and the result of the CBLAL test. The questionnaire scale was adjusted to make it easier for the participants to grasp. The level of multidimensional literacy was represented by strongly agree (five), and illiteracy became strongly disagree (one). The PD needs were measured in a similar manner. The need starts from 'no need any training' (one) to 'extremely need training' (five). Because the elements in the questionnaire were in the form of statements saying that the participants have literacy in the topics indicated, the usage of the scale is thought to be more acceptable. Following the quantitative findings of the questionnaire, qualitative data was gathered through interviews. The interview was performed in various ways, depending on the availability of the participants. WhatsApp's voice notes function was utilized for NT2 and NT4, a face-to-face interview was done with ET1 and NT4, and a Zoom meeting was done with ET3.

Data analysis

A descriptive statistic was used to analyse the quantitative data. Each descriptor was calculated using means and standard deviations. The level of teachers' CBLAL and PD was determined by the quantitative data obtained from the 32 items questionnaire and 15 items of the CBLAL test. For the categorization, the participants were grouped into five levels of CBLAL

using equal cut-off points on the scale: illiteracy (0.0-1.0), nominal literacy (1.1-2.0), functional literacy (2.1-3.0), procedural-conceptual literacy (3.1-4.0), and multidimensional literacy (4.1-5.0). It is also applied for the categorization of PD in CBLAL.

The description for each level of CBLAL used in this investigation are: (1) illiteracy: teachers lack familiarity with language assessment concepts and methodologies; (2) nominal literacy: teachers may have misconceptions regarding terminology and concepts related to language evaluation; (3) functional literacy: teachers are well-versed in a wide range of basic language assessment concepts and vocabulary; (4) procedural-conceptual literacy: teachers understand basic language assessment ideas and use what they have learned in the classroom; (5) multidimensional literacy: instructors understand philosophical, historical, and social components of assessment knowledge that goes beyond common conceptions.

In the second phase, various measures were taken to analyse the interview data. The data were subjected to deductive thematic analysis, following the transcription of the interviews. The codes and themes were examined to see which best characterized and addressed the CBLAL. Finally, after extensively analysing the transcript, the final pieces linked to the study topic were picked and presented as the study findings.

FINDINGS

The findings of this study are in the form of mean of EFL teachers' CBLAL level between ETs and NTs, their voices of each CBLAL aspect and their needs of CBLAL training. The qualitative finding was used to address the outcome from the questionnaire, such as how instructors perceived their knowledge, technical skills, and principles on language evaluations.

Teachers' Perceived Level of CBLAL

The questionnaire on the CBLAL revealed that the teachers' CBLAL level was categorized between functional literacy and procedural-conceptual literacy level as shown in Table 1, standing at 3.737. It means that EFL teachers grasped fundamental principles in language assessment and could use what they have learned in the classroom. The items asked teachers' knowledge of whether they understand the components in the CBLA in ELT. The findings revealed that the difference between the self-perceived level of CBLAL by ETs and NTs was not significant. Table 1 shows that although both groups were relatively at the same level, the mean showed that ETs had better CBLAL levels (3.683) with training needs at 3.956 than NTs (3.650) whose training need stands at 4.093. As an alternative way to investigate EFL teachers' CBLAL, teachers' knowledge was measured using the objective CBLAL test. The test result indicated that all of the teachers were at functional

level of CBLAL (Table 1). The results showed that their overall mean score is 46.5. However, ETs had a higher level of CBLA knowledge at 49.9 than NTs at 39.1 score means. The result of CBLAL test was utilized to support the conclusions of the perceived CBLAL level of the teachers. The CBLAL level was also further described by the result of the interviews based on the dimension of CBLAL.

