

## Classroom Assessment Practices of EFL Lecturers with Current Curriculum Implementation: Where Policy Meets Practice

Alfian Alfian<sup>1\*</sup>, M. Nur Akbar Rasyid<sup>2</sup>, Akhmad Habibi<sup>3</sup>,  
Noprival Noprival<sup>4</sup> & Mohammad Yusuf<sup>1</sup>

<sup>1</sup> Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi, Jambi, Indonesia

<sup>2</sup> Universitas Islam Negeri Alauddin, Makassar

<sup>3</sup> Universitas Jambi, Jambi, Indonesia

<sup>4</sup> Sekolah Tinggi Ilmu Kesehatan Harapan Ibu Jambi, Jambi, Indonesia

alfian@uinjambi.ac.id

### ARTICLE HISTORY

Received : 2022-08-15

Revised : 2022-11-22

Accepted : 2022-12-04

### KEYWORDS

Assessment practice

Curriculum

IQF

Attitude

Knowledge

KKNI

EFL Students

### ABSTRACT

Within the context of Indonesian higher education, the assessment focuses on the three domains of students' competencies, namely knowledge, skills, and attitude. Therefore, this study aims to explore the assessment practice of EFL lecturers, especially the methods used in assessing the three students' competencies. It also explores the general assessment practice and examines differences by educational degree and length of teaching experience. This quantitative study relies on survey data from 71 participants joining the EFL lecturers' association in Indonesia. Furthermore, the convenient sampling technique was used to determine the sample, and the data were analyzed using SPSS version 24. The results showed that the lecturers employed various methods in assessing the three domains of students' competencies. It was also found that the lecturers mostly employed observation (66.2%) for assessing attitudes domain, tests (33.8%) for assessing knowledge, and work performance (29.6%) for skills. The findings demonstrated that the lecturers have followed the assessment policy of the government related to the method, purpose, instrumentation, procedure, and reporting grade. The results also showed no differences in the assessment practice between lecturers with Master and Doctoral degrees. Additionally, there was no difference among the four categories of teaching experiences in the assessment practice. These findings are expected to become a reference for lecturers and curriculum development in conducting and designing appropriate assessments to guarantee teaching and learning quality.

## 1. Introduction

Student assessment is an important component of the curriculum because it provides both information and evidence of learning. The assessment results could be used to modify teaching, develop curricula, and create a program to improve learning quality (Angelo, 2012; Black & Wiliam, 2010; Nitko & Brookhart, 2007; Linn & Miller, 2005). Furthermore, it determines the learner's achievement and the success of teaching and learning goals (Hamodi et al., 2017; Widiastuti, 2016; Otaya, 2017). The purposes of assessment can be achieved with proper methods, instruments, and techniques, as well as procedures that are properly situated in well-founded theory. Within the Indonesian context of higher education, the theory has been adopted by the government and universities through assessment guidelines practice (see DIKTI, 2019;

DIKTIS, 2019), which the students also use. However, previous studies have shown that several university teachers use assessment practices they are familiar with (Reynolds-Keefers, 2010; Rohl, 1999). Therefore, new assessment methods, tools, and activities were learned through intuitions (Scarino, 2013). This point implies that some teachers may not use the guide written in their assessment policy, which may be true in an Indonesian context. The lecturers' assessment practices could be further understood by exploring the current practice in implementing the new curriculum.

Assessment practice has been the priority of studies for decades. Several growing reports have been made on the area of assessment (Alemi & Khanlarzadeh, 2016; Box et al., 2015; Hamodi et al., 2017; Ibrahim et

al., 2017; Lake & Olson, 2020; Martha et al., 2021; Panadero et al., 2019; Sathasivam et al., 2019). For example, Panadero et al. (2019) studied the Spanish university assessment practices, focusing specifically on situations in higher education through library research. The teachers used various assessment instruments, although the final score of the student's grades was mostly obtained from a final examination. The study showed that the teachers rarely used formative assessments by peers or self-assessments. Despite the variations in assessment, they were unlikely to change the practices for their first to fourth year.

Several studies on the assessment practice have been conducted within the Indonesian context (Arrafi, 2021; Retnawati, Hadi, & Nugraha, 2016; Puad & Ashton, 2021; Zaim & Arsyad, 2020). Most of these studies focused on and examined the assessment model at the secondary school level. For example, Puad and Ashton (2021) studied an Islamic boarding school focusing on implementing formative and summative assessments. Until recently, there were only several reports on assessment practices in higher educational levels after the implementation of the Indonesian Qualification Framework (IQF) based curriculum (referred to here as KKNi) (e.g., Nasution, Sudrajat, & Jahro, 2019; Yustitia & Wardani, 2017). These studies mostly address authentic assessment practices. However, little attention has been given to the practices of EFL teachers that focus on the methods and procedures used for assessing the three students' competencies and practice generally demanded in the current curriculum implementation within the Indonesian context.

This study aims to fill out this empirical gap by investigating how lecturers of English as a Foreign Language (EFL) conducted an assessment, particularly the method used in assessing the three-domain competencies as guided by KKNi. It also examines the relationship between the assessment practice and the educational background (Master's and Doctoral degree) of lecturers, as well as their teaching experiences. More specifically, this study was conducted to answer the following research questions: (1) What methods are mostly used in assessing students' competencies? (2) How are the lecturer's assessment practices regarding the method, purpose, instrumentation, procedure, reporting grade, and barrier? (3) Is there a difference in assessment practice between Master and Doctoral degree lecturers? (4) Are assessment practice differences among the length of teaching experiences?

This study provides an overview of how EFL lecturers conducted an assessment of student learning. It also emphasizes the essential assessment point in English teaching, which covers students' competencies of attitude, skills, and knowledge. This means the method commonly used in assessing each competency

was provided. In addition, the results consist of the latest analysis and insights into assessment practice, such as method, purpose, instrumentation, procedure, assessment score, and barrier. This can serve as a reference for the lecturers when conducting assessments in the future. This study is expected to contribute to the assessment data in EFL related to faculty practice, which the lecturers require in the practice. Furthermore, it is significant because it provides relevant information on the assessment practices for universities, education stakeholders, and curriculum designers in formulating policies for language learning.

## 2. Literature Review

### 2.1 Theory of Assessment

Assessment is defined as a systematic process that is useful for gathering information in the form of material to conclude a process (Reynolds et al. (2010). This is similar to Reynolds (2010) and Ghafar (2011), who defined assessment as a systematic procedure that includes collecting, analyzing, and translating evidence achieved by learners, as well as the extent to which the learning objectives were met. It is a step for obtaining various information to determine the learning process policy (Uno & Koni, 2012; Custer et al., 2000). Meanwhile, in the classroom context, it refers to the process of assigning or determining values to students based on specific criteria (Harun & Iqbal, 2018).

Assessment has a very important role in the context of educational institutions, one of which is to analyze whether the learning has achieved the intended outcome (William, 2011). Furthermore, it helps monitor, improve, or ensure learning success and serves as material for the accountability of an educational institution (Earl, 2003; Nitko & Brookhart, 2012). Assessment is a way to ensure and improve the quality of educational provision (Abera et al., 2017). Based on the purpose and the time, the experts in the field classified assessment into two kinds, namely the formative and summative. This division was initially made by Bloom, Hastings, and Madaus in 1971 (Arrafi, 2021). Summative assessment is used to measure learning outcomes, often for judgemental purposes, which are conducted at the end of a program for validation and selection. This kind of assessment is still widely used by educational institutions, although it is often combined with formative.

