TEACHERS LANGUAGE ASSESSMENT LITERACY: A SYSTEMATIC REVIEW

Indah Puspawati
English Language Education Department, Faculty of Education,
Universitas Muhammadiyah Yogyakarta, Indonesia
Email: indahpuspawati@umy.ac.id

Abstract: As an integral part of teaching, assessment is critical since the information from the assessment is essential for many decisions in teaching and learning. Therefore, teachers must hold sufficient knowledge of assessment literacy. The interest in developing Language Assessment Literacy (LAL) has grown rapidly in the last few years. This growing interest for teachers LAL has sparked discussions related to the topics. This article aims to review research conducted related to the topic of teachers' LAL. This research used a systematic review as the research methodology where 21 articles from various peer-reviewed and reputable journals related to language assessment literacy in the past five years were reviewed to understand LAL comprehensively. Five steps of the systematic review were done in this research, namely framing the questions, identifying relevant works, assessing the quality of the studies, summarizing the evidence, and interpreting the findings. The findings of the reviews reported that the research of LAL in the past five years were teachers' LAL beliefs, the LAL beliefs and their assessment practices, teachers' LAL training needs, and developing teachers' LAL. The findings also suggested that teachers were still in needs of continuous support to develop teachers' LAL because they lack of LAL training both for pre-service and in-service training. This research recommended that LAL training should be carefully planned to cater the needs of teachers assessment practices.
Keywords: *Language Assessment Literacy; teachers' language assessment literacy; LAL beliefs, assessment practices, systematic review*

INTRODUCTION
Language Assessment Literacy (henceforth LAL) has been used for more than twenty years to describe the teachers' ability to use their knowledge and skills of designing assessments, administering them, and interpreting the assessment result (Sevimel-Sahin & Subasi, 2019). Sevimel-Sahin and Subasi (2019) added that LAL also referred to teachers' ability to understand trends in assessments and use them to evaluate their assessment practices in their teaching context. Giraldo (2020) explained that LAL represented teachers' level of knowledge, skills, and principles required to involve in language assessment. The LAL included designing and evaluating language assessment, understanding and using scores from assessment to make decisions about students' language abilities (Giraldo, 2020).

In attempts to further define LAL, researchers had offered descriptions of what LAL entails. One of the more detailed description of LAL was offered by Giraldo (2018). He argued that four sets of skills are important for teachers LAL. Those are instructional skills, design skills for designing language assessment, skills in educational measurements, and technological skills needed for language measurements such as using software, programs, and the internet needed for assessment purposes. For knowledge components, knowledge of applied linguistics, theories and concepts of language assessments, and teachers' language assessment context are three knowledge that teachers should have (Giraldo, 2018). Finally, those principles measure the teachers' awareness and action toward critical issues in language assessment, such as using the assessment results to decide about the learners, critically analyzing the impact of standardized tests, and awareness of applying fair and transparent language assessments. Giraldo (2018) also developed descriptors for every component, resulting in sixty-six descriptors to assess teachers' LAL.

Research on teachers' LAL revealed that teachers' LAL certainly played important roles in teachers' assessment practices. Teachers with

adequate LAL proved to be able to assess in their teaching. In contrast, teachers with inadequate LAL had problems with their assessment practices. Giraldo (2021) suggested that there is an urgent need for making sure that teachers had appropriate levels of LAL because they are the one who constantly involved in collecting data about their students language abilities. The needs for improving teachers' LAL also constantly reported in many research. However, in order to provide teachers with LAL training, there is also a need for understanding LAL in various contexts and how LAL is interpreted in different educational contexts (Coombe et al., 2020). Coombe, et. al. (2020) also suggested that more recent research of LAL can identify recent construct of teachers LAL knowledge and the conception on how the knowledge is put into practice. Thus, understanding these recent issues may provide clear pictures on how to support teachers' LAL. In addition, Sevimel-Sahin and Subasi (2019) suggested that exploring recent trends in LAL can give information on what has been known, what gaps and challenges are and what needs to be done in the future related to teachers LAL.

Therefore, this research aims to determine the topics included in teachers' LAL research in the last five years. By investigating the trends and topics in teachers' LAL, future research on teachers' LAL can be determined to fill the gaps in research in teachers' LAL. Thus, the needs of teachers' LAL can be clearly identified through reviewing the research that has been done in the past five years. Thus, this research question is formulated as follows. What has been studied on the empirical research on LAL in the past five years?

This study used a systematic review as its methodology. A systematic review is a detailed and comprehensive review of relevant studies on a particular topic to address formulated questions or problems (Uman, 2011). A systematic review usually involves identifying and evaluating all relevant studies and presenting the conclusions as cohesive (Ungvarsky, 2020). This research methodology is particularly suitable for research that investigates a specific topic's research trends to determine their implications for future research (Sevimel-Sahin & Subasi, 2019). The systematic review in this study was done in five steps: framing the questions, identifying relevant works, assessing the quality of the studies, summarizing the evidence, and interpreting the findings (Khan et al., 2003). The research question used for collecting the relevant study was what has been studied on the empirical research on LAL in the past five years to determine the implication of these topics for future research on LAL. With this research question in mind, the step of relevant research article was done. The researcher searched for articles from reputable databases using the keywords "language assessment literacy" and "teachers assessment literacy", the researcher searched for articles from reputable databases. Database from e-resources.perpusnas.go.id was mainly used for collecting relevant research. The researcher set the criteria to identify relevant works included in this research. They should be empirical research on teachers' LAL, going through peer review for publication, published in the last five years (January 2017 – April 2021), published in English, and accessible to the researcher. With the criteria, there were 22 articles included in this research. Below is the profile of the articles.

