EFL Writing Teachers' Practices and Values of Assessment for and as Learning in A Constrained Context

Ima Fitriyah (imafitria@iainkediri.ac.id)
Department of English, Faculty of Letters, Universitas Negeri Malang, Indonesia
State Islamic Institute of Kediri (IAIN Kediri), Indonesia

Imelda Gozali (imelda.gozali@ukwms.ac.id)
Department of English, Faculty of Letters, Universitas Negeri Malang, Indonesia
Universitas Katolik Widya Mandala Surabaya, Indonesia

Utami Widiati (utami.widiati.fs@um.ac.id) *Corresponding author
Department of English, Faculty of Letters, Universitas Negeri Malang, Indonesia

Niamika El Khoiri (niamika.el.fs@um.ac.id)
Department of English, Faculty of Letters, Universitas Negeri Malang, Indonesia

Amreet Kaur Jageer Singh (amreet@pbmpu.upsi.edu.my)
Centre for Languages and General Studies, Sultan Idris Education University, Malaysia

Abstract

Assessment for Learning (AfL) and Assessment as Learning (AaL) are gaining increasing prominence especially in university EFL writing lessons. This study sheds light on the practices and values of Indonesian EFL writing teachers towards AfL and AaL strategies during the constrained context of online learning setting because of the COVID-19 pandemic. By using descriptive statistics and thematic analysis, explanatory sequential mixed method was used in conjunction with a case study research design. Quantitative data were obtained from 54 teachers who responded to the AfL and AaL Strategy Questionnaire (AfL-AaLSQ), indicating their practices and values of AfL-AaL strategies. Meanwhile, the qualitative data were the results of interview sessions with seven selected participants, who described their practices and values in details. The findings show that all respondents accorded high importance to all AfL-AaL strategies and rated themselves as frequently implementing the AfL-AaL strategies. In both practice and values, the highest mean value was obtained for "providing assessment criteria", and the lowest for the "peer- and self-assessment". Interestingly, the charting of the mean scores shows a persistent gap between lower-rated practices and the higher-rated values. Finally, several reasons, including the effect of the constrained context, and ramifications of this discovery are examined in depth.

Keywords: AfL; AaL; constrained context; practices; values; and writing assessment

Introduction

The outbreak of the COVID-19 pandemic has dramatically transformed the educational landscape, forcing teachers to adapt to a constrained context of remote teaching and learning.

As a result, the process of assessment has become increasingly complex (Zou et al., 2021), with educators having to navigate the challenges of online assessment while maintaining a belief in its efficacy. In this new environment, there is a growing need for assessment to focus on process rather than just outcomes, utilizing approaches such as Assessment for Learning (AfL) and Assessment as Learning (AaL). However, implementing these strategies in the practice of online assessment requires a deeper understanding of EFL writing teacher practice and the values of AfL and AaL. This paper aims to explore these issues, examining the benefits and challenges of online assessment and the need to incorporate the principles of AfL and AaL into teaching practices in order to enhance student learning outcomes.

AfL and AaL have gained prominence in assessment policy statements worldwide due to their potential to improve students' learning and self-regulation skills (Alderson et al., 2017; Baird et al., 2017). As a subset of AfL, the pedagogical principles for AfL are applicable to AaL. Accordingly, the term AfL-AaL in this study is used to represent the integrated principles of both AfL and AaL. The widespread adoption of AfL-AaL strategies in classrooms has been initiated by some countries. While these approaches have been promoted at the policy level in many university EFL programs, there remains a lack of research on their implementation in writing courses, particularly in Indonesian language classes (Fitriyah et al., 2022; Nurhayati, 2020).

The shift to online learning because of COVID-19 pandemic has presented challenges for teachers seeking to implement effective AfL-AaL practices, as research has shown that many teachers lack sufficient assessment literacy and understanding of AfL-AaL principles (Ashraf & Zolfaghari, 2018; Fitriyah et al., 2022). In many parts of the world, including Indonesia, teachers struggle to access adequate resources and training to implement effective assessment practices (Fitriyah et al., 2022; Anam & Putri, 2021) during the constrained of online learning. As a result, teachers' assessment literacy is often inadequate, leading to ineffective implementation of AfL-AaL in classrooms. For example, some teachers are not fully aware of their own and their students' roles in the assessment process, and there are limited opportunities for students to collaborate and rate their peers' work (Fitriyah et al., 2022, Hentasmaka & Cahyono, 2021). In addition, some teachers may not understand certain AfL-AaL techniques as well as summative assessment techniques (Lee & Coniam, 2013; Zhao, 2018), further hindering their attempts to apply AfL-AaL strategies. This is particularly relevant in writing classes, where AfL-AaL principles are critical. The constraints of online learning are part of a broader context of limited support and facilities for assessment practices. For instance, teachers in Spain who practiced self- and peer-assessment reported technical difficulties in creating and monitoring group assessments remotely, leading to a decline in the use of such assessments during the pandemic (Panadero et al., 2022). Similarly, students in Turkey expressed dissatisfaction with the frequent use of "rapid assessment" by their instructors during the pandemic, due to the non-discriminatory nature of the test scores and the increased workload of remote teaching (Senel & Senel, 2021). Iranian teachers also faced challenges with online assessment, including the absence of directives from institutions on organizing assessments and the lack of financial support for necessary technology and connectivity (Ghanbari & Nowroozi, 2021).

While the world has transitioned to the endemic phase, online learning and assessment remain considerable alternatives in post-Covid19 educational landscape (Ewing, 2021). Thus, the challenges mentioned earlier still need to be considered when examining how AfL-AaL are being implemented in writing classes context. This study aims to investigate the practices and values of AfL-AaL as enacted by Indonesian EFL writing teachers during a moment of constraint in online learning. By using Lee's (2007) AfL-AaL framework to construct the questionnaire, this study seeks to shed light on the extent to which Indonesian university EFL writing teachers employ AfL-AaL techniques and how these strategies are perceived as

essential. This research is intended to portray the writing teachers' belief of the critical role of AfL-AaL in writing instruction. Investigating the practices and values in Afl-AaL is deemed critical due to the need to identify potential gaps between teachers' assessment approaches and implementation, and to adopt measures to narrow such gaps. This is particularly urgent in Indonesia, which stands in need of an assessment reform (Arrafii, 2021). In sum, the study intends to fill the research gap led by the following questions:

1. Do Indonesian university EFL writing teachers regularly employ AfL-AaL techniques, specifically in the constrained context?

2. To what extent are these AfL-AaL strategies seen as essential by Indonesian university EFL writing teachers?

3. Is there a gap between Indonesian EFL writing teachers' practices and values in implementing AfL-AaL and why is that so?

Literature Review

Implementation of AfL-AaL Principles in Writing Classes

Assessment encompasses the actions taken by educators and learners to evaluate and provide feedback in order to enhance teaching and learning (Black & Wiliam, 1998). AfL, a type of assessment, is particularly valuable in reinforcing learning and improving instruction. Studies have shown that AfL has a positive impact on students' motivation and performance (Ghaffar et al., 2020; Lee & Coniam, 2013; Nurhayati, 2020). AfL emphasizes the active involvement of both teachers and students in the assessment process through feedback, discussion, and observation, thereby promoting autonomy and self-regulation (Chong, 2018). In writing classes, Lee (2007, 2017) developed five AfL principles, which serve as the framework for this study. These principles include sharing learning objectives, helping students understand standards through assessment criteria, involving them in peer- and self-assessment, providing constructive feedback, and creating a safe learning environment so that mistakes are seen as a natural learning process. For easy reference, the AfL framework is shown in diagrammatic form as Figure 1.