On the other hand, formative is conducted during and throughout the learning process and focuses on learning improvement. It assesses the students in the process of forming their competencies and skills, therefore helping them to continue the growth process (Brown & Abeywickrama, 2010, p. 7). Most higher educational institutions have employed both types of assessment to make final decisions and improve learning (Black & Wiliam, 2018).

Black and William (2018) emphasized that assessment is classified into three kinds based on purposes: “assessment *as* learning, assessment for learning, and assessment of learning.” Assessment as learning is used in a more specific sense, that is, to optimize learning (Earl, 2012). Meanwhile, “assessment for learning” refers to the process by which teachers use assessment evidence to inform their teaching. The assessment determines the extent to which students have achieved intended learning outcomes. Previously, Earl (2003) stated that summative is carried out at the end of a learning process to monitor and guarantee learning outcomes. In this case, it is referred to as an assessment of learning. In contrast, both assessments for and as learning aim to improve the process and are carried out during the formative (Black & Wiliam, 2006; Newton, 2007).

## 2.2 Assessment Policy in the Indonesian Higher Educational Context

Within the university context of teaching and learning, the implementation of the Indonesian Qualification Framework (IQF) based Curriculum (referred to as KKNI) has recommended assessment to measure the standards of competency, such as knowledge, skills, and attitude of the learners (government regulation/ Permenristekdikti 44 the Year of 2015 number 5). These domains are graduate attributes that need to be developed by higher education institutions (Chalmers & Partridge, 2013). Therefore, the Indonesian government, through the Directorate of Higher Education (DIKTI) and Directorate of Islamic Higher Education (DIKTIS), has provided a policy for university lecturers to conduct student assessments. This assessment is aimed at ensuring similar or fair standards and contributing to the quality of higher education (Panadero et al., 2019).

The policy has provided guidelines for lecturers to conduct assessments, which are written in the curriculum document of the university. Furthermore, the guidelines consist of assessment principles, purposes, methods, instruments, assessment procedures, and procedures for giving final grades or grade appeals, etc. (DIKTI, 2018; DIKTIS, 2019). They also contain all the information about how students’ performance in the course will be assessed, including the type of assessment instruments, namely exams and exercises, the weight of each instrument, how the final grade will be calculated, and others. (Panadero et al., 2019). For example, the assessment techniques include observation, participation, work performance, written and oral tests, and questionnaires. The existence of various kinds of techniques aims to show the effectiveness of the assessment and the achievement of the students (Brown & Race, 2013). The guidelines provide the instrument used in assessing students, which is relevant to the curriculum change and response to the learning outcome (Yahiji et al., 2019). It also informs about the techniques for

assessing the three domains of students’ competencies. For example, observational assessment techniques and peer and self-assessment techniques are used to examine learners’ attitudes. According to Hattie and Timperley (2007), peers and self-assessment are very effective for assessing the aspects of attitude because they are very helpful in developing student learning in terms of self-regulation. Moreover, Willey and Gardner (2007) emphasized that both assessments have a positive effect. Therefore, they can improve learning outcomes and further motivate students to learn.

The knowledge domain of the students can be assessed by using written or oral tests, presentations, observations, and assessment techniques which refer to higher-level thinking-oriented assessments, case studies, and problem-solving tasks that require learners to think critically (Yuwono & Pasani, 2018). The assessment of knowledge should involve various techniques. However, written tests (examinations) in the form of summative are usually the main choice for teachers in various educational institutions (see, Kartono, 2011; Panadero et al., 2019). Meanwhile, formative often have little proportion in determining students’ achievement and are mostly neglected (Salirawati, 1998). Regarding the assessment of the skills domain, the teachers use the techniques of students’ performance, practices, simulations, and others. The performance-based assessment gives tasks that can show skills and encourage students to perform or think effectively and efficiently. This is aimed at helping students compete in the work world and solve their problems (Otaya, 2017).

Since the implementation of the assessment policy by the government, several growing studies have examined the assessment i.e Gahara, 2017; Nikmah & Makhshun, 2021; Nasution et al., 2019; Puad & Ashton, 2021; Saftari & Fajriah, 2019; Umami 2018; Saefurrohman, 2018; Zaim, & Arsyad, 2020; Yustitia & Wardani, 2017. For example, Puad and Ashton (2021) investigated Secondary EFL teachers’ views on classroom-based assessment at an Islamic boarding school, and the results showed that the teacher tends to view assessment from summative rather than formative perspectives. It was also demonstrated that there is an emphasis on assessing students’ attitudes and behaviour in classrooms alongside their academic ability. This is supported by Zulaiha et al. (2020), stating that teachers had the necessary knowledge of assessment principles and used that information in their instruction.

However, there was a gap between instructors’ knowledge and its application in classroom activities, notably during the implementation and monitoring phases. Some influences were also found on teachers’ practice of classroom assessment, such as the local or school policy, the use of nonachievement factors (e.g., students’ attendance and attitudes), and parents’ involvement in their children’s education.

Other studies focusing on authentic assessment (assessing students' attitudes, skills, and knowledge domains) have also been conducted i.e., Gahara, 2017; Nikmah & Makhshun, 2021; Saftari, & Fajriah, 2019; Umami, 2018. For example, Gahara (2017) examined the use of authentic assessment for Islamic religious education courses in junior high school. A qualitative case study approach was conducted by interviewing four teachers and two principals, while the data were obtained through observations and documentation. The results demonstrated that the success of authentic assessment could be seen in the integration of students' attitudes, knowledge and skills competencies. Therefore, the character competencies of students are being productive, creative, with attitude, skilled, and knowledgeable. The assessment of the three domains was also supported by Saftari & Fajriah (2019), who found that the techniques and instruments used in the 2013 Curriculum include the assessment of attitudes, knowledge, and skills competencies. Furthermore, Gahara (2017) showed that authentic assessments are carried out using various methods: tests, portfolios, projects, peers, and self-assessment.

Umami (2018) emphasized that assessment can be conducted in religious courses using various methods following the student's competencies. For example, the attitude domain can be assessed through observation, journal, peer, and self-assessment. Furthermore, the knowledge domain can be assessed by written tests, observation, and tasks, while skills are assessed by work performance, project, product, and portfolio. These results are supported by Nikmah & Makhshun (2021), finding that assessment techniques and instruments used by Islamic education teachers are a.) Attitude aspect, using observation techniques, peer-to-peer, journal, and self-assessments. b.) Knowledge aspect, using assignments and written and oral tests techniques. c.) The skills aspect is using practice test techniques, project appraisal, performance, and portfolio assessment.

Several studies have assessed students' outcomes in the three domains (attitude, skills, and knowledge). Several reports were also made at the secondary school level, and the majority of the subject were Islamic religious teachers. These studies have provided theoretical and practical contributions to the particular subject. However, only a few reports have been made at the university level on the subject of EFL lecturers. Additionally, none of the studies particularly explore the dominant method used by EFL lecturers in assessing students' competencies. Therefore, this study aims to fill this gap by exploring the dominant methods used by EFL lecturers in assessing students' competencies. It also explores the different assessment practices from 5 constructs concerning the competencies.

