TEACHERS' UNDERSTANDING OF FORMATIVE ASSESSMENT

Mohammad Arsyad Arrafii[1]; Baiq Sumarni[2]

[1,2]Departement of English Language Education, Faculty of Education for Language and Arts, IKIP Mataram
Jln. Pemuda No.59A, Mataram, Nusa Tenggara Barat 83125, Indonesia
[1]moh_arrafi@yahoo.com; [2]meniq84@yahoo.com

How to Cite: Arrafi, M. A., & Sumarni, B. (2018). Teachers' understanding of formative assessment. *Lingua Cultura, 12*(1), 45-52. https://doi.org/10.21512/lc.v12i1.2113

ABSTRACT

The research explored teachers' understanding of formative assessment in the secondary school level, particularly in the context of English teaching and examined factors related to the teacher assessment in the literacy level. To achieve the intended purposes, a self-designed instrument named Teacher Formative Assessment Literacy Questionnaire (TFALTQ) was emploted. Cronbach alpha measured the internal consistency of items that measuring the latent constructs were adequately accepted (α= 0,67), and exploratory the factor analysis using rotation matric revealed a robust factor loading of the variance explained with KMO statistic of 0,72. The questionnaire was distributed to all English teachers who taught in the secondary high school in central Lombok. There were 243 teachers from public and private high schools were participated in the study. The participants came from both public school and religious school, regardless of their school types and level. Using the case of 243 English teachers, the analysis reveals that the teachers' understanding of formative assessment is inadequate and that four dimensions of teachers' formative assessment literacy and understanding are revealed. These include an assessment to serve the accountability purpose, examination/test driven learning, the procedural approach to learning and assessment, and the receptive role of students. The multiple hierarchical regression analysis is performed to identify athe factors that are influencing the teachers' formative assessment literacy. The analysis suggests that the teachers have a poor understanding of formative assessment, insufficient training in assessment and gender is a strong predictors of teachers understanding of formative assessment.

Keywords: formative assessment, assessment literacy, secondary school teachers, English lesson

INTRODUCTION

The role of assessment as a part of teaching and learning is essential to gather reliable information for teachers. It draws the informed judgments about the learner progress against the task criteria as well as to provide the feedback for both teachers about how they teach and the students about how they learn. The assessment also allows teachers to make decisions about further stages of teaching and learning process. To be effective, assessment should be designed and developed based on the principles of effective assessment practices. Developing effective assessment involves several steps, including identifying the learning goals and targets, planning how to assess, writing good quality items, using marking practices, and selecting an assessment forms (Nitko & Brookhart, 2011). Selecting a form of assessment should be based on their impact on expected student's learning outcomes, feasibility, validity, and reliability.

The classroom assessment cannot be separated from its improvement purposes. It is used to gather information about students' learning through collecting evidence of their performance and use this evidence for quality improvement. Such assessment with the purpose of providing feedback to student learning as well as to teachers' teaching performance is best known as formative assessment.

Research has indicated that the use of formative assessment as the classroom assessment is beneficial to improve students' academic performance (Black & Wiliam, 1998). It is also considered that formative assessment is an alternative solution for meaningful instruction which is likely able to improve the quality of the education. However, in the Indonesian context, the formative assessment seems to have a peripheral rule in assessing the students' performance. Instead, the summative assessment seems to dominate the assessment practices in the classroom. In primary and secondary education level, a large stakes assessment takes dominant rule for assessing the student learning. Arrafii and Suhaili (2015) have indicated that the common practice of assessing the students' learning in Indonesian secondary classroom is mainly summative.

In these regards, despite its beneficial effects on learning, the formative assessment also encounters some barriers for the application (OECD, 2005). The literature

argues that heavy workloads for the teachers' for applying the formative assessment are one of the barriers, especially when it is applied in the large classroom size which is commonly found in Indonesian classroom. At the same time, the students from the Indonesian perspective, they are required by the education system to be prepared for the national examination. Thus, instead of giving the attention to achieve intended learning goals described in the curriculum and syllabus, the teaching and learning process emphasizes on testing the preparation in which the teachers are busy preparing strategies to help the students to answer and pass the national examination. This also may impede the teachers from using the formative assessment. In addition, the teachers' understanding about formative assessment may be insufficient which prevents them implementing the formative assessment in the classroom (Bennet, 2011; Arrafii & Suhaili, 2015).

