

**SKILL-BASED LANGUAGE ASSESSMENT KNOWLEDGE: A STUDY OF SENIOR HIGH SCHOOL TEACHERS IN A REGENCY OF JAMBI PROVINCE**

Armiwati<sup>1</sup>  
Hustarna<sup>2</sup>  
Susanah<sup>3</sup>  
Melati<sup>4</sup>

<sup>1,2,3,4</sup>Universitas Jambi, Jambi, Indonesia

Corresponding author: [armiwati@unja.ac.id](mailto:armiwati@unja.ac.id)

**Abstract**

This study explores the English teachers' level of knowledge of skill-based language assessment among some high schools situated in a regency in Jambi province. The questionnaire is used to obtain quantitative data and cover several aspects of learning assessment. It consists of 60 items across four constructs of the Language Assessment Knowledge Scale (LAKS), developed by Ölmezer-Öztürk and Aydın (2019), and has been completed by 24 respondents. The results showed that the highest percentage of correct answers in assessing reading was for item 13 (100%), listening was for item 30 (100%), writing was for item 32 (91.7%), speaking was for item 52 (95.8%), and do know was for item 50 (20.8%). The highest percentage of incorrect answers in assessing reading was for item 11 (83.3%); in listening, item 22 (100%); in writing, item 41 (91.7%); and in speaking, item 46 (83.3%). The mean correct answers were 9 out of 15 items for reading, 7 out of 15 for listening, 7 out of 15 for writing, and 9 out of 15 for speaking. Participants' knowledge levels in both reading and speaking were categorized as moderate, showing slightly better results than in listening and writing, which were categorized as low. Based on the descriptive statistical analysis, it is concluded that teachers' knowledge of language assessment was moderate to low across the assessed skills. These results suggest a need for continuous teacher development programs and ongoing professional support for language assessment literacy to help teachers refine their assessment practices.

**Keywords:** English Teacher, Language Assessment, Skill Assesment, Teacher Knowledge

**INTRODUCTION**

Language teachers should conduct assessments not only to determine students' language proficiency and track their progress, but also to gain insights into how their own teaching practice impacts student learning. By conducting learning assessments, teachers can identify students' language proficiency levels, monitor their learning progress, evaluate the effectiveness of instructional practices, and provide valuable insights to improve teaching quality and learning outcomes. To ensure the effectiveness of these assessments, teachers are required to possess sufficient knowledge and competence in learning assessment and demonstrate language assessment literacy. Therefore, teachers must develop adequate Language Assessment Literacy (LAL).

Language Assessment Literacy (LAL) is defined as the essential knowledge, skills, and principles of assessment that teachers should master in order to implement effective language learning assessments (Tajeddin et al., 2022). Language Assessment Literacy is teacher cognition, referring to what teachers know, think, and believe about the teaching process. A teacher demonstrates language assessment literacy if they can select appropriate assessment methods, conduct assessments effectively, and use the results to enhance language teaching and learning in the classroom.

Assessment serves as a crucial reference for teachers to enhance the quality of their instructional practice. Stiggins (1995, p. 240) emphasized that assessment literate teachers know “what they are assessing, why they are doing it, how best to assess the skill, knowledge of interest, how to generate good examples of student performance, what can potentially go wrong with the assessment, and how to prevent that from happening”. Through regular assessment, language teachers can recognize and evaluate how

teaching strategies contribute to meaningful learning experiences for all. Lam (2015, p.240) also defined it as “teachers’ understandings and mastery of assessment concepts, measurement knowledge, test construction skills, principles about test impact, and assessment procedures which can influence significant educational decisions within a wider social context” (p. 172). Language teaching practice in schools is closely related to language learning assessment for two reasons. First, teaching and learning activities require evaluation to determine the effectiveness of teaching for teachers and students, namely: how teachers' teaching facilitates the improvement of students' language skills. Second, evaluation or assessment of learning provides real data for teachers to improve the quality of their teaching. Therefore, teacher cognition about language assessment literacy is how teachers use their beliefs and knowledge about learning assessment to improve student and teacher performance in achieving language learning goals.