### 3. Method

This study aims to explore the practice of EFL lecturers in assessing the implementation of the current curriculum in the Indonesian higher education context. It also specifically examines the method used to assess the three EFL student competencies domains: attitude, knowledge, and skill. Furthermore, it aims to investigate the assessment practice related to the instrumentation, purpose, procedure, and reporting grade. A quantitative survey design was employed to have a more comprehensive understanding and clear picture of the assessment practice.

#### 3.1 Population and Sample

The population of this study is English lecturers joining the ELITE and LITA, two English Lecturer Associations in Indonesia, with 153 LITA and 253 ELITE members, respectively. Furthermore, a random sampling technique was used to recruit 71 English lecturers who responded to the questionnaire as the study sample. Table 3.1 presents the demographic information of the population.

**Table 3.1.** Demographic information

No	Characteristics	Variables	N	Percentage/%
1	Gender	Male	28	39.4
		Female	43	60.6
2	Degree	Master	48	67.6
		Doctoral	23	32.4
3	Teaching Experience	1-5 year	7	9.9
		6-10 year	22	31.0
		11 – 15 year	15	21.1
		➤ 16 year	27	38.0

It was found that two-thirds or 43 respondents were females, and more than one-third, or 28 were males. Most respondents have completed Master's degrees (48 people), and only 23 have Doctorate. Furthermore, most had more than 16 years of teaching experience, and only a few had 1 to 5 years. This implies the lecturers have much experience conducting an assessment.

### 3.2 Instrumentation and Data Collection

The data were obtained from the respondents using the Google form-based questionnaire. The items on the questionnaire were adapted from two available resources designed by Bouchaib (2016) and Onyefulu (2018). Furthermore, the questionnaire was originally designed to investigate the assessment practice and was modified to match the current context of the study. The questionnaire consists of three parts: The first has seven closed-ended items describing the respondents' informational background. The second part consists of seven open-ended statements measuring the method used to assess the student's competencies domain. Finally, the last part consists of 29 questions in the form of statements using the Five Likert Scale, namely Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree, that focus on the methods, purpose, instrumentation, procedures assessing students, reporting the final grade, and the barrier in conducting an assessment. Moreover, the questionnaire was validated by expert judgment before sending the Google Form link (<https://bit.ly/AngketAssessment>) to the respondents through a WhatsApp Group (WAG). Two lecturers with doctoral degrees in Language Education and Educational evaluation major gave their opinion about the questionnaire. In line with the data collection, the purposes of the study were explained in the questionnaire, and participation was voluntary.

### 3.3 Reliability and Validity

Cronbach's alpha ( $\alpha$ ) test was conducted to measure the instrument's reliability using the SPSS version 24. It is generally accepted that a questionnaire is reliable when the Cronbach alpha value is  $> 0.60$ . Based on the data processing results using the SPSS version 24, the overall reliability coefficient was 0.851, indicating that the questionnaire was reliable for this study. Meanwhile, the instrument's validity was measured using the expert's opinions. A practical approach to content validity is sought by experts' judgment (Furwana, 2019). Content validations are used to check whether a test was correctly designed or to validate the translation and standardization of an instrument used in a different culture (Fernández-Gómez et al., 2020). Two experts were asked to give their opinions on the instruments, specifically the questionnaire, for clarity, accuracy, and relevance to classroom assessment practices (Onyefulu, 2018). The instrument can examine student assessment practices with several improvements based on the process.

### 3.4 Data Analysis

Before analyzing the data, the responses to the questionnaire were downloaded from Google form as a Microsoft excel sheet, coded, and imported to the SPSS application version 24 for further analysis. Furthermore, descriptive statistics were used to find the frequency, mean, standard deviation, and other numeric details. Inferential statistics of independent sample t-test was employed to determine whether there were differences between educational degrees. Additionally, Analysis of Variance (ANOVA) was applied to analyze whether there were differences in teaching lengths, while the significant level of 0.05 determined the variances.

**Table 3.2** Mean Score Interpretation

Range of mean score	Practices
1.00 – 1.90	Very low
2.00- 2.90	Low
3.00-3.50	Moderate
3.60-4.49	High
4.50- 5.00	Very high

According to Table 3.2, the mean score interpretation of the assessment practice of the participants ranges from 1 to 5. Therefore, the mean score between 1.00 to 1.90 indicates a very low practice of assessment, while 4.50 to 5.00 shows a very high frequency. This mean score interpretation was adopted and slightly modified from several studies (including Tuan, 2021).

### 4. Findings

Even though there are several studies on classroom assessment, only a few reported on the assessment practice of EFL lecturers concerning the method of assessing learners' competencies, including attitude, skills, knowledge, and other related constructs covering method, purpose, and instrumentation, procedure, assessment score, and barrier. Therefore, this study aims to fill this gap by exploring the dominant assessment methods used by EFL lecturers in assessing students' competencies. It also aims to explore the assessment practices from five constructs (method, purpose and instrumentation, procedure, assessment score, as well as barrier) concerning the competencies. Four questions guide this study, namely 1) What methods are mostly used in assessing students' competencies? 2) How are the lecturer's assessment practices concerning the method, purpose, instrumentation, procedure, reporting grade, and barrier? 3) Is there a difference in assessment practice between master's and doctoral degree lecturers? 4) Are assessment practice differences among the length of teaching experiences? A survey design was used to

answer these questions by distributing questionnaires to participants while the data were obtained and analyzed using SPSS version 24. The analysis results of the questionnaire are presented based on the research question.

#### 4.1 Methods Used in Assessing Student Competencies

Descriptive analysis was employed to find out the methods that are mostly used in assessing students' competencies by indicating the percentage of the participants who responded to the questionnaire. The assessment method used by the respondents in the three domains of student competencies was identified using descriptive statistics, such as Frequency and Percentage (%). The three students' competencies of the assessment method are presented in table 4.1.

**Table 4.1.** Frequency of assessment Method of attitude, knowledge, and skills.

Methods	Attitude		Knowledge		Skills	
	Freq.	%	Freq	%	Freq.	%
Observation	47	66.2	6	8.5	18	25.4
Participation	7	9.9	12	16.9	10	14.1
Performance	2	2.8	19	26.8	21	29.6
Questionnaire	10	14.1	8	14.1	2	2.8
Self-assessment	3	4.2	0	4.2	0	0
Presentation	0	0	2	2.8	3	4.2
Fieldwork	0	0	0	0	2	2.8
Test	2	2.8	24	33.8	15	21.1
<b>Total</b>	<b>71</b>	<b>100.0</b>	<b>71</b>	<b>71</b>	<b>100.0</b>	<b>100.0</b>

According to table 4.1 above, the lecturers use various methods to assess the students' attitudes. However, the method that was mostly employed was observation (66.2% or 47). This is followed sequentially by questionnaire (14.1%), participation (9.9%), and self-assessment (4.2%). Interestingly, two methods of assessing students' attitudes (performance and written test) domain were unlikely chosen by the lecturers. Only 2 lecturers (2.8%) used each of the written tests and performance methods to assess students' attitudes toward EFL teaching and learning. This indicates that various methods were used to assess the students' competencies. However, the method mostly employed by the lecturers in assessing the attitude domain is observation.