Until recently, little is known about the teachers' literacy level and practice of formative assessment in the secondary school, particularly in Central Lombok. Arrafii and Suhaili's (2015) qualitative study indicates that the teachers' understanding of formative assessment is inadequate. However, this finding could not be generalized to the larger population due to the nature of qualitative research design. For this reason, the quantitative research to uncover and generalize the teachers' understanding of formative assessment with representative teacher sample in the central Lombok is urgently needed. The purposes of this research are to find out whether TFALQ serves as a useful instrument for evaluating in-service teachers' formative assessment literacy in Indonesia, particularly in Central Lombok. It aims to investigate the teachers' literacy of formative assessment, and to find out the factors that related to teachers' literacy of formative assessment.

Defining the formative assessment can be broadly defined as a form of assessment that focuses on the process of learning for improvement of the students' performance. However, until recently, the definition of formative assessment is still vague, and there is no consensus as to what the term of formative assessment means. Very often, the formative assessment may be understood to have similar ideas to diagnostic the assessment. However, some of the authors consider these two assessments as two separate practices (Nitko & Brookharts, 2011) with different purposes. While the diagnostic assessment is designed before the instruction and merely for the sake of planning instruction, the formative assessment seeks an understanding of the status of the learning with the purpose of identifying next steps for the improvement (Assessment Reform Group, 2010).

Black and Wiliam (1998) have defined the formative assessment as, "All those activities undertaken by teachers, and/or by their students, which provide feedback to modify the teaching and learning activities in which they are engaged". Black and Wiliam (1998) have highlighted the power of assessment to provide feedback, the importance of the social context of learning, and some issues related to the distinction between the formative and summative purpose of assessment. They also find the beneficial effects of the formative assessment towards the student learning have been well documented in the literature.

However, Hattie and Jaeger (1998) have criticized Black and Wiliam's review because their position strongly supported the importance of assessment and feedback for teaching while neglecting the learning. Hattie (2005) has suggested the assessment is about teaching as much as it is about learning. For formative assessment to deliver its promise, feedback must relate to how the students learn. According to Hattie (2005), a comprehensive theory of formative assessment must include a theory of learning because it is not necessarily for the learner who has a problem. It may include the teacher or teaching methods that need remediation.

Whereas, Popham (2008) has defined the formative assessment as, "A planned process during which the teacher or students use the assessment-based on the evidence to adjust ongoing learning and instruction". Further, Hattie and Jaeger (1998) have proposed five postulates of formative assessment. The first assessment is the achievement enhanced to the extent that teachers' set challenging goals relative to the present student competencies. The second assessment is the achievement enhanced through feedback (of certain types). The third assessment is the achievement enhanced if students are taught to accept feedback to verify, rather than enhance, their sense of self-efficacy. The fourth assessment is the achievement enhanced to the extent that teachers become more automatics in key teaching competencies, so spend more time on feedback. The fifth assessment is the achievement enhanced by re-conceptualizing existing information as well as acquiring new information.

This explicitly states that the formative assessment should include the goal setting (and standards), assessment design, performance evaluation (against standards), informative feedback, and the action of both teachers' (and possibly peers). For an assessment to be formative, it requires feedback which indicates the existence of a gap between the actual level of work being assessed and the required standard (Taras, 2005). Responding to the critics, Black and Wiliam (2009) have revisited a theory of formative assessment. They argue that practice in a classroom is formative to the extent that the evidence about student achievement is elicited, interpreted, and used by the teachers and learners or their peers to make decisions about the next steps in the instruction. They also refer to the classroom practices as being formative or not only if they are intended to provide feedback.

If a so-called formative assessment activity is not used to provide feedback, it is hardly formative. However, feedback should be understood at the same circumstance by both teachers'. Otherwise, the feedback remains worthless for learning. For this reason, Black and William (2009) have identified the need for communication between teacher and student about the interpretation, the feedback given to the students, and the next step because the feedback could not be formative unless it is understood. In addition, peer as an element of formative practices could be necessarily involved (Black & William, 2009).

Based on the cases studies, the research related to formative assessment, OECD (2005) has identified the six key elements of formative assessment. It includes the establishment of a classroom culture that encourages interaction and the use of assessment tools, the establishment of learning goals and tracking of individual student progress toward those goals, the use of varied instruction methods to meet diverse student needs, the use of varied approaches to assessing student understanding, the feedback on student performance and adaptation of instruction to meet identified needs, and the active involvement of students in the learning process.