In the context of education in Indonesia, every formal educational institution is required to conduct learning assessments to evaluate its teaching and learning activities over time. Therefore, at every level of school education in Indonesia, teachers are required to conduct end-of-semester assessments. The results of these assessments serve as benchmarks for schools and teachers to improve the quality of their teaching and learning. To ensure optimal results in language assessments, teachers need adequate knowledge and skills in language learning assessment. Teachers are required to understand the principles of language assessment as a literacy in the context of language learning.

Since Language learning assessment literacy plays a crucial role in the success of language teaching in schools, research on this literacy continues to grow rapidly. Many studies have attempted to identify the literacy levels of language teachers in learning assessment, especially in a foreign-language context. These studies indicate that teachers still have inadequate levels of assessment literacy, requiring guidance on conducting effective assessments (Anam & Putri, 2021; Tajeddin et al, 2022). In the Indonesian context, research on LAL remains very limited. Several recent studies on LAL in Indonesian schools indicate that teacher literacy levels are still inadequate, with their teaching practices (Anam & Putri, 2021; Mirizon, 2021). These studies examined teachers working in urban areas in Indonesia (Anam & Putri, 2021; Mirizon, 2021). Unfortunately, there has been no research targeting language teachers in schools outside urban areas.

Given the limited research on schools in urban areas, the researchers were interested in conducting the study in schools outside urban areas to determine whether their previous findings align with teachers' expectations in the field. Therefore, this study will explore language teachers' cognition and literacy in language learning assessment at the senior high school level in one regency in Jambi. It aimed to find answers to the following research question through the study: what are the general skill-based language assessment knowledge (LAK) levels of EFL teachers among High School English Teachers in a regency of Jambi province?

### **METHODS**

This study used a Convergent Parallel Mixed Methods research design. This research design was chosen because quantitative data served as a baseline for subsequent research. In contrast, qualitative data were collected and compared to yield more valid, comprehensive, and mutually reinforcing findings (Creswell & Plano Clark, 2018; Fraenkel, Wallen, & Hyun, 2015). This mixed-method approach is expected to provide a comprehensive picture of teachers’ knowledge and beliefs about learning assessment literacy and its application in schools.

The population consisted of English teachers who are members of the MGMP English Subject Teachers' Council in one of the regencies of Jambi Province. To obtain quantitative data, the sample was selected using the total population technique; all high school English teachers in the regency were used as the research sample. Twenty-four senior high school language teachers participated in the language assessment and responded to the items. The participants’ academic qualifications included three master’s graduates and twenty-one bachelor’s graduates. Demographic features and participant numbers are shown in the following table.

Table 1. Demographic features and the number of participants

No.	School	Number of Participants
1	SMAN I	3
2	SMAN 2	3
3	SMAN 3	1
4	SMAN 4	1
5	SMAN 6	5
6	SMAN 7	2
7	SMAN 8	2
8	SMAN I2	1
9	SMAN I3	1
10	SMAN I6	2
11	SMAN I7	1
12	SMAN I9	1
13	SMAN Terpadu Nurulssalam	1
Total	13	24

The questionnaire was used to obtain quantitative data and cover several aspects of learning assessment. It consists of 60 items with four constructs of the Language Assessment Knowledge Scale (LAKS) developed by Ölmezer Öztürk and Aydın (2019) as follows on table 2.

Table 2. Language Assessment Knowledge Scale (LAKS)