METHOD

Table 1. *The profile of LAL research articles*

	Author(S)/Year	Design	Participants
1.	Öz & Attay (2017)	Qualitative	12 Turkish university teachers
2.	Baker and Riches (2017)	Action research	120 Highschool EFL teachers in Haiti
3.	Janatifar and Marandi (2018)	Mixed	280 Iranian English language teachers
4.	Razavipour and Rezagah (2018)	Qualitative	14 Iranian English teachers
5.	Yastıbaş and Takkaç (2018)	Qualitative	8 Turkish teachers
6.	Giraldo and Murcia (2018)	Action research	Professors, language education experts and pre-service teachers
7.	Kremmel et al. (2018)	Action research	102 test items writers from Austria
8.	Giraldo (2019)	Qualitative	5 Columbian teachers
9.	Babaii and Asadnia (2019)	Qualitative	5 Iranian EFL teachers
10.	Firoozi et al. (2019)	Qualitative	15 Iranian EFL headteachers

11.	Ölmezer-Öztürk & Aydin (2019)	Quantitative	542 Iranian university language teachers
12.	Sultana (2019)	Qualitative	10 Bangladeshi secondary school teachers
13.	Levi and Inbar-Lourie's (2019)	Action research	16 language teachers
14.	Giraldo and Murcia (2019)	Action research	33 pre-service teachers in Columbia
15.	Puspawati (2019)	Mixed	30 Indonesian university teachers
16.	Zulaiha et al. (2020)	Mixed	22 Indonesian EFL junior high school teachers
17.	Liu and Li (2020)	Mixed	65 primary school English teachers in China
18.	Ahmadi and Ketabi (2020)	Mixed	309 Iranian teachers
19.	Bahtiar and Purnawarman (2020)	Mixed	40 Indonesian English teachers
20.	Saputra et al. (2020)	Qualitative	25 Indonesian secondary school teachers
21.	Yan and Fan (2020)	Qualitative	Language testers, EFL teachers, and graduate students in China
22.	Bøhn and Tsagari (2021)	Qualitative	Five teacher educators in Norway

From the article profile above, there were variations in the research methodology used, such as qualitative, action research, and mixed-method. It could be considered an improvement because the earlier research about LAL relied heavily on qualitative design (Sevimel-Sahin & Subasi, 2019). There were also variations in the participants involved in the studies where more stakeholders were involved. These variations in research design and participants could contribute to a more comprehensive understanding of LAL. The research in LAL had developed in complexity and scope to understand LAL better. Finally, the last step for the systematic reviews was summarizing the evidence and interpreting the findings. The evidence summary and the findings are presented in the next section.

RESULTS AND DISCUSSION

The research on teachers' LAL included in this study was analyzed to interpret the findings to answer the earlier research question. After careful analysis, four main topics emerged from the topics being discussed in the articles. Those topics are teachers' LAL beliefs, teachers' LAL beliefs and practices, teachers' LAL training needs, and developing teachers' LAL. Therefore, this research will discuss those four topics.

Results
Teachers' LAL beliefs
Research on teachers' LAL beliefs discussed how teachers perceived their ability in assessment, as found in Ölmezer-Öztürk & Aydin's (2019) study that investigated language assessment knowledge (LAK) of EFL teachers in higher education context in Turkey using the LAK scale developed and validated by the researchers. This research attempted to answer these research questions: 1) What are the general and skill-based LAK levels of the teachers?; 2) Is there any correlation among the skill-based LAK levels?; 3) Does LAK level change according to demographic features such as years of teaching experience, educational background, workplace, having an assessment course in their undergraduate program, training in assessment, and being a testing office member?, and 4) Does the teachers' LAL level change corresponding to their perceived self-competency in assessing each language skills? Five hundred forty-two (542) language teachers from 53 Iranian universities participated in the data collection for this study. Participants filled out a 60-item language assessment knowledge scale (LAKS), and the data were analyzed statistically. The data showed that teachers had more knowledge of assessing reading comprehension than assessing other language skills. It was also revealed that all types of skill-based knowledge were significantly and positively correlated with general LAKS. Also, there was no significant difference among the demographic features and their LAKS except for the testing office members whose LAKS were significantly higher than the non-members. Finally, there was a mismatch between teachers' self-perceived assessment competence and their actual LAL level.

There were also studies related to teachers' LAL level, conducted using a self-rating questionnaire where teachers rated their ability in assessment. One research about teachers' LAL level was

conducted by Puspawati (2019). Her research studied teachers' knowledge, skills, and principles of Indonesian teachers of higher education. The research investigated the teachers' LAL levels and their perceptions toward their LAL. Thirty (30) language teachers in higher education in Indonesia participated in this study. Surveys and interviews were conducted to collect the data. The survey was a self-rating questionnaire where participants rated their knowledge, skills, and principles in language assessment. The interviews were carried out with five teachers to collect data on their LAL perceptions. The data from the questionnaire revealed that teachers generally had a 'fair' LAL level, which means that teachers had a wide range of basic concepts and terms related to language assessment but were not yet capable of applying them in their practice. The interview data revealed that teachers were more confident in their teaching skills than their assessment skills. Another research on teachers' LAL levels was conducted by Bahtiar and Purnawarman (2020). Their research explored the Indonesian junior and senior high school English teachers' LAL levels and investigated their perception of LAL and their experiences in assessing the students.

The participants of this research were 40 Indonesian English teachers. A 66-item questionnaire was used as a primary instrument in this study, and interviews were also used for collecting the data. The questionnaire comprised four parts such as LAL knowledge, skills, principles, and practices. The interviews were done as a follow-up to the questionnaire. The data revealed that even though most teachers rate themselves 'good' and 'fair' in the LAL surveys, they claimed they were not familiar with LAL. Therefore, these teachers still needed training in LAL, preferably training that covered topics such as choosing the right assessment method for their students, writing test items, and conducting assessment validity and reliability. This study emphasizes the need for continued LAL training for teachers.