Table 4.1 presents the frequency and percentage of the assessment methods for the knowledge domain. The table also illustrates the various methods used in assessing students' knowledge domain. However, most lecturers choose the test method (33.8%). This test technique means that they either used written tests or oral. The second method used in assessing knowledge is asking students to perform their work (26.8%). Another method employed by the lecturer is participation (16.9%), followed by questionnaire (11.3%), and observation (8.5%). The last method is presentation with 2.8%. Therefore, the test was mostly employed, which indicates that the lecturers consider it an appropriate method for assessing the knowledge

domain. Besides, it follows the guidelines of the methods for assessing knowledge.

The frequency and the percentage of the methods for assessing the skills domain are also shown in Table 4.1. The table illustrates that lecturers also employ various kinds of methods in assessing student skills. Most of the lecturers chose performance (29.6%) and observation (25.4%) in assessing this domain. This is followed by the test (21.1%), and participation (14.1%), while the least methods were questionnaires and fieldwork (2.8% of each). Interestingly, the presentation was only chosen by 3 or 4.2% participants, whereas it is one of the methods for assessing students' skills. This result indicates that the most favourable method for assessing this domain is work performance, which is recommended by the assessment guidelines and policy.

#### 4.2 Lecturer's Assessment Practice

Descriptive statistics were conducted to find the mean scores of the assessment practice in the six constructs, including method, purpose, instrumentation, procedure, reporting grade, and barrier. Table 4.2 presents the mean scores and standard deviation of each construct of assessment practice based on the 5-point Likert scale. It is also shown that mean scores ranged from 3.63 to 4.50.

**Table 4.2** The profile of assessment practice

Constructs	Minimum	Maximum	Mean	Std. Deviation
Method	3.33	5.00	4.22	.41252
Purpose	3.00	5.00	4.40	.52475
Instrumentation	2.00	5.00	3.64	.65388
Procedure	3.13	5.00	4.43	.46572
Reporting	3.75	5.00	4.50	.42989
Barrier	2.67	5.00	4.19	.71207
			4.23	

According to Table 4.2, the overall mean score of the lecturer's responses to the questionnaire was 4.23. This point indicates that the assessment practice related to the method, purposes, procedures, and reporting, and the barrier was between "agree and very agree". Therefore, the overall mean score was categorized as high in accordance with the criteria given by Tuan (2021). The method subscale means a score of 4.22, indicates that the respondents used various methods to assess students' learning. Furthermore, the purposes subscale mean score of 4.40 reveals that the respondent had a clear purpose in conducting the assessment. The instrumentation subscale means the score was the lowest (3.60), indicating that the responses to this scale were "rather agree". This shows that the respondents

used one or two instruments in assessing student learning. The procedure subscale score was 4.42, which indicates that the respondents followed the procedures on the assessment guidelines. Meanwhile, the grade scale had the highest mean score ( $M=4.50$ ), indicating that the respondents prioritized the midterm and final examinations (summative assessment).

### 4.3. The Differences in Assessment Practice by Educational Degree and Teaching Experience

According to the statistical analysis using an independent sample t-test, the results of the different assessment practice based on the educational background (degree) is presented in Table 4.3.

**Table 4.3.** Assessment practice based on the educational degrees of the lecturers

Degree	N	Mean	Std. Deviation	F	Sig.
S2	48	4.17	.38199	.099	.754
S3	23	4.13	.38155		
Total	71	4.15	.38177		

The result of the independent sample t-test showed that educational background (degree) did not have any significant effect on the overall practice of the assessment ( $F(0.099) = .099, p < 0.05$ ). Accordingly, the two variables have similar assessment practices, indicating that one's degree level cannot significantly

influence the other's assessment practice. This is supported by the mean score of the two educational backgrounds, which indicated no statistically significant differences in the mean score of S2 degree (4.22) and S3 degree (4.24) in assessment practice.

The One-Way ANOVA was used to analyze the differences in assessment practice by four teaching experiences. The statistical analysis of the difference in

the assessment practice based on the teaching experience is presented in Table 4.4.

**Table 4.4.** Assessment practice based on teaching experiences

N	Mean	Std. Deviation	F	Sig.
7	4.33	.32264	.981	.407
22	4.25	.37962		
15	4.30	.30906		
27	4.15	.33456		
71	4.23	.34321		

The ANOVA results indicated that teaching experience did not have a significant effect on the assessment practice ( $F(2, 281) = 981, p < 0.05$ ). Accordingly, there were no differences in assessment practice among the four teaching experiences. Therefore, one's teaching experience cannot significantly influence the other's assessment practice.

This is also supported by the mean score of each teaching experience, which indicated no statistically significant differences among the four teaching experiences. Further, the Tukey-HSD post Hoc was conducted to identify the significant differences between groups accurately, and the results are summarised in Table 4.5.

**Table 4.5.** Tukey-HSD test and overall assessment practice.

(I) Teaching experience	(J) Teaching experiences	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1-5 year	6-10 year	.08189	.14899	.946	-.3107	.4745
	11-15 year	.03089	.15716	.997	-.3832	.4450
	>16 year	.18574	.14563	.582	-.1979	.5694
6-10 year	1-5 year	-.08189	.14899	.946	-.4745	.3107
	11-15 year	-.05101	.11497	.971	-.3539	.2519
	>16 year	.10385	.09861	.719	-.1560	.3637
11-15 year	1-5 year	-.03089	.15716	.997	-.4450	.3832
	6-10 year	.05101	.11497	.971	-.2519	.3539
	>16 year	.15485	.11057	.503	-.1365	.4462
>16 year	1-5 year	-.18574	.14563	.582	-.5694	.1979
	6-10 year	-.10385	.09861	.719	-.3637	.1560
11-15 year		-.15485	.11057	.503	-.4462	.1365

According to Table 4.5, the results of Tukey-HSD Post Hoc indicate that there were no differences between 1-5 years ( $p = 0.582$ ), 6-10 ( $p = 0.719$ ), 11-15 ( $p = 0.503$ ), and more than 16 years of experiences

( $p = 0.503$ ). In terms of the mean score differences, it is noticeable that the score was above 4.00 for all the lengths of teaching experiences. The highest mean score was obtained in 1-5 years of teaching experience

(M4.33, SD= 0.32), while the lowest was found in 16 years or more (M=4.15, SD=0.33). This indicates no relationship between the length of teaching experiences and the assessment practice. In other words, the assessment conducted across the teaching experiences had similar practices, which means that the lecturers in this current study followed the guideline in the assessment policy document.

In summary, the results showed that lecturers used various assessment methods to evaluate students' competencies across three domains: attitude, skills, and knowledge. The main conclusion drawn from the research questions on these issues is that each student's competency was assessed using a particular methodology. For instance, the attitude domain was assessed mainly by observation, skill by work performance, and the knowledge domain by the testing method. Another significant result is related to the assessment practice in general, which showed that the lecturers highly implemented the policy.