Meanwhile, Bennet (2011) has proposed two ideas for the meaningful definition of formative assessment. He asserts that the formative assessment should have a theory of

action and a concrete instantiation. Theory of action should specify the entities of formative assessment, including the characteristics and components that provide rationales for each characteristic and components. It should also hint the synergy of both entities to work together to produce the intended learning outcomes (Bennet, 2010).

Meanwhile, the concrete of instantiation illustrates how the formative assessment works in a real setting and how it reflects the theory. As part of this idea, Bennet (2011) has presented "Keep Learning on Track" (KLT or ETS) program as an adequate example of the theory of action and concrete instantiation of formative assessment. The theory of action of KLT is based on one big idea and five key strategies. The big idea is teachers use evidence to adapt teaching and learning to meet immediate learning needs minute-by-minute and day by day (ETS, 2010).

The five keys strategy are Sharing Learning Expectation (i.e., clarifying and sharing learning intentions and criteria for success), Questioning (i.e., engineering effective classroom discussions, questions and learning tasks that elicit evidence of learning), Feedback, Self-assessment (i.e., activating students as ownership of their own learning), and Peer Assessment (i.e., activating students as instructional resources for one another). This theory of action and instantiation can guide the teacher to implement formative assessment in the classroom and meet the beneficial effect of formative assessment on student learning (Bennet, 2011).

Assessment development should consider the principles of effective assessment. The principles of assessment include a variety of principles which classified and described in many ways depending on the frame of reference that used. According to Linn & Miller (2005), to be effective, there are five general principles of assessment to be taken into consideration; the first principle is that clear specification of what to be assessed (learning targets) in the assessment process. This principle requires a specific description of the characteristics of learning targets to be assessed should be in advance before the selection or development of the assessment procedures. Learning targets could be a basic cognitive process dimension of revised Bloom's taxonomy; remember and understand or higher cognitive dimensions such as evaluate and create or both.

The second principle is an assessment method that should be selected because of its relevance to the characteristics or performance to be measured. Frequently, the assessment methods are selected based on their objectivity, accuracy, or convenience. However, the assessment methods are appropriate for some user and inappropriate for the others. In assessing the student achievement, it is necessary to match the desired learning goals and the types of assessment method that is used. When a teacher is eager to find the students' ability to develop and organize ideas and write a well-integrated composition as learning goals, the multiple-choice test would be a poor choice, and another form such as essay writing would be better one.

The third principle is that the comprehensive assessment requires a variety of the assessment method. This principle indicates that none of the single types of assessment can test a vast array of learning and development outcomes. The achievement tests such as multiple choice and short answer test are useful to measure the knowledge, understanding, and application outcomes, but essay tests and other written projects are more appropriate to assess the capability to express ideas and organize them. The quality assessment uses several different tasks, such as samples of writing, student retelling, records of independent reading, self-evaluations, and checklists. In making these choices, the teacher should have a trust on his/her own intuition based on the knowledge and observations of the students. Besides, the formal types of assessments which have proclaimed their validity and reliability using various statistical procedures, the use of informal assessment such as questioning, student-teacher dialogues, and teacher observations is essential.

The fourth principle is that the proper use of assessment methods requires the awareness of their limitation. This principle indicates that every assessment procedure from a well-developed measurement instrument, such as the standardized achievement test, to rather some crude assessment devices, self-report technique, the subject to various types of measurement errors. However, a keen awareness of assessment instruments makes it possible to use them effectively than the cruder instrument. The greater its limitation, the more caution it is required in its use. The last principle is that the assessment is a means to an end, not an end itself. This principle emphasizes the role of assessment as a process of obtaining the evidence and information about the instructional decisions that are made on, not as the purpose itself.

From the assessment principles discussed, it can be drawn some of the characteristics of quality assessment, such as clear learning targets and appropriate assessment methods, which are useful to involve when constructing the high-quality assessment. Stiggins (1995) has listed some of the requirements for quality assessment. He also believes that a high-quality assessment should have some clear targets. When designing the assessment task, the teacher should determine the clear targets of assessment, includes specific competence and capability that should be demonstrated by the students. Teachers should understand what they are assessing whether the skills, knowledge, product, or reasoning. They also should set the focus purposes of assessment that they construct. In addition, a quality assessment should have proper methods to examine student achievement. The different learning targets request different assessment method. Besides that, the assessment should include sound samplings that allow the teacher generates the confident inferences about how the respondents would have given all the possible exercises. Finally, the accurate assessment is free to bias and distortion.