The research about teachers' LAL level revealed how teachers perceived their ability to assess, and both studies reported that teachers had inadequate LAL to carry out assessments in their teaching context. The inadequate LAL to carry out assessment also reflected on research conducted by Sultana (2019) that defined and measured the English teachers in Bangladesh LAL. First, the research objectives were to determine the teachers' readiness to execute assessment-related tasks and examine to what extent and how the teachers perceived their LAL in their teaching practices. Ten teachers at secondary schools in Dhaka, Bangladesh, participated in this study. The participants were interviewed to obtain the data, then recorded, transcribed, and analyzed using inductive thematic analysis. The findings revealed that teachers hardly had academic qualifications related to language testing because the training programs in undergraduate studies mainly focused on teaching and placed little to no attention on language testing. Therefore, teachers relied on ready-to-use tests to assess their students and professional suggestions from their senior colleagues to improve their assessment practices. Teachers also rarely used alternative assessments such as portfolio, peer assessment, or any other assessment form due to their lack of knowledge on such assessments. Critical examinations on their assessment practices were also never conducted by the teachers.

Babaii and Asadnia (2019) also conducted research investigating the teachers' reflective practices on the LAL. This research answered the following questions: 1) what teachers reflected on their assessment practices; 2) what LAL dimensions the teacher reflected on their current assessment practices; 3) what the teacher recommendations for improving their LAL; and 4) how their reflections affected their future language assessment planning. The data in this study were collected through online discussions, reflective narratives, interviews, and scenarios. A constant cooperative method analyzed and identified the emerging themes, categorized the codes and determined the categories. The findings revealed that teachers did not put as much attention to language assessment as to language teaching methodology in their practice since they were not professionally trained about assessment. They relied on colleagues' consultations, mentorships, and published materials to develop their LAL. Therefore, teachers suggested that language assessment communities where teachers could share knowledge and practices would be the best platform for improving their LAL. Lastly, teacher professional development programs devoted to improving their LAL should be transformational rather than transmissive.

Similarly, Ahmadi and Ketabi (2020) conducted research investigating the Iranian teachers' LAL status. The objective of this research was to explore the features that characterized Iranian English teachers' perceptions of LAL. The data collection methods used in this study were questionnaires and semi-structured interviews. Three hundred and nine (309) Iranian teachers answered the close-ended questionnaire, and 24 teachers were interviewed using semi-structured interviews. The data showed that LAL was still of concern to the teachers, and teachers agreed that their LAL knowledge and practice were not ideal and needed to be improved. However, they agreed that knowledge of assessment concepts and principles, understanding of local assessment practices, and the ability to use assessment results to make decisions were features participants deemed important for language teachers to have. Therefore, this research suggested that LAL be emphasized in pre-service and in-service teacher training programs.

Teachers' LAL beliefs and assessment practices
Some studies investigated teachers' beliefs on their LAL and how it was reflected in their assessment practices. The research about teachers' LAL beliefs and their assessment practices had two different results. Research reported discrepancies in the teachers' LAL beliefs and their practices (Liu & Li, 2020; Öz & Attay, 2017; Zulaiha et al., 2020), while research conducted by Giraldo's (2019) and Yastıbaş and Takkaç (2018) revealed that teachers' LAL beliefs were reflected in their assessment practices. The studies are summarized below.

Liu and Li's (2020) study aimed at investigating primary school English teachers' assessment literacy. The research attempted to answer the LAL level of those teachers, find the teachers' problems to design assessments for young learners and solve those problems. Sixty-five (65) primary school English teachers participated in this study, and the data were collected through a survey of language assessment inventory, recorded classroom observations, and interviews. The survey data revealed that teachers who participated in this study showed a low level of assessment literacy, especially in understanding principles of language assessments, administering, scoring, and interpreting the results of assessments, and communicating assessment results. Although teachers claimed that they were well-prepared for their assessment practices, no evidence showed in the observations. Teachers also claimed that they lacked language assessment training when they were both pre-service and in-service teachers. Therefore, training on language assessment should be included in pre-service teacher training programs, and professional development programs to improve in-service teachers' assessment literacy should also be emphasized to help teachers solve problems in assessment.

Oz and Attay (2017) investigated the teachers' literacy on in-class assessments and its reflection on the literacy on their in-class assessments practice. Specifically, this study examined Turkish EFL instructors' perceptions of in-class language assessment and how they reflect their knowledge of language assessment in their in-class assessment practice. This study involved twelve teachers working in a Turkish university. Teachers who participated in this study consisted of novice teachers with less than five years of teaching experience and expert teachers with more than five years of experience. They conducted qualitative data with a semi-structured interview as a data collection method. Each interview was recorded, transcribed, and analyzed, but no article on how the transcribed interviews were analyzed. The result showed a mismatch between teachers' assessment literacy knowledge and their in-class assessment practices. Teachers in this research were knowledgeable in the importance of in-class assessment, and they were also familiar with the theory of principles of a good language assessment. However, their practice in in-class language assessment did not reflect their knowledge. Thus, this study recommended giving teachers continuous professional development programs to raise their awareness and improve their assessment practices.

Zulaiha et al. (2020) investigated teachers' perceptions and their practices in assessment. This study also examined how teachers' assessments perception are reflected in their assessment practices. Twenty-two (22) Indonesian EFL junior high school teachers participated in this study. A survey was developed to collect the data on teachers' perceptions and practices of assessment, and interviews were also conducted with five teachers. The data from questionnaires were analyzed quantitatively by calculating means, frequency, and standard deviation, while the data from the interviews were transcribed and coded to

find emerging themes. The data from both the questionnaire and the interview suggested a discrepancy between the teachers' perception of assessment and their assessment practices. The teachers reported good assessment knowledge, but they failed to implement it in their assessment practices.

