

HOW LITERATE AM I ABOUT ASSESSMENT: EVIDENCE FROM INDONESIAN EFL PRE-SERVICE AND IN-SERVICE TEACHERS

Syafi'ul Anam

*English Department, Faculty of Languages and Arts
State University of Surabaya, Surabaya, Indonesia
Email: syafiul.anam@unesa.ac.id*

Nanin Verina Widya Putri

*Language and Literature Study Program, Postgraduate School
State University of Surabaya, Surabaya, Indonesia
Email: nanin.19029@mhs.unesa.ac.id*

APA Citation: Anam, S. & Putri, N. V. W. (2021). How literate am I about assessment: Evidence from Indonesian EFL pre-service and in-service teachers. *English Review: Journal of English Education*, 9(2), pp. 377-388. <https://doi.org/10.25134/erjee.v9i2.4374>

Received: 15-02-2021

Accepted: 02-04-2021

Published: 15-06-2021

Abstract: Language Assessment Literacy (LAL) for teachers is a pivotal content area which teachers must have for their professional development. Despite increasing interest in language assessment, research focusing on LAL of pre-service and in-service teachers, especially in Indonesian EFL context, has yet to receive due attention. Nevertheless, understanding pre-service and in-service teachers' literacy of classroom assessment is essential for the quality of classroom assessment practice, hence the improvement of language learning and teaching. This study aimed to investigate Indonesian EFL pre-service and in-service teachers' literacy of classroom assessment. Using a quantitative approach with ex-post facto research design, this study examined the extent of assessment literacy among EFL pre-service and in-service teachers. Participants were 60 English department students who have done teaching practicum courses and 41 in-service teachers in the Indonesian province of East Java. The finding revealed that the pre-service and in-service teachers perceived a moderate level of assessment literacy. Further, the two groups of participants demonstrated a statistically significant difference in overall assessment literacy, as well as in two of its dimensions, namely administering and scoring assessment results and alternative assessment. The findings of this study could shed light to the understanding of EFL pre-service and in-service teachers' literacy of classroom assessment.

Keywords: *assessment literacy; pre-service teachers; in-service teachers; classroom assessment*

INTRODUCTION

Assessment is acknowledged as an inseparable part of the teaching and learning process (Ashraf & Zolfaghari, 2018; Bijsterbosch, Béneker, Kuiper, & van der Schee, 2019; Cheng & Fox, 2017; Popham, 2009). Without assessment, teachers will be difficult to identify the extent of students' abilities, and students have no information about their current performance. It implies that assessment results enable teachers and students to understand how teaching and learning take place. However, to implement assessment practice successfully, teachers need to understand the basic principles of classroom-based assessment. More importantly, they need to know how to employ their knowledge and principles of assessment in classroom practice. To this end, teachers across the educational context should be aware of their assessment ability and knowledge in order to develop students' mastery of designated lessons.

In language classrooms, English language teachers' ability to implement assessment is considered as one of the most crucial skills. The way teachers practice assessment in the classroom can influence students' learning quality. As Umam and Indah (2020) pointed out, how teachers assess students' performance has a meaningful impact on everything in the classroom. Similarly, Zulaiha and Mulyono (2020) have highlighted that the success of assessment practice could influence students' achievement. For this reason, teachers need to have adequate knowledge and skill of assessment to support classroom assessment practice (Jeong, 2013; Koh, Burke, Luke, Gong, & Tan, 2018; Popham, 2009). The skill and knowledge of assessment is typically called assessment literacy, which is recognized as a crucial part of teachers' professional development (Abell & Siegel, 2011; Engelsen & Smith, 2014; Inbar-Lourie, 2013; Xu & Brown, 2016).

Language Assessment Literacy (LAL) is generally defined as 'basic understanding of sound assessment practice and the ability to apply that knowledge to measure language learning in different contexts' (Yan, Zhang, & Fan, 2018, p. 158). Teachers identified as assessment-literate are able to determine *what* assessment method they have to implement, *how* they assess, and *when* they have to assess so that teachers can gain information about students' performance (Jeong, 2013; Stiggins, 1999). Specifically, language teachers who are assessment-literate refer to those who have knowledge and abilities of assessment integrated with language-specific competencies (Inbar-Lourie, 2008). They have general assessment knowledge, understand the purposes of language assessment, know appropriate methods, and are capable of interpreting assessment practice (Inbar-Lourie, 2008; Zulaiha, Mulyono, & Ambarsari, 2020).