Similarly, Butler and McMunn (2006) have proposed the quality assessment should include clear purposes of assessment tasks which are generally divided into three; summative, formative, and diagnostic purposes. However, Airasian (2001) have argued that among the characteristics of quality assessment mentioned, the most important quality of assessment is whether it is valid and reliable. Assessment validity refers to whether the information being gathered by assessment tasks is relevant to the decision that needs to be made. Without validity, the assessment data would not lead to the correct decision. Assessment reliability indicates the consistency of assessment information gathered by the assessment tasks. Inconsistent information does not help the teacher to make decisions (Linn & Miller, 2005). Nevertheless, Moss (2003) has warned that a single assessment does not reflect the holistic picture of the student performance, even though the level of the assessment validity and reliability seem to be strong. There is a need to consider several different means of assessment to enhance the appropriateness of the decisions that are made.

Further, Linn & Miller (2005) have suggested some

considerations in preparing the relevant assessment tasks. Their suggestions include matching between the tasks and desired outcomes, gaining a representative sample of tasks, exclude irrelevant barriers to the performance, and avoiding unintended clues in the objective test items. In addition, to its validity and reliability, selecting the assessment method should be based on the kinds of performance to be assessed, whether it is designed to assess knowledge, reasoning, skills, products, and dispositions (Butler & McMunn, 2006).

These five areas need different types of assessment methods. When the intention is to assess knowledge, reasoning, and problem-solving, then selected response test items can be used to test the student mastery of subject matter knowledge (Stiggins, 1995). The selected response tests the items that include multiple-choice, true-false, matching, and fill-in test items. If the teacher wants to assess their skills, they could use a performance method of assessment. Nitko and Brookhart (2011) have differed 10 types of performance assessment, namely structured, on-demand tasks, paper-pencil tasks, tasks requiring other equipment and resources, demonstration, typical performance tasks, long-term project, individual and combining group project, experiments, oral presentations and dramatizations, and simulations.

The research has indicated that there are some factors that are affecting teachers' assessment literacy, such as gender and major, education and assessment training, and teaching experience (Mertler, 2004). Alkharusi (2011) has found that measurement and testing knowledge of pre-service teachers' assessment literacy tend to vary according to the gender and major. The research has reported that the male teachers tend to have a higher level of measurement and testing knowledge than the female teachers. And those pre-service teachers specializing in areas such as English language, math, and science, tend to possess a higher level of measurement and testing knowledge than those who are specializing in performance areas such as art education and physical education.

The assessment training and teaching experience are also considered as important factors that influencing teacher' assessment literacy. When comparing the assessment literacy of pre-service and in-service teachers, the studies indicate that the assessment literacy level of pre-service teachers tends to be lower than the in-service teachers (Mertler, 2004). Meanwhile, the assessment training is considered to have the positive impact on the teachers' feeling regarding assessment and confidence in using assessment. It is indicated by the teacher performance on a post-test is on average higher than their performance on the pre-test (Mertler, 2009). Other factors which may influence the teachers' assessment literacy include the school resources, socio-economic status, and education level. Although, the former studies that are addressing these assume that the factors are scarcely available. These related factors are considered important and may influence the teachers' assessment literacy. Therefore, they are included in the research.

The massive research studies have been conducted to address the issues in formative assessment. Most of the researchers have focused on the impact of formative assessment on the students' learning. However, little has been conducted to investigate the teachers' knowledge and practice of formative assessment. Work of Crossouard, Pryor, and Torrance (2004), Keen (2005), Wylie and Lyon (2015), and Knight (2003) are few exceptions. Crossouard, Pryor, and Torrance (2004) have found that the participants in their research and the educators as the students in an educational doctorate program has little theoretical understanding of formative assessment, which may affect their ability to promote learning in their students. However, this study is not intended to investigate the teachers' knowledge about formative assessment; rather it seeks the teachers that live in experience with formative assessment and has been conducted at the higher education level.