This can be seen from the mean score of the assessment practice constructs. In addition, the teacher's adoption of the assessment practice, which indicated that there were no discrepancies across levels of educational degrees or among four teaching experiences, is evidence that the assessment policy was implemented.

## 5. Discussion

This study aims to explore the assessment practice of EFL lecturers in terms of the methods used in assessing three students' competencies covering attitude, skills, and knowledge as demanded by the curriculum at the Indonesian English lecturers' community, ELITE and LITA. Furthermore, it aims to explore the assessment practice related to method, purpose, instrumentation, procedure, reporting grade, and barrier based on the lecturer's educational background and the length of teaching experiences in relation to the student's competencies.

The results demonstrated various methods employed by the lecturer to assess the students' three domains. It was also found that every domain has its assessment method that the lecturers favour. For example, the attitude domain was mainly assessed by observation techniques, skills by work performance, and knowledge by testing techniques. The findings indicated that the lecturer's degree or educational background and the length of teaching experience did not show any differences in assessment practice. Moreover, the lecturers followed the guideline in assessing students' competencies.

Concerning the results, several important points were discussed based on the proposed research questions to provide new insight for EFL lecturers in assessing students learning.

### 5.1. Method used in assessing attitude, knowledge, and skills

The results showed that various methods were used for research question 1, which is related to the method used by EFL lecturers in assessing the three students' competencies domains. It was also found that specific techniques mainly assessed certain competencies. For example, the attitude domain was mainly assessed by observation, skills by work performance, and knowledge by testing techniques. This indicates that the lecturers were knowledgeable about the assessment policy and guidelines provided by the Directorate of Islamic Higher Education (DIKTIS) (2019).

The results related to the assessment of student attitude, in which observation was the most used, supported the previous studies in a similar context i.e., Kusaeri, 2019; Maba & Mantra, 2017; Yang, 2006). For example, Kusaeri (2019) emphasized that students' attitudes could be primarily assessed through observation. Similarly, Maba and Mantra (2017), concerning the assessment models employed by Indonesian elementary school teachers, reported that the majority of student attitudes might be evaluated through observation. This is in line with Kasmah (2018), which stated that lecturers apply observation and participation methods in assessing the student learning process. Meanwhile, in this current context of the study assessment of attitude, the domain emphasized spiritual assessment, including religious or moral issues, and social attitudes, such as discipline, self-confidence, responsibility, courtesy, honesty, and tolerance (Kurniati & Khaliq, 2019). The similarity between these results and the previous ones may be due to the deep understanding of EFL lecturers about the assessment guidelines and policy of the university, which is under the ministry of religious affairs. Although the assessment policy is from the Directorate of Islamic Higher Education, the directorate works by adapting from Indonesian Higher Education (Ministry of Education, culture, research, and technology).

Based on the assessment of the knowledge domain, this study showed that the test (written and oral tests) was commonly employed by EFL lecturers. According to previous studies, teachers commonly used testing techniques in assessing the knowledge domain (Alfiriani, 2016; Ermawati & Hidayat, 2017; Gan & Davison, 2017; Kurniati & Khaliq, 2019). For example, Alfiriani (2016) found that teachers generally choose written tests to assess students' knowledge. Similarly, Kurniati and Khaliq (2019) emphasized that written and oral tests were among the four types selected to assess students' knowledge. Ermawati and Hidayat (2017) found that the knowledge domain was assessed through examinations divided into daily, mid-term, and final tests. The test was employed in the current study for the assessment of the knowledge domain and is also recommended by the assessment policy through guidelines.

Besides this recommendation, there are several reasons for using the technique in assessing knowledge. For example, multiple choice or other tests can assess learners' engagement in higher levels of cognitive reasoning, such as application and synthesis of knowledge or higher-order thinking skills (HOTS) (Singh, & De Villiers, 2012; Zaidi et al., 2018). However, it was argued that using tests, such as multiple-choice, does not reflect current theories of learning and cognition nor measure students' learning achievement and the abilities they acquire for future success (Yani, 2017).

The skills domain is one of the competencies recommended to be assessed in the current KKNi-based curriculum implementation. This study showed that work performance was one of the most frequent assessment methods used by EFL lecturers in assessing students' skills. This indicates that the lecturers used the assessment guidelines by the Ministry of Religion (2019). This report aligns with previous studies (Kurniati & Khaliq, 2019; Maba & Mantra, 2017; Rukmini & Saputri, 2017). For example, Kurniati and Khaliq (2019) found that performance was used to assess students' skills, especially in English.

Similarly, Listiyawati et al. (2021) found that performance-based and peer assessments, as well as portfolios, were mainly used to assess students' skills. Furthermore, Rukmini and Saputri (2017), on the authentic assessment to measure students' productive English skills based on the 2013 curriculum, found that English skills, such as speaking, were assessed through performance with the grading system using a rubric. Work performance can be categorized as an authentic assessment since it requires the students to perform a task (Maba & Mantra, 2017) which shows the students' actual skills. Additionally, Ermawati and Hidayat (2017) emphasized that the attitude and skills domains have a significant role in showing the student's actual ability after they have learned the theory.

The work performance can be assessed individually, in groups, or in pairs. Performance-based assessment has a significant role in the teaching and learning process. According to Norris (2009), performance-based assessment helps provide diagnostic feedback, which assists students in improving their learning. It is used for making summative decisions about the student's skills. Moreover, it can improve their learning awareness because it requires the students to prepare and practice for their performance. VanTassel-Baska (2013) emphasized that performance-based assessment helps the lecturers obtain information about students' strengths and weaknesses, their needs, and what they know to make decisions for future instruction. This study showed that most lecturers use performance-based assessment, although some select different methods for assessing skills domain. One possible reason performance-based assessment is ineffective is

that the lecturers have limited knowledge, competence, and experience in assessing skill domains. Second, the lecturers have limited knowledge in choosing the assessment instrument.

## **5.2. Assessment practice on assessing attitude, knowledge, and skills**

Research question two concerning the assessment practice includes a method, purposes, instrumentation, procedure, reporting the grade, and barriers in assessing students. The results showed that the respondents indicated "agree and very agree" with the statements written on the questionnaire. This indicates that the lecturer's assessment practice was very high. In another word, the lecturers commonly conducted the assessment based on the guidelines recommended by the university and The Directorate of Islamic Higher Education (DIKTIS, 2019). However, the significant finding that needs to be discussed in this research question is the assessment instrumentation and reporting of the grade. This study showed that the lecturers responded with "rather agree" to the statement on the instrumentation variable's item. This indicates that the lecturers may have difficulty designing the instrument for their assessment practice, especially rubrics not used by the teacher (Rahmawati et al., 2019). This may be related to the instrument requirement, which must cover validity and reliability (Bariah, 2019). Moreover, the difficulty in developing instruments for assessment is also found in the previous studies (i.e. Azizah, 2018; Ermawati & Hidayat, 2017). Ermawati and Hidayat (2017) found that the lecturers have limited knowledge about the instrument in assessment.