LAL has three main components, including skills, knowledge, and principles (Davies, 2008; Deygers & Malone, 2019; Fulcher, 2012). Consequently, language teachers are expected to master the concept of assessment theory which consists of knowledge of classroom assessment combined with language pedagogy (Inbar-Lourie, 2008; Yan et al., 2018; Zolfaghari & Ahmadi, 2016). It is supported by Harding and Kremmel's (2016) statement that language teachers, who frequently conduct classroom assessment, need to be "conversant and competent in the principles and practice of language assessment" (p.415). When they are assessment-literate, they are able to monitor students' progress, know students' current improvement, and promote learning (Earl, 2013). To help students gain their learning goals, teachers should be aware of improving their assessment literacy so that they can implement appropriate methods in classroom assessment (Xu & Brown, 2016). Moreover, since they have learned about language assessment in pre-service education, they are expected to create meaningful assessment for the improvement of students' learning.

However, recent studies have showed that second/foreign language teachers still lack LAL (Lam, 2019; Nemati, Alavi, Mohebbi, & Masjedlou, 2017; Popham, 2001; Qian, 2014; Tsagari & Vogt, 2017). For example, a study by Nemati et al., (2017), investigating the assessment ability of Iranian English teachers, demonstrated that teachers have inadequate assessment knowledge and training in writing skill. Another study in Hong Kong context with

similar results comes from Qian (2014). It examined school-based English language assessment, and the result revealed English teachers' low ability in marking skill when assessing learners' speaking ability. The reasons why teachers have inadequate LAL vary, such as contextual factors (i.e. assessment policy, stakeholders, resources, and constraints) and experiential factors (i.e. assessment development and use) (Crusan, Plakans, & Gebril, 2016; Vogt & Tsagari, 2014; Yan et al., 2018). The studies underscore that language teachers still face many challenges in assessment practice.

Despite the significance of assessment literacy for language teachers, it has yet to receive much attention from EFL Indonesian researchers (Zulaiha, et al., 2020). A recent study by Zulaiha et al (2020), one among very few, investigated EFL teachers' assessment literacy through their perception in the Indonesian context. They asserted that EFL teachers in Indonesia were assessment-literate and aware of assessment principles. However, they asserted that they did not explore the quality of teachers' assessment practice. Another study by Zulaiha and Mulyono (2020) surveyed training needs of assessment literacy among 147 junior high school EFL teachers. The finding showed that teachers expected to have abilities to select tests for use, develop test specifications, and develop test tasks and items. The other study of in-service teachers by Umam and Indah (2020) revealed that in-service English teachers have poor level of assessment literacy, which contradicts with the finding of Luthfiyyah and Basyari's (2020) study, revealing a moderate literacy level of EFL secondary teachers.

The previous studies of LAL conducted in Indonesian and overseas are predominantly focused on in-service English language teachers (e.g. Koh et al., 2018; Lam, 2019; Nemati et al., 2017; Tsagari & Vogt, 2017; Yan et al., 2018; Zulaiha & Mulyono, 2020; Zulaiha et al., 2020). Consequently, empirical evidence about pre-service teachers' knowledge and skills of assessment is very limited, let alone from Indonesian context. The evidence is important as the information of their level of assessment literacy. When they have poor literacy of assessment, they will 'be less likely to help students attain higher levels of academic achievement' (Herrera & Macías, 2015, p. 304). They should have more opportunities in assessment training during their education program to increase their professional

development (Lopez & Bernal, 2009). Pre-service teachers also need to improve their knowledge and skills to allow them design the task and implement authentic assessment to prepare real-world practice properly. With adequate assessment literacy, they can utilize assessment data to support instructional practices and employ classroom-based assessment effectively (Popham, 2009).

Regarding the previous studies already reviewed, examining assessment literacy of pre-service and in-service teachers in Indonesian EFL classroom was worth doing. Extending the research focus to Indonesian EFL classroom benefits assessment practice as well as teaching learning processes (Edwards, 2017; Willis, Adie, & Klenowski, 2013). For these reasons, the present study aimed to investigate the extent to which pre-service and in-service Indonesian EFL teachers rated their level of classroom assessment literacy, and whether they differed in classroom assessment literacy.

METHOD

The study employed a quantitative approach with ex-post facto research. A total of 101 participants (42 male and 59 female) were recruited through convenience sampling, meaning that all the participants took part on the basis of their availability (Weathington, Cunningham, & Pittenger, 2010). The participants comprised 60 English department students from one state university and two private universities who have done teaching practicum courses and 41 in-service teachers in East Java. All participants reported having taken assessment courses when they did their undergraduate degree in English education, and in-service teachers also reported learning assessment when attending teacher trainings. All the in-service teachers had a range of teaching experience between 1 to 18 years. Most of the in-service teachers taught in state secondary schools (64%) and the remainders taught in private secondary schools.

