EXPLORING LANGUAGE ASSESSMENT LITERACY OF EFL TEACHERS IN THE CONTEXT OF INDONESIAN HIGHER EDUCATION

Andri Suherman

Universitas Hamzanwadi, Selong, Indonesia
Email: andrisuherman@hamzanwadi.ac.id

Abstract: The present study aims at examining EFL teachers' perceptions of Language Assessment Literacy (hereafter LAL) in the context of Indonesian higher education. To this aim, purposive sampling technique was employed to select sixteen EFL university teachers to participate in an online questionnaire and semi-structured interviews. The finding showed that the participants had applied six types of LAL in their teaching practices with formative assessment methods and test validity and reliability being the two most applied (more than 80%) by the participants. Following this, between 75% and 80% participants had applied two other types of LAL, linguistics competence and test results. The last two types of LAL, digital literacy and grading criteria, were applied by less than 70% participants. Another finding from qualitative data revealed that the majority of participants (over 70%) ranked their level of LAL knowledge and skills as low. Despite some limitations, the present study provides some implications including the necessity for the EFL teachers to be equipped with relevant knowledge and skills to make the way for quality assessment. It is therefore university authorities should organize professional development programs with the purpose to help teachers improve their assessment literacy. In addition, all educational parties should collaborate to consistently conduct literacy assessment projects throughout teachers' professional career. Some suggestions for future research in the field of language assessment are also presented.
Keywords: *EFL teachers; language assessment; university teachers.*

INTRODUCTION

Assessment can be assumed as one of the most important aspects of instructional contexts as it may significantly affect the quality of teaching and students' learning (Aryadoust *et al.*, 2021). Researchers argued that assessment should be utilized to boost deep learning, to enhance students' learning motivation, to help students recognize self-concept within themselves, and to help students understand the quality of assessment (McNamara *et al.*, 2019). It is therefore, assessment should be properly implemented as it can provide all parties (students, teachers, and assessors) with valid information about students' learning achievement (Salmani-Nodoushan, 2020). In fact, there is a strong relationship between assessment and teaching in a way that assessment can help teachers improve their teaching quality, and teaching can help teachers enhance their assessment quality (Coombe *et al.,* 2020). Nevertheless, it is highly unlikely that such relationship can be developed if teachers are not equipped with adequate assessment knowledge. The necessity of assessment knowledge for teachers triggered some scholars (Giraldo, 2018; McNamara *et al.,* 2019; Pastore & Andrade, 2019) to introduce and expand the paramount concept of "assessment literacy".

In the field of foreign language or second language education, assessment literacy is mainly focusing on the testing stakeholders' knowledge of the principles of assessment, and the implementation of the knowledge (Coombe *et al.,* 2020). Although the concept of assessment literacy is lacking of consensus, in consistent with the Standards of Teacher Competence in Educational Assessment of Students, Giraldo (2018) argued that language teachers need to possess a high level of language assessment literacy (hereafter LAL) to be able to develop appropriate testing methods, to administer tests, to score tests and use them to aid instruction, to communicate the test results to stakeholders, and to become aware of the ethical use of tests. Similarly, Sultana (2019) highlighted the pivotal role of LAL which could help teachers to collect accurate information about their students' learning achievement and to use this information to improve their teaching practices. It is therefore, as Wind and Peterson (2019) noted, language teachers are required to possess a high level of LAL in all level of education, including in the context of higher education. With regards to this, Kremmel and Harding (2020) mentioned the significance of examining how LAL is perceived

by university language teachers. The results can be used by university officials to determine whether or not teachers are equipped with required LAL. If not, the university officials need to organize relevant training courses to help the teachers enhance the level of their LAL.