The research on the teachers' understanding of formative assessment has documented that many of the teachers do not understand the formative assessment comprehensively after investigating the English trainee students' understanding the formative assessment in the writing development topic. Keen (2005) has concluded that trainees of English teaching have lack ability to identify the strengths, achievements, and shortcomings in the students' writing. Their assessment also lacks the accuracy which avoids them being able to suggest the directions for students' writing development and to inform their own planning and teaching. Similarly, Knight (2003) has reported that the teachers in two primary schools in New Zealand are unclear about what constitutes and the effective formative assessment, especially to the effective feedback. They struggle to define effective feedback with any details.

Moreover, as part of formative assessment, the feedback gains adequate attention. Research on students' perception of effective feedback often finds the students that do not think the feedback is important for their learning. Duncan (2007) has claimed that the students do not pay attention nor make sense to their teacher feedbacks and comments. One of the reason is they might not understand the purposes of the feedback process. In addition, it is well documented in the literature that the teachers see the feedback as disintegrated part of learning process and is considered as a teacher that owned endeavor (Taras, 2005). From the teachers' perspective, Hattie and Timperley (2007) have assumed that many of the teachers tend to focus on the correctional aspects rather than the instructional aspects of feedback. In the Indonesian context where this research is conducted, to extend to the writers' knowledge, the issue of the teachers' formative literacy assessment remains largely unexplored.

METHODS

The researchers have administered the questionnaire to all English teachers who teach in the secondary high school in Central Lombok. There are 243 teachers from public and private high schools in Central Lombok that are participated in the study. The participants come from both public school and religious school, regardless of their school types and level.

The research employs the survey design. According to Creswell (2011), the survey design is a procedure in quantitative research by which the investigators administer a survey to a sample or to an entire population of people to describe attitudes, opinions, behaviors or characteristics of the population. Particularly, the research uses the cross-sectional survey design by which the investigators collect the data at one point in time.

The main instrument that is used to collect the data is the Teacher Formative Assessment Literacy Questionnaire (TFALQ) which is self-designed and developed by the team of the researchers. The questionnaire is intended to measure the teachers' knowledge and understanding of the basic principles of the formative classroom assessment practices.

It consists of 29 items with four options of response that reflects the extent to which teachers understand formative assessment. The items are scored from 1-4 with a high total score reflects the advanced understanding of formative assessment literacy. The design and development of TFALQ are informed by several assessment theories underpinning formative assessment. These include the distinction between the assessment of learning and assessment for learning, the performance versus mastery learning goal orientation (Watkins et al., 2000). The notions of convergent and divergent approaches are to assessment (Torrance & Pryor, 2001), assessment of learning strategies and practices (Black & Wiliam, 2009; Assessment Reform Group, 2010). Table 1 provides a sample of items in the TFALQ.

Table 1 The Samples of Items in the TFALQ

Based on your current knowledge about assessment, to what extent do the following statements constitute formative assessment?	Very little	Little	Some	A lot
Classroom questionings in formative assessment are not mainly used to elicit factual knowledge from students	☐	☐	☐	☐
Teachers ensure that students have shared similar understanding of learning targets with their teachers	☐	☐	☐	☐

The measure of internal consistency of Cronbach alpha is used to test the reliability of the instrument. Meanwhile, the validity of the instrument is measured using Principal Component Analysis (PCA) with orthogonal rotation matrix on the items of the teacher assessment literacy questionnaire determined how many constructs are included in the instrument. Applying these procedures are answered by the first research question.

To answer the second research question, since the items in TFALQ are scaled with four options that scored between 1-4 with a high total score reflects a high level of assessment literacy, the accumulation of the response gained by the teachers are used to indicate the teachers' formative assessment literacy. To answer the third research question, the multiple linear regressions are employed. The regression analysis is a set of statistical techniques that allow one to assess the relationship between one dependent variable and several independent variables (Tabachnick & Fidell, 2007). Further, regression analysis could suggest the evidence for which of the predicting factors exerts the most influence on the dependent variable. Regression analysis in this study investigates the factors that are influencing the teachers' formative assessment literacy at the teachers' demographic factors such as sex, socioeconomic status, education level, teaching experience, and further assessment training.

Prior to the multiple regression analysis, an exploratory analysis is conducted, including the screening for missing data, means, variances, skewness, kurtosis, normality, outliers, and correlation analysis. The exploratory analysis confirms the variables that included in the regression model from those excluded. Only variables with less than 5% of missing value are normally distributed, linearly related to dependent variables are included in the regression model.