Giraldo's (2019) research examined teachers' practices and beliefs in LAL. Specifically, the research investigated what language assessment practices teachers had, what beliefs about language assessment those teachers had, and what the implications can be derived from the teachers' assessment practices and beliefs for teachers LAL. This study was a qualitative case study involving five Columbian teachers. The data were collected through online interviews and online reflective journals. The grounded approach was used to analyze the data both from the interviews and the reflective journals. The findings in this study revealed that teachers' practices and beliefs had implied that they had good LAL as they exhibited the knowledge, skills, and principles in LAL in both their practice and beliefs. It was also indicated that teachers' knowledge, skills, and principles in language assessment originated from their experience designing and using assessment for their teaching rather than formal training. It implied that training for developing teachers' LAL should include the teachers' assessment context. With similar findings, Yastıbaş and Takkaç (2018) explored the Turkish EFL teachers' strategies in developing their assessments to understand the implementation of LAL in their assessment practice. Eight Turkish teachers participated in this study, and the think-aloud protocol was used as the data collection method "to describe the cognitive processes the participants used in developing their exams" (p. 181). The data analysis was carried out using content analysis involving a process of coding, theming, organizing, and interpreting. The finding reported that the teachers in this study showed an appropriate implementation of LAL in their assessment practices. When designing and developing assessments, teachers considered content validity, student-centeredness, washback effects, and critical attitude toward their developed assessment. The findings of this study were certainly different from other studies, which reported that teachers had inadequate LAL.

Other research on teachers' LAL beliefs and assessment practice also revealed that teachers' LAL had a big influence on their practices. The research indicated that teachers who had inadequate LAL could not respond to the policy changes in their teaching context to carry out the assessment practices. They were reported to fail to adapt their assessment practices in response to policy changes regarding the curriculum in their teaching context. For example, Firoozi et al. (2019) explored the LAL essential element needs of Iranian EFL secondary school teachers to adapt to the recently changing curriculum policy. The research studied how teachers adapted to the assessment policy recently reinforced in their assessment practices. It also investigated what LAL components could bridge the assessment policy to their practices. Fifteen (15) EFL headteachers in Iran participated in this study, and interviews were used as a data collection method. The study suggested that the participants still had trouble adapting their assessment practices to the new curriculum reform. The headteachers needed transparent and practical guidelines of language assessment. More language assessment knowledge and changes in teachers' beliefs in the assessment were also needed to implement the assessment policy better into their assessment practices. The gap between the assessment policy and assessment practice can be a bridge by aligning the policies with the norms of assessment practice and improving teachers' LAL. This research also suggested that for teachers to do an effective assessment, they needed sufficient knowledge of the curriculum, course content, and subject matter knowledge.

Razavipour and Rezagah (2018) explored the challenges Iranian teachers faced dealing with language assessment reform in Iran and investigated whether the teachers possess the professional competencies (LAL) to adjust their assessment practice as demanded by the reform. The data in this study were collected through focus group interviews and content analysis on the recently designed and used teachers' assessments. Fourteen (14) teachers participated in this study. The data revealed that participants believed that the assessment reform in their context was pretty much forcefully imposed on them, so that they faced many challenges in adapting their assessment practices to the guidelines of the assessment reform. Difficulties included the lack of infrastructure and

resources to adapt to the assessment reforms and the high demand for the assessment results that made teachers hesitant in adapting to the reform. This research reported that teachers did not have adequate LAL to meet their LAL reform demands. For example, teachers' skills in designing tests were still inadequate. It was proven from the tests or assessments that they created or used recently lacked construct validity. Teachers were also reported to have inadequate basic knowledge of language assessment. The researchers suggested giving more training to improve teachers' LAL to adapt to the assessment reform.

Teachers' LAL training needs
The findings on the teachers' beliefs, teachers' beliefs and assessment practices concluded an emergence need to develop LAL through training. Before deciding what training would be effective for both pre-service and in-service teachers to develop teachers' LAL, studies were conducted to serve as a needs analysis to find out what aspects of LAL should be included. The research included in this study investigated the training needs from teachers' perspectives (Giraldo & Murcia, 2018; Janatifar & Marandi, 2018) and other stakeholders' perceptions (Bøhn & Tsagari, 2021; Giraldo, 2018; Yan & Fan, 2020). The needs analysis on LAL was usually done to identify what LAL aspects teachers and other stakeholders thought needed to be developed to enable teachers to carry out their assessment practices. The summaries of their research are presented below.

Janatifar and Marandi (2018) defined the nature of LAL in the Iranian context based on EFL teachers' beliefs. This research aimed to determine the components included in a language testing or assessment course based on the teachers' perceptions. The study employed a mix-method design and involved 280 Iranian English language teachers. The data in this study were collected using a survey that consists of three parts. In the first part of the survey, participants were asked to rate components of LAL on a 5-point scale ranging from *unimportant* to *essentials*. The data were then analyzed quantitatively using confirmatory and exploratory factor analyses and cross-validation. The second part of the survey was open-ended questions where the participants were asked to offer their opinion on some language assessment issues, particularly focusing on components they believed

should be included in the language test/assessment course. The data from the open-ended questions were analyzed qualitatively. The results suggested that the participants in this study believed that four components should be included in LA, such as "test design and development, large-scale standardized testing and classroom assessment, beyond-the-test aspects (which mainly includes social and ethical aspects of language testing or assessment), and reliability and validity" (p. 369). Thus, the researchers suggested that these four components be considered when designing courses to improve LAL teachers.