Figure 1
The theoretical framework of this study following Lee's (2007, 2017) five AfL principles. The abbreviation is ours.
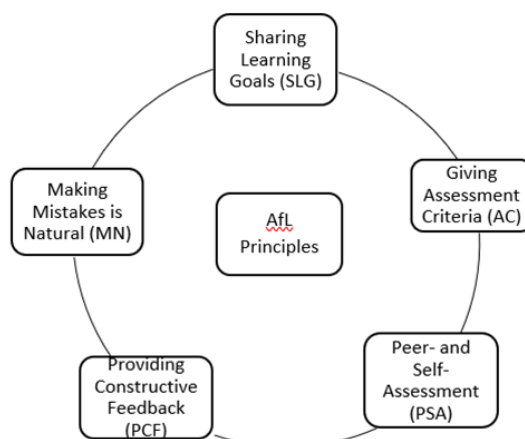

By following these principles, educators can effectively engage students in the learning process, providing them with the necessary support to succeed. Furthermore, assessment as learning (AaL), a subset of AfL, emphasizes the use of assessment to develop and support

students' metacognition (Earl, 2013), making students the critical connector between assessment and their own learning (Zhang et al., 2022). While AfL focuses on teachers as active participants in enhancing student learning and instruction, AaL places the emphasis on students themselves (Dann, 2014). AaL involves students in assessing their own learning and using that information to guide their future learning. The principles are students' creating personal learning objectives, being involved in assessment, being learning resources for one another, and taking responsibility for their own writing. These are all parts of the AaL principles featured in the AfL strategies (Lee, 2016). Figure 2 depicted herein elucidates the integrative relationship between AaL and AfL, presenting them as constitutive components within a comprehensive framework for the writing assessment.

Figure 2
AaL as part of the AfL principles (Lee, 2017)
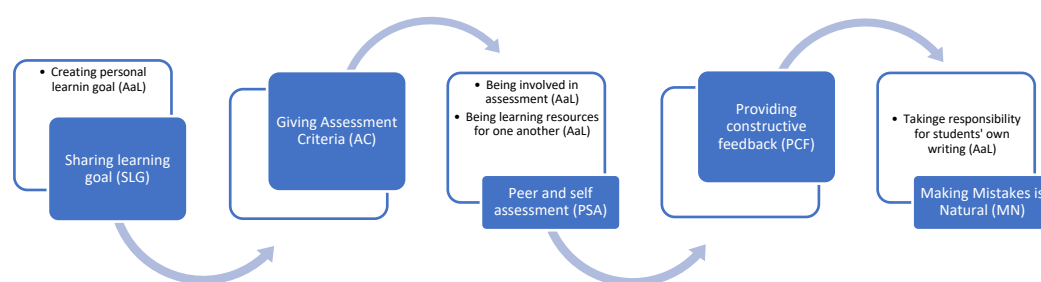
While there is growing interest in the application of AfL-AaL principles in writing classes and their potential to enhance students' motivation and performance, it is crucial to critically examine the evidence supporting these claims.. The existing research provides valuable insights into the principles of AfL-AaL. These studies have emphasized the importance of clear learning objectives, as demonstrated by Lee (2017), in promoting effective learning processes and assessments. Similarly, the utilization of rubrics or assessment criteria (AC), as indicated by Greenberg (2015), has been shown to elevate the quality of student writing. However, it is crucial to approach these findings with a critical lens, recognizing the need for further investigation to establish the generalizability of these principles and their potential applicability to diverse writing genres.

Nonetheless, the landscape of AfL-AaL principles is not without its complexities, as indicated by recent research. For instance, Panadero et al.'s (2022) survey of 936 language teachers reveals a concerning trend where certain assessment practices, such as involving students in assessment (the fourth principle according to Lee, 2017), deteriorated during the pandemic. This observation underscores the dynamic nature of education and the impact of external factors on the effectiveness of assessment practices. Furthermore, Wulandari et al. (2021) shed light on the nuanced perspectives of novice Indonesian EFL teachers regarding the implementation of peer feedback. Their study highlights the variability in perceptions regarding the utility and feasibility of this practice, demonstrating the need for a more comprehensive understanding of its impact on student learning. In an effort to synthesize these findings, Vuogan and Li's (2022) meta-analysis of 26 empirical studies offers a comprehensive perspective on peer feedback. They found that peer feedback appeared to be more effective for content than for language aspects. This suggests that educators may need to reevaluate their strategies for implementing peer feedback in the classroom, taking into account the specific learning objectives and subject matter at hand.

In summary, while the initial principles of AfL-AaL, such as clear learning objectives and the use of rubrics, have shown promise in promoting effective learning and assessment, it is evident that the landscape is far from uniform. The dynamic nature of assessment, coupled

with the evolving challenges posed by external factors like the pandemic, necessitates ongoing critical inquiry into these principles. Researchers and writing teachers alike must engage in further exploration to determine their generalizability and adaptability across various contexts, ensuring that assessment practices continue to align with the evolving needs of students and the education landscape. Thus, this research is necessary to investigate the implementation of AfL and AaL in writing classes.

Teachers' Belief of AfL-AaL Principles in Writing Classes

Numerous studies have examined the application of AfL-AaL principles by writing teachers (Fitriyah et al., 2022; Wu et al., 2021; Widiastuti et al., 2020). However, these practices are not supported by sufficient research on teachers' beliefs about the importance of AfL-AaL principles. Without a deep understanding of the value of these principles, their implementation may lack coherence and effectiveness (Wu et al., 2021). To have a more comprehensive discussion of AfL-AaL practices in writing classrooms, it is necessary to explore teachers' beliefs in the significance of applying AfL-AaL principles. When a teacher can apply the principles effectively, he also highly values them, leading to an ideal practice in AfL-AaL.

Research has demonstrated that different AfL-AaL principles are associated with distinct teacher beliefs. For instance, Jiang and Yu (2021) found that feedback was a crucial principle valued by writing teachers. Xiang et al. (2022) reported positive attitudes toward adopting AfL-AaL strategies in writing, while Wang and Lee (2021) revealed that teachers valued AfL-AaL for enhancing students' agentic engagement. Similarly, in Wang et al.'s (2020) study of Chinese EFL teachers, they ascribed greater value to AfL strategies that make learning explicit and promote students' autonomy compared to Assessment of Learning (AoL) approach, which mainly focuses on grading. However, Wu et al.'s (2021) survey-based research found that teachers in the Chinese context did not place high value on self-led AfL strategies and had low belief in self and peer assessment. Nevertheless, involving students in self-assessment is a critical component of AfL-AaL (Lee, 2017), and teachers should trust students and provide appropriate training for these strategies. Xu and Harfitt (2019) underscored the importance of involving students in self- and peer-evaluation and self-reflection, as AfL-AaL helps develop students' self-assessment skills, enabling them to become reflective and autonomous learners. Given the importance of AfL-AaL principles for both teachers and students, this research is needed to reconcile the potential differences in how much value teachers attribute to each of the AfL-AaL strategies.

AfL-AaL Practice in a Constrained Context

The term "constrained context" refers to the challenging conditions that hinder effective teaching and learning processes. Al-Seghayer (2014) identified four key constraints that could impede EFL students' performance: beliefs, curricular, pedagogical, and administrative constraints. For teachers, limitations arising from the teacher preparation program and teaching method constitute pedagogical constraints, while instructional time and teaching resources are considered curricular constraints. In recent times, the sudden shift to remote online learning during the COVID-19 pandemic can also be considered an additional constraint to the already challenging teaching and learning of L2 writing. This study focuses on teachers' practices and values of AfL-AaL in the emergency remote online writing class, which we regard as under a constrained context.