The participants' assessment literacy was measured using Language Assessment Literacy Scale (LALS). The instrument was administered in English through Google form. It gauged the participants' perceived literacy in English classroom assessment, comprising the dimensions: (1) *designing test/assessment*, (2) *administering and scoring assessment results*, (3) *alternative assessment*, (4) *validity and reliability issues*, (5) *using assessment results to make decisions*. In total, the instrument had 25 items,

with a six-point Likert scale ranging from 1 (not at all skilled) to 6 (highly skilled). The items were adapted from those developed by Zhang and Burry-Stock (1997) and Jarr (2012), as well as prompted by important literature on assessment literacy (Coombe, Vafadar, & Mohebbi, 2020; Mertler, 2003; National Council on Measurement in Education (NCME), American Federation of Teachers (AFT), 1990). To ensure content validity, the scale then was reviewed by two experts in language assessment, and necessary revisions were made accordingly. The researchers calculated reliability coefficients, for the 25-item LALS (.93) and its five dimensions: designing test/assessment instrument (.80), administering and scoring assessment results (.86), alternative assessment (.84), validity and reliability issues (.88), and using assessment results to make decisions (.90). The instrument' reliability coefficients suggested satisfactory consistency for research purposes.

The survey data were then analysed quantitatively using descriptive and multivariate analysis of variance (MANOVA). The descriptive statistics comprising mean and standard deviation as the summary statistics of the data set to determine the extent of pre-service and in-service English teachers' assessment literacy. The descriptive statistics of the individual items in each subscale and score of the five dimensions were analysed and presented to display the level of assessment literacy the in-service and pre-service teachers had, whether low, moderate, or high. To allow for such categorization, the participants' responses were grouped into three levels of assessment literacy using equal cut-off points on the scale: low (1.00-2.70), moderate (2.71-4.40) and high (4.41-6.00). Additionally, the one-way MANOVA was employed to examine whether there were differences between the pre-service and in-service teachers in assessment literacy, as well as its five dimensions. This multivariate test was preferred instead of multiple independent sample t-tests as this could reduce the possibility of Type 1 error (Pallant, 2016).

FINDINGS AND DISCUSSION

Level of literacy in classroom assessment among pre-service and in-service teachers

In order to address the first research objective about the level of literacy in EFL classroom assessment for Indonesian pre-service and in-service teachers, means and standard deviations the LALS were calculated. The participants'

literacy in EFL classroom assessment was analysed from five dimensions, as displayed in Figure 1.

The figure displayed a moderate level of assessment literacy for all participants, regardless of whether they were in-service teachers or pre-service EFL teachers ($M=4.37$ and $M=4.03$ respectively), suggesting that the participants had good understanding and skills related to EFL classroom assessment. Specifically, in-service teachers, on average, perceived themselves highly literate in administering and scoring assessment results ($M=4.55$) and using the results to make

decisions ($M= 4.60$). On the other hands, pre-service teachers only showed lower literacy of the two aspects of assessment, $M= 4.12$ and $M= 4.12$ respectively. Meanwhile, both pre-service and in-service teachers only reported moderate level of literacy in the three other aspects of assessment, designing test/assessment instrument, alternative assessment, and validity and reliability issues, with the mean scores ranging from 3.96 to 4.38. Compared to other aspects of assessment, the participants seem to be less confident with their ability in determining validity and reliability to ensure test quality.