Giraldo and Murcia (2018) attempted to expand the construct of LAL by investigating the pre-service teachers' and language teacher educators' perceptions on the components that should be included in a language assessment course. It was a part of an action research study conducted in an English language teaching program in a state university in Columbia. The action research was conducted in a course entitled Seminar in Language assessments. This research was the diagnostic stage of the action research, which aimed to collect data from professors, a language teacher education expert, and pre-service teachers who enrolled in the course. The data collected were about the components included in the course. In this diagnostic, the data were collected through open questionnaires, a semi-structured interview, multiple-choice questionnaires, and researchers' notes. The data were analyzed qualitatively, employing open, axial, and selective coding, and the multiple-choice questionnaire was analyzed quantitatively by calculating percentages on language assessment topics included in the course. The findings indicated that designing assessments was the most important component included in an assessment course. Participants in this study all agreed that practical issues on designing assessments were needed for teachers to help them carry out the task in their teaching. Secondly, knowledge of assessment methods, the validity of the assessment, and bilingualism issues should also be included in an assessment course. Finally, the study suggested that fairness issues in assessment should also be included in a language assessment course to develop teachers' LAL comprehensively.

Bøhn and Tsagari (2021) explored the teachers' LAL development. This research investigated the point of view of teacher educators on aspects of

LAL Norwegian teachers' need to carry out sound assessment practices. The research also scrutinized teacher educators' perception of whether Norwegian teachers' LAL was compatible with eight dimensions of LAL. Five teacher educators from five different universities in Norway were recruited to participate in the study. Interviews were used for collecting the data in this research. The interview consisted of open-ended questions which collected the data about the participants' perceptions about skills and knowledge Norwegian teachers needed to carry out good assessments and close-ended questions which collected the data on the relevance of the LAL eight dimension for Norwegian teachers. According to the participants, the findings reported that to carry out good assessments, teachers need at least four competence, namely disciplinary competence assessment-specific, pedagogical and collaboration competence. Principles and concepts of assessment, local assessment practices, scores, and decision-making were three dimensions that teacher educators found relevant to teachers' LAL. The findings implied that collaborative professional community activities where teachers could communicate and interact with colleagues might be good for improving teachers' LAL.

Yan and Fan (2020) investigated the perceptions of different stakeholders on profiles and the development of LAL. This research involved language testers, EFL teachers, and graduate students of language studies program to explore their interests, needs, and expectations in language assessment. Semi-structured interviews were used for the data collection, and twenty stakeholders participated in this study. The study revealed a major difference in LAL profiles across the stakeholders. Language testers and graduate students were more familiar with LAL terminologies and assessment research, more capable of assessing their LAL profiles and describing their LAL development.

Meanwhile, EFL teachers' LAL profiles revealed that they were less familiar with LAL, but they showed a good level of assessment intuitions. The teachers claimed they develop their LAL through a range of experiences of the assessment practices in their teaching context. With the different profiles and development of LAL across stakeholders, the research recommended that apprenticeship and the local assessment context be two important elements to consider when designing assessment training programs.

Developing teachers' LAL
The last finding of the systematic reviews on the articles revealed some attempts to improve or develop teachers' LAL. Most of the research reported the professional development program conducted to develop in-service teachers' LAL (Baker & Riches, 2017; Kremmel et al., 2018; Levi & Inbar-Lourie, 2019; Saputra et al., 2020), and one research reported a course to develop pre-service teachers' LAL (Giraldo & Murcia, 2019). All research reported positive development on teachers' LAL. The summaries of the research are presented below:

Baker and Riches (2017) studied a series of workshops to develop teachers' and language assessment professionals' language assessment literacy. This research aimed to investigate how teachers' and language assessment specialists' LAL developed during the collaborative process of critiquing and revising Haitian EFL national examinations and examine what successes and challenges arise for teachers during this collaboration process. This study was qualitative research involving 120 high school EFL teachers in Haiti. The data collection methods in this study involved teachers' feedback on drafts of the revised Haitian EFL national exam where teachers demonstrate their knowledge on language assessments, an open-ended survey with teachers' participants (N= 92), which was conducted after the workshop, and semi-structured interviews to 11 teachers who are willing to participate after the workshop. The data were analyzed through open coding and iterative analysis for determining categories of findings. The findings revealed that the workshop had developed teachers' LAL. Teachers reportedly improved their understanding of designing assessments, such as creating reading comprehension questions, evaluating integrated skills and language components, understanding the importance of assessment in teaching, understanding the construct of language ability, and being aware of a more student-centered approach in assessment.

Similarly, the language assessment specialist who became the facilitators in the workshop also showed development in their LAL. They showed increasing awareness of cultural issues, political

issues in language assessment, practical constraints in assessment administration, and exam washback issues. Finally, building a good rapport between teachers and facilitators was considered a success in the workshop, and communicating and cooperative decision-making with colleagues was proven challenging.

Kremmel et al. (2018) investigated the effects of a project to improve the teachers' LAL. In the project, teachers developed high-stake tests as a part of their professional development program. In this project, teachers were involved in activities to develop high-stake tests such as writing test specification, developing test items, developing tasks for productive skills, developing rating scales, giving feedback on the developed tests, tasks, and rating scale, revising the tests, tasks, and rating scale, and participating in standard-setting. The data in this research were obtained from a LAL questionnaire that collected data about the teachers' motivation to be involved in the project, their motivation to stay involved, and their attitudes toward their involvement in the project. The last part of the survey collected the data about whether the training fostered their LAL and what areas of the project best fostered their LAL. One hundred two (102) item writers participated in the project completed the survey. The teachers' LAL findings indicated that the project helped the teachers gain their expertise in an item and task development, selecting tests for their teaching context, understanding concepts of validity, reliability, and practicality. The study also suggested that teachers involved in the project could apply the skills and knowledge they got from the project into their small-scale classroom context, indicating that hands-on assessment training could develop teachers' LAL.