In the context of Indonesia, Arrafii (2021) conducted a study on Indonesian teachers' readiness to implement AfL as an assessment reform strategy and identified several barriers to its implementation. Making use of Kozma's (2003) framework of micro, meso, and macro

levels, he divided the constraints for successful assessment reform into internal and external-to-school factors. Internal factors include students, teachers, resources at school, and admission issues, while the government and parents' lack of ongoing involvement and support comprise the external causes. He concluded that, overall, there were more indications of challenges than of possibilities, leading to the perception that change was a daunting task to undertake.

During the COVID-19 pandemic, some studies have documented both success stories and challenges faced by instructors in practicing AfL-AaL as embodied in online assessment and peer feedback (Almossa & Alzahrani, 2022; Ghanbari & Nowroozi, 2021). For example, Almossa and Alzahrani (2022) studied the assessment practices used by teaching staff in Saudi universities during the COVID-19 pandemic. The results showed that the instructors implemented mostly feedback giving, linking assessments to learning objectives and learning outcomes, using scoring guides, and monitoring and revising assessment approaches. On the other hand, Ghanbari and Nowroozi (2021), who interviewed and elicited the narratives of 20 Iranian EFL university lecturers, unveiled the initial challenges that the teachers faced in implementing online assessment, starting from pedagogical to technical, administrative, and affective in nature. While ultimately the teachers were able to adapt to the online assessment process, the absence of directives and financial support from their institution for organizing online assessments remained.

Other studies conducted within the context of the pandemic investigated the extent of the adjustment that EFL teachers made in their assessment practices and the factors influencing such adjustment. Zou et al. (2021) identified three types of Chinese EFL teachers' engagement with online assessment: disturbing, auxiliary, and integral. In the disturbing engagement type, the teacher in their case study was resistant to the use of technology, resulting in a decrease in peer assessment practice during online learning. For the auxiliary type of engagement, the teacher showed mixed feelings about the use of technology and used it simply as an instrument to substitute face-to-face practice. However, the teacher who exhibited an integral engagement was excited about employing technology to promote independent learning in the students, as shown in peer-feedback training. Similarly, Zhang et al. (2021) studied Chinese EFL university teachers' changes in online assessment practices during the pandemic. They revealed that the changes can be divided into improvised and planned types, which are mediated by both contextual and experiential factors, with the latter exerting greater influence on the variability of the teachers' online assessment practices. While all of the aforementioned studies provide valuable insights on EFL teachers' changes in online assessment strategies in the context of the pandemic, research juxtaposing teachers' practice and values of AfL-AaL within a constrained context of the emergency remote online learning is still relatively scarce, and hence becomes the focus of this present research.

Method

Research Design

This study is designed primarily as mixed method research, in order to provide a comprehensive picture of a case by comparing quantitative and qualitative evidence from several angles (Creswell, 2012). The objective of using both quantitative and qualitative data was to also demonstrate the reliability of the study by collecting data from many sources, allowing for in-depth perspectives on the AfL-AaL practices and values. In this study, both quantitative data from a questionnaire and qualitative data from a set of interview questions are used to inform the result. In particular, the explanatory sequential design (Creswell & Clark, 2017) is employed. Essentially, it is a two-phase design in which quantitative data were first collected, followed by qualitative data that provide in-depth explanation of the quantitative

data analysis result. For this study, the application of the explanatory sequential mixed method approach is depicted as Figure 3. At the same time, this study is also framed as a case study that examined Indonesian EFL writing teachers' AfL and AaL practices and values in a constrained context.

Figure 3

Research Steps according to the Explanatory Sequential Mixed Method Design (adapted from Creswell and Clark (2017)).
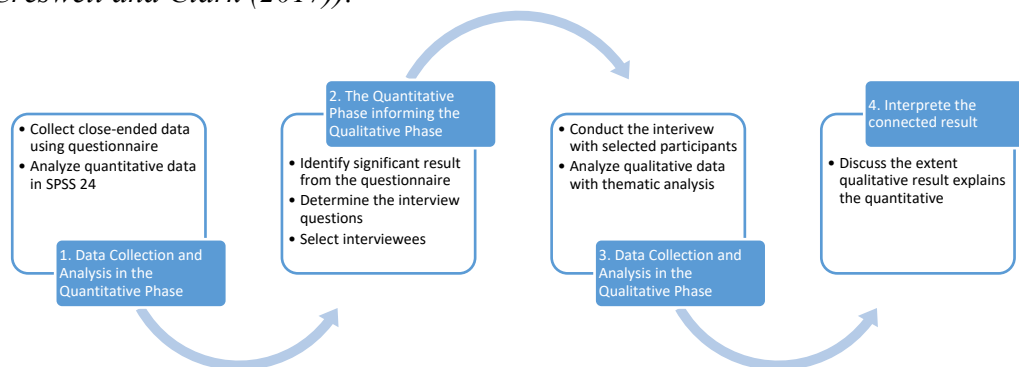
Instruments

In the quantitative part of the research, a questionnaire-based data was obtained to determine how frequently and how much Indonesian university EFL writing teachers employ and value fundamental AfL-AaL approaches, as reported by themselves. The Assessment for Learning and Assessment as Learning Strategy Questionnaire (AfL-AaLSQ) was utilized to gather data which answers all research questions. We developed our instrument by adapting the existing questionnaire AfL-SQ (Wu et al., 2021) and added Lee's (2007, 2017) principles of AfL and AaL in the questionnaire. Accordingly, the 60-item dual Likert-type scale design is employed to evaluate both teachers' judgments of the frequency and values they attributed to their AfL-AaL practices by utilizing the same questions.

To ascertain the validity of the instrument, the first draft of the questionnaire was forwarded to two experts on research instrument development for review and verification to ensure content validity. Based on their input, several items were rewritten and others deleted. Following the content validation check of the instrument, the second draft of the questionnaire was tried out to 30 respondents. Next, a repeated-measures ANOVA in SPSS Version 24 was performed to determine the validity and reliability measurement of the questionnaire. The result of the tryout shows that one item is considered as invalid, and thus the invalid item was eliminated. Hence, the final set for both practices and values contain 29 items each (58 in total). The constructs of the items, as well as our proposed abbreviations for the constructs to facilitate analysis, can be seen from Table 1.

Table 1

Constructs of AfL-AaL SQ

The AfL-AaL constructs	Researchers' Abbreviations	Items
Sharing Learning Goals with students	SLG	3
Helping students understand the standards (Assessment Criteria) they are working towards	AC	4
Involving students in assessment (Peer and Self-Assessment)	PSA	10
Providing Constructive Feedback that helps	PCF	7
Creating a classroom culture where Mistakes are a Natural part of learning and where everyone can improve	MN	5

The statistic that is used to assess the extent to which the scales display consistency is Cronbach's alpha, with 0.6 being the alpha level which indicates a good level of conceptual relatedness among items. Analysis on the AfL-AaL practice questions produced Cronbach alpha of 0.953 and the AfL-AaL values questions is 0.971. These values are greater than 0.6 so it can be concluded that the questionnaire is reliable. A copy of the questionnaire is given in Appendix A.

For the subsequent, qualitative data gathering process, a set of interview questions was used to gain deeper understanding on the result of the questionnaire's data analysis and to answer the question on the reason why a gap exists between EFL writing teachers values and practices of AfL-AaL. Following the steps of explanatory sequential mixed method design as shown in Figure 1, the questions were posteriorly determined upon identifying the quantitative results which were deemed significant. The list of questions is given in Appendix B.

Participants

The participants are Indonesian EFL teachers who taught writing in a higher education level during the full online learning. Table 2 consists of the information of 54 participants who have voluntarily filled out the questionnaire.