Table 1
Summary of CBLAL level and PD need

CBLAL	All participants		ETs		NTs	
	Mean	SD	Mean	SD	Mean	SD
Questionnaire	3.737	.826	3.693	1.033	3.650	.875
Test	46.5	16.69	49.9	14.80	39.1	17.67
PD	4.056	.813	3.956	.827	4.093	.928

Teachers' Perceptions of CBLAL Dimensions
Teachers' Literacy on Technical skill of CBLA

In line with the quantitative data, all respondents confirmed that they were well versed in preparing CBLA. Teachers believed and experienced that the first thing to be done before constructing the assessment were deciding the objectives of the assessment. A respondent mentioned that she literary referred to lesson's goals to make a test. In addition, the teachers viewed the use of a table of specifications (ToS) as helpful in CBLA. A respondent stated, *'I controlled my test based on the blueprint I already made before.'* ET1 noted that the advantages of the ToS were to measure how the test appropriately assesses students' language achievement based on the lesson.

Table 2
Literacy on Technical Skill

Items	Experienced				Novice			
	CBLAL		PD		CBLAL		PD	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
QA1	3.906	.892	4.250	.803	3.434	.843	4.304	.634
QA2	3.906	.817	4.031	.782	3.739	.915	4.173	.716
QA3	4.000	.915	4.062	.715	4.130	.868	4.391	.583
QA4	3.968	.932	3.906	.817	3.782	.735	4.043	.824
QA5	3.687	.859	3.593	.945	3.608	1.033	3.956	1.021
QA6	3.562	.840	3.906	.817	3.304	.822	3.869	.757

Furthermore, both ETs and NTs asserted that they were aware of how to construct tests by adopting or adapting from the existing test, as shown in Table 2, showing the mean of 3.906. They argued that the use of existing tests should be carefully applied. Adjusting the test allows them to be critical of the sources of the test. They did not want their assessment to be inappropriate for their students' level. As pointed out by ET3: *"I construct my own test. The item and the questions were my hand-made adjusted to students' level. However, the texts or dialogue were taken from trusted sources on the internet"*.

The teachers thought that students were more motivated when they could do the test which is appropriate for their level. ET1 mentioned how she

was regretful to experiment by giving her students a higher-level test: “if they are still at a low level, giving them many HOTS questions will be wasteful.”

According to ET1, students were also interested in getting feedback after the test was held. It was easier for them to learn from their mistakes and thereby foster successful English learning. When it comes to descriptor no. QA3 (Appendix 1), the participants knew how to give feedback on student assessments (4.072; Appendix 2). However, ET1 and ET3 reported that they had difficulty providing timely feedback to students due to teachers' workload. Based on the interview data, the instructors were seen to be competent at assessing students' language growth, as two participants stated that they utilized both tests and other alternative assessments such as observation and interview to see the students' language progress.

Teachers' literacy on scoring and decision making of CBLA

Different from ET1, NT2 was unsure whether or not she gave an appropriate score. ET1 stated that the objective test must have a score of 1, but for subjective tests and essay tests, she would consider the level of difficulty of the questions. Thus, she did not necessarily give a score of 1 for questions that were considered cognitively demanding. In her own words: “I had no problem in scoring. I knew what to do with objective tests, subjective tests and informal tests. The scoring should follow to the way the assessment is delivered. Strict scoring, observation, and scoring rubrics are always helpful for any situation” (ET1).

Table 3
Literacy on Scoring and Decision Making

Items	Experienced				Novice			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
QA7	3.937	.840	4.000	.718	3.869	.814	4.217	.735
QA8	3.875	.751	4.031	.782	3.826	.777	4.304	.702

ET1 also confirmed the result of QA8 on how she knew what to do after noticing students' scores. On the other hand, NT2, several times, found different ways to give scores from one teacher to another. Therefore, she was still adapting herself to the existing teaching and assessment model in her institution. After some time, she had gotten to know the character of her students and tried to give an appropriate assessment according to their level. One thing that created a dilemma for her is that the EFL ability of NT2's students was still low, but her seniors gave high scores on questions that were considered not cognitively demanding. Such action was taken so that the grades do not look much different from the standard grades set by the school. As a new teacher, NT2 highly valued her idealism. In another story, NT4 understood well how to score (QA7). She said: “I do not have difficulties in scoring, however when students' scores were under the standard (KKM),

teachers needed to work harder for the improvement, no matter the ways to do so.” However, both the NTs have the same dilemma. They were in the phase of adjusting themselves into the rules of the institution without losing their ideals.