The previous studies emphasized that the lecturers use their ways of designing instruments because they do not have formal assessment training. Therefore, they primarily draw on students' experiences or work within the faculty/department/area instructional style (Rodríguez-Gómez et al., 2013). Several studies have recommended that the difficulty in designing instruments can be solved by reading literature about assessment and upgrading knowledge through training for students. In this current context, the lecturers could also use the assessment guidelines provided by the government and the university.

In terms of the student's final grade (reporting variable), the significant results of this subscale showed that the lecturers reported the students' final grade by accommodating four components of grade, namely Assignment, Participation, and Midterm and final tests. This indicates that summative assessment was the priority of the lecturers in assessing the students since the final grade was mainly taken from the test type in the form of Mid and Final tests. Both tests also emphasized the student's knowledge domain which may neglect their attitude and skills. This does not correlate with the students' learning outcome as

mandated by the curriculum (see. DIKTIS, 2019; Permendikbud, 2020). Additionally, the results showed that the summative assessment contradicts the principle of authentic assessment, which emphasizes assessment *for* and *as* learning. The assessment *for*, which was carried out during the learning process, focuses on assessing the process, which aims to improve itself through formative (Black & Wiliam, 2006; Newton, 2007).

Formative assessment is conducted by obtaining data on improving students' learning outcomes, understanding the competencies or teaching materials studied and formulating the information. Subsequently, the most effective learning activities are decided for students in order to effectively understand the competencies given (Adinda, et al, 2022). Additionally, Earl (2003) emphasized that learning assessment generally aims to monitor and guarantee learning outcomes. The focus on the summative assessment using test techniques found in this study is similar to the previous reports (i.e. Friantary & Martina, 2018). For example, Friantary and Martina (2018) evaluated EFL teachers in assessing students' outcomes of the implementation of the 2013 curriculum (K13). It was found that the written test techniques were dominantly employed to assess the knowledge domain of the students. They also emphasized that the student learning outcomes assessment implemented by the Indonesian EFL teachers in their study slightly differs from the standard provided by the guidelines.

### **5.3. Assessment practice based on the educational degree and teaching experience**

Concerning the research questions related to the differences in the assessment practice for assessing students' competencies, as well as the educational degree and the length of teaching experiences, the results showed there was no statistical difference in the assessment practice between master's and doctoral degree lecturers. This is similar to the assessment practice based on the length of teaching experiences. The analysis results also showed no statistical differences in assessment practice among the four categories of teaching experience. This indicates no relationship between the educational degree and the length of teaching English with the assessment practice.

The possible reason for this is that the lecturers are in the same department; therefore, they can share assessment experiences. Another potential explanation is that the lecturers have the same convenient ways of assessment, regardless of their educational degree and the length of teaching experiences. This is because they follow the policy of the Directorate of Islamic Higher Education on assessing students concerning the method, purpose, instrumentation, procedure, and reporting grade. This result is in line with previous studies. For example, Beenstock and Feldman (2018) stated that lecturer in the same department tends to have a similar

scoring method (assessment). This may apply to the current study, whose sample is from the English department. Therefore, lecturers learn and imitate each other in conducting assessments. This is in line with Scarino (2013), stating that lecturers learn new assessment methods, tools, and assessment activities on their job.

This study discussed the critical point in relation to the assessment method used by EFL lecturers and the practice generally used in assessing the three competencies. As it was presented, the results have been compared and contrasted with the previous studies. It was also found that the lecturers followed the government's guidelines in assessing the student's competencies. Two implications can be drawn from this study. Firstly, lecturers should consider using the appropriate methods in assessing students learning in the current curriculum. Secondly, they must use the policy guideline the university and the government provide. This study is limited because it only examines the method of assessing students' competencies. Therefore, a further survey is recommended to be conducted in relation to the problem and barriers to implementing assessment policy.

## **6. Conclusion**

This study contributes to the understanding of the assessment practice carried out by EFL lecturers in assessing the three domains of competencies, namely attitude, skills, and knowledge at the university level. The results significantly showed that EFL lecturers employed various methods to assess students' competencies. Furthermore, they followed the assessment guidelines provided by the government and university. The other key novelty finding is that the assessment of students learning in the three domains should consider the method, purpose, instrumentation, and procedure for conducting the assessment in order to have accurate results and also achieve the learning goals. The following recommendations are put forward to lecturers, quality assurance institutions, curriculum developers, and the university. First, the lecturer is recommended to improve the quality of knowledge and skill in evaluating and assessing learning by following the guidelines issued by the university, DIKTIS, and DIKTI. Second, they should use these results as a source of assessment literature. Third, they are advised to attend seminars, workshops, or training related to learning and assessment in the current curriculum. Fourth, the university should provide all facilities for implementing reliable and valid assessments.

## **7. Acknowledgment**

This study was fully funded by Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi with grant Number B- /L.II/PP.00.9/05/2021. The authors are grateful to the Rector of Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi for the grant.

## References

- Adinda, D., & Denami, M. (2022, February). "Emergency course" During the covid 19 pandemic: How to support students' competencies development?. In *ICoNvET 2021: Proceedings of the 4th International Conference on Vocational Education and Technology, ICoNvET 2021, 27 November 2021, Singaraja, Bali, Indonesia* (p. 350). European Alliance for Innovation.
- Alfiriani, A. (2016). *Evaluasi pembelajaran dan implementasinya*. Padang Sukabina Press
- Beenstock, M., & Feldman, D. (2018). Decomposing university grades: a longitudinal study of students and their instructors. *Studies in Higher Education*, 43(1), 114-133. <https://doi.org/10.1080/03075079.2016.1157858>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in education: Principles, policy & practice*, 25(6), 551-575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Chalmers, D. & Partridge, L. (2013). Teaching graduate attributes and academic skills. In L. Hunt & D. Chalmers (Eds.), *University teaching in focus: A learning-centered approach* (pp. 56-73). London: Routledge. <https://doi.org/10.4324/9780203079690-4>
- Custer, R. L., & et al. (2000). *Using authentic assessment in vocational education. clearinghouse on adults, career, and vocational education*. The Ohio State University.
- Diktis (2018, June 23). *The PTKI curriculum development guide refers to the KKNI and SN-Dikti*. Ministry of Religion of the Republic of Indonesia. <https://fe.uin-malang.ac.id/panduan-pengembangan-kurikulum-ptki-mengacu-pada-kkni-dan-sn-dikti/>
- DIKTIS . (2019, June 23). Learning and assessment guidelines for PTKI. the ministry of religion of the republic of Indonesia. <https://pdfcoffee.com/pedoman-pembelajaran-dan-penilaian-pada-ptkipdf-pdf-free.html>
- Direktorat Jenderal Pendidikan Tinggi. (2012, June 22). Law of the republic of Indonesia number 12 of 2012 concerning higher education. <https://peraturan.bpk.go.id/Home/Details/39063/uu-no-12-year-2012>
- Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.
- Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.
- Ermawati, S., & Hidayat, T. (2017). Penilaian autentik dan relevansinya dengan kualitas hasil pembelajaran (persepsi dosen dan mahasiswa IKIP PGRI Bojonegoro). *Jurnal Pendidikan Ilmu Sosial*, 27(1), 92-103. <https://doi.org/10.30734/jpe.v5i1.145>
- Fernández-Gómez, E., Martín-Salvador, A., Luque-Vara, T., Sánchez-Ojeda, M. A., Navarro-Prado, S., & Enrique-Mirón, C. (2020). Content validation through expert judgement of an instrument on the nutritional knowledge, beliefs, and habits of pregnant women. *Nutrients*, 12(4), 1136. <https://doi.org/10.3390/nu12041136>
- Friantary, H., & Martina, F. (2018). Evaluasi implementasi penilaian hasil belajar berdasarkan kurikulum 2013 oleh Guru bahasa Inggris dan bahasa Indonesia di MTS Ja-Alhaq Kota Bengkulu. *Silampari Bisa: Jurnal Penelitian Pendidikan Bahasa Indonesia, Daerah, Dan Asing*, 1(2), 264-283. <https://doi.org/10.31540/silamparibisa.v1i2.202>
- Furwana, D. (2019). Validity and reliability of teacher-made English summative test at second grade of vocational high school 2 Palopo. *Language Circle: Journal of Language and Literature*, 13(2). <https://doi.org/10.15294/lc.v13i2.18967>
- Gahara, B. (2017). Implementasi penilaian autentik pada pembelajaran pendidikan agama Islam kurikulum 2013. *Tanzhim*, 1(1), 93-109. <https://doi.org/10.24952/fitrah.v2i2.470>
- Gan, Z., Oon, E. P. T., & Davison, C. (2017). ESL students' oral performance in English language school-based assessment: results of an empirical study. *Language Testing in Asia*, 7(1), 1-21.
- García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande-de-Prado, M. (2021). Recommendations for mandatory online assessment in higher education during the COVID-19 pandemic. In *radical solutions for education in a Crisis Context* (pp. 85-98). Springer, Singapore. [https://doi.org/10.1007/978-981-15-7869-4\\_6](https://doi.org/10.1007/978-981-15-7869-4_6)
- Ghafar, M. N. A. (2011). *Pembinaan & analisis ujian bilik darjah. Edisi Kedua*. Penerbit UTM Press.
- Hamodi, C., López-Pastor, V. M., & López-Pastor, A. T. (2017). If I experience formative assessment whilst studying at university, will I put it into practice later as a lecturer? Formative and shared assessment in Initial Lecturer Education (ITE). *European Journal of Lecturer Education*, 40(2), 171-190. <https://doi.org/10.1080/02619768.2017.1281909>