Table 2
Demographic information of participants

Information	Categories	Frequency	Percentage
Gender	Female	36	66.7
	Male	18	33.3
Teaching experience	1-3 years	14	25.9
	3-5 years	16	29.6
	>7 years	24	44.4

In line with the procedure of the explanatory sequential mixed method design, the participants for the interview were selected through stratified random sampling from the pool of teachers who filled out the questionnaire and who had also indicated their willingness to be interviewed. Table 3 contains the demographic data of the interviewees.

Table 3
Profile of interviewed participants

Participant	Gender	Institution type	Teaching writing experience	Educational background	Students Level
P1	Male	Private	5 years	Master (Ph.D candidate)	Undergraduate students
P2	Female	Public	10 years	Doctor	Graduate students
P3	Male	Private	4 years	Master	Undergraduate students
P4	Female	Public	3 years	Master	Undergraduate students
P5	Male	Public	5 years	Master (Ph.D candidate)	Undergraduate students
P6	Female	Private	3 years	Master (Ph.D candidate)	Undergraduate students
P7	Female	Public	3 years	Master (Ph.D candidate)	Undergraduate students

Data Collection

As depicted in Figure 1, the first step in this research was collecting the quantitative data. The Questionnaire was transcribed in a Google Form format and the link was sent to the EFL writing teachers in WhatsApp groups or social media platforms for EFL university teachers. Teacherparticipants were given the necessary information to make an informed decision about whether or not to participate in the study, and to know that participation was optional. They were also informed that the data obtained were kept discreetly and in a manner that protected the identity of the respondents. When the deadline to complete the Questionnaire was reached, 54 responses were obtained.

Teachers who had completed the survey and agreed to be interviewed were the subjects of the second, qualitative phase. Of the 54 teachers who completed the questionnaire, 32 agreed to participate in the interview. Stratified random sampling (Creswell, 2012) was used to choose the interviewees. The selection criteria were based on the demographic data such as teaching experience, gender and educational background, in order to garner a proportional balance. In light of our limited resources, we were only able to interview seven of them. The seven was regarded as sufficiently diverse to capture the population of EFL writing teachers. While care was taken to avoid bias by ensuring that the selection of the interview participant in each stratum (teaching experience, gender, and educational background) was completely random, we acknowledge a small proportion of convenient sampling since not all who indicated their agreement to be interviewed returned our communication when contacted. To those who responded, an interview consent form was sent in advance, and they expressed their agreement by signing and returning the form. Then, approximately 60-minute semi-structured interviews were conducted and recorded. All the interviews took place via video conference, except one which was conducted face-to-face.

To triangulate the qualitative data, the interviewees were asked to provide the recordings of their online teaching, and two (P5 and P6) agreed. Two authors observed the videos independently in order to verify the interview data with the teachers' practice as shown in the videos.

Data Analysis

The closed questionnaire's responses were analyzed statistically using SPSS 24. Descriptive statistics (i.e. frequency counts and percentages) were calculated for all questions. The result of the calculation is in the form of Means, which were then interpreted based on the ranges of the frequency and the value of AfL and AaL as seen in Table 4. Inferential statistics were also used to examine the differences between the variables.

Table 4
The interpretation of the means obtained in terms of frequency or value

Means	Practice frequency	Level of values
<1	Never	Not important at all
1.01-2	Almost never	Not important
2.01-3	Sometimes	Neutral
3.01-4	Usually	Important
4.01-5	Always	Very important

The interview data (after they had been transcribed in full) were categorized through a process of qualitative thematic analysis (Miles et al., 2018). This process involves reading the data carefully, identifying key issues, and then organizing these issues into a set of broader categories. For example, one of the interview questions asked teachers about their assessment

practice in writing class. The question itself thus constituted the broad category within which answers (i.e. about the practice of assessment and the ways in providing feedback) were then analyzed. As suggested by Creswell and Creswell (2018), the data findings were validated by a validator in term of coding the result of interview. After the interview recording was transcribed, the relevant segments of the transcription were color-coded based on the broader criteria in accordance with the findings of the quantitative phase. The criteria used were "AfL-AaL practice", "AfL-AaL value", and "gap". In accordance with the forth step of the explanatory sequential mixed method design process (ref. Figure 1), data analysis also involved a comparison of the questionnaire and interview data; this allowed us to corroborate particular conclusions from two perspectives, to illustrate quantitative findings with qualitative examples, and to obtain a more meaningful understanding of why teachers answered particular questionnaire items in the ways they did. As for the video observation, the authors took note of the assessment practices of P5 and P6 as shown in the recording. For instance, both P5 and P6 were seen using Google Docs to conduct peer-assessment with their students.

Findings

In line with the explanatory sequential design research design and the research questions, the quantitative findings are presented first, followed by the qualitative ones, which were dependent on the quantitative results.

Teachers AfL-AaL Practices, Values and Their Gap from the Quantitative phase

Table 5 summarizes the descriptive statistics for the result of the first part of the Assessment for Learning and Assessment as Learning Strategy Questionnaire (AfL-AaLSQ) questionnaire, which deals specifically with the practice of Afl-AaL techniques of the university EFL writing teachers in their lessons during the time of constraint. In particular, the Means for the questions in each construct were aggregated to reveal the Mean per construct.

Table 5

Descriptive statistics for the Afl-AaLSQ questionnaire: Practice

No.	Construct	Mean	SD
1	Sharing learning goals with students (SLG)	4.12	0.60
2	Giving assessment criteria (AC)	4.21	0.54
3	Involving students in peer- and self-assessment (PSA)	3.83	0.65
4	Providing constructive feedback that helps (PCF)	4.00	0.53
5	Creating a classroom culture where making mistake is natural (MN)	3.99	0.54

Based on the interpretation given in Table 4, Table 5 shows that the participants generally practice AfL-AaL techniques, with the Means approaching 4 or slightly above 4 (interpreted as "always"). It can also be seen that AC is practiced the most (M = 4.21, SD = 0.54), and PSA the least (M = 3.83, SD = 0.65).

Next, the descriptive statistics for the second part of the AfL-AaLSQ questionnaire, which particularly assesses the teachers' values of the AfL-AaL techniques, is shown in Table 6. The complete result of the Means of each indicator could be seen in Appendix A.

Table 6

Descriptive statistics for the Afl-AaLSQ questionnaire: Value

No.	Construct	Mean	SD

1	Sharing learning goals with students (SLG)	4.40	0.61
2	Giving assessment criteria (AC)	4.47	0.52
3	Involving students in peer- and self-assessment (PSA)	4.28	0.59
4	Providing constructive feedback that helps (PCF)	4.32	0.59
5	Creating a classroom culture where making mistake is natural (MN)	4.43	0.57

From Table 6 and the interpretation given by Table 4, the Means for the Afl-AaLSQ values are all above 4 (interpreted as "very important") with the highest value being placed on AC (M = 4.47, SD = 0.52), and the least on PSA (M = 4.28, SD = 0.59). Hence, this result appears to mirror that of part 1 as shown in Table 4, which also reflects teachers' AfL-AaL practice being most frequent for AC and least frequent for PSA.

In order to represent graphically the comparison between the Means for the part 1 ("practice") and part 2 ("value") of the Afl-AaLSQ questionnaire result, the Means were charted on a line graph as shown in Figure 4.