As for the research of LAL, there have been many LAL-related studies over the last few decades, which were conducted by many researchers across different contexts. For example, in Chinese educational context, Yan and Fan (2021) investigated contextual and experiential factors which had affected assessment literacy of three different groups of people (EFL teachers, graduate students, and language assessors). Their finding revealed that different LAL profiles were demonstrated by the participants both at individual and group categories. At the group category, graduate students and language assessors reported a higher level of LAL than did the EFL teachers. At the individual category, however, each participant displayed a different process of LAL development although they shared similar patterns. Meanwhile, Lam (2019) examined the assessment literacy of 66 EFL secondary teachers in Hong Kong. He revealed that the majority of the teachers had a required level of assessment literacy, yet found difficulty to differentiate between two major elements in assessment, "assessment of learning" and "assessment for learning". Other than in China, some other researchers have also conducted relevant studies in different countries. In Iran, for example, Watmani *et al. (*2020) compared assessment literacy between EFL teachers and non-EFL teachers by employing assessment literacy scale. Their finding showed that both of the group of participants had a limited knowledge of assessment concept along with the different perceptions of LAL among the two groups. Meanwhile, Bøhn and Tsagari (2021) examined teacher educators' perception of LAL in the context of Norwegian education. Their finding showed that the participants claimed four different competences of LAL that teachers should have; "disciplinary competence, assessment-specific competence, pedagogical competence, and collaboration competence" (Bøhn & Tsagari, 2021, p.226).

In the contexts of Indonesian education, LAL has also been an interesting topic of research for the past few years. For example, Zulaiha *et al.* (2020) examined 22 secondary school EFL teachers' perception of LAL by employing a survey as their research instrument. Their finding showed that although the teachers demonstrated a high level of assessment knowledge, there was a huge discrepancy between their assessment knowledge and assessment practices. The discrepancy was particularly identified in two major stages, implementation and monitoring. Meanwhile, Mirizon (2021) investigated the level of EFL teachers' (*n=6*) assessment literacy at two different Indonesian secondary schools. By adapting the Classroom Assessment Literacy Inventory (CALL), his finding revealed that the teachers were found to have a fairly literate of LAL. In this regard, the teachers demonstrated difficulties when assessing students' learning because they failed to interpret the basic competence in the curriculum, and they rarely employed formative assessment methods due to lack of the ability to manage the classroom time. While the studies above were conducted at secondary school level, Marhaeni *et al. (*2020) examined 144 teachers' perception of LAL at elementary school level. By using the 30-item Classroom Assessment Literacy Questionnaire (CALT), their finding showed that the teachers' assessment literacy was perceived as "fair". In this sense, there was no significant different of LAL found between lower-grade teachers and upper-grade teachers.

Based on the review of studies above, it can be assumed that relatively not many researchers have explored the dimension of LAL in the context of Indonesian higher education. Furthermore, not many scholars have qualitatively explored the EFL teachers' perceptions of LAL at the university level in Indonesia. Thus, it is significantly important that the current study help expanding our understanding of the dimension of LAL by exploring EFL teachers' conceptions of LAL in the context of Indonesian higher education.

The followings are two questions that need to reveal in this study: (1) What types of LAL have EFL university teachers applied? (2) What do EFL university teachers think about their LAL?

It is hoped that the findings of this study can Improve the assessment quality by providing the language assessors with some valuable insight into teachers' perception of LAL.

METHOD

This study was conducted at Mataram University, Indonesia. To select EFL university teachers to participate in this study, purposive sampling method was employed. It yielded 16 EFL teachers specialized in Applied Linguistics (*n=6*), TESOL-Teaching English to Speakers of Other Languages (*n=8*), and Translation (*n=2*). It should be noted that the data collection was continually conducted until data saturation emerged. It means that the

data gained was repetitive and the participants shared no new things. Out of those 16 participants, nine were EFL male teachers (56%) and seven were females EFL teachers (44%). Their length of teaching experiences ranged from 1 to 5 (50%), 6 to 10 (25%), and 11 to 15 (25%). The information of participated teachers is displayed in Table 1 below.

This study employed a mixed method to collect data which consists of two instruments, an online questionnaire and semi-structured interviews. In regards with online questionnaire, it was designed by adapting the Qualtrics questionnaire platform. The questionnaire consists of 6 dichotomous questions regarding types of LAL that the university teachers had applied in their teaching performances. Based on their 'yes' and 'no' answers, they were required to answer multiple choice questions regarding the level of their LAL. The questions consisted of 3 options; high, medium, and low.