For the categorical variables, the categories with adequate frequency are included in the regression model. According Green as cited in Field (2009), to indicate the sample size for each of the nominal predictors in regression analysis, it depends on the objective of regression analysis, whether to test the overall fit of the regression model or to test the contribution the individual predictor within the model. When the objective is to test the individual predictor, which is the objective of this study. Green has suggested using a minimum sample size of $104 + k$, where k is the number of predictors in the regression model. If the predictors are 5, the minimum sample size would be 109. Hence, the category with less than 109 cases emerges to the category that meets this rule of thumb.

Further, the correlation analysis is conducted to ensure that the continuous variables correlate with the dependent variable. If an independent variable correlates with the dependent variable with the minimum of $r = 0,05$, it will be included in the model (Cohen et al., 2003). The regression analysis will suggest which factors that predicting the teachers' assessment literacy of formative assessment and propose the strength of each factor affecting the outcome variable.

RESULTS AND DISCUSSIONS

Results of the research presented here derive from 243 teachers in Lombok. There are 359 copies of questionnaires that are distributed to the public and private schools/madrasah in Central Lombok, and the researchers receive 280 returned questionnaires. After screening the questionnaire, many of them are incomplete and contained massive missing, and therefore removed for analysis. Hence, only 243 cases are carried out for further analysis. The demographic characteristics of respondents are shown in Table 2. The male participants and non-public servant teachers dominate to participate. From the highest education qualification group, the numbers of diploma graduates outweigh those with master degree holders. The study is conducted from August 2nd to 31st 2016.

Table 2 The Demographic Information of the Participants in the Study (N=243)

Sex		Occupational status		Qualification	
Male	female	PNS	Non-PNS	S1	S2
163	80	108	138	235	8

TFALQ consists of 29 scale items which measure five latent constructs that related to formative assessment. These constructs and items are four items of the purposes of formative assessment, five items of sharing learning intention and criteria for success, seven items of strategic questioning and classroom dialogue, six items of feedback that bring learning forward, and seven items of self and peer assessment. Based on the reliability analysis using Cronbach alpha, the reliability of the scale for each construct in the TFALQ is adequate. Similarly, the analysis

of the instrument validity using Keiser-Mayor-Olkin (KMO) coefficient reveals a robust factor solution for all constructs that measured, yielding six latent factors, above 1 eigenvalue. However, one factor lies on the cross point (inflection). Thus, it is considered beyond the construct of TFALQ that can be seen in Table 3.

Table 3 Coefficient of Validity and Reliability of TFALQ

No	Constructs	α	KMO
1	Purposes of formative assessment	0,65	0,77
2	Sharing learning intention and criteria for success	0,75	
3	Strategic questioning and class-room dialogue	0,53	
4	Feedback that bring learning forward	0,72	
5	Self and peer assessment	0,58	

The analysis of dimensions of teachers' formative assessment understanding using exploratory factor analysis (EFA) reveals that TFALQ has comprised of four dimensions of teachers understanding of formative assessment. These dimensions are assessment serves accountability purpose that reflected in items 2, 7, 23, 28; examination/test driven learning (items 1, 10, 16, 21); procedural approach to learning reflected in items 4, 5, 9, 11, 27; and receptive role of students revealed in items 6, 19, 25.

The analysis of the teachers' understanding of formative assessment reveals that teachers' understanding of formative assessment is relatively poor. The mean score achieved by the teachers in the research is 47,3, out of 100 (M= 47,3, SD= 0,53). This replicates roughly similar evidence that is reported by other studies. Investigating the teachers' assessment literacy, Plake, Impara, and Fager (1993) have concluded that teachers' knowledge of classroom assessment is insufficient, and they are not adequately prepared to assess student learning. It shows as evidenced by only 66 percent of the correct answer provided by the teachers in Assessment Literacy Inventory Test. The mean score of the teachers' response is 23 out of 35 items. Similarly, Campbell, Murphy, and Holt (2002)'s study that reported the pre-service teachers (M=21) average the two fewer questions answered correctly than the in-service teachers (M=23).

Mertler (2004) has studied the assessment literacy of both pre-service and in-service teachers, and then statistically compares the two groups. Using a slightly modified version of the Teacher Assessment Literacy Questionnaire, he obtains similar results to both Plake, Impara, and Fager (1993) and Campbell, Murphy, and Holt (2002)'s studies. The average score for in-service teachers is equal to 22 items answered correctly. The score is quite like that obtained by Plake, Impara, and Fager (1993) who gained the average of 23 points.