Figure 4

Comparison of the Means for each construct between the "practice" and the "value" of AfL-AaL techniques
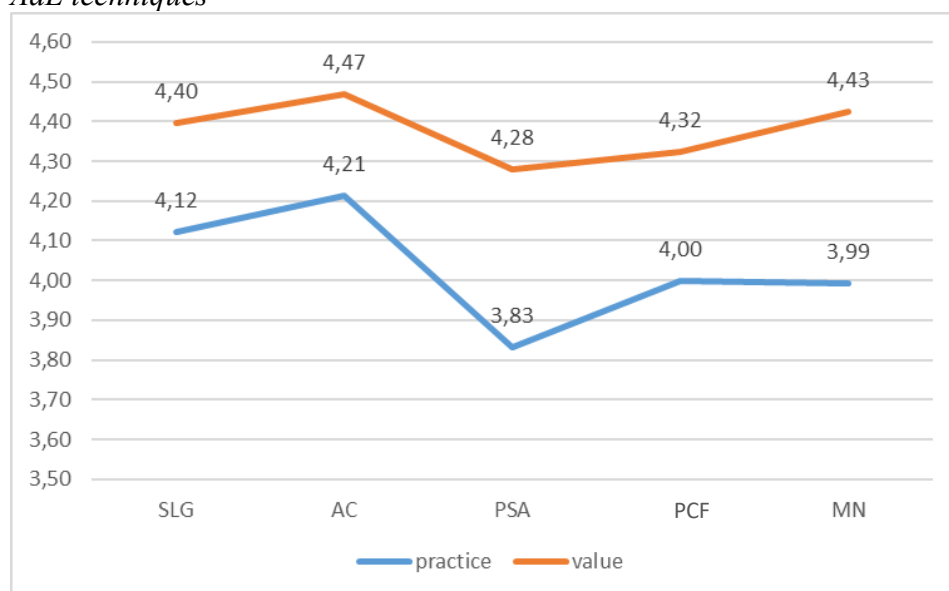

Figure 4 displays the gap between the Means for the teachers' AfL-AaL values as compared to the practice, with the values generally being higher than the practice. In order to ascertain whether or not the "practice vs value" difference between the Means for each construct is statistically significant, the data were first tested for the normality of its distribution, in order for the paired t-test computation to be validly performed. The result of the normality test using the Kolmogorov-Smirnov method is shown in Table 7.

Table 7

Normality test result

	Kolmogorov-Smirnov		
	Statistic	df	Sig.
Practice	.129	54	.026
Value	.115	54	.072

As in in Table 7, both the p values for "Practice" and "Value" data are smaller than 0.05, which means that the data are not normally-distributed and, therefore, parametric test cannot be used. Therefore, the Wilcoxon Signed Rank test for non-parametric data was used instead. The result shows that Z score is -5.494 with the Asymptotic Significance (2-tailed) of .000 (p < 0.05). Hence, it can be concluded that there exists a significant difference between the Means of the "Value" and "Practice", with the positive sign indicating that the Means of the "Value" are greater than "Practice". Finally, a summary of the gap between "Value" and "Practice" is given in Table 8.

Table 8

The gap between the Means of the "Practice" and "Value" of AfL-AaL

No.	Construct	M1 (Practice)	M2 (Value)	M2-M1
1	Sharing learning goals with students (SLG)	4.12	4.40	0.28
2	Giving assessment criteria (AC)	4.21	4.47	0.26
3	Involving students in peer- and self-assessment (PSA)	3.83	4.28	0.45
4	Providing constructive feedback that helps (PCF)	4.00	4.32	0.32
5	Creating a classroom culture where making mistake is natural (MN)	3.99	4.43	0.44

From Table 8, it can be seen that the greatest gap (M2 – M1) occurs in the PSA category (0.45), and the smallest in AC (0.26).

Teachers' AfL-AaL Practices, Values and Their Gap from the Qualitative phase

The interview questions (Appendix B) were drafted based on the light shed by the quantitative data. Since the quantitative data demonstrated the greatest gap in PSA and the smallest in AC, the interview findings were focused on these two constructs.

 a. **AfL-AaL Practice**

From the interview result, it can be concluded that all the EFL writing teachers in this study "helped students understand the standards they are working towards" ("AC"), albeit in a varied manner. For example, P1 used writing template so that students will not be lost in writing their essays. P2, P4 and P3 used writing guideline and modelling, although P4 also provided simple rubric in point forms. P5 and P6 used scoring rubric for the writing tasks. Meanwhile, P7 utilized questions in providing the guideline for the students. Hence, the interview results seem to corroborate the high means of AC practice in the quantitative result. Relevant extracts from some participants are shown here:

> P5: "After describing the learning objective I also provide the scoring rubrics or description, so before we start writing they know the areas that they have to focus on throughout the writing process".

> P2: "I first of all invited my students know the course outline, the goal, the model, and also all the criteria of academic writing that we would like to learn in this semester, all clearly stated in the course outline".

When it comes to "involving students in assessment" ("PSA"), the interview result demonstrated the differing practices of the participants. P5, P6, and P7 conducted peer

assessment but not self-assessment. P2 practiced both peer- and self-assessment using Automated Writing Evaluation (AWE) tools, although she expressed her preference for peer-assessment. In contrast, P4 encouraged her students to use Grammarly, plagiarism checker, and paraphrasing tools, and she hardly asked the students to do peer-assessment. P3 practiced both peer- and self-assessment but in an unstructured manner; he used codes in giving feedback to the students, and they will have to self-correct or ask their friends to interpret the feedback. Lastly, P1 did both peer and self-assessment by following an academic writing textbook and using the materials provided there. Therefore, the low means of PSA practice in the survey result appear to reflect the apparent inconsistencies in the participants in practicing PSA; they might have rated themselves low because they prefer peer rather than self-assessment, or vice versa. The variety of the participants' PSA practice and some representative quotes are presented in Table 9.

Table 9
The variety of PSA practice among the participants

No.	PSA Practice	Participants	Representative Quotes
1	Peer-assessment only	P5, P6, P7	P5: No, I have never asked my students to do self-assessment. Because from my point of view, when they first learn writing, which is introductory paragraph writing, the first level, they have only little background of the writing process
2	Self-assessment only	P4	P4: I have not done peer feedback … I am afraid that it won't get done, because only certain students were active. Besides, if a student is paired with another who is not competent, the student won't gain much.
3	Peer- and self-assessments	P1, P2, P3	P1: I used a textbook, which has self- and peer-assessment activities. So, I first asked them to do self-assessment following the book, and then they did peer-assessment by exchanging their works.

In addition, the constrained context of having to assess students online did produce some negative impacts on most of the participants, although P5 was able to employ technology in order to overcome online assessment barriers. Some of the problems are related to, for example, giving feedback during synchronous meetings; P3 lamented that it was not so effective due to unstable connection. Doing peer feedback is also seen to be harder in online setting for P4, since she was unable to supervise the process closely, as she herself said:

> P4: "It should be theoretically important huh. It's just that, maybe it just took the right time, meaning the right media to conduct peer assessments. If offline it will be easy to monitor and so on".

During online peer feedback session, P1 resorted to asking students to swap their works through WhatsApp group, and so the process became slower and rather unstructured. P6 also put forward the problem of plagiarism during online learning, in which students produced works that seem to exceed their capability.

> P6: "When it is onsite, I got my students to write on the spot, on a piece of paper. When it is an online, I found that some students whom I know are average in competence suddenly produced good quality essay".

P5, on the other hand, was able to leverage on the technological affordances by, getting his students to write synchronously on Google Docs and activating the "live edit" feature, and so he was able to see whether or not the students truly write their own works or copy-paste from other sources. He also thinks that he could give feedback faster in online mode than offline, making use of the automated writing evaluation tools in Microsoft Word. In general, teachers continue to attempt to execute the AfL-AaL technique, despite the fact that technical hurdles make it realistically suboptimal.

b. AfL-AaL Values

The interview result unveiled the belief of most participants, who put high premium on the value of giving assessment criteria ("AC") at the beginning of the course. In their opinions, it is important for the students to be familiar with the assessment criteria upfront so that they are aware of the success criteria and to plan the appropriate strategy to achieve the learning goals. In P2's own words: 'Having equal understanding [of the learning goals] is very important. If not all the students understand the final goal of the course, then they would not know what strategy that they will do to achieve the goal'. Interestingly, P1, drawing from his own personal experience, is of the opinion that providing a detailed rubric has the potential of increasing students' anxiety. However, the overall tenor of the interview result for this part is still in line with the corresponding quantitative part, namely the high means for the participants' values of AC.