Teachers' Literacy on Language pedagogy of CBLA

Table 4 shows the level of CBLAL in term of assessing language skills. Teachers understood the advantages of integrative assessment.

Table 4
Language Pedagogy

Items	Experienced				Novice			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
QA9	3.875	.751	4.093	.734	3.521	.897	4.260	.864
QA10	3.781	.750	4.125	.832	3.391	.940	4.347	.831
QA11	3.593	.910	3.812	.780	3.695	.875	4.217	.795
QA12	3.875	.832	4.093	.777	3.478	.845	4.130	.814
QA13	3.625	.832	3.937	.840	3.913	.792	4.304	.822

They coordinated the language skills test to encompass the nature of language in evaluation. On the other hand, ET3 emphasized that integration cannot involve too many skills. The limitation would help her in giving a valid score. She should also have a clear scoring rubric to separate the skill. However, discrete point test was still used in some of the teachers' institutions. The compelling part is that most teachers did not know the term discrete-point (as seen from test result; this question only got three correct answers), but they knew how to test language components separately. As ET1 revealed: “I assess their language skills separately (QA12) by using performance-based and portfolio-based. However, objective tests for summative tests integrate many components, structures, vocabularies, reading, etc.”

ET1 was able to empower students to study collaboratively by giving them project-based assessments. Concomitantly, all of the interviewed teachers agreed that cultural aspect, students' attitudes, and integration were essential components to be assessed as seen in QA13. The result was high for both NTs and ETs (3.625 and 3.913; Table 4). They believe that students' attitude was more important than their knowledge. While evaluating students' cognitive aptitude was necessary, evaluating affective factors was crucial. NT2 said: “I would prefer to have a good attitude and high morality students to students who are cognitively sound or have good English skills but do not have good manners.” Finally, cultural integration was in two forms, observed during students' performance and integrated into the test, i.e., reading, vocabularies, structures, and writing test.

Teachers' Literacy on Theories and Principles of CBLA

Table 5 reveals that teachers understand how to make their tests valid and make them as accurate and reliable as possible (QA15). However, due to ETs

workload, they rarely measure their test statistically. Similar to ETs, NTs understood how to make their test acceptable.

Table 5
Literacy on Theories and Principles

	Experienced				Novice			
	CBLAL		PD		CBLAL		PD	
Items	Mean	SD	Mean	SD	Mean	SD	Mean	SD
QA14	3.718	.840	4.093	.777	3.782	.795	4.130	.757
QA15	3.750	.751	4.125	.793	3.869	.757	4.260	.619
QA16	3.371	.760	4.062	.759	3.652	.884	4.087	.733

In regard to this, NT4 pointed out: “My test was valid since it was based on the learned lesson and content. It was also reliable because I can use it more than once.”

For item QA16, (on the use statistics to study the quality of tests) 55 participants were in functional literacy level in the skill of using statistical analysis of the CBLAL (3.371), as seen from Table 5. Their answers included psychometric analysis, item problems, item discrimination, and statistically calculating test validity and reliability. Although the participants stated that they were concerned about assessment validity and reliability, they claimed they had never statistically calculated them. However, they knew and reflected if their students make a lot of mistakes. NT1 confessed: “I got a valuable lesson after the assessment; my reflection was; was it my mistakes? were they too lazy? And was the test too hard for them?”

The findings on how teachers reflected their knowledge and ability in using CBLA allow future researchers to add more alternatives regarding teachers' CBLAL.