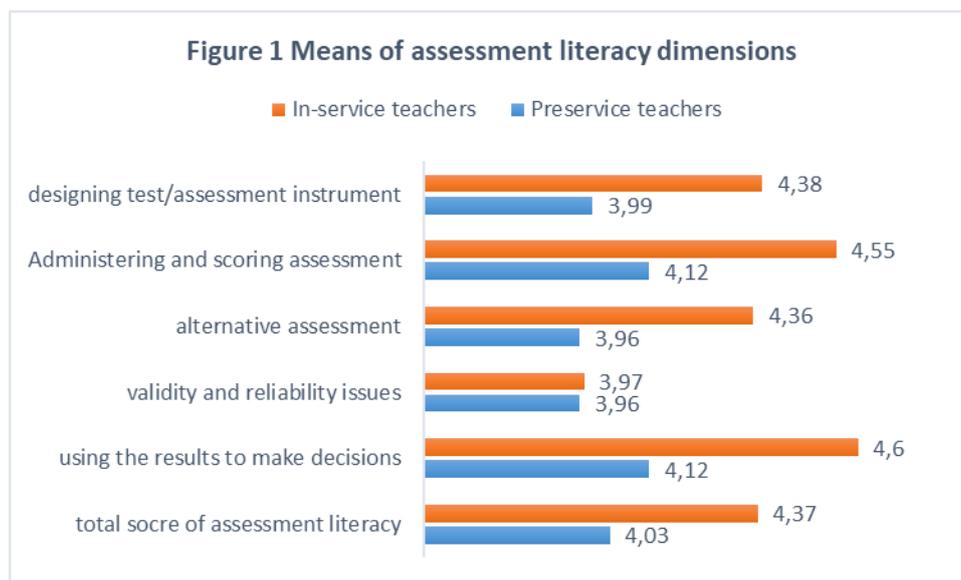


Figure 1. Means of assessment literacy dimation

A closer inspection of the individual items for LALS indicated that pre-service and in-service EFL teachers felt capable of doing classroom assessment, with a mean range for all items measuring literacy assessment from 3.68 to 4.55 for pre-service teachers and from 3.71 to 4.74 for in-service teachers (Table 1). The pre-service and in-service teachers shared similarities in two of the four highest rated items on the ability in using assessment results to improve teaching-learning process ($M= 4.25$; $M=4.74$ respectively) and the ability in ensuring fairness in assessing students ($M= 4.55$; $M= 4.62$ respectively). The two groups also rated their ability lowest on two items measuring literacy in determining the validity of objective tests ($M= 3.72$ for pre-service teachers; $M=3.74$ for in-service teachers) and performance assessment ($M=3.70$ for pre-service; and $M=3.79$ for in-service teachers). However, the two groups of participants also differed in other highest and

lowest rated items. The two other items rated highest by the pre-service teachers were related to the ability to give feedback to students ($M=4.20$) and the ability to score students' responses in listening and reading assessment ($M=4.18$). Meanwhile, other highest ratings by in-service teachers were found on two items on the perceived ability to use formative assessment to monitor student learning ($M=4.58$) and the ability to use various test items for listening and reading ($M=4.67$). Further inspection of the items also showed that the pre-service teachers rated their literacy lowest on the items measuring ability to develop blueprint/test specifications and using portfolio while their in-service teacher counterparts rated their literacy lowest on the items measuring ability to use reflective journal/learning log and to analyse items for better quality test.

Table 1. Means and standard deviation of the assessment literacy items

No	Item	Pre-service		In-service	
		Mean	SD	Mean	SD
Designing test/assessment					
1	Developing blueprint/test specifications	3.68	.79	4.33	.61
2	Developing various performance assessment based on learning objectives	4.08	.94	4.50	.67
3	Developing various test items for receptive skills based on learning objectives	4.10	.99	4.43	.67
4	Constructing a model answer for scoring essay questions	4.03	.94	4.38	.62
5	Developing analytic and holistic scoring rubrics	4.05	1.03	4.29	.63
Administering and scoring assessment results					
6	Implementing various performance assessment of speaking & writing skills	4.03	.99	4.48	.63
7	Implementing various test items for listening & reading	4.12	.97	4.67	.61
8	Scoring students' task performance	4.17	.99	4.52	.63
9	Scoring students' responses of receptive skills	4.18	.99	4.55	.70
Alternative assessment					
10	Using portfolio assessments	3.80	1.13	4.38	.54
11	Using self-assessment	4.13	1.06	4.50	.59
12	Using peer-assessment in English class	4.03	.66	4.50	.59
13	Using formal observation in English class	3.87	1.16	4.45	.63
14	Using informal, continuous, non-test type of assessment	3.93	.97	4.52	.63
15	Using reflective journal/learning log	4.02	1.20	3.81	.71
Validity and reliability issues					
16	Ensuring fairness in assessing students	4.55	1.11	4.62	.62
17	Determining the reliability of objective tests	4.02	1.02	4.12	.67
18	Determining the validity of objective tests	3.72	.89	3.74	.77
19	Determining the reliability of performance assessment	3.92	1.11	3.83	.76
20	Determining the validity of performance assessment	3.70	.79	3.79	.87
21	Conducting item analysis	3.92	1.06	3.71	.80
Using assessment results to make decisions					
22	Using assessment results when making decisions about individual students	4.03	1.01	4.57	.63
23	Using formative classroom assessments to monitor student learning	4.00	1.09	4.58	.70
24	Using assessment results to improve teaching-learning process	4.25	1.03	4.74	.66
25	Giving feedback to students based on information from tests/assessment to improve students' performance	4.20	1.02	4.52	.67

Differences in literacy assessment between pre-service and in-service teachers

One-way Manova was conducted to determine whether there were statistically significant differences in the mean scores of five dimensions of assessment literacy for pre-service and in-service teachers. Prior to running the Manova, preliminary analyses of the assumption and outlier were conducted. Using the outlier labelling rule of Hoaglin, Iglewicz, and Tukey (1986), there were no univariate outliers identified in either pre-service and in-service group in the scores of five assessment literacy dimensions. Multivariate

outlier detection was also performed using Mahalanobis D^2 at $p < .001$. Case no. 62 was spotted as multivariate outlier as its p value was less than .001 and then deleted accordingly. The visual assessment of histograms also suggested approximately normal distribution of the scores. As indicated in scatterplots, no curvilinear shapes were found and each pair of the scores in the two groups were linearly related. The assumption of homogeneity of variance-covariance was violated, the Box's M test value of 66.79, at $p = .00$ which was significant. However, with the group sizes are over 30, Manova is robust to such violation. Due

to the violation, Pillai's trace statistic was reported instead of other statistics (Tabachnick & Fidell, 2001, p. 252).