The interview results also provided a glimpse of why the participants' values on PSA is lowest in the quantitative part; it seems that not all appreciate the utility and benefits of PSA. While some believe in the necessity of PSA for students' learning and teachers' workload reduction, others think that PSA is time-consuming, cognitively demanding, and unsuitable for the students, considering their English proficiency and mentality, as illustrated in the subsequent paragraphs. In term of conducting peer assessment online, three participants (P4, P5, and P6) agree that making students to be assessors to their friends is not easy and should be done carefully especially when their English proficiency is low. The followings are the interview excerpts in this regard.

> P5: "This is somewhat confusing from my perspective, to make sure that they actually know how to do the peer and self-assessment. So, it is not only practically demanding but also cognitively demanding for both sides, I mean for the students and also the teacher".

> P6: "PSA might be difficult to do in practice. In my class, some of the students lack English competence, so they don't know how to comment on their friends' works, or sometimes I observed that their comments were not so accurate. … On the other hand, I also think that the criteria of PSA (after reading the questionnaire again) are too idealistic. It will be difficult to achieve all those goals in the reality of day-to-day EFL teaching".

Therefore, P4, for example, opted to do peer feedback only on selected writing features which she deems easier for the students, such as only for grammatical mistakes or errors at sentence level, as shown here by her comment:

> P4: "I have the heart to do peer-assessment on grammar, because the rights and wrongs are clear; the answers are evident. As long as it doesn't follow the theory,

then it's wrong. I don't have the heart to ask them to do peer feedback on the content".

P5 and P6 added more points that, in their opinions, make PSA difficult to achieve in practice, namely adding to teachers' workload and being unsuitable for the habit of Indonesian students. In addition, P6 thought that self-assessment could only be done through automated writing evaluation tool which is only accessible through paid subscription.

P5: "When we do peer and self-assessment that means more workload, like having many students doing self-assessment, we check their self-assessment, and then we check their final assessment, [This is] very demanding for me".

P6: "Perhaps in Indonesia we are not so used to being reflective, … Once I tried to conduct a self-reflection activity, and then later some students gave me feedback saying that I should skip this part".

During online assessment, some teachers (P1, P4) were hesitant to over-burden their students with work, so peer feedback task was not set as compulsory, and in the end only the few diligent students who did the peer feedback. This is what P4 said in this regard:

P4: "Yes, [I did not do peer assessment], because I am afraid that it won't get done, because, during online learning, only certain students were active. If I ask them [to do error analysis], those who reply are the same students".

In contrast with the previously-mentioned three participants, four of them (P1, P2, P3, and P7) believe that giving a chance for the peer to provide feedback would be good for both the students and the teachers. The students could make reflection from their friends' work, and the teachers save time off correcting mistakes. The followings are statements from P1, P2, and P7:

P1: "In SLA [Second Language Acquisition], students will do better academically if they study with friends because they will be more comfortable making errors, will be more open to constructive feedback, and will be less likely to feel embarrassed by their failures…".

P2: "I do not specifically focus on giving feedback on their grammatical use, I ask them to do peer feedback concerning grammar, …, then I focus more on the content and organization".

P7: "At paragraph level, I believe [the students] can do [peer-assessment]. By reading their friends' work, they can see whether their works are better or worse. Anyway, I provided them with guiding questions [for peer-assessment] and I gave the final feedback".

In sum, there is a great disparity in the participants' perceived value of peer- and self-assessment, which might have explained the low means score of PSA values in the quantitative result. A summary of the differing PSA values and the resulting practice among the participants is given in Table 10.

Table 10
The differing PSA values and the resulting practice among the participants

No.	PSA Values Participants	Reasons	Resulting Practice

1	Peer-assessment (negative)	P4, P5, P6	1. Practically and cognitively demanding for students and teachers (P5, P6) 2. Difficulty imposed by online learning (P4)	Infrequent peer feedback (P5, P6) or peer feedback only on simple grammatical feature (P4)
2	Self-assessment (negative)	P5, P6	1. Extra workload for teachers in conducting both self- and peer-assessment (P5) 2. Unsuitable for Indonesian students' mentality (P6) 3. Can only be done through expensive online tool (P6)	Did not do self-assessment
3	Peer-assessment (positive)	P1, P2, P3, P7	1. Feedback from friends are less intimidating (P1) 2. Teachers can focus on global aspect when giving feedback (P2) 3. Students are capable of generating feedback for simple grammatical features (P3, P7) 4. Students can learn from one another (P7)	1. Peer feedback on simple grammatical feature (P2, P3, P7) 2. Online peer feedback is casual and not compulsory (P1) 3. Provide scaffolding and final feedback (P7)

c. Gap between AfL-AaL Practice and Values

Lastly, the interview result and all that have been mentioned earlier gave an inkling on the reason behind the gap between the practice and value of Afl-AaL in constrained context as indicated by the survey participants. The biggest gap occurs in PSA due to the participants' mixed sentiments about the values and practice of PSA. As previously mentioned, four interviewees gave high rating on PSA values since they are convinced of its benefits and effectiveness for students' learning. However, the remaining participants were more skeptical in terms of the suitability of PSA gauging from students' English competence, mentality, and teachers' workload. All of these is exacerbated by the online setting, which affected most of the teachers in performing the Afl-AaL due to their perceived insufficiency in technological skills, lack of students' academic integrity, and psychological burden to the students and the teachers themselves.

Discussion

EFL Writing Teachers' Practice of AfL-AaL in The Constrained Context

The findings reveal that EFL writing teachers practiced AfL-AaL principles in the constrained context to a substantial extent as seen by the result of the questionnaire, with the highest means score occurring in "helping students understand the standards they are working towards" or "AC" (M = 4.21), lowest in "involving students in assessment" or "PSA" (M = 3.83). In other words, the AfL-AaL principles were applied in the "always" category except for the PSA principle (only "usually").

The high score on AC practice means that the AfL-AaL principles that are implemented most frequently are giving clear learning goal and providing assessment criteria for the students. This result is consistent with the teachers surveyed by Panadero et al. (2022), who similarly reported that the use of rubric remained unaffected during the pandemic. This stability in utilizing assessment criteria suggests that teachers recognize the significance of providing

clear learning goals and assessment criteria to students. It aligns with the concept of pre-writing instructional scaffolding proposed by Lee (2017), which emphasizes the role of these strategies in supporting student learning and connecting learning outcomes to assessment. Lee (2017) further argues that effective AfL practice entails collaborative assessment planning between teachers and students, emphasizing the integration of instructional scaffolding with assessment. Therefore, the high score on "AC" practice implies that teachers in this study recognize the importance of not only setting learning goals but also considering how to assess the writing assignments and using assessment criteria to inform their teaching. This holistic approach to teaching and assessment promotes a deeper understanding of the subject matter and enhances the overall learning experience. Teachers who did not use rubrics may need to give it more consideration, since students who used rubric while preparing their reports were shown to write a higher quality report than students who did not (Greenberg, 2015). Students also improved the quality of their own reports after using the rubric to grade a report written by one of their classmates.

The results of this study demonstrate that teachers did not consider themselves always practicing the PSA principle in AfL-AaL, either because of the online learning condition, or partly because of individual preferences to peer- or self-assessment. Likewise, the teachers in Panadero et al.'s (2022) study admitted that they found it difficult to carry out both peer- and self-assessment during the pandemic, thus reducing the frequency of peer-assessment by 50%. On teachers' perception on the feasibility of peer- and self-assessment in the Indonesian context, this study is strongly aligned with that of Wulandari et al. (2021). Just like the teachers in their study, our participants are split into one group who firmly believe in and practice self- and peer-assessment, and another group who express doubts on students' ability to carry out effective peer assessment. Hence, we would like to echo the call of Panadero et al. (2022) for a greater emphasis to assessment literacy in teachers' professional development and a clear institutional regulation with regards to formative assessment practice (peer- and self-assessment).