Teachers’ Perceived Needs of CBLAL Training

The teachers believe that their CBLAL is between functional literacy and procedural-conceptual literacy. However, they rated themselves at 4.056 as shown in Table 1 for CBLA training. This finding indicates that they strongly felt that they needed language assessment training; This is a reasonable perception. Even though most of them rated themselves to be at procedural-conceptual literacy level, the results of the CBLA knowledge test show that they were at functional literacy. They were aware of their inadequacies as well as the necessity for PD. NTs perceived higher training needs; at 4.093, than ETs (3.956) as seen in Table 1. This score is not unexpected since the NTs had a lower CBLAL level than ETs. Appendix 2 illustrates a high demand for training on creating tests and evaluations that incorporate all skills, such as QB12 mean score of 4.109.

Interestingly, most participants perceived that they required training to increase their capacity to design questions which are more legitimate, trustworthy, and statistically know the level of difficulty of the questions. The result of qualitative

data revealed that the interviewed participants need training in constructing test items (QB1 and QB2), giving score (QB7), measuring the validity and reliability of test items (QB14, QB15, and QB16; Appendix 1). ET3 stated: “The assessment methodology is affected by changes in the curriculum, thus training on how to develop questions in line with the requirements of the curriculum must be updated.” Furthermore, NT4 asserted her expectation to have training in giving appropriate score. Related to training in measuring reliability and validity of a test, ET1 said: “I want to know how to measure the validity and reliability of the test; I have never gotten this kind of skill previously.” As a result, even if they had functional and procedural-conceptual literacy level, they were conscious of the need to continue enhancing their skills in CBLA.

DISCUSSION

The purposes of this research were to find out how much CBLAL level teachers had and how they perceive the need of PD about it. This study evinces that EFL teachers considered their CBLAL was somewhere between functional literacy and procedural-conceptual literacy, implying that they had the necessary knowledge, abilities, and ideas to execute CBLA. However, the test results suggested that the participants required professional development in terms of CBLAL, since they only had a functional degree of CBLAL. This level is insufficient because their assessment duties are crucial in determining students' learning results. The PD needs questionnaire reflects how they need to learn more to provide better assessment practices. The findings of this study are comparable to the study conducted by Xu and Brown (2016). They argued that a fair level of assessment literacy for higher education teachers was insufficient because their assessment responsibilities were immense. Lam (2020) found the same thing, with participants reporting a moderate LAL level in his research. In addition, Puspawati's (2019) study indicated that the LAL of her subjects was fairly low. Even though the current study's findings cannot be generalized due to the limited sample size, the study adds to evidence that teachers' low CBLAL is still a global concern (Fulcher, 2012; Gan et al., 2019; Lan & Fan, 2019).

Furthermore, the interview revealed that they are knowledgeable for classroom assessment, yet they still need the training to enhance their CBLAL. Their need for further CBLAL instruction has been expressed in various stakeholders (Anam & Putri, 2021; Yan & Fan, 2021) and educational settings (Coombe et al., 2020; Vogt et al., 2020). Regarding different CBLAL levels between NTs and ETs, the current study differs from Yan and Fan's (2021) findings. They found that language testing researchers and graduate students demonstrated greater familiarity with assessment concepts than did EFL teachers. However, EFL

teachers could understand and link those concepts with assessment practices. The current study reveals different findings. Experienced teachers appeared to be more knowledgeable than the novice teachers.

In terms of teachers' perceptions of their CBLAL dimensions, the study discovered that both groups reported having a conceptual and procedural literacy of all four dimensions including technical skills, scoring and decision making, language pedagogy, and principles of assessment. The quantitative and qualitative evidence both point to the same conclusion. Teachers' knowledge of scoring and decision making and language pedagogy was scored best among the four categories, while knowledge of theories and principles of language assessment was rated lowest. This result might imply that the teachers were more confident in their understanding of broadly applied linguistics, such as language assessment methods, and foreign language acquisition patterns, than in their knowledge of CBLAL principles. This conclusion is consistent with the findings of Yan et al.'s (2018) study, which found that participants were uninterested in evaluation theories and principles.