- Harun, C. Z., & Iqbal, M. (2018). Performance assessment of state senior high school teachers aged 56 years and above. *International Journal of Instruction*, 11(1), 33-46.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 71(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Ibrahim, M. M., Baharin, M. N., Mohamad, M. M., & Yukasof, Y. (2017). Innovative approaches to assessment: Develop a sense of direction to promote students learning. *Pertanika Journal of Social Sciences & Humanities*. Special Issue (25S), p149-155
- Kartomo (2011). *Efektifitas penilaian diri dan teman sejawat untuk penilaian formative dan sumative pada pembelajaran mata kuliah analisis komplek. Prosiding seminar Nasional Matematika*. Prodi matematika Universitas Muhammadiyah Surakarta. <https://doi.org/10.24269/js.v2i2.802>
- Kasmah, K. A. S. M. A. H. (2018). Konsep dasar evaluasi dalam kurikulum berbasis kkn. *Adaara: Jurnal Manajemen Pendidikan Islam*, 6(1), 555-570
- Kurniati, N., & Khaliq, A. (2019, September). Penilaian sikap, pengetahuan dan keterampilan dalam pembelajaran bahasa Inggris yang berorientasi kurikulum 2013. In Seminar Nasional Taman Siswa Bima (Vol. 1, No. 1, pp. 309-316). <https://doi.org/10.30738/caraka.v5i2.4832>
- Kusaeri, K. (2019). Penilaian sikap dalam pembelajaran Matematika [Assessment of attitudes in learning Mathematics]. *Jurnal Pendidikan Matematika (JPM)*, 5(2), 61-70. <https://doi.org/10.33474/jpm.v5i2.1588>
- Lake, R., & Olson, L. (2020). *Learning as we go: principles for effective assessment during the COVID-19 pandemic*. Center on Reinventing Public Education.
- Linn, R. L., & Miller, M.D. (2005). *Measurement and assessment in teaching (9th ed.)*. Upper Saddle River, NJ: Pearson Education, Inc.
- Listiyawati, L., Anggita, A., Wahyuni, R., & Heriyanto, H. (2021). Penilaian alternatif pada keterampilan berbicara bahasa Inggris di sekolah menengah kejuruan kota Pontianak [Alternative assessment on English speaking skills in Pontianak City Vocational High School]. *Eksos*, 17(1), 62-73. <https://doi.org/10.31573/eksos.v17i1.342>
- Maba, W., & Mantra, I. B. N. (2017). An analysis of assessment models employed by the Indonesian elementary school lecturers. *International journal of social sciences and humanities*, 1(1), 39-45. <https://doi.org/10.29332/ijssh.v1n1.38>
- Martha, A. S. D., Junus, K., Santoso, H. B., & Suhartanto, H. (2021). Assessing undergraduate students-learning competencies: A case study of higher education context in Indonesia. *Education Sciences*, 11(4), 189. <https://doi.org/10.3390/educsci11040189>
- Menristekdikti (2015, June 22). *Standar Nasional Pendidikan Tinggi (Number 44 of 2015 concerning SNPT)*. <https://peraturan.bpk.go.id/Home/Details/140595/permen-ristekdikti-no-44-year-2015>
- Nasution, H. F., Sudrajat, A., & Jahro, I. S. (2019). Analysis of affective assessment material in the textbook evaluation of Chemistry learning outcomes based on SNPT and Curriculum KKN. In 4th Annual International Seminar on Transformative Education and Educational Leadership (AISTEEL 2019) (pp. 727-729). Atlantis Press. <https://doi.org/10.2991/aisteel-19.2019.133>
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education*, 14(2), 149-170. <http://dx.doi.org/10.1080/09695940701478321>
- Nikmah, L., & Makhshun, T. (2021). Implementasi penilaian autentik dalam pembelajaran pendidikan agama Islam. *Prosiding Konstelasi Ilmiah Mahasiswa Unissula (KIMU) Klaster Humanoira*. <https://doi.org/10.19105/rjpai.v3i2.6358>
- Nitko, A. J., & Brookhart, S. M. (2012). *Educational assessment of students (6th ed.)*. Pearson.
- Norris, J. M. (2009). Task-based teaching and testing. In M. Long and C. Doughty (Eds.), *Handbook of language teaching* (pp. 578-594). Cambridge, MA: Blackwell. <https://doi.org/10.1002/9781444315783.ch30>
- Onyefulu, C. (2018). Assessment practices of teachers in selected primary and secondary schools in Jamaica. *Open Access Library Journal*, 5(12), 1-25. <https://doi.org/10.4236/oalib.1105038>
- Otaya, L. G. (2017). Konstruksi penilaian berbasis kinerja (Performance based assessment) dalam meningkatkan kemampuan penalaran mahasiswa pada mata kuliah statistika [Construction of performance based assessment in improving students' reasoning ability in statistics course]. *Tadbir: Jurnal Manajemen Pendidikan Islam*, 5(1), 28-51. <https://doi.org/10.24127/ja.v5i1.850>
- Panadero, E., Fraile, J., Fernández Ruiz, J., Castilla-Estévez, D., & Ruiz, M. A. (2019). Spanish university assessment practices: examination tradition with diversity by faculty. *Assessment &*