The second instrument, semi-structured interviews, were administered with the participants. The interviews began with an open-ended question (ex: how do you define language assessment literacy?) and continued with specific questions written in the checklist that had been developed before. If needed, additional or relevant questions were raised to elicit elaborate answers. It should be noted that the interviews were carried out in the participants' first language, Bahasa Indonesia, to allow them express their opinions with a greater ease without any struggles with second language barriers that might occur during the interviews. Each interview lasted for about 60 minutes and was audio recorded to be analyzed later. It is also worthy to note that the interviews were carried out in diverse temporal and spatial settings with the purpose to allow the participants deliver their thought without being affected by the time and place of the interviews.

Table 1. *The information of the participants*

Teachers	Gender	Level	Major	Teaching Experience (in years)
T1	F	Lecturer	TESOL	7
T2	M	Senior Lecturer	TESOL	12
T3	M	Senior Lecturer	Applied Linguistics	11
T4	F	Lecturer	TESOL	4
T5	F	Lecturer	Translation	7
T6	M	Senior Lecturer	TESOL	8
T7	M	Senior Lecturer	Applied Linguistics	14
T8	M	Senior Lecturer	Applied Linguistics	12
T9	F	Senior Lecturer	TESOL	10
T10	M	Associate Professor	Translation	13
T11	F	Lecturer	Applied Linguistics	5
T12	F	Senior Lecturer	TESOL	15
T13	M	Senior Lecturer	Applied Linguistics	10
T14	M	Associate Professor	TESOL	14
T15	F	Lecturer	TESOL	3
T16	M	Senior Lecturer	Applied Linguistics	13

From the online questionnaire, the quantitative data which consists of dichotomous and multiple-choice questions were statistically analyzed. On the other hand, the data gained from interview transcriptions were first coded and qualitatively analyzed. With regards to coding scheme, the procedure used was open coding by examining similarities and differences of all actions or events that emerged in the interview transcriptions. In this sense, the same actions were grouped together and put into broad categories. For instance, several actions or events that seemed to relate to what teachers know fell into the category "knowledge". Then, the actions or events that appeared to relate to technological competence fell into the sub-category "digital literacy". Last, the actions or events that seemed to relate to virtual assessment techniques fell into the theme "online test procedure". Once the identification analysis had finished, these three elements (category, sub-category and theme) generated codes which were then applied to all the interview transcripts.

Table 2. *Coding scheme for interview data*

Categories	Sub-Categories	Themes	Codes

A.	**Knowledge**	1.	Digital literacy	a.	Online test procedure	A-1-a
				b.	Electronic devices	A-1-b
		2.	Test validity and reliability	a.	Test design	A-2-a
				b.	Statistical formula	A-2-b
		3.	Linguistic Competence	a.	Teaching methods	A-3-a
				b.	Assessment techniques	A-3-b
				c.	Teaching media	A-3-c
B.	**Skill**	1.	Formative Assessment Methods	a.	Peer/self-assessment	B-1-a
				b.	Students' portfolio	B-1-b
		2.	Grading Criteria	a.	Productive skills criteria	B-2-a
				b.	Receptive skills criteria	B-2-b
		3.	Test Results	a.	Teachers' direct Feedback	B-3-a
				b.	Teachers' written report	B-3-b

RESULTS AND DISCUSSION

What types of LAL have EFL university teachers applied?

Based on the result of online questionnaire (see figure 1 below), it was found that the EFL university teachers have applied six types of LAL in their teachings during covid-19 pandemic. It is clear that two types of LAL, formative assessment methods and test validity and reliability were the most popular with 84% and 82% participants respectively. Following these, other two types of LAL, linguistics competence and test result, had been applied by between 75% and 80% teachers in their teaching practices. Meanwhile, digital literacy and grading criteria were the two least favorite types of LAL (less than 70%) which had been applied by the teachers.