The findings for the technical skills dimension were similar to the scoring and decision-making in that teachers reported having conceptual and procedural literacy level such as developing or using a ready-made test, multiple methods in language assessment, and assessment results to improve teaching. However, there is a discrepancy between the quantitative and qualitative data in the description regarding giving students feedback. The score in this dimension was high, according to the poll, but the interviews revealed that giving feedback was one of the obstacles they faced. ET 2 said "I understand that students need feedback from assessments for their learning, and I know how to give feedback," one participant said. "However, I frequently miss the deadline because it takes me forever to read the students' works." This discrepancy might imply a disconnection between teachers' beliefs and their assessing approach. This result is closely related to Cheng and Wang's (2007) study, where the results demonstrate how teachers make day-to-day decisions in the assessment contexts.

Furthermore, the dimension of theories and principles in CBLAL had the lowest mean score for the component of CBLAL. When questioned why the score was low, the majority of them stated that their previous training did not include statistics and that their present assessment context did not need it. The lack of competence in statistics abilities is comparable to the findings of Anam and Putri (2021) and Ogan-Bekiroglu and Suzuk (2014). They found that participants had trouble with educational measurement tasks such as item generation and statistical calculations on validity and reliability. It is also in line with Coombe et al. (2020), who found that teachers lacked knowledge in quality test development and valid assessment procedures. However, having this

literacy is important because the basic criterion of language evaluation is validity (Brown, 2010). The lack of a criteria for validity will have an impact on the quality of the instrument used. As a result, increasing teachers' understanding of validity evaluation is critical for assuring the correctness of assessments and inferences drawn by instructors.

According to the findings, ETs have a better understanding of assessment theory and how to combine it with language pedagogy so that they can administer and score tests as well as use various alternative assessments, such as portfolios, learning logs, peer assessment, or self-assessment, that are appropriate for their students. This conclusion corroborates Plake et al.'s (1993) finding that preservice instructors had somewhat worse classroom assessment abilities than in-service teachers. Higher literacy among ETs may be due to experiential variables such as designing and implementing assessment in their classrooms, teaching experience, teacher training, and learning from peers in the English teacher organization (Vogt & Tsagari, 2014; Widiastuti et al., 2020; Yan et al., 2018). Although such factors are not available to new teachers, such experience factors may enhance their awareness of the importance of increasing their assessment literacy to create and implement appropriate classroom assessment techniques.

CONCLUSION

This study has shown that EFL middle school teachers have functional and procedural-conceptual CBLAL levels. Qualitative data demonstrates that their perception of CBLAL is congruent with the self-rating results of the questionnaire. The dimensions of theories and principles of CBLAL are the lowest literacy perceived and require the most training. However, there is a slightly different level between ETs and NTs. ETs have better knowledge and practice than NTs. Even though NTs still have a relatively recent understanding of CBLAL, sufficient teaching and assessing experience will form broader assessment literacy and expertise. Both groups also admit that they need PD after recognizing their limited assessment abilities. However, generalizing the results of this study should be done with caution, considering the limited sample. In addition, further research might use a more comprehensive test to determine the level of teachers' CBLAL by involving more participants. Regardless of the limitations of this study, the results still provide an overview of the literacy level of teachers and enriches the knowledge on classroom language assessment. The current research findings have pedagogical consequences for both secondary school ETs and NTs. Teachers should participate actively in various professional development activities, focusing on classroom evaluation. Such evaluation would assist students in attaining their learning goals and enhance their language abilities.