An inadequate theoretical understanding of teachers about formative assessment has also been revealed by other studies by Crossouard, Pryor, and Torrance (2004), Keen (2005), and Knight (2003). The low level of teachers' literacy of formative assessment in the Indonesian context may be explained by lack of further training on assessment literacy and practice. Usually, the training program that is provided for secondary English teachers mainly focused on the methodological aspect of teaching with lesser attention to the assessment process and of student work or performance. For instance, English language training for *madrasah* English teachers in three regions of Lombok; West Lombok, Central Lombok, and Mataram train in-service teachers to be able to understand and apply communicative English language teaching (ELTIS, 2010). The training program strongly emphasizes on didactic approaches of teaching, enhancing the teachers' knowledge and skills of teaching methodology without strong attention on assessment.

In addition, the teachers in Central Lombok have been settled and engaged in a long tradition of high stake evaluative system of assessment which drives their current knowledge and understanding of assessment. Their knowledge of assessment is within the boundaries of summative assessment rather than the formative assessment. In other words, formative assessment is beyond the black box (Black & Wiliam, 1998) of teachers' classroom assessment practice. It is likely that the teachers may have an implicit understanding of formative assessment. They are still far legging in the theoretical concept of formative assessment because of the existing tradition or culture of high stake summative testing.

Table 4 illustrates the factors that may influence the teachers' formative assessment literacy. It is shown that in the model 1, the gender has been proven to be a significant factor that is predicting formative assessment literacy. The female teachers are more likely to have a better understanding of formative assessment than male teachers do. However, the other two predictors in the model 1 (occupational status as being public servant/PNS or Non-PNS and education qualification as being either diploma/S1 or master degree/S2 graduates) do not significantly predict the teachers' understanding of formative assessment.

This may be explained by the fact that the teachers' exposure on the assessment during their candidature of the teacher in the tertiary education level and the master degree training does not provide sufficient training on assessment. Most universities and teacher training institutes teach the only single subject that related to the assessment which is offered in two hours in a weekly meeting. For instance, English department students are required to complete a single Language Testing module during their academic years. Moreover, this module strongly focuses on testing and statistical calculation rather than on the teaching and learning improvement aspect of assessment.

However, in the model 2, none of the demographic predictors are proven to have a significant impact on teachers' formative assessment literacy. Whereas, the other factor, such as professional training on assessment, is the significant predictor. It has suggested that the significant effect of the gender is probably mediated by the professional training. The teachers who are reported to receive further professional training on assessment are more likely to outperform those who are reported to have a lack of professional training regardless of their gender. It is therefore recommended that the professional training on the assessment should be integrated into formal education and the assessment that focused on training should be widely offered for in-service teachers across Indonesia.

Table 4 The Results of Hierarchical Regression Analysis of Teachers' Demographic Factors on Formative Assessment Literacy

Predictors	Model 1			Model 2	
	β	B	Sig	B	B
Female (Male as reference)	0,31	0,13	0,03	0,03	0,01
Non PNS (PNS as reference)	-0,09	-0,04	0,09	-0,47	-0,22
S2 (S1 as reference)	0,12	0,05	0,69	0,60	0,27
Professional training				0,40	0,02

R2 for Model 1 = 3,3% R2 for Model 2 = 18,9%; Sig at α = 0,05

CONCLUSIONS

TFALTQ which has been validated with 243 English teachers in Lombok, has been used to obtain the data about secondary high school teachers' understanding of the formative assessment. The analysis of the survey data has revealed four dimensions of teachers' formative assessment literacy and understanding. These include the assessment that serves an accountability purpose, examination or test-driven learning, procedural approach to learning and assessment, and receptive role of students. In addition, the research has identified some related factors that influencing the teachers' formative assessment literacy and the analysis suggests that the teachers have poor understanding of formative assessment, insufficient training on assessment and gender are strong predictors of teachers understanding of formative assessment.

The research has recommended that a further professional training program to improve in-service teacher theoretical understanding of formative assessment needs to be urgently conducted. And the teacher training institute needs to offer more assessment related to the courses, addressing issues on formative assessment literacy and practices, and providing more time allocated for such courses.

REFERENCES

Airasian, P. W. (2001). *Classroom assessment.* New York: McGraw-Hill companies.