The results of the analysis revealed a statistically significant difference in overall assessment literacy scores of the pre-service and in-service teachers, $F(5, 93) = 3.39, p < .005$; Pillai's Trace = 3.388; partial $\eta^2 = .154$. The partial eta squared suggested that the effect of teacher category (pre- service and in-service) on the combined scores of the assessment literacy was only 15%, which was small. The result means that in-service teachers were perceived more literate in classroom assessment than their pre-service teacher counterparts. Following up the significant difference, multiple F-tests were performed to determine where the difference lay between the two groups of participants. A Bonferroni correction was used to account for multiple tests being run; therefore, statistical significance at $p \leq .01$ is accepted (see Tabachnick & Fidell 2007, p. 270). As Table 2 displays, a significant difference existed in two dimensions of literacy assessment, namely

administering and scoring assessment results ($F(1, 97) = 6.664; p = .01$) and alternative assessment ($F(1, 97) = 6.575; p = .01$). The effect size of teacher category on *administering and scoring assessment results* and *alternative assessment* was relatively small, partial $\eta^2 = .064$ and partial $\eta^2 = .063$ respectively. The mean scores suggest that in-service teachers reported higher perceived literacy in *administering and scoring assessment* ($M = 4.55, SD = .54$) and *alternative assessment* ($M = 4.36, SD = .45$) than pre-service teachers ($M = 4.12, SD = .82; M = 3.96, SD = .77$) did. However, no significant difference was observed in the other dimensions of assessment literacy: using assessment results to make decisions ($F(1, 97) = 4.347; p = .04$; partial $\eta^2 = .043$), designing test/assessment instrument ($F(1, 97) = 3.452; p = .06$; partial $\eta^2 = .034$), and validity and reliability issues ($F(1, 97) = 0.42; p = .83$; partial $\eta^2 = .00$). The nonsignificant differences suggest that the pre-service and in-service teachers didn't differ in the three areas, and the effect of the teacher category was very small.

Table 2. *Differences in assessment literacy between pre-service and in-service teachers*

Variable	Pre-service	In-service	Partial eta squared (η^2)		
	<i>M(SD)</i>	<i>M(SD)</i>	<i>F(1, 99)</i>	<i>p</i>	
Designing test/assessment	3.99 (.73)	4.38 (.47)	3.452	.06	.034
Administering and scoring assessment results	4.12 (.82)	4.55 (.54)	6.664	.01	.064*
Alternative assessment	3.96 (.77)	4.36 (.45)	6.575	.01	.063*
Validity and reliability issues	3.96 (.77)	3.97 (.61)	.042	.84	.000
Using assessment results to make decisions	4.12 (.94)	4.60 (.51)	4.347	.04	.043

Note: *M* (mean scores), *SD* (standard deviation), * $p < .01$

In view of less evidence about assessment literacy in EFL classrooms, this study attempted to examine how EFL Indonesian pre-service and in-service teachers perceived their knowledge and skills of assessment in language classroom. In general, the findings reveal that the two groups of teachers claimed having quite good understanding and skills related to EFL classroom language assessment. This suggests that the teachers are quite conversant and competent in practicing EFL language assessment. This empirical evidence supports the finding of a study by Luthfiyyah and Basyari's (2020), showing a moderate degree of assessment literacy among Indonesian secondary teachers. A similar finding of a study by Wise et.al (1991) also indicated pre-service teachers were quite skilled in assessment and

measurement. The perception of being knowledgeable and skilled in classroom assessment can possibly be attributed to university coursework/training the participants have taken or trial and error in the classroom. The finding of the present study is certainly worth noting as teachers are required to possess adequate knowledge and skills of assessment to support assessment practice in EFL classroom (Deluca & Klingerb, 2010; Jeong, 2013; Koh et al., 2018; Popham, 2009). With such a literacy level, the participants seem to know various assessment methods they can use appropriately, to be capable of monitoring students' learning progress, and make use of the results for the improvement of students' performance in classrooms (Jeong, 2013; Stiggins, 1999).

Another finding shows that both pre-service and in-service teachers rated their literacy highest on the item about the ability in ensuring fairness in assessing students. This implies that they have conviction and strong intention to be fair in their assessment practice. The finding accords with that of studies conducted by Phillips (2002) and Tierney (2010) about teachers' high awareness of students' perception of fairness in classroom assessment practice. The awareness is crucial for teachers to have sound assessment practices and demonstrate school accountability for teaching-learning processes. In addition, with the increasing demands for recognizing diverse students and data-driven decision making, maintaining this principle of assessment is inevitable (Campbell, 2013; Tierney, 2013). Teachers need to be aware of the importance of assessing their students fairly, especially in performance-based or authentic assessment. Such awareness helps teachers conduct assessment without any forms of favouritism or bias. Bias is likely to happen in classroom assessment which frequently uses subjective measurement and involves personal contacts. For example, teachers might favour students who are likeable or well-behaved but mark down those who have negative or annoying traits. Teachers' conviction of being fair, as found in this study, can prevent such favouritism, while at the same time the teachers, as well as the school, need to have measures, like using the same assessment tasks or well-defined criteria, to ensure fairness in assessment.