ACKNOWLEDGMENTS

The first and second author of this study would like to thank LPDP (Indonesia Endowment Fund for Education), Ministry of Finance, Republic Indonesia for Doctoral scholarship 2021-2023. The authors are also grateful to the Chief of English Teachers Association in East Java for his assistance during the data collection.

REFERENCES

- Alderson, J. C., Brunfaut, T., & Harding, L. (2017). Bridging assessment and learning: A view from second and foreign language assessment. *Assessment in Education: Principles, Policy & Practice, 24*(3), 379–387. <https://doi.org/10.1080/0969594x.2017.1331201>
- Anam, S., & Putri, N. V. W. (2021). How literate am I about assessment: Evidence from Indonesian EFL pre-service and in-service teachers. *English Review: Journal of English Education, 9*(2), 377–388.
- Ashraf, H., & Zolfaghari, S. (2018). EFL teachers' assessment literacy and their reflective teaching. *International Journal of Instruction, 11*(1), 425–436.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Brindley, G. (2001). Assessment. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (1st ed., pp. 137–143). Cambridge University Press. <https://doi.org/10.1017/CBO9780511667206.021>
- Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly, 4*(1), 85–107. <https://doi.org/10.1080/15434300701348409>
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia, 10*(1), 3. <https://doi.org/10.1186/s40468-020-00101-6>
- Creswell, J. W. (2012). Collecting qualitative data. *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.
- Dysthe, O., Engelsen, K. S., & Lima, I. A. Å. (2007). Variations in portfolio assessment in higher education: Discussion of quality issues based on a Norwegian survey across institutions and disciplines. *Assessing Writing, 12*(2), 129–148. <https://doi.org/10.1016/j.asw.2007.10.002>
- Elshawa, N. R. M., Heng, C. S., & Abdullah, A. N. (2016). Teachers' assessment literacy and washback effect of assessment. *International Journal of Applied Linguistics and English Literature, 5*(4), 135–141. <https://doi.org/10.7575/aiac.ijalel.v.5n.4p.135>
- Fitriyah, I., & Jannah, M. (2021). Online assessment effect in EFL classroom: An investigation on students and teachers' perceptions. *Indonesian Journal of English Language Teaching and Applied Linguistics (IJELTAL), 5*(2), 265–284.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Fulcher, G. (2021). 3 language assessment literacy in a learning-oriented assessment framework. In *Learning-oriented language assessment: putting theory into practice* (pp. 34–48). Routledge. <https://doi.org/10.4324/9781003014102-4>
- Gan, Z., Leung, C., He, J., & Nang, H. (2019). Classroom assessment practices and learning motivation: A case study of Chinese EFL students. *TESOL Quarterly, 53*(2), 514–529. <https://doi.org/10.1002/tesq.476>
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development, 20*(1), 179–195. <https://doi.org/10.15446/profile.v20n1.62089>
- Green, A. (2018). Assessment for learning in language education. *Iranian Journal of Language Teaching Research, 6*(3), 9–18.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385–402. <https://doi.org/10.1177/0265532208090158>
- Ingersoll, R. M., & Smith, T. M. (2003). The wrong solution to the teacher shortage. *Educational Leadership, 60*, 30–35.
- Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation, 61*, 112–122. <https://doi.org/10.1016/j.stueduc.2019.03.003>
- Latif, M. W. (2021). Exploring tertiary EFL practitioners' knowledge base component of assessment literacy: Implications for teacher professional development. *Language Testing in Asia, 11*(1), 1–22. <https://doi.org/10.1186/s40468-021-00130-9>
- Luthfiyyah, R., Basyari, I. W., & Dwiniasih, D. (2020). EFL secondary teachers' assessment literacy: Assessment conceptions and practices. *Journal on English as a Foreign Language, 10*(2), 402–421. <https://doi.org/10.23971/jefl.v10i2.2101>
- Noori, A., Shafie, N. H., Mashwani, H. U., & Tareen, H. (2017). Afghan EFL lecturers' assessment practices in the classroom. *Imperial Journal of Interdisciplinary Research, 3*(10), 130–143.
- Nurdiana, N. N. (2021). Language teacher assessment literacy: A current review. *Journal of English Language and Culture, 11*(1). <https://doi.org/10.30813/jelc.v11i1.2291>