Levi and Inbar-Lourie (2019) examined the assessment knowledge base or LAL of prospective and practicing teachers who were participated in a formative classroom assessment course. After a formative classroom assessment course, their research examined the language teachers' LAL through assessment tools and reflection. Besides, the research also intended to find out whether the teachers' LAL could be associated with the current constructs of LAL. Sixteen language teachers participated in this study. They attended the course and created language tests and performance-based tasks focused mainly on reading comprehension

with essay writing and listening comprehension. The data in this research were collected through the thematic content analysis of the teacher-produced tests and tasks and compared with the current LAL construct. Reflective comments from the teachers were also used for collecting the data. The findings reported that after the course, teachers attested that they could plan, design and construct test items and constructs to match their assessment objectives claiming that they implemented assessment knowledge and skills presented in the course to their assessment practices. However, only a little evidence of their claim was shown in their assessment artifacts, further verifying the gap between teachers' understanding of LAL and their assessment practices.

Saputra et al. (2020) reported a case study studying the teachers engaged in a professional learning community (PLC) project to develop their LAL. This research. The project was designed to familiarize the teachers with the policies, theories, and practices of authentic language assessment in the curriculum currently enforced in the teaching context. The study investigated the changes that teachers experienced in their assessment beliefs and practices. 25 English teachers in secondary schools in Indonesia were invited to join the project, where they had weekly discussions and observations spanning one semester for one semester. Multiple interviews and observations of six out of 25 teachers were done to collect the data analyzed thematically. The study revealed that the PLC project contributed to the changes in participants' beliefs and practices of authentic assessment in their class. The study suggested the use of collaborative projects to improve in-service teachers' LAL.

Giraldo and Murcia (2019) conducted action research to see the effects of a classroom language assessment course toward pre-service teachers LAL. Its objectives were to investigate how pre-service teachers LAL developed during the course, what factors of the course impacted their LAL, and what instructional recommendations this study could offer. The course ran for one semester and met for four hours a week. Thirty-three (33) pre-service teachers and one-course instructors participated in this study. Data were collected through questionnaires, semi-structured interviews, and instructor and researcher's journals. The data revealed that the pre-service teachers enrolled in the course developed their LAL in two areas:

improvement in their perception of language assessment from limited views to a more professionalized process-oriented assessment. Besides, they were able to review assessments and assessment practices critically. The act of designing language assessment and how the course was carefully designed based on needs analysis and taught engagingly and strategically to improve LAL were two factors that contributed to the pre-service teachers' LAL improvements. This study recommended that to design an effective language assessment course for pre-service teachers needs analysis from multi-stakeholders should be conducted before designing the course. Besides, the course should involve pre-service teachers actively designing assessments tasks to trigger informed decisions in assessment.

Discussion

After carefully reviewing each article in this study, four main topics were discussed in the empirical research regarding teachers' LAL in the past five years. The first topic discussed teachers' beliefs of LAL. The research methodology used in this topic was mostly mixed methods using quantitative design using survey or questionnaire of self-rating LAL level and knowledge, and qualitative design using interviews, reflections, and scenarios. The findings showed that teachers in the studies recognized that knowledge and skills in language assessment were important aspects of their teaching (Ahmadi & Ketabi, 2020), but teachers also claimed that they had concerns about their LAL (Ahmadi & Ketabi, 2020; Bahtiar & Purnawarman, 2020; Puspawati, 2019; Sultana, 2019). Teachers claimed that they were more confident in their teaching skills than their assessment skills (Puspawati, 2019; Sultana, 2019), and they even claimed that they somewhat neglected their LAL and put more emphasis on language teaching methodologies (Babaii & Asadnia, 2019; Bahtiar & Purnawarman, 2020; Sultana, 2019). One of the causes of that practice was because teachers lacked training on LAL both when they were pre-service teachers and in-service teachers (Babaii & Asadnia, 2019; Sultana, 2019). In addition, the studies investigated on the teachers' LAL levels revealed that teachers often reported their perceived ability as not ideal (Ahmadi & Ketabi, 2020), fair (Puspawati, 2019) or good (Bahtiar & Purnawarman, 2020). Similarly, teachers in Ölmezer-Öztürk & Aydin's (2019)

research also reported lower language assessment knowledge than testing office members. The teachers were reported to rely on ready-made assessment tools instead of developing their assessment, and they also relied on mentorship from more senior teachers or colleagues to help them improve their LAL. Thus, the studies suggested that training on LAL for both pre-service and in-service teachers is vital to equip teachers better to do assessment practices.

The next emergence themes from the studies included in this review were the teachers' LAL beliefs and assessment practices. The research on this topic explored the teachers' beliefs on their LAL and how they reflect on their assessment practices. The research methodologies used in the studies exploring this issue were mostly interviews, reflective journals for their beliefs, content analysis of the teachers' assessment tools, and observations to compare the assessment beliefs and practices. There were two different findings reported for this issue. The first findings reported teacher beliefs and assessment practices (Liu & Li, 2020; Öz & Attay, 2017; Zulaiha et al., 2020). Teachers in those studies claimed they were well prepared for their assessment practices, but there was no evidence on their assessment practices (Liu & Li, 2020). Teachers were also claimed that they had good knowledge in LAL (Zulaiha et al., 2020), and they also believed that they had good knowledge of in-class assessment and theories of principles of language assessments (Öz & Attay, 2017). However, their LAL knowledge was not reflected in their assessment practices (Öz & Attay, 2017; Zulaiha et al., 2020).