1. ASSESSING READING	
1	Asking learners to summarize the reading text is a way of assessing their reading skills.
2	When asking several questions about a reading text, all the questions are independent of each
3	Cloze test is used for assessing the main idea of the text.
4	In a reading exam, using a text learners have encountered before is not a problem.
5	One reading text is enough to be included in a reading exam
6	The language of the questions is simpler than the text itself.
7	Errors of spelling are penalized while scoring.
8	Taking vocabulary difficulty into consideration is necessary in assessing reading skills.
9	items.
10	The more items a reading text is followed, the more reliable it becomes
11	Using the same words in the correct option as in the text is not a problem.
12	Simplification of reading texts is avoided.
13	Reading texts in a reading exam include various genres (essay, article, etc.).
14	In top-down approach, assessment is on overall comprehension of the reading text.
15	Using ungrammatical distractors in multiple choice questions in a reading exam is a problem.
2. ASSESSING LISTENING	
16	Using reading texts for listening purposes poses a problem.
17	Including redundancy (e.g. what I mean to say is that ....) in a listening text poses a problem.
18	Any type of listening text is used for note-taking.
19	Spelling errors are ignored in scoring the dictation.
20	In discrete-point testing, comprehension is at the literal/local level.
21	Errors of grammar or spelling are penalized while scoring.
22	Using dictation diagnostically in assessing listening skills does not pose a problem.
23	A listening cloze test is a way of selective listening
24	Giving learners a transcript of the listening text is a valid way of assessing listening skills.
25	Phonemic discrimination tasks (e.g. minimal pairs such as sheep-ship) are examples of integrative
26	Dictation is a kind of discrete-point testing.
27	Intensive questions using lead-in questions are avoided in listening tests.
28	Asking learners to listen to names or numbers is called intensive listening.
3. ASSESSING WRITING	
29	In selective listening, learners are expected to look for certain information.
30	Giving two options to learners and asking them to write about one ensure reliable and valid
31	scoring.
32	Analytic scoring is used to see the strengths and weaknesses of learners.
33	The parts of a scoring scale and the scores in each part do not change for different levels of
34	learners.
35	When there is a disagreement between the scores of the two raters, they score the written work
36	again.
37	Learners are required to write about at least two tasks in the exam rather than one task.
38	Giving restrictive prompts/guidelines to learners for the writing task is avoided.
39	Giving learners an opinion and asking them to discuss it is a valid way of assessing their writing
40	skills.
41	Using visuals which guide learners for writing poses a problem.
42	Holistic scoring is used to see whether the learner is proficient or not at the end of the term

### Assessing Speaking

46	When the interlocutor does not understand the learner, giving that feeling or saying it poses a problem.
47	Giving learners one task is enough to assess speaking skills.
48	Interlocutors' showing interest by verbal and non verbal signals poses a problem.
49	When it becomes apparent that the learner cannot reach the criterion level, the task is ended.
50	Using holistic and analytic scales at the same time poses a problem.
51	Reading aloud is a technique used to assess speaking skills.
52	In interlocutor-learner interviews, the teacher has the chance to adapt the questions being asked.
53	In interactive tasks, more than two learners pose a problem.
54	The interlocutor gives the score when the learner is in the exam room.
55	In a speaking exam, production and comprehension are assessed together .
56	Asking learners to repeat a word, phrase or a sentence is a way of assessing speaking skills.
57	Discussion among learners is a way of assessing speaking skills.
58	A checklist is a means of scoring oral presentations in in-class assessment.
59	When the focus is to assess discourse, role plays are used.
60	In peer interaction, random matching is avoided.

Those 60 items with four constructs of the Language Assessment Knowledge Scale (LAKS) developed by Ölmezer Öztürk and Aydın (2019) have been completed by 24 participants. The data collection process preceded by a pilot test of the questionnaires and interview protocol. Once the pilot test and interview protocol results were valid, the questionnaires were distributed to all school teachers who were members of MGMP in one of the regencies of Jambi province. The quantitative data were analyzed statistically. The analysis included frequencies, Means, and comparisons of the results of the various questionnaire questions. This method generated an overview and descriptive statistics regarding the level of literacy in language learning assessments as perceived by teachers.

### FINDINGS AND DISCUSSION

The findings from the participants presented in Table 3 showed that the highest percentage of correct answers in assessing reading was for item 13 (100%): *Reading texts in a reading exam includes various genres (essay, article, etc.)*. It showed that the respondents were knowledgeable about this item and categorized it very well. The lowest percentage of correct answers was for item 11 (12.5%): *"Using the same words in the correct option as in the text is not a problem."* Only 3 participants answered correctly that using identical lexical items from the source text in the correct choice or option does not constitute an error. Most participants confirmed that it was acceptable to use the same words in the correct option/choice, and using identical wording from the text in the correct answer is permissible.