The concern regarding the students' low English proficiency caused some teachers to practice selective peer feedback, namely limiting the scope of peer feedback to vocabulary and grammar, which the teachers think the students are more capable of. This is in contrast with the evidence put forward by Vuogan and Li (2022), who indicated that peer feedback was more effective on content rather than language (lexis and grammar). Hence, teachers in the present study might need to reconsider how they can apply a more comprehensive peer feedback in the classroom (Winarti et al., 2021).

The constrained context experienced by the participants in this study posed additional challenges in implementing AfL-AaL, although some were able to maximize technological affordances to their advantage. Similar difficulties in enacting online assessment have been reported in several studies (Almossa & Alzahrani, 2022; Ghanbari & Nowroozi, 2021; Payant & Zuniga, 2022). However, teachers have been shown to adapt their instructional and assessment strategies to address the pedagogical, technical, administrative, and affective challenges associated with online assessment as the course progressed (Fitriyah & Jannah, 2021; Ghanbari & Nowroozi, 2021; Zou et al., 2021). Unlike the findings of Zou et al. (2021), none of the participants in our study seems to fall under the "disturbing" type of online assessment engagement, since all our teachers were open and able to operate one or several digital platforms. Hence, they fall somewhere between the "auxiliary" and "integral" mode, as evidenced by their use of Google Docs for peer-assessment and Grammarly, plagiarism checker, and paraphrasing tools for self-assessment. Our findings also support those of Zhang et al. (2021) in their analysis of the contextual factors affecting Chinese EFL teachers' change in assessment practice during the pandemic. In particular, teachers' conception towards assessment and its impact to students' lives resulted in the unwillingness to cause assessment-

induced stress to the students during the pandemic, which was also reported by the participants in our study.

EFL Writing Teachers' Value of AfL-AaL Principles

The result shows that teachers assigned high value to all of the principles of AfL-AaL. The quantitative findings reveal that the means scores for all principles are above 4.00, which indicates the category of "very important." Among the high scores, teachers' AfL-AaL value was highest for AC (4.47) and lowest for PSA (4.28). Hence, teachers reported that providing writing assessment criteria (AC) seems to be the most important thing that teachers have to apply in assessing writing.

The high regard that teachers have on giving assessment criteria is consistent with the literature (Wang & Lee, 2021; Xiang et al., 2022), which demonstrate that making assessment rubrics transparent to students can help them improve their writing quality and self-regulation. The participants in this study opined that this practice is a must in writing classes, whether online or offline. However, teachers in this study should consider the finding of Xiang et al. (2022), who shows that their students were involved in the process of co-constructing assessment criteria and developing their own learning goal. This kind of practice may seem like a utopian dream to be implemented in the lower students' proficiency context of this present study, as lamented by many of the participants. Nonetheless, Lee (2016) points out that putting students at the center of assessment and learning does not mean that teachers entirely withdraw from students' learning. Instead, writing teachers need to take specific and focused pedagogical actions to gradually expand students' learning opportunities for collaboration and reflection.

Meanwhile, PSA or involving students in assessment was highly-valued but rated the least among the five AfL-AaL principles. The result of the interview supports this finding. Three teachers believe that this practice is quite important, but they regarded their students' level of writing ability to be limited in order to conduct peer or self-assessment effectively. This finding accords with studies conducted in China (Wang et al., 2020; Wu et al., 2021), who similarly revealed the low value of student-led AfL-AaL strategies (self- and peer-assessment) ascribed by EFL teachers there. Regrettably, the scant belief in the value of promoting students' autonomy through AfL-AaL might have dire consequence to students' motivation and achievement, specifically in L2 writing (Wang et al., 2020). Thus, similar to the EFL teachers in China, some paradigm shift in the concept of assessment, coupled with tackling challenges at various contextual levels, is needed to boost teachers' PSA values.

Wu's et al.'s (2021) study also reveals the contextual factor behind the teachers' low score of PSA value. Due to a culture that requires hierarchy in teaching and tends to be top-down, there is still room for improvement in the teachers' trust in students (Wu et al., 2021). A similar situation occurred in our research. Some teachers have not had the courage to allow students to provide peer criticism or self-assessment, while others have ventured to do so based on a number of criteria. This study reports that the greatest concern regarding peer feedback is the low English proficiency of the students, which was perceived by the teachers to have a negative effect on the validity and reliability of feedback. The lack of English competence prevents students from providing comprehensive and constructive corrective feedback, and teachers limit the scope of peer feedback to vocabulary, grammar, and syntax. There is also another possibility that the students in our context prefer to have teacher feedback than those coming from their peer, as also happened in other studies (Alharbi, 2022; Cheng et al., 2021). All of the above-mentioned factors may have contributed to the teachers' lower belief in PSA.

In their study, Xiang et al. (2022) shed light on the challenge teachers faced in applying AaL in the writing class, particularly in conducting peer review based on the established

criteria. They found that some students could not grasp each item's grading degree in the rubric, and it was also challenging for them to match the criteria with specific parts of the essay because of their limited writing and assessment ability. The students with low language proficiency admitted that they could not identify deep-seated problems and produce constructive and in-depth feedback after one round of review. Hence, teachers in this current study might need to provide more illustration and modeling of peer reviews from the teachers (or competent peers) to the lower-proficiency students before they can independently engage in peer and self-assessment (Wang & Lee, 2021).

Gap between Practice and Values in Implementing AfL-AaL

The quantitative findings indicated significant differences between teacher's practice and value on the AfL-AaL with the values being consistently higher than the practice. The gap is shown by the difference between the means of practice (M1) and the means of value (M2). This gap is smallest in AC (M2-M1 = 0.26), and biggest in PSA (M2 – M1 = 0.45). In short, teachers do not fully implement what they believe. The findings suggest a need to reconcile the theory-practice gap.

From the five principles of the AfL-AaL, the smaller gap in AC indicates that AC is generally both highly valued and frequently practiced by the participants. As mentioned earlier, the teachers in this study are convinced of the necessity of AC and consider it as a standard practice for EFL writing teachers. On the contrary, PSA is quite highly valued but least practiced by the participants. This is consistent with the findings of Wu et al. (2021), who similarly reported incongruence between Chinese EFL teachers' ideals and practice on AfL, with the largest disparity occurring in peer- and self-assessment. As previously-mentioned, the possible causes of such gap range from teachers' workload and time constraints, as well as power differential between teachers and students influenced by cultural context, which in turn engendered a certain amount of distrust between the two. The present study reflected similar situation, wherein some participants voiced out their concern regarding the lack of time and heavy workload. However, our findings add to Wu et al.'s (2021) by manifesting teachers' misgivings that students with lower English proficiency are unable to provide effective peer feedback. In addition, students' lack of ability for self-reflection was also cited by our participants, and thus adding to Wu et al.'s (2021) "teaching and learning culture" factor, in that teachers see themselves as the only credible assessors, so students hardly practice self-assessment.

The gap between AfL-AaL practice and value in our study is further widened by the constrained context that the teachers found themselves in. The participants in this study admitted to lack the technological know-how necessary to conduct effective online assessment in the emergency remote learning. They were also wary of students' potential dishonesty when completing their written works, and were aware of the possible psychological burden experienced by their students when studying from home. Thus, the teachers in this study refrained from carrying out peer- and self-assessment, thinking that this practice lacks practicality, utility, and empathy in the emergency remote online context. Our study thus adds further to the explanations of the gap by Wu et al. (2021) by including another contextual factor influencing the lack of alignment in teachers' AfL-AaL value and practice.