The other finding that both in-service and pre-service teachers rated items lowest on the perceived ability to determine validity of performance and objective tests and to conduct item analysis to ensure test quality is no surprise. A previous study by Brookhart (2001) demonstrated a similar finding that teachers lacked expertise in quality test construction and valid assessment procedure. It is likely that assessment-related trainings or courses the participants took didn't give adequate attention to issues related to validities and item analysis. The participants might also be less interested in validity evaluation because it is considered complicated and only dealing with psychometric properties of assessment. This finding is certainly concerning as validity is the main principle of language assessment (Bonner, 2013; Brown, 2010). The absence of the condition for validity will affect the quality of the instrument designed and used by teachers. Therefore, raising teachers' awareness of validity evaluation is crucial in

ensuring the accuracy of an assessment and inference made by teachers based on the results of assessment. This is because teachers and schools currently have greater autonomy making decisions about students based on assessment, such as grades, promotions, and graduation (Cirocky & Anam, 2021), following the Ministry of Education and Culture's recent policy of "*Merdeka Belajar*" (Freedom of Learning) (Circular no 1, 2020). Through trainings or independent learning, teachers should be made more aware of the basic tenets of validity when constructing classroom assessment. For example, the teachers might ask their peers to check the content validity of the instrument in classroom assessment. They do not need to conduct any complex statistical analyses for validation process, typically required in that of high stake testing. The teachers also have the knowledge and ability of designing a rubric with well-defined score categories based on instructional objectives in their language classroom. Additionally, the increasing attention to English as a lingua franca (ELF) demands more effort from language teachers also need to learn more about validity and test constructs (Lazaraton, 2017).

The result of the Manova demonstrates that in-service teachers had higher overall assessment literacy score, as well as the scores of administering and scoring assessment and alternative assessment dimensions, than their pre-service teacher counterparts. This suggests that in-service teachers have better mastery of assessment theory and combine it with language pedagogy so that they are conversant in administering and scoring, as well as using various alternative assessment, like portfolios, learning log, peer assessment or self-assessment, suitable for their students. This finding lends support to that of Plake, Impara and Fager's (1993) study that pre-service teachers exhibited somewhat weaker skills in classroom assessment than in-service teachers did. In-service teachers' higher literacy found in this study might result from experiential factors, like experiences of developing and using assessment in their own classroom, teaching experience, teacher trainings, learning from peers in English teacher association (Crusan et al., 2016; Vogt & Tsagari, 2014; Yan et al., 2018). Such experiential factors might raise their awareness of improving their assessment literacy so that they design and use appropriate methods of classroom assessment. Pre-service teachers do not have such experiential factors. Contextual factors, like the policy of assessment in *Merdeka*

Belajar and resources provided by schools, might also make the difference in the degree of assessment literacy between in-service and pre-service teachers. With the policy in place, teachers inevitably have to develop their knowledge and skills in classroom assessment and to implement them at school.

The current study's findings have pedagogical implications for secondary school teachers, school, and teacher educators. Firstly, teachers have to be actively engaged in various activities of professional development specifically focusing on classroom assessment. Such engagement will contribute to improving their knowledge and skills in assessment, which will then help students achieve learning goals and improve language skills. With their autonomy in hand, the teachers will be able to improve their literacy assessment through self-initiated or school-mandated professional development. Secondly, school management should facilitate teachers to be actively engaged in teacher trainings or other activities which can enhance their knowledge in assessment. Requiring teachers to be active in English teacher association or providing them with rich resources of classroom assessment can be effective forms of developing their assessment literacy. They also need to be ensured that school supports their autonomy to exercise their skills of classroom assessment because they have extensive knowledge of their students and classroom. School can also create collaborative environments in which teachers can share ideas and make joint work related to designing and administering assessment. Thirdly, teacher educators need to review the contents of assessment coursework by giving adequate attention to validity related issues and awarding more semester hours. They also should provide student teachers more practical experiences in developing quality assessment instruments and ask them to research real practices of classroom-based assessment so that the course content is relevant to the assessment realities of the classroom.