Alkharusi, H. (2011). Psychometric properties of the teacher assessment literacy questionnaire for pre-service teachers in Oman. *Social and Behavioral Science, 29*, 1614-1624. doi: 10.1016/j.sbspro.2011.11.404

Arrafii, M. A., & Suhaili, M. (2015). *Formative Assessment literacy and practice: A study of English teaching in Indonesian Secondary School.* Report paper prepared for Directorate General of Higher Education, Ministry of National Education.

Assessment Reform Group. (2010). *Assessment is for learning: 10 principles.* Retrieved from http://www.assessment-reform-group.org.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice. 5*(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of Formative Assessment. *Educational Assessment, Evaluation, and Accountability, 21*(1), 5-31.

Bennet, R. E. (2010). Cognitive base assessment of, for and as learning: A preliminary theory of action for Summative and Formative Assessment. *Measurement: Interdisciplinary Research and Perspectives, 8*(2-3), 389-407.

Bennet, R. E. (2011). Formative Assessment: A critical review. *Assessment in Education: Principles, Policy and Practice, 18*(1), 5-15.

Butler, S. M., & McMunn. (2006). *A teacher's guide to classroom assessment: Understanding and using assessment to improve student learning.* San Francisco: Jossey-Bass.

Campbell, C., Murphy, J. A., & Holt, J. K. (2002). *Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers.* Paper presented at the meeting of the Mid-Western Educational Research Association, Columbus, OH.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral science.* London: Lawrence Erlbaum Associates, Publishers.

Creswell, J. W. (2011). *Educational research: Planning, conducting and evaluation quantitative and Qualitative Research* (4th Eds.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Crossouard, B., Pryor, T., & Torrance, H. (2004). *Creating an alternative assessment regime with Online FA: Developing a researcher identity.* Paper presented at European Conference on Education Research, in Crete, Greece, 21–22 September.

Duncan, N. (2007). 'Feed-forward': Improving student's use of tutor comments. *Assessment & Evaluation in Higher Education, 32*(3), 571-581.

ELTIS (English Language training for Islamic School). (2010). *A program completion report.* Retrieved from www.lapis-eltis.org.

ETS (Educational Testing Service). (2010). *About the KLT program.* Princeton: NJ. Retrieved on March 10th, 2014 from http://www.ets.org/Media/Campaign/12652/about.html.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thausand Oaks, California: SAGE Publication Inc.

Hattie, J., & Jaeger, R. (1998). Assessment and classroom learning: A deductive approach. *Assessment in Education: Principles, Policy, and Practice, 5*(1), 111-122.

Hattie, J. (2005). *What is the nature of evidence that makes a difference to learning?* Paper presented at the ACER research conference, Melbourne.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Keen, J. (2005). Assessment for writing development: Trainee English teachers' understanding of Formative Assessment. *Teacher Development, 9*(2), 237-253.

Knight, P. (2003). *An evaluation of the quality of teacher feedback to students: A study of numeracy teaching*

in the Primary Education sector. Paper presented at the AARE/NZARE conference, Auckland, New Zealand.

Linn, R. L. & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th Eds.) Upple Saddle River, New Jersey: Pearson Merrill Prentice Hall.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference. *American Secondary Education, 33*(1), 49 – 64.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools, 12*(2), 101 – 113.

Moss, P. A. (2003). Conceptualizing validity for the classroom. *Educational Measurement: Issues and Practices, 22*(4), 13-25.

Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Boston, MA: Pearson.

OECD. (2005). *FA: Improving learning in secondary classroom.* New York: OECD Publishing.

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: a national survey. *Educational Measurement: Issues and Practice, 12*(4), 1210 – 1239.

Popham, W. J. (2008). *Transformative assessment in action: An Inside look at applying the process.* Alexandria, VA: Association for Supervision and Curriculum Development.

Stiggins, R. J. (1995). Assessment literacy for 21st century. *Phil Delta Kappan, 77*(3), 238-245.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics.* Boston: Allyn & Bacon.

Taras, M. (2005). Assessment, summative and formative: Some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466–78.

Torrance, H., & Pryor, J. (2001). Developing Formative Assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal, 27*(5), 615-631.

Watkins, C., Carnell, E., Lodge, L., Wagner, P., & Whalley, C. (2000) *Learning about learning.* London: Routledge.

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice, 22*(1), 140-160.