Besides, participants received 16.7% "do not know" responses for item 14, *the top-down approach, which assesses overall comprehension of the reading text*, indicating that respondents were still unsure about it. The mean or average of correct answers was 58.89%, and the mean of incorrect answers was 36.67%. These percentages show that, in general, participants performed better on this reading skill, with correct answers exceeding incorrect answers, and the 4.44% "do not know" responses indicate only a small number of uncertain responses.

Based on the statistical analysis, the median, or middle value, of the data arranged from smallest to largest shows that correct answers accounted for 58.3% of the median, and incorrect answers accounted for 37.5%. Since the correct answers are higher. Then, incorrect answers confirm that the majority of participants demonstrated a satisfactory understanding of this language skill. The mode of correct answers was 95.8; the most frequent score among participants was very high, and some participants performed excellently. The standard deviation (STDEV) of 32.80 for correct answers and 32.89 for incorrect answers indicates that many participants did well, others still struggle, and some are incorrect. Both correct and incorrect answers indicate high variability in the dataset, with values farther from the mean.



Figure 1. Assessing Reading

Table 3. Assessing Reading

Statistic	Correct answer	Incorrect answer	Do not know
AVERAGE	58.89	36.67	4.44
MEDIAN	58.3	37.5	4.1
MODE	95.8	4.2	4.2
STDEV	32.80	32.89	5.10

The results showed that the highest percentage of correct answers in the listening assessment was for item 30 (100%): *“In selective listening, learners are expected to look for certain information.”* All participants answered *True* to this item, indicating they achieved a relatively perfect score and demonstrating clear responses in accuracy and comprehension. The highest percentage of incorrect answers in the listening assessment was for item 22 (100%). *‘Phonemic discrimination tasks (e.g., minimal pairs such as sheep ship) are examples of integrative testing.’* The correct answer is *False*, but the participant chose this option, and all participants confirmed it was true (e.g., essay, role-playing, etc.). The participants performed very poorly on this item. The highest percentage of *‘Do not know’* responses was for item 28 (16.7%), which had the most uncertain answers.

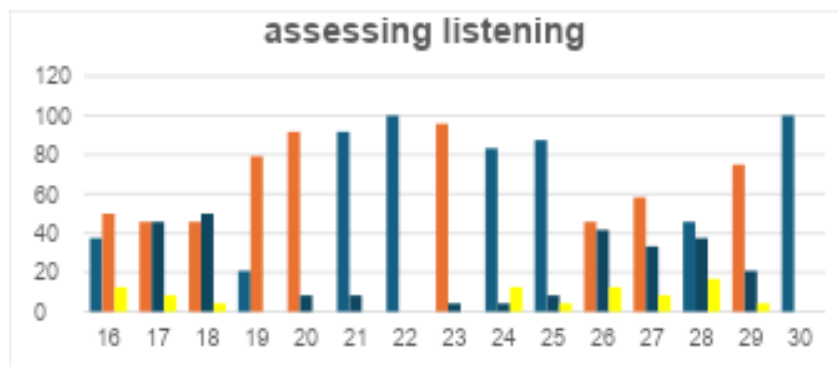


Figure 2. Assessing Listening

The mean of correct answers is slightly lower than that of incorrect answers,  $44.67 < 45.1$ , and 5.6 for do not know. On average, participants struggled with listening assessment tasks, as evidenced by their scores below 50%. The median, as the middle value of the correct answers, was also lower than the incorrect answers,  $41.7 < 47.9$ , and 4.2 for do not know. The most frequently occurring value for the correct answer mode was 45.8, and for the incorrect answer mode, 50. It showed that the most frequent pattern among participants was still dominated by incorrect answers and zero 'do not know' responses, as participants did not choose this option.

Table 4. Assessing Listening

Statistic	Correct answer	Incorrect answer	Do not know
AVERAGE	44.67	46.1	5.6
MEDIAN	41.7	47.9	4.2
MODE	45.8	50	0
STDEV	32.64	35.17	5.82

The standard deviation for correct answers was 32.64, and that for incorrect answers was 35.17, both of which were categorized as quite high, indicating a wide variation in participants' listening assessment performance. It implies that the data points were very different from each other; they had greater distances between them.