This finding indicated that not all the EFL university teachers in this study have applied six types of LAL in their teaching practices. As Rasooli *et al. (*2018) suggested, because of limited experience in relevant trainings, many EFL teachers have lost their opportunity to develop their LAL. As a result, they tend to simply skip applying certain types of LAL in their teachings (Levi & Inbar-Lourie, 2020). For example, in the case of formative assessment methods (84%) in this study, it can be assumed that remaining participants (16%) have no sufficient competence to assess their students formatively. It is therefore, as argued by many scholars (Giraldo, 2018; Lee *et al.,* 2020; Ozan & Kincal, 2018), teachers need to be trained and facilitated with relevant trainings with the purpose to develop their LAL and become more competence in assessing their students' learning.
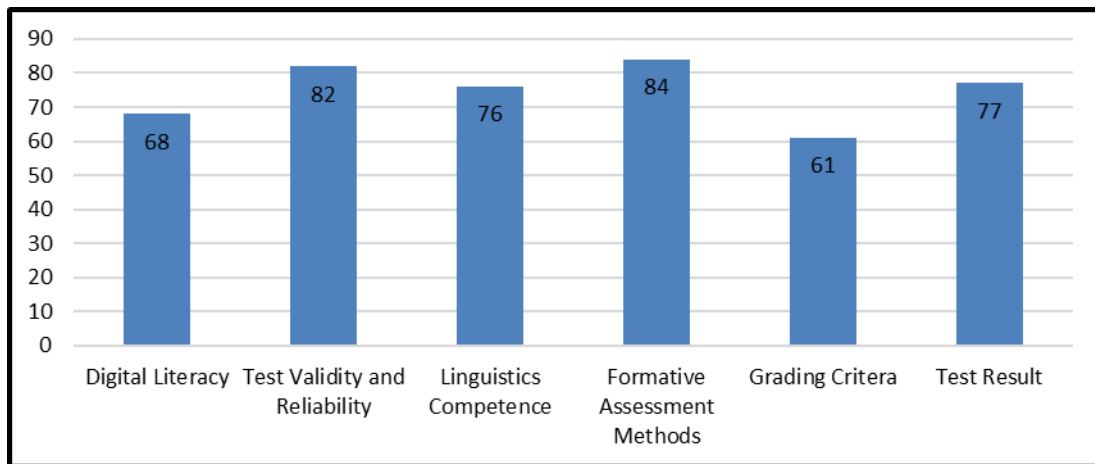

Figure 1. *Types of LAL for EFL university teachers*

What do EFL teachers think about their LAL?
Knowledge

Based on the data gained from online questionnaire (see figure 2 below), it was found that more than 70% participants ranked their level of LAL as low in each of the six types of LAL. In contrast, the level of medium and high were ranked less than 65% by the participants. As shown in the figure 2 below, it is clear that the first theme in the category of knowledge is related to digital literacy with 74% of participants thought that they had a low level of digital literacy. In this sense, although the teachers recognized the significance of technology in assessment practices, they were doubting their knowledge of the use of technology in virtual education. For instance, T4 commented in the interview session:

"I realize the essence of digital platforms to assess students' English skills, especially during this Covid-19 pandemic. But I honestly don't

know much how to administer online tests using Adobe Connect, for example. I often use the simple application like WhatsApps."
(Code: A-1-a, A-1-b)

The interview excerpt above indicates that virtual education is perceived as an integral element of LAL. Despite this, the participants doubted their ability when assessing their students virtually as they regarded their digital literacy was limited. This finding aligns with that of Gomez-Trigueros *et al.* (2019) which reported that teachers recognized the value of digital platforms in language assessment, yet lacked of digital literacy. The second theme from the data questionnaire is regarding test validity and reliability with 75% of participants ranked their knowledge of test validity and reliability as low. In this sense, the participants asserted that teachers should acknowledge that the tests they administered were valid and reliable. In relation to test validity, for example, T8 responded in the interview process:

"I always make my best effort to design tests which represent the course outline. However, designing a quality test is not easy since it requires adequate knowledge of test validation and I don't know much about statistical formula for making reliable tests. This is the reason why I often use ready-made tests provided by my institution which mostly lack of validity, I guess." (Code: A-2-a, A-2-b)