- Evaluation in Higher Education*, 44(3), 379-397. <https://doi.org/10.1080/02602938.2018.1512553>
- Pemerintah Indonesia (2020, January 29). Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 3 tahun 2020 Tentang Standar Nasional Pendidikan Tinggi. <https://ldikti13.kemdikbud.go.id/2020/01/29/peraturan-menteri-pendidikan-dan-kebudayaan-republik-indonesia-nomor-3-tahun-2020-tentang-standar-nasional-pendidikan-tinggi/>
- Puad, L. M. A. Z., & Ashton, K. (2021). Teachers' views on classroom-based assessment: an exploratory study at an Islamic boarding school in Indonesia. *Asia Pacific Journal of Education*, 41(2), 253-265
- Rahmawati, L. E., Suwandi, S., Saddhono, K., & Setiawan, B. (2019). Construction of test instrument to assess foreign student's competence of Indonesian language through objective test. *International Journal of Instruction*, 12(4), 35-48. <https://doi.org/10.29333/iji.2019.1243a>
- Retnawati, H. (2015). Hambatan guru matematika sekolah menengah pertama dalam menerapkan kurikulum baru. *Cakrawala Pendidikan*, 34(3), 390-403. <https://doi.org/10.21831/cp.v3i3.7694>
- Retnawati, H., Hadi, S., & Nugraha, A. C. (2016). Vocational high school Lecturers' difficulties in implementing the assessment in curriculum 2013 in Yogyakarta province of Indonesia. *International Journal of Instruction*, 9(1), 33-48. <https://doi.org/10.12973/iji.2016.914a>
- Reynold, C. R., Livingstone, R. B. & Wilson, V. (2010). *Measurement and assessment in Education*. Pearson.
- Rodríguez-Gómez, G. R., Ibarra-Sáiz, M. S. I., & García-Jiménez, E. G. (2013). Autoevaluación, evaluación entre iguales y coevaluación: conceptualización y práctica en las universidades españolas. *Revista de investigación en educación*, 2(11), 198-210. <https://doi.org/10.1344/reyd2019.19.29097>
- Rohl, M. (1999). Profiling ESL children: How teachers interpret and use national and state assessment frameworks. *Queensland Journal of Educational Research*, 15(1), 113-122.
- Rukmini, D., & Saputri, L. A. D. E. (2017). The authentic assessment to measure students' English productive skills based on 2013 curriculum. *Indonesian Journal of Applied Linguistics*, 7(2), 263-273.
- Saefurrohman, S. (2018, December). EFL teachers assessment methods in oral Communications. In *5th Asia Pasific Education Conference (AECON 2018)* (pp. 268-272). Atlantis Press.
- Saftari, M., & Fajriah, N. (2019). Penilaian ranah afektif dalam bentuk penilaian skala sikap untuk menilai hasil belajar. *Edutainment: Jurnal Ilmu Pendidikan dan Kependidikan*, 7(1), 71-81. <https://doi.org/10.35438/e.v7i1.164>
- Salirawati, D. 1998. Perlunya tes formatif dalam upaya peningkatan kualitas pendidikan di jurusan Kimia dan di jurusan lain pada umumnya. *Cakrawala Pendidikan*, Anniversary special edition, 191-201. <https://doi.org/10.31227/osf.io/nh3ea>
- Sathasivam, R. V., Samuel, M., Norjoharudden, M. N., Tee, M. Y., & Leong, K. E. (2019). Assessment for learning: Espoused and enacted practices of Malaysian teachers. *Pertanika Journal of Social Sciences and Humanifies*, 27, 47-62.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309-327. <https://doi.org/10.1177/0265532213480128>
- Singh, U. G., & De Villiers, M. R. (2012). The use of different types of multiple-choice questions in electronic assessment. *Progressio*, 34(3), 125-143.
- Thorn, D.W. and Deitz, J.C. (1989) Examining Content Validity through the Use of Content Experts. *Occupation, Participation, and Health*, 9, 334-346 <https://doi.org/10.1177/153944928900900602>
- Tuan, D. M. (2021). Vietnamese EFL teachers' perceptions and practices of reflective teaching as a tool for professional development. *REiLA: Journal of Research and Innovation in Language*, 3(3), 170-180. <https://doi.org/10.31849/reila.v3i3.7961>
- Umami, M. (2018). Penilaian autentik pembelajaran pendidikan agama Islam dan budi pekerti dalam kurikulum 2013. *Jurnal Kependidikan*, 6(2), 222-232. <https://doi.org/10.24090/jk.v6i2.2259>
- Uno, H. B., & Koni, S. (2012). *Assesment pembelajaran*. Bumi Aksara.
- VanTassel-Baska, J. (2014). Performance-based assessment. *Gifted Child Today*, 37(1), 41-47. <https://doi.org/10.1177/1076217513509618>
- Wiliam, D. (2011). What is assessment for learning?. *Studies in educational evaluation*, 37(1), 3-14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Yahiji, K., Otaya, L. G., & Anwar, O. (2019). Assessment model of student field practice at Faculty of Tarbiyah and Teaching Training in Indonesia: A reality and expectation. *International Journal of Instruction*, 12(1), 251-268. <https://doi.org/10.29333/iji.2019.12117a>

- Yang, H. (2006). A report of an ESL classroom observation in two language schools in Auckland. *TESL Canada Journal*, 23(2), 1-11. <https://doi.org/10.18806/tesl.v23i2.52>
- Yani, G. (2017). *Performance -based assessments on Students` speaking skill practiced by Senior High School English teachers in West Java* [Doctoral dissertation], Universitas Pendidikan Indonesia).
- Yustitia, V., & Wardani, I. S. (2017). Authentic assessment analysis based on the KKNI curriculum in applied statistics learning. In Ideas for 21st Century Education: *Proceedings of the Asian Education Symposium (AES 2016)*, November 22-23, 2016, Bandung, Indonesia (p. 55). Routledge. <https://doi.org/10.1201/9781315166575-9>
- Yuwono, I., & Pasani, C. F. (2018). The evaluation of higher order thinking skills assessment of special needs education students with guided inquiry method. *Journal of ICSAR*, 2(1), 28-31. <https://doi.org/10.17977/um005v2i12018p028>
- Zaidi, N. L. B., Grob, K. L., Monrad, S. M., Kurtz, J. B., Tai, A., Ahmed, A. Z., & Santen, S. A. (2018). Pushing critical thinking skills with multiple-choice questions: does Bloom's taxonomy work?. *Academic Medicine*, 93(6), 856-859. <https://doi.org/10.1097/acm.0000000000002087>
- Zaim, M., & Arsyad, S. (2020). Authentic Assessment for speaking skills: Problem and solution for English secondary school teachers in Indonesia. *International Journal of Instruction*, 13(3), 587-604. <https://doi.org/10.29333/iji.2020.13340a>
- 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.