On the other hand, different findings of teachers' LAL beliefs and their assessment practices were reported in the studies conducted by Yastıbaş and Takkaç (2018) and Giraldo (2019). These two studies reported that teachers showed knowledge skills and principles in both their beliefs and practices (Giraldo, 2019), and they also showed an appropriate implementation of LAL in their assessment practices. Their assessment practices also included elements needed to present in LAL, such as content validity, students-centeredness, washback effects, and critical attitude. These two different findings certainly bring enrichment to the LAL research, which might indicate that LAL had finally started to get recognized for its importance in the language education field that stakeholders

started to focus on developing teachers' LAL. LAL's importance is also shown in studies conducted by Razavipour and Rezagah (2018) and Firoozi et al. (2019). Their research investigated how teachers' LAL allowed them to respond to the curriculum changes in their respective teaching contexts. Both studies reported that teachers with inadequate LAL struggled to adapt their assessment practices to the new curriculum demands. Their studies suggested that teachers need more training to respond to the new curriculum reforms. All studies suggested LAL training for better assessment practices regardless of the teachers' LAL beliefs and assessment practices. However, the kinds of training given to teachers should be chosen carefully to facilitate LAL development. Thus, careful needs analysis for teachers' training for developing teachers' LAL needs to be carried out.

The next emergent finding from this review was about needs analysis on what components of LAL teachers should develop. Studies included in this review suggested some components of LAL that needed to be incorporated into LAL training. The components for LAL training included LAL knowledge such as knowledge of different assessment methods (Giraldo & Murcia, 2018), principles and concepts of assessments designing and developing assessments (Bøhn & Tsagari, 2021), knowledge on large-scale standardized testing and classroom assessment (Janatifar & Marandi, 2018), and knowledge on local assessment practices (Bøhn & Tsagari, 2021). Besides, skills of LAL should also be included, such as the skill to design and develop assessments (Janatifar & Marandi, 2018), to test assessments' validity and reliability (Giraldo & Murcia, 2018; Janatifar & Marandi, 2018), and to score and make a decision based on the assessment (Bøhn & Tsagari, 2021). Other aspects such as social and ethical aspects of language assessment (Janatifar & Marandi, 2018) and issues on bilingualism and fairness in assessments (Giraldo & Murcia, 2018) also needed to be included in LAL training. In addition, the main consideration when designing the training for teachers' LAL was that training should be designed in such a way that teachers had hands-on experience on the assessment, it should also consider the collaborative nature of the training, and it should be contextualized in the teachers' teaching, or this sense, assessment context (Bøhn & Tsagari, 2021; Giraldo & Murcia, 2018; Janatifar & Marandi,

2018; Yan & Fan, 2020). These components could be valuable to consider when designing the LAL training for pre-service and in-service teachers.

The fourth topic discussed in the studies was about developing teachers' LAL. Studies have reported professional development programs for developing in-service teachers' LAL and training programs for developing LAL pre-service teachers. Various programs were conducted to develop teachers LAL, such as a workshop to collaboratively critique and revise standardized testing (Baker & Riches, 2017), a project to develop a high-stake test where participants develop, revise, and rate tests, tasks, and rating scales of a standardized high-stakes test (Kremmel et al., 2018), a formative assessment course where teachers created language test and performance-based tasks (Levi & Inbar-Lourie, 2019), and a project of a professional learning community (PLC) where teachers collaboratively learn to design authentic language assessment (Saputra et al., 2020). Besides, a study reported action research to develop pre-service teachers LAL through a classroom language assessment course designed to involve pre-service teachers to gain knowledge, skills, and principles in LAL (Giraldo & Murcia, 2019). All the studies reported positive effects on the teachers who participated in the programs they developed and improved their LAL. It is proven from the studies that teachers' active involvement in assessment starting from planning, designing, administering, scoring, giving feedback, critiquing, and revising assessment was necessary for improving teachers' LAL. Besides, collaboration was also important for teachers' LAL since it could help teachers support one another in their assessment practices.

CONCLUSION

This systematic review investigated the topics discussed in the empirical research of LAL to see what has been discussed and what future research should be directed. The findings of this review indicated that the topics discussed covered teachers' LAL beliefs, teachers' LAL beliefs and their assessment practices, teachers' LAL training needs, and developing teachers' LAL. The discussion on teachers' LAL beliefs reported that although teachers acknowledge the importance of LAL, LAL has been somewhat neglected by teachers who emphasize language teaching methodologies in their

teaching. Teachers also believed that they were not professionally trained to carry out assessment tasks in their teaching contexts, so they relied so much on the mentorship from their senior peers to develop their LAL. From this finding, the teachers believed that continuous development programs were needed to help teachers improve their LAL. The next topics discussed in the studies included in this review were how teachers' beliefs are reflected in their assessment practices. The findings suggested two different results where. The first result indicated a mismatch between teachers' LAL beliefs and their assessment practices, where teachers' assessment practices did not reflect their LAL beliefs. However, other studies indicated that teachers had adequate, even good, LAL, which was shown clearly in their assessment practices. Despite the different findings, the studies about teachers' LAL beliefs and assessment practices revealed that LAL influenced teachers' practice. Teachers who had inadequate LAL failed to carry out assessments practices as mandated in the curriculum reforms because teachers did not have the assessment knowledge and skills to respond to the policy changes. The third topic discussed in the studies was teachers' LAL training needs. The needs analysis showed what LAL components need to be included in the training and how effective LAL training should be conducted. The last topic discussed was related to programs and courses to develop teachers' LAL. The programs and courses developed mainly offered teachers hands-on experience on assessment and involved collaborative work in assessment.