Figure 3. Assessing Writing

Table 5, which presents the results of the writing assessment, showed that the highest percentage of correct answers was for item 32 (91.7%): “Analytic scoring is used to see the strengths and weaknesses of learners.” Nearly all participants understood that the analytic scoring approach uses a rubric that breaks a task into individual components, providing detailed feedback on both strengths and weaknesses in assessing writing. The highest percentage of incorrect answers in assessing writing was for item 41 (91.7%): “In controlled writing, learners have the chance to convey new information.” It indicated that most participants did not know that controlled writing activities are structured exercises that guide learners to practice specific grammar and writing skills, focusing on form and accuracy. It changes to focus on sentence structure, grammar, punctuation, rewriting, and imitation. The highest percentage of "do not know" responses was for item 40 (16.75%): “Analytic scoring leads to greater reliability than holistic scoring in writing”. Even though nearly all participants knew that, to understand learners' strengths and weaknesses, analytic scoring is appropriate for language teachers, the percentage of participants who were unsure about this item was the highest response on this assessment of writing skills.

Table 5. Assessing Writing

Statistic	Correct answer	Incorrect answer	Do not know
AVERAGE	48.05	45.85	6.10
MEDIAN	50.00	50.00	4.10
MODE	20.80	8.30	0.00
STDEV	27.20	29.97	7.70

The statistic showed that the mean score for correct answers in writing is slightly lower than the median, and the mode is considerably smaller ( $48.05 < 50.00 < 20.80$ ), indicating a negative skew. Most participants achieved moderate to high scores, but some had low scores. The difference between the highest and lowest performance values is very significant, as the standard deviation is relatively large 27.20 indicating a wide range of performance levels among participants. Table 6. Assessing Speaking showed that the highest percentage of correct answers in assessing speaking was item 52 (95.8%): *In interlocutor-learner interviews, the teacher has the chance to adapt the questions being asked.*

Most participants confirmed that, in interviews, the interlocutor (teacher) can adjust the questions to the learner as needed. The highest percentage of incorrect answers in assessing speaking was item 46 (83.3%): *When the interlocutor does not understand the learner, giving that feeling or saying it poses a problem.* For this item, most participants gave incorrect responses. The correct answer is false because the teacher can create new words, simplify, or switch languages. The highest 'do not know' percentage in the speaking assessment was item 50 (20.8%): *Using holistic and analytic scales at the same time poses a problem.* The participants hesitated to give the correct answer to this item, as using both would provide a complete result.

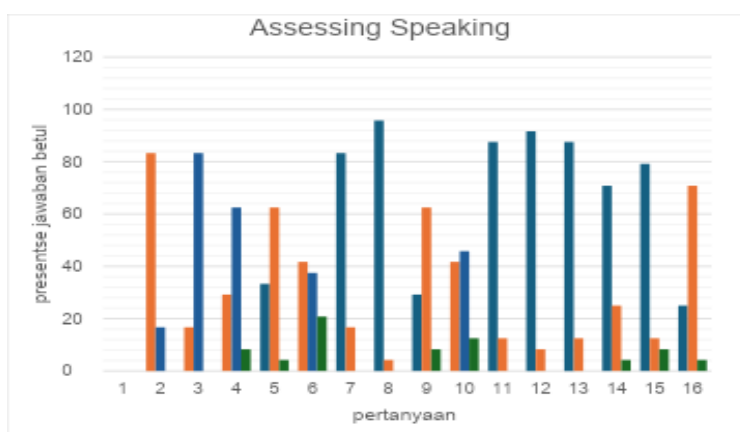


Figure 4. Assessing Speaking

The descriptive statistics presented in the table for the three data sets include the mean, median, mode, and standard deviation, indicating that most scores are relatively high. In contrast, a few low scores reduce the overall mean. The mean of correct was lower than both the median and the mode; the distribution of scores in this data set is negatively skewed.  $61.75 < 75 < 83.30$ , indicating that most scores were high, while a few low scores reduced the overall mean. It concluded that participants generally performed well, though several outliers performed worse.