The statement above shows that both validity and reliability are highly valued by the teachers. Nevertheless, they were skeptical of their own knowledge of test validity and reliability. Consequently, they were dependent on ready-made tests provided by their institution. This finding supports what Rezai *et al.* (2021) have reported that although EFL teachers recognized how a quality test should be, they seemed to have inadequate knowledge of test validity and reliability.
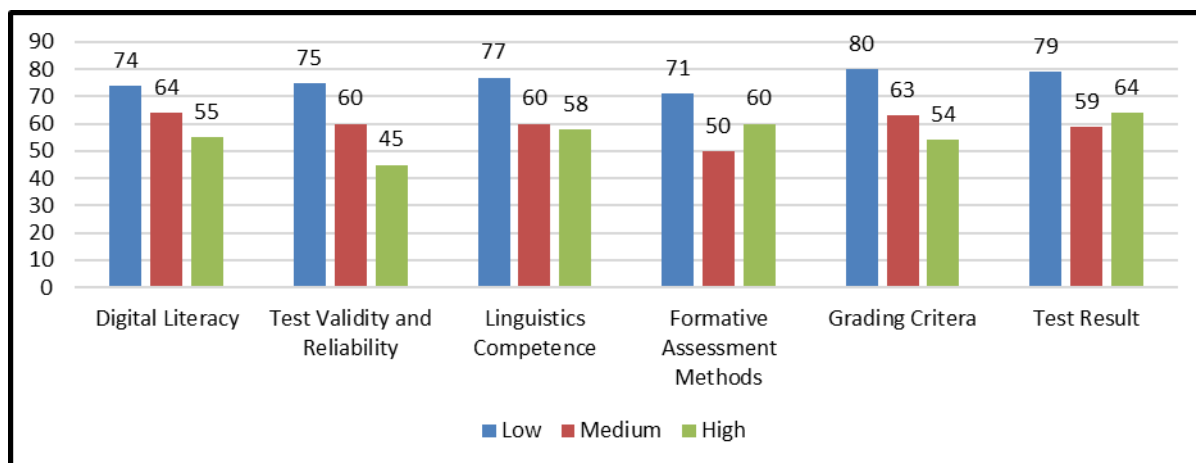

Figure 2. *Teachers' Perceptions of their level of LAL*

The last theme within knowledge category is related to teachers' linguistics competence with 77% of EFL teachers argued that their level of linguistics competence was low. The respondents recognized how teachers' linguistics competence affects their quality assessment. In the interview process, T2, for instance, reflected:

"I believe that if a teacher has adequate linguistics competence (ex: teaching methodologies, content knowledge, and teaching media), the testing practices will be found meaningful and effective by his/her students. That is the reasons why I always try to develop my linguistic competence by attending educational seminars and reading relevant books or journals." (Code: A-3-a, A-3-b, A-3-c)

The excerpts above shows that the teachers acknowledged the essence of linguistics competence as it might affect their assessment practices. Thus, they made their best efforts to familiarize themselves with basic knowledge of linguistics. This finding is in accordance with that of Abdulrahman and Ayyash (2019) which revealed that students often relate low-quality testing practices with teachers' lack of linguistics knowledge.

In sum, in the domain of knowledge, the first theme that emerged from the data analysis was related to digital literacy. The participants were aware of the significance of technology in assessment practices, particularly during this Covid-19 pandemic period. They found the need to be equipped with such digital knowledge as they possessed a very low-level literacy in the technological things. This finding may be explained from the perspective that the increasing interest in making use of electronic devices in

assessment practices is linked to the efficiency they provide for language assessors (Kuimova, 2018). In alignment with Khoiriyah (2020), this finding could be argued that the prominence of Computer-Assisted Language Learning (CALL) has benefitted many testing stakeholders. Thus, it is reasonable that teachers need to update their digital literacy to become proficient assessors in the digital world (Kahveci, 2021). While the first theme was related to teacher literacy in digital technology, the second theme was regarding teacher literacy in test validity and reliability. The respondents realized that a good test should be valid and reliable. Nevertheless, they had no sufficient knowledge of how to design a test which has a high level of validity and reliability. As a consequent, they prefer to use ready-made tests provided by their institution which were relatively not valid and reliable. This finding may be discussed from this respect that if a test does not satisfy validity requirements, it fails to assess what it is intended to assess, and if a test does not meet reliability requirements, it fails to consistently assess the intended competence (Tosuncuoglu, 2018). Another possible reason for this finding, as Hamid *et al.* (2019) noted, is that test validity increases the probability rate of making correct decisions about students' life, and reliability provides teachers with consistent information about students' competence. Thus, knowledge of test validity and reliability should be developed by teachers to become more expert in testing practices. The last theme in the category of knowledge was regarding teacher linguistics competence. The respondents recognized the importance of linguistics competence as it might affect their test quality. This finding may be explained from this viewpoint that teachers with high-level of linguistics competence relatively know how to implement effective teaching methods and design meaningful testing formats (Rasooli *et al.*, 2018). In alignment with Abdulrahman and Ayyash (2019), it could be assumed that if proficient teachers know how to linguistically deliver the lesson topics, it provides students with a great opportunity to learn and showcase their language skills. Therefore, linguistics competence should be highly valued by teachers as it may affect the quality of their teaching and testing practices. In brief, the data analysis revealed three main themes under knowledge domain namely digital literacy, test validity and reliability, and linguistics competence. Each of them was perceived as essential by the participants and required to develop.