- Ogan-Bekiroglu, F., & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice. *The Curriculum Journal, 25*(3), 344–371.
<https://doi.org/10.1080/09585176.2014.899916>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing, 30*(3), 381–402.
<https://doi.org/10.1177/0265532213480337>
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12*(4), 10–12.
<https://doi.org/doi:10.1111/j.1745-3992.1993.tb00548.x>
- Poehner, M. E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly, 43*(3), 471–491.
<https://doi.org/10.1002/j.1545-7249.2009.tb00245.x>
- Prasetyo, A. H. (2018). Language assessment literacy development: A student-teacher's experiences in teaching practice program. *LLT Journal: A Journal on Language and Language Teaching, 21*(2), 199–206.
- Puspawati, I. (2019). Understanding teachers' knowledge, skills, and principles on language assessment: a survey on teachers' language assessment literacy. *Proceedings of the Third International Conference on Sustainable Innovation 2019 – Humanity, Education and Social Sciences (IcoSIHESS 2019), 353*, 70–75.
<https://doi.org/10.2991/icosihess-19.2019.11>
- Qian, D. D., & Cumming, A. (2017). Researching English language assessment in China: Focusing on high-stakes testing. *Language Assessment Quarterly, 14*(2), 97–100.
<https://doi.org/10.1080/15434303.2017.1295969>
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing, 21*(3), 249–258.
<https://doi.org/10.1191/0265532204lt283ed>
- Stephen, S. (2012). Looking backward and forward at classroom-based language assessment. *ELT Journal, 66*(4), 523–532.
<https://doi.org/10.1093/elt/ccs041>
- Stiggins, R. J. (1991). Assessment literacy. *The Phi Delta Kappan, 72*(7), 534–539.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing, 30*(3), 403–412.
<https://doi.org/10.1177/0265532213480338>
- Trisanti, N. (2019, June). Assessment literacy analysis on designing classroom language test. In *UNNES International Conference on English Language Teaching, Literature, and Translation (ELTLT 2018)* (pp. 126–130). Atlantis Press.
- Viengsang, R. (2016). Exploring pre-service English teachers' language assessment literacy. *Modern Journal of Language Teaching Methods, 6*(5), 432.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly, 11*(4), 374–402.
<https://doi.org/10.1080/15434303.2014.960046>
- Vogt, K., Tsagari, D., Csépes, I., Green, A., & Sifakis, N. (2020). Linking learners' perspectives on language assessment practices to teachers' assessment literacy enhancement (tale): Insights from four European Countries. *Language Assessment Quarterly, 17*(4), 410–433.
<https://doi.org/10.1080/15434303.2020.1776714>
- Vogt, K., Tsagari, D., & Spanoudis, G. (2020). What do teachers think they want? A comparative study of in-service language teachers' beliefs on lal training needs. *Language Assessment Quarterly, 17*(4), 386–409.
<https://doi.org/10.1080/15434303.2020.1781128>
- Widiastuti, I. A. M. S., Mukminatien, N., Prayogo, J. A., & Irawati, E., (2020). Dissonances between teachers' beliefs and practices of formative assessment in EFL classes. *International Journal of Instruction, 13*(1), 71–84.
<https://doi.org/10.29333/iji.2020.1315a>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149–162.
<https://doi.org/10.1016/j.tate.2016.05.010>
- Yan, X., & Fan, J. (2021). “Am I qualified to be a language tester?”: Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing, 38*(2), 219–246.
<https://doi.org/10.1177/0265532220929924>
- Yan, X., Zhang, C., & Fan, J. J. (2018). “Assessment knowledge is important, but ...”: How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System, 74*, 158–168.
<https://doi.org/10.1016/j.system.2018.03.003>