Mode is the most frequently occurring value in the data. The most common score for correct answers was 83.30, and the most common score for incorrect answers was 12.50. It indicated that the most frequent pattern among participants was still significantly dominated by correct answers. The mode of 0 for the "do not know" response confirmed that no participant hesitated to answer those items. The large standard deviation confirmed that the scores were widely dispersed. It indicated that the data points do not cluster tightly around the mean but rather spread over a wide range of values. The statistical results for Productive skills revealed notable differences in participants' performance across writing and speaking. The assessment writing data show an average score of 48.05, a median of 50.00, and a mode of 20.80, indicating that the scores are moderate but with some low outliers, as reflected in the negative skewness. In contrast, the speaking data show higher performance, with an average of 61.94, a median of

75.00, and a mode of 83.30, suggesting that most participants achieved strong results. However, a few lower scores caused a positive skew. The standard deviation values for both writing (27.20) and speaking (27.76) indicate wide variability among participants.

Overall, Participants performed considerably better at assessing speaking than at assessing writing; both assessment skills show large variability, indicating that participants differ widely in ability. Some performed well, while others had significant difficulties. The difference between the highest and lowest performance values is very significant, and the most frequent pattern among participants was still slightly dominated by correct answers. The data show that the participants performed better on the reading assessment than on the listening assessment. The reading scores show both higher mean values and higher peak distributions, suggesting that participants possess stronger knowledge and greater familiarity with assessing reading skills. The high mode in the reading assessment results further suggests that a considerable number of participants demonstrated strong proficiency in this area. In contrast, the listening scores were lower and more evenly distributed, suggesting that this skill posed greater challenges for the participants. Despite these differences, both reading and listening assessment results showed wide variability, indicating that the participants' proficiency levels are not homogeneous across the group.

**Table 6. Assessing Speaking**

<b>Statistic</b>	<b>Correct answer</b>	<b>Incorrect answer</b>	<b>Do not know</b>
AVERAGE	61.94	33.34	4.72
MEDIAN	75.00	25.00	4.20
MODE	83.30	12.50	0.00
STDEV	27.76	25.58	6.06

## CONCLUSION

The findings reveal that participants performed better at assessing productive skills, especially speaking, than at assessing receptive skills. The higher scores in speaking and reading than in writing and listening suggest that participants are relatively more proficient in productive skills, while writing and listening are their weaker areas. Pedagogically, these outcomes emphasize the need for a more balanced instructional approach that gives equal attention to both receptive and productive aspects of language assessment knowledge. There is a need for continuous teacher training and development programs, along with ongoing professional support for language assessment literacy, to help teachers refine their assessment practices. A sustainable program that strengthens knowledge in assessing language skills through a process-oriented approach could be one of the topics to be discussed and emphasized in MGMP.

## REFERENCES

- Anam, S., Putri, N.V.W. How literate am I about assessment: Evidence from Indonesian EFL pre-service and in-service teachers? *English Review: Journal of English Education*, 9(2), 377-388. 2021. <https://doi.org/10.25134/erjee.v9i2.4347>
- Creswell, J. W., dan Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research. Third Edition*. Thousand Oaks, California: SAGE Publications, Inc. 2018
- Fraenkel, J. R., Wallen, N. E., dan Hyun, H. H. (2015) *How to Design and Evaluate Research in Education: Ninth edition*. New York: McGraw Hill Education.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32 (2), 169-197.
- Mirizon, S. Teachers of English assessment literacy: unveiling what they know and do. *IRJE: Indonesian Research Journal in Education*, 5(1), 2021, 124-141. <https://doi.org/10.22437/irje.v5i1.12834>
- Ölmezer-Öztürk, E., & Aydın, B. (2019). Investigating language assessment knowledge of EFL teachers. *Hacettepe University Journal of Education*, 34(3), 602-620. doi: 10.16986/HUJE.2018043465. <https://dergipark.org.tr/tr/download/article-file/774576>
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77 (3), 238-245.
- Tajeddin, Z., Khatib, M., Mahdavi, M. Critical language assessment literacy of EFL teachers: scale construction and validation. *Language Testing*, 1-30, 2022. <https://doi.org/10.1177/02655322211057040>