CONCLUSION

The present study concludes that the EFL pre-service and in-service teachers' literacy of classroom assessment were at moderate level, suggesting quite good understanding and skills in classroom assessment. Specifically, the participants perceived themselves less capable in developing test specification, determining validity of test instrument, and performing item analysis,

but they felt more capable of using assessment results to improve English teaching and of giving feedback to students' work or performance. In addition, and there is a significant difference between assessment literacy scores of the pre-service and in-service teachers. Those findings should be read with caution in terms of generalizability because the participants of this study were selected through convenience sample. Consequently, future research should involve more participants randomly to allow for generalizability of the results to other contexts. In addition, the findings of the study were drawn from self-report data only indicating perceived assessment literacy, which is although useful for diagnostic information. So, further studies are expected to measure the knowledge and skills of classroom assessment through a test which is able to allow for a behavioural measure of literacy. The use of such an instrument measures actual assessment literacy within a classroom-based assessment context. Regardless of these limitations, the results of this study provide insights into the notion of language classroom assessment, as well as direct implications for curriculum development for pre-service and in-service teacher education and trainings in similar contexts.

REFERENCES

- Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In *The professional knowledge base of science teaching* (pp. 205–221). Springer.
- Ashraf, H., & Zolfaghari, S. (2018). EFL teachers' assessment literacy and their reflective teaching. *International Journal of Instruction*, 11(1), 425–436.
<https://doi.org/10.12973/iji.2018.11129a>
- Bijsterbosch, E., Béneker, T., Kuiper, W., & van der Schee, J. (2019). Teacher Professional Growth on Assessment Literacy: A Case Study of Prevocational Geography Education in the Netherlands. *Teacher Educator*, 54(4), 420–445.
<https://doi.org/10.1080/08878730.2019.1606373>
- Bonner, S. M. (2013). Validity in classroom assessment: Purposes, properties, and principles. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 87–106). London: Sage.
<https://doi.org/10.4135/9781452218649.n6>
- Brookhart, S. M. (2001). The Standards and Classroom Assessment Research. In *Paper presented at the annual meeting of the American Association of*

- Colleges for Teacher Education* (p. 15). Dallas, TX.: ERIC Document Reproduction Service No. ED451189.
- Brown, H. D. (2018). *Language Assessment Principles and Classroom Practice* (Third). New York: Pearson Education.
- Campbell, C. (2013). Research on teacher competency in classroom assessment. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 71–84). <https://doi.org/10.4135/9781452218649.n5>
- Cheng, L., & Fox, J. (2017). *Assessment in the Language Classroom. Assessment in the Language Classroom*. <https://doi.org/10.1057/978-1-137-46484-2>
- Cirocki, A., & Anam, S. (2021). “How much freedom do we have?” The perceived autonomy of secondary school EFL teachers in Indonesia. *Language Teaching Research*. <https://doi.org/10.1177/13621688211007472>
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Correction to: Language assessment literacy: what do we need to learn, unlearn, and relearn? (*Language Testing in Asia*, (2020), 10, 1, (3), 10.1186/s40468-020-00101-6). *Language Testing in Asia*, 10(1), <https://doi.org/10.1186/s40468-020-00114-1>
- Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers’ knowledge, beliefs, and practices. *Assessing Writing*, 28, 43–56. <https://doi.org/10.1016/j.asw.2016.03.001>
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
- Deygers, B., & Malone, M. E. (2019). Language assessment literacy in university admission policies, or the dialogue that isn’t. *Language Testing*, 36(3), 347–368. <https://doi.org/10.1177/0265532219826390>
- Earl, L. M. (2013). Using Assessment to Motivate Learning. In *Assessment as learning: Using classroom assessment to maximize student learning* (pp. 77–86). Corwin Press.
- Edwards, F. (2017). *The development of summative assessment literacy: An exploration of the experiences of beginner secondary science teachers in New Zealand*. The University of Waikato.
- Engelsen, K. S., & Smith, K. (2014). Assessment literacy. In *Designing assessment for quality learning* (pp. 91–107). Springer.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In *Handbook of Second Language Assessment* (pp. 413–427). Germany: De Gruyter Mouton. <https://doi.org/10.1515/9781614513827-027>
- Herrera, L., & Macías, D. F. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colomb. Appl. Linguist. J*, 17(2), 302–312. <https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09>
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396), 991–999. <https://doi.org/10.1080/01621459.1986.10478363>
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402. <https://doi.org/10.1177/0265532208090158>
- Inbar-Lourie, O. (2013). Guest Editorial to the special issue on language assessment literacy. *Language Testing*, 30(3), 301–307. <https://doi.org/10.1177/0265532213480126>
- Jarr, K. A. (2012). *Education Practitioners’ Interpretation and Use of Assessment Results*. Iowa Research Online. Theses and Dissertations, University of Iowa. <https://doi.org/10.17077/etd.35vh2oc1>
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, 30(3), 345–362. <https://doi.org/10.1177/0265532213480334>
- Koh, K., Burke, L. E. C. A., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 22(3), 264–288. <https://doi.org/10.1177/1362168816684366>
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81(1), 78–89. <https://doi.org/10.1016/j.system.2019.01.006>
- Lazaraton, A. (2017). Qualitative methods of validation. In E. Shohamy, I. G. Or, & S. May (Eds.), *Encyclopedia of language and education: Language testing and assessment* (pp. 211–224). Cham, Switzerland: Springer.
- Lopez, A., & Bernal, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile Issues in Teachers Professional Development*, 11(2), 55–70. Retrieved from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-07902009000200005
- Luthfiyyah, R., & Basyari, I. W. (2020). EFL secondary teachers’ assessment literacy: Assessment conceptions and practices. *Journal*