APPENDICIES

APPENDIX 1

Questionnaire of CBLAL adapted from Lan and Fan (2019)

CBLAL Aspect	Number	Statement
Technical skills	QA1	I have enough knowledge in preparing classroom test
	QA2	I have sufficient knowledge using ready-made test from textbook packages or from other sources
	QA3	I understand how to give feedback to students based on information from tests/assessment
	QA4	I am aware of how to use self- or peer-assessment.
	QA5	I am knowledgeable about using informal, continuous, non-test type of assessment is well formed in my mind
	QA6	I am well-versed in using the Language Portfolio, an adaptation of it or some other portfolio.
Scores and decision making	QA7	I am well-informed about giving grades.
Language pedagogy	QA8	I have sufficient knowledge how to find out what needs to be taught/learned.
	QA9	I am well-versed in testing/Assessing Receptive skills (reading/listening).
	QA10	I have enough knowledge in testing/Assessing Productive skills (speaking/writing).
	QA11	Testing/Assessing microlinguistic aspects (grammar/vocabulary) is well formed in my mind
	QA12	I am knowledgeable about testing/Assessing Integrated language skills.
	QA13	I am knowledgeable about testing/assessing aspects of culture (students' attitudes)
Theories and principles	QA14	I am aware of how to establishing reliability of tests/assessment.
	QA15	I am well-versed in establishing validity of tests/assessment.
	QA16	I understand how to use statistics to study the quality of tests/assessment

Questionnaire of professional development need

CBLAL Training need	Number	Statement
Technical skills	QB1	training in preparing classroom test
	QB2	instruction in using ready-made test from textbook packages or from other sources
	QB3	working out in giving feedback to students based on information from tests/assessment
	QB4	training in using self- or peer-assessment.
	QB5	instruction in using informal, continuous, non-test type of assessment is well formed in my mind
	QB6	training in using the Language Portfolio, an adaptation of it or some other portfolio.
Scores and decision making	QB7	pointing in giving grades.
Language pedagogy	QB8	training in finding out what needs to be taught/learned.
	QB9	training in testing/Assessing Receptive skills (reading/listening).
	QB10	training in testing/Assessing Productive skills (speaking/writing).
	QB11	training in testing/assessing microlinguistic aspects (grammar/vocabulary) is well formed in my mind
	QB12	training in testing/Assessing Integrated language skills.
	QB13	training in testing/assessing aspects of culture (students' attitudes)
Theories and principles	QB14	Training in establishing reliability of tests/assessment.
	QB15	Training in establishing validity of tests/assessment.
	QB16	Training in using statistics to study the quality of tests/assessment

APPENDIX 2

The result of the questionnaires of all participants, both NTs and ETs

Components	CBLAL			PD		
	Items	Mean	SD	Item	Mean	SD
Technical skill	QA1	3.727	.911	QB1	4.272	.826
	QA2	3.854	.848	QB2	4.109	.785
	QA3	4.072	.899	QB3	4.072	.813
	QA4	3.872	.861	QB4	3.909	.928
	QA5	3.636	.949	QB5	3.709	.993
	QA6	3.436	.855	QB6	3.890	.853
Score and decision making	QA7	3.927	.813	QB7	3.981	.912
	QA8	3.890	.761	QB8	4.145	.869
Language pedagogy	QA9	3.727	.826	QB9	4.090	.928
	QA10	3.600	.873	QB10	4.272	.870
	QA11	3.690	.879	QB11	3.963	.881
	QA12	3.672	.840	QB12	4.109	.853
	QA13	3.727	.870	QB13	4.090	.928
	QA14	3.781	.809	QB14	4.127	.861
Theories and principles	QA15	3.781	.853	QB15	4.145	.869
	QA16	3.400	.914	QB16	4.018	.827