Skills

In the domain of skills, as displayed in the figure 1 above, the first theme emerged from the result is regarding formative assessment methods with 71% of the EFL teachers ranked their skill level of formative assessment as low, although some of them (60%) thought that they had a high-level skill of applying formative assessment methods. In the interview process, one of the participants (T1) mentioned the benefits of students' self-assessment:

> "Students' self-assessment and peer-assessment are two of my favorite types of formative assessment methods. These help students to reflect upon their previous performances and improve their future learning. However, I feel that such -assessment methods are very time-consuming and portfolio assessment requires diverse skills to undertake. Therefore, I still prefer summative assessment techniques (multiple choice tests, short-answer essays, gap-fill questions) for practicality and economical reason." (Code: B-1-a, B-1-b)

The statement above indicates that formative assessment methods are positively treated by the teachers due to the fact that they can facilitate student learning. Nonetheless, some shortcomings (ex: consuming much time and requiring certain skills) embedded on such methods lead the teachers to prefer summative assessment methods due to their practical use and inexpensive cost. This finding is consistent with that of Leenknecht *et al.* (2021) which revealed that EFL teachers acknowledged the advantages of formative assessment techniques, yet rarely implemented them due to several drawbacks. As a result, they favored summative assessment techniques because of practicality and economic reasons.

The second theme from the questionnaire data was related to grading criteria with 80% participants thought that they had a low-level skill of designing grading criteria. In the interview session, for instance, T7 reflected:

> "It is extremely important that teachers tell their students about grading procedures at least one week before the test is administered. It helps students to focus on certain parts of the lessons. However, it is not easy to design grading criteria for speaking skills particularly. I need to adjust grading criteria I found in the internet with my students' learning purposes." (Code: B-2-a, B-2-b)

The interview quote above shows that the respondents highly valued grading criteria by

communicating them with their students in advance. Although the teachers faced some challenges when designing assessment criteria for speaking skills, they made their best effort to adjust the criteria with their teaching contexts. This finding supports what Mai (2019) have found that clear and transparent assessment criteria have a significant impact not only on students learning, but also on teachers' testing quality.

The last theme in the category of skill was regarding test results. Although 79% of participants argued that their skill level of test result was low, some of them (64%) thought that they had a high-level of reporting test results. In this sense, the EFL university teachers argued that students' test results should be reported along with relevant feedback. For example, T16 commented during interview session:

> "In my opinion, it is extremely essential that teachers provide students with some feedback after taking a test. Feedback can help students to see and rectify their mistakes for their future learning. But teachers need to be mindful of sharing test results to the low-proficient students as they may lose self-confidence and learning motivation."

The excerpt above indicates that communicating test results with students is significantly essential for the teachers. This is because they believe that effective feedback could help students with their future learning. However, they still worried about announcing tests score as it might adversely affect self-confidence of the low-proficient students. This finding aligns with that of Rasooli *et al.* (2018) which revealed that teachers found it important to announce students' scores, but they seemed to become resistant as it potentially lowered students' learning motivation.

In sum, in the domain of skill, three different themes were also found from the data analysis. The first one was regarding formative assessment techniques. Although the teachers recognized the benefits of assessing their students learning formatively, they still loved assessing their students learning summatively. This finding may be explained from the perspective that summative assessment methods are easy to design and are flexible to use in different contexts for different purposes (Tursunboevna, 2022). In alignment with Giraldo (2018), it could be argued that formative assessment techniques required certain skills to undertake which eventually hindered teachers to implement them. However, considering pedagogical advantages that formative assessment methods have to offer, as noted by previous studies