- on *English as a Foreign Language*, 10(2), 402–421. <https://doi.org/10.23971/jefl.v10i2.2101>
- Mertler, C. A. (2003). Pre-service versus in-services teachers' assessment literacy: Does classroom experience make a difference? *Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH*, 2–27. <https://doi.org/10.1.1.556.4217&rep=rep1&type=pdf>
- Ministry of Education and Culture. (2020). *Freedom of Learning in determination of student graduation and student admission* (Circular of No. 1, 2020). Ministry of Education and Culture.
- National Council on Measurement in Education (NCME), American Federation of Teachers (AFT), National Education Association (NEA). (1990). Standards for Teacher Competence in Educational Assessment of Students. *Educational Measurement: Issues and Practice*, 9(4), 30–30. <https://doi.org/10.1111/j.1745-3992.1990.tb00391.x>
- Nemati, M., Alavi, S. M., Mohebbi, H., & Masjedlou, A. P. (2017). Teachers' writing proficiency and assessment ability: the missing link in teachers' written corrective feedback practice in an Iranian EFL context. *Language Testing in Asia*, 7(1). <https://doi.org/10.1186/s40468-017-0053-0>
- Pallant, J. (2016). *Survival manual: A step by step guide to data analysis using SPSS program*. Crows Nest: NSW: Allen & Unwin.
- Phillips, M. (2002). Is "same" treatment "fair" treatment? In J. Shulman, A. K. Whittaker, & M. Lew (Eds.), *Using assessments to teach for understanding: A casebook for educators*. New York: Teachers College Press.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10–12.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. ASCD (Association for Supervision and Curriculum Development).
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>.
- Qian, D. D. (2014). School-based English language assessment as a high-stakes examination component in Hong Kong: Insights of frontline assessors. *Assessment in Education: Principles, Policy & Practice*, 21(3), 251–270. <https://doi.org/10.1080/0969594X.2014.915207>
- Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, 81(3), 191–198.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics, ed. NH*. Boston: Allyn & Bacon, MA.
- Tierney, R. D. (2010). *Insights into fairness in classroom assessment: Experienced English teachers share their practical wisdom*. Doctoral Dissertation University of Ottawa (Canada).
- Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 125–144). London: Sage.
- Tsagari, D., & Vogt, K. (2017). Assessment literacy of Foreign Language teachers around Europe: Research, challenges and future prospects. *Language Testing and Assessment*, 6(1), 41–61.
- Umam, A., & Indah, Y. A. (2020). Exploring In-Service TEYL Teachers' Assessment Literacy: Implication for Continuing Professional Development. *Journal of English Educators Society*, 5(1), 47–51. <https://doi.org/10.21070/jees.v5i1.364>
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- Weathington, B. L., Cunningham, C. J. L., & Pittenger, D. J. (2010). *Research methods for the behavioral and social sciences*. John Wiley & Sons.
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *Australian Educational Researcher*, 40(2), 241–256. <https://doi.org/10.1007/s13384-013-0089-9>
- Wise, S. L., Lukin, L. E., & Roos, L. L. (1991). Teacher Beliefs About Training in Testing and Measurement. *Journal of Teacher Education*, 42(1), 37–42. <https://doi.org/10.1177/002248719104200106>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Yan, X., Zhang, C., & Fan, J. J. (2018). "Assessment knowledge is important, but ...": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158–168. <https://doi.org/10.1016/j.system.2018.03.003>
- Zhang, Z., & Burry-Stock, J. (1997). *Assessment Practices Inventory: A Multivariate Analysis of Teachers' Perceived Assessment Competency*. Paper presented at the Annual Meeting of the National Council on Measurement in Education. Chicago, Illinois: ERIC. Retrieved from <https://files.eric.ed.gov/fulltext/ED408333.pdf>
- Zolfaghari, F., & Ahmadi, A. (2016). Assessment literacy components across subject matters. *Cogent Education*, 3(1). <https://doi.org/10.1080/2331186X.2016.1252561>

- Zulaiha, S., & Mulyono, H. (2020). Exploring junior high school EFL teachers' training needs of assessment literacy. *Cogent Education*, 7(1). <https://doi.org/10.1080/2331186X.2020.1772943>
- 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>

Syafi'ul Anam & Nanin Verina Widya Putri

How literate am I about assessment: Evidence from Indonesian EFL pre-service and in-service teachers