REFERENCES

Ahmadi, M. R. S., & Ketabi, S. (2020). *Features of Language Assessment Literacy in Iranian English Language Teachers' Perceptions and Practices*. *38*(1), 191–223. https://doi.org/10.22099/jtls.2020.34843.2739

Babaii, E., & Asadnia, F. (2019). A long walk to language assessment literacy: EFL teachers' reflection on language assessment research and practice. *Reflective Practice*, *20*(6), 745–760. https://doi.org/10.1080/14623943.2019.1688779

Bahtiar, I., & Purnawarman, P. (2020). *Investigating English Teachers' Comprehension in Language Assessment Literacy (LAL)*. *508*(Icite), 303–310. https://doi.org/10.2991/assehr.k.201214.253

Baker, B. A., & Riches, C. (2017). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 1–25. https://doi.org/10.1177/0265532217716732

Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in norway. *Journal of Language Teaching and Research*, *12*(2), 222–233. https://doi.org/10.17507/jltr.1202.02

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). https://doi.org/10.1186/s40468-020-00101-6

Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, *9*(1). https://doi.org/10.1186/s40468-019-0078-7

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

Giraldo, F. (2019). Language Assessment Practices and Beliefs: Implications for Language Assessment Literacy. *How*, *26*(1), 35–61. https://doi.org/10.19183/how.26.1.481

Giraldo, F. (2020). A post-positivist and interpretive approach to researching teachers' language assessment literacy. *Profile: Issues in Teachers' Professional Development*, *22*(1), 189–200. https://doi.org/10.15446/profile.v22n1.78188

Giraldo, F. (2021). A reflection on initiatives for teachers' professional development through language assessment literacy. *Profile: Issues in Teachers' Professional Development*, *23*(1), 197–213. https://doi.org/10.15446/profile.v23n1.83094

Giraldo, F., & Murcia, D. (2018). Language Assessment Literacy for Pre-service Teachers: Course Expectations from Different Stakeholders. *GiST Education and Learning Research Journal*, *16*(16), 56–77. https://doi.org/10.26817/16925777.425

Giraldo, F., & Murcia, D. Q. (2019). Language Assessment Literacy and the Professional Development of Pre-Service Language Teachers. *Colombian Applied Linguistics Journal*, *21*(2), 243–259. https://doi.org/10.14483/22487085.14514

Janatifar, M., & Marandi, S. S. (2018). *Iranian EFL Teachers' Language Assessment Literacy (LAL) under an Assessing Lens*. 307–328. https://doi.org/10.22108/are.2018.108818.1221

Khan, K. S., Khunz, R., Kleijnen, J., & Antes, G. (2003). five step for systematic review.pdf. *Journal of the Royal Society of Medicine*, *96*, 118–121.

Kremmel, B., Eberharter, K., Holzknecht, F., & Konrad, E. (2018). Teacher Involvement in High-Stakes Language Testing. *Teacher Involvement in High-Stakes Language Testing*. https://doi.org/10.1007/978-3-319-77177-9

Levi, T., & Inbar-Lourie, O. (2019). Assessment Literacy or Language Assessment Literacy: Learning from the Teachers. *Language Assessment Quarterly*, 168–182. https://doi.org/10.1080/15434303.2019.1692347

Liu, J., & Li, X. (2020). Assessing Young English Learners: Language Assessment Literacy of Chinese Primary School English Teachers. *International Journal of TESOL Studies*, *2*, 36–49. https://doi.org/10.46451//ijts.2020.12.05

Ölmezer-Öztürk, E., & Aydin, B. (2019). Investigating language assessment knowledge of efl teachers. *Hacettepe Egitim Dergisi*, *34*(3), 602–620. https://doi.org/10.16986/HUJE.2018043465

Öz, S., & Attay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, *6*(1), 25–44.

Puspawati, I. (2019). *Understanding Teachers' Knowledge, Skills, and Principles on Language Assessment: A Survey on Teachers' Language Assessment Literacy*. *353*(IcoSIHESS), 70–75. https://doi.org/10.2991/icosihess-19.2019.11

Razavipour, K., & Rezagah, K. (2018). Language assessment in the new English curriculum in Iran: managerial, institutional, and professional barriers. *Language Testing in Asia*, *8*(1). https://doi.org/10.1186/s40468-018-0061-8

Saputra, E. R., Hamied, F. A., & Suherdi, D. (2020). The development of beliefs and practices of language assessment literacy: does a professional learning community help? *Journal of Education for Teaching*, *46*(3), 414–416. https://doi.org/10.1080/02607476.2020.1761250

Sevimel-Sahin, A., & Subasi, G. (2019). An overview of language assessment literacy research within English language education context. *Journal of Theoretical Educational Science*, *12*(4), 1340–1364.

Sultana, N. (2019). Language assessment literacy: an uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, *9*(1), 1–14. https://doi.org/10.1186/s40468-019-0077-8

Uman, L. S. (2011). Information Management For The Busy Practitioner Systematic Reviews and Meta-Analyses Information Management for the Busy Practitioner. *J Can Acad Child Adolesc Psychiatry*, *20*(1), 57. www.cochrane.org

Ungvarsky, J. (2020). *Systematic Review*. Salem Press Encyclopedia.

Yan, X., & Fan, J. (2020). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 1–28. https://doi.org/10.1177/0265532220929924

Yastibas, A. E., & Takkaç, M. (2018). Understanding Language Assessment Literacy: Developing Language Assessments. *Journal of Language and Linguistic Studies*, *14*(1), 178–193.

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. https://doi.org/10.13187/ejced.2020.1.189