(Lee *et al.,* 2020; Ozan & Kincal, 2018), it is essential that teachers need to be trained with relevant skills to become proficient in assessing their students' learning formatively. The second theme elicited from the data analysis was related to grading criteria. The respondents realized that clear and transparent grading criteria had a positive relation with students' learning outcome. When assessing students' communicative skills, however, the teachers found it difficult to design appropriate grading procedures. This finding could be discussed from the view that valid grading criteria helped teachers to generate students' test scores without being contaminated by irrelevant constructs (Von-Wangenheim, 2018). Another possible reason for this finding could be students would consider their test score as fair when transparent and unbiased grading procedures were informed in advance (Quinn, 2020). Given the fact that clear grading criteria may lead to positive testing results, it is reasonably important that teachers need to develop their grading skills especially when it comes to speaking abilities. The last theme in the category of skill was regarding test results. The respondents were aware of providing students with positive feedback after taking a test could benefit their learning in the future. Nonetheless, they found it unethical to inform students about their test scores as it might negatively affect self-confidence of the low-proficient students. This finding may be explained from this viewpoint that positive feedback could help students to consolidate their learning, yet might destroy their self-efficacy (Ryan & Henderson, 2018). In contrast, Rios and Ihlenfeldt (2021) assumed that when test results were informed appropriately, along with positive feedback, students would not lose their self-confidence but become more motivated instead. To make it more meaningful and fairer, Rasoolie *et al.* (2018) even argued that teachers may present some positive feedback but keep the students' scores confidential. Thus, it is worthy to note that teachers need to familiarize themselves with some skills to help them deal with students' test scores. To close, three main themes under skill domain emerged from the data analysis namely formative assessment methods, grading criteria, and test results. Three of these elements were conceived as important by the respondent and thus need to develop.

CONCLUSION

The current study aimed at investigating the perception of Indonesian EFL teachers regarding the fundamental of LAL. Based on the data

analysis, two types of LAL domains emerged. The first one is the domain of knowledge which encompassed three different themes (e.g. digital literacy, test validity and reliability, and linguistics competence), and the second one is the domain of skill which also comprised three types of themes (e.g. formative assessment methods, grading criteria, and test results). As the finding revealed, it can be assumed that EFL teachers should be equipped with relevant knowledge and skill to make the way for quality assessment.

Some pedagogical implications can be seen from the study's findings. First, school authorities should organize professional development programs with the purpose to help teachers improve their assessment literacy. Second, formative assessment methods (e.g. peer assessment, self-assessment, portfolio assessment) should be more implemented. Such methods can be beneficial for student learning if the implementation is adjusted with teaching contexts and students' learning purposes. Another implication is teachers' assessment practices are likely to be perceived as effective if they consider contextual factors (e.g. students' interests, learning styles, culture, age and gender). The last implication is intended for the researchers in the field of language assessment. In this regard, they should be mindful of LAL elements which may change over time as the consequence of change in the views on L2 teaching.

Despite the implications above, it should be noted that the current study is not without limitations. Aside from small sample of participants, this study examined the fundamental of LAL from the teachers' perspectives. To reach a more comprehensive framework, future research can examine the fundamental of LAL from the students' perspectives. Furthermore, this study provided some meaningful insight into the effects of linguistics competence on teachers' assessment quality. More research needs to explore how teachers' linguistics competence affects students' perspectives of quality assessment. While this study revealed digital LAL from the teachers' perspectives, future studies can also examine digital LAL from the perspectives of both EFL teachers and students. In addition to this, future studies can research what grading criteria as perceived efficient by both EFL teachers and students. Last but not least, interested researchers can investigate the perceptions of both EFL teachers and students regarding the effects of announcing test results on students learning.

REFERENCES

Abdulrahman, N. C., & Ayyash, E. A. S. A. (2019). Linguistic competence, communicative competence, and interactional competence. *Journal of Advances in Linguistics*, *10*(1), 1600-1616.

Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, *38*(1), 6-40.

Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research, 12*(2), 222–233.

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), 1-16.

Giraldo, F. (2018). Language assessment literacy implications for language teachers. *HOW Journal*, *28*(3), 78-92. https://doi.org/10.19183/how.28.3.673

Gómez-Trigueros, I. M., Ruiz-Bañuls, M., & Ortega-Sánchez, D. (2019). Digital literacy of teachers in training: Moving from ICTS (information and communication technologies) to LKTs (learning and knowledge technologies). *Education Sciences*, *9*(4), 274-284.

Hamid, M. O., Hardy, I., & Reyes, V. (2019). Test-takers' perspectives on a global test of English: questions of fairness, justice and validity. *Language testing in Asia*, *9*(1), 1-20.

Kahveci, P. (2021). Language teachers' digital literacy and self-efficacy: Are they related? *ELT Research Journal*, *10*(2), 123-139.

Khoiriyah, K. (2020). CALL and SLA theory: developing A framework to analyze web-based materials for teaching listening skills. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, *8*(1), 80-92.

Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, *17*(1), 100-120.

Kuimova, M., Burleigh, D., Uzunboylu, H., & Bazhenov, R. (2018). Positive effects of mobile learning on foreign language learning. *TEM journal*, *7*(4), 837-841.

Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System, 81*(1), 78–89.

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, *33*(2), 124-140.

Leenknecht, M., Wijnia, L., Köhlen, M., Fryer, L., Rikers, R., & Loyens, S. (2021). Formative

assessment as practice: The role of students' motivation. *Assessment & Evaluation in Higher Education*, *46*(2), 236-255.

Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, *17*(2), 168-182.

Mai, D. T. (2019). A review of theories and research into second language writing assessment criteria. *VNU Journal of Foreign Studies*, *35*(3), 104-126.

Marhaeni, A. N., Paramartha, A. G. Y., Wirabrata, D. G. F., Dantes, N., Susilawan, I. A., & Handayani, D. A. P. (2020). Uncovering assessment literacy of elementary teachers in Singaraja, Bali. In *3rd International Conference on Innovative Research Across Disciplines (ICIRAD 2019)* (pp. 252-257). Atlantis Press.

McNamara, T., Knoch, U., Fan, J., & Rossner, R. (2019). *Fairness, justice & language assessment*. Oxford University Press.

Mirizon, S. (2021). Teachers of English assessment literacy: Unveiling what they know and do. *Indonesian Research Journal in Educational*, *5*(1), 124-141.

Ozan, C., & Kıncal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory & Practice*, *18*(1), 85-118.

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional mode. *Teaching and Teacher Education, 84,* 128-138.

Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, *42*(3), 375-392.

Rasooli, A., Zandi, H., & DeLuca, C. (2018). Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond. *Studies in Educational Evaluation, 56*(1), 164–181.

Rezai, A., Alibakhshi, G., Farokhipour, S., Miri, M. (2021). A phonomyography study on language assessment literacy: hearing from Iranian university teachers. *Language Testing in Asia, 11*(1), 1–26.

Rios, J. A., & Ihlenfeldt, S. D. (2021). State assessment score reporting practices for English learner parents. *Educational Measurement: Issues and Practice*, *40*(3), 31-41.

Ryan, T., & Henderson, M. (2018). Feeling feedback: Students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education*, *43*(6), 880-892.

Salmani-Nodoushan, M. A. (2020). Language assessment: Lessons learnt from the existing literature. *Online Submission*, *14*(2), 135-146.

Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia, 9*(1), 1–14.

Tosuncuoglu, I. (2018). Importance of assessment in ELT. *Journal of Education and Training Studies*, *6*(9), 163-167.

Tursunboevna, N. Z. (2022). Various types of assessment in language teaching and learning. *Eurasian journal of social sciences, philosophy and culture*, *2*(3), 140-145.

Von-Wangenheim, C. G., Hauck, J. C., Demetrio, M. F., Pelle, R., da Cruz Alves, N., Barbosa, H., & Azevedo, L. F. (2018). Code master--automatic assessment and grading of app inventor and snap programs. *Informatics in Education*, *17*(1), 117-150.

Watmani, R., Asadollahfam, H., & Behin, B. (2020). Demystifying language assessment literacy among high school teachers of English as a foreign language in Iran: Implications for teacher education reforms. *International Journal of Language Testing, 10*(2), 129–144.

Wind, S. A., & Peterson, M. E. (2018). A systematic review of methods for evaluating rating quality in language assessment. *Language Testing*, *35*(2), 161-192.

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.

Zulaiha, S., Mulyono, H., Ambarsari, L. (2020). An investigation into EFL teachers' assessment literacy: Indonesian teachers' perceptions and classroom practice. *European Journal of Contemporary Education, 9*(1), 189-201.