The factors influencing the gap between values and practice in AfL-AaL strategies can also be analyzed from the contextual factors at micro, meso, and macro level, as was done by Wang et al. (2020), who similarly revealed a gap between their teachers' belief in promoting learners' autonomy in assessment and their self-perceived practice of the same. Our results corroborate theirs at the micro level, namely student attribute (low English proficiency), as well as at the meso level, namely the school factor, which resulted in heavy workload of the teachers.

However, at the macro level, the cultural influence alluded to by our participants refers more to the Indonesian students' mindset, not very much accustomed to reflective thinking and self-evaluation. The macro level factor of Wang et al. (2020) pertains to Chinese educational policy of placing great emphasis on high-stake tests. Besides, our results also point to other micro-level factors conditioned by the constrained context of the emergency remote online teaching, namely teachers' perceived insufficiency in technological skill, their apprehension of students' academic dishonesty, and unwillingness to overload the students. Thus, we add more nuances to the multi-level contextual factors that explicate the value-practice gap in the teachers' AfL-AaL implementation.

The existence of the incongruence between AfL-AaL practice and value as evidenced in our study, as well as our postulate of the causes of such gap, carries several implications. Firstly, the fact that the gap was identified within a constrained context implies that the gap could potentially be narrowed outside of the said context. Hence, now that almost all classes have reverted to face-to-face mode, it is imperative that teachers, who supposedly believe in the importance of AfL-AaL principles, strive to put them into practice, specifically the peer- and self-assessment, which has been shown to be least deployed in this study. As previously-mentioned, technological skill acquired during the pandemic can fruitfully be leveraged in offline setting, such as the use of Google Docs for collaborative, peer-assessment. AI-powered tools such as Grammarly, Quillbot, and ChatGPT can potentially be effective platforms for students' self-evaluation.

Secondly, while the study focused on how teachers practiced AfL-AaL in EFL writing classes in Indonesia, the principles of AfL-AaL are highly relevant and applicable to other contexts and subjects as well. The development of autonomous and self-regulated learners is a goal of education systems worldwide in the 21st century, and AfL-AaL can be a valuable tool in achieving this objective. This research has the potential to inform and improve teaching practices in other contexts and subjects, and highlights the importance of implementing formative assessment strategies to support students' learning and development.

By way of a recommendation, we would like to reiterate the call made in previous studies (Panadero et al., 2022; Wang et al., 2020; Wu et al., 2021) for greater emphasis in assessment literacy training for pre-service teachers and in-service teachers' professional development. Besides responding to the factors at the micro level, such as training teachers in using technology for online assessment, the cultural barriers at the macro level will specifically need to be addressed. In particular, teachers in culture similar to Indonesia, where students are perceived to be passive and obedient, will need to learn to relinquish "their power" and allow students to take control of their own learning, which is the ultimate aim of the AfL-AaL principles. This might be the first step in the incremental approach to assessment reform in Indonesia, and countries of similar culture, as exhorted by Arrafii (2021).

Limitations and Recommendations

This study is not without its limitation. Firstly, the small sample size (54 participants for the quantitative phase, and seven for the qualitative phase) might have made the findings less generalizable. The collection of data through narratives and survey may not reflect the most evident practice of AfL-AaL. Nonetheless, this study combines quantitative and qualitative data that are complimentary. Observations on how AfL-AaL approaches may be used in writing classrooms in diverse circumstances can be supplemented by future studies.

Conclusion

The present study provides valuable insights into the perceived value of AfL-AaL principles among teachers, while highlighting the challenges associated with their effective implementation in practice. Although one of the principles was found to be difficult to implement due to students' capabilities, the overall positive attitude towards AfL-AaL ideals is encouraging. It is important to note, however, that the shift to online learning has created additional hurdles for instructors attempting to implement these principles effectively. While the current findings suggest that the use of AfL-AaL may be more feasible in a face-to-face learning environment, it is essential to acknowledge the ongoing role of online learning in post-pandemic education. Therefore, this research holds important implications for language teachers and educational administrators seeking to maximize the potential of digital affordances to enhance teaching and learning. By recognizing the challenges and opportunities associated with the use of AfL-AaL principles in online learning, educators can work towards developing effective pedagogical strategies that facilitate student acquisition of writing skills in a range of contexts.

References

Alderson, J. C., Brunfaut, T., & Harding, L. (2017). Bridging assessment and learning: A view from second and foreign language assessment. *Assessment in Education: Principles, Policy & Practice, 24*(3), 379–387. https://doi.org/10.1080/0969594x.2017.1331201

Alharbi, M. A. (2022). Exploring the impact of teacher feedback modes and features on students' text revisions in writing. *Assessing Writing*, *52*, 100610. https://doi.org/10.1016/j.asw.2022.100610

Almossa, S. Y., & Alzahrani, S. M. (2022). Assessment practices in Saudi higher education during the COVID-19 pandemic. *Humanities and Social Sciences Communications*, *9*(1), 1–9. https://doi.org/10.1057/s41599-021-01025-z

Al-Seghayer, K. (2014). The four most common constraints affecting English teaching in Saudi Arabia. *International Journal of English Linguistics*, *4*(5), 17-26. https://doi.org/10.5539/ijel.v4n5p17

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

Arsyad Arrafii, M. (2021). Assessment reform in Indonesia: Contextual barriers and opportunities for implementation. *Asia Pacific Journal of Education*, *43*(1), 79-94. https://doi.org/10.1080/02188791.2021.1898931

Ashraf, H., & Zolfaghari, S. (2018). EFL teachers' assessment literacy and their reflective teaching. *International Journal of Instruction*, *11*(1), 425-436. https://doi.org/10.12973/iji.2018.11129a

Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice*, *24*(3), 317–350. https://doi.org/10.1080/0969594X.2017.1319337

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Cheng, X., Zhang, L. J., & Yan, Q. (2021). Exploring teacher written feedback in EFL writing classrooms: Beliefs and practices in interaction. *Language Teaching Research*, 1-31 136216882110576. https://doi.org/10.1177/13621688211057665

Chong, S. W. (2018). Three paradigms of classroom assessment: Implications for written feedback research. *Language Assessment Quarterly*, *15*(4), 330–347. https://doi.org/10.1080/15434303.2017.1405423

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed). Pearson. http://nuir.nkumbauniversity.ac.ug/handle/20.500.12383/985

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. 780.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications, 453.

Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice, 21*(2), 149–166. https://doi.org/10.1080/0969594X.2014.898128

Earl, L. M. (2013). Assessment for learning; assessment as learning: Changing practices means changing beliefs. *Assessment*, *80*, 63–71. https://wlts.edb.hkedcity.net/filemanager/file/EBook_Assessment_and_Learning_Issue_2_2013.pdf

Ewing, L. A. (2021). Rethinking higher education post COVID-19. *The Future of Service Post-COVID-19 Pandemic, Volume 1: Rapid Adoption of Digital Service Technology*, 37-54. https://library.oapen.org/bitstream/handle/20.500.12657/46838/2021_Book_TheFutureOfServicePost-COVID-1.pdf?sequence=1#page=48

Fitriyah, I., Basthomi, Y., Khotimah, K., & Gozali, I. (2022). Implementation of assessment for learning in online EFL writing class: A case of novice undergraduate teachers. *LEARN Journal: Language Education and Acquisition Research Network*, *15*(2), 129-159. https://so04.tci-thaijo.org/index.php/LEARN/index

Fitriyah, I., & Jannah, M. (2021). Online assessment effect in EFL classroom: An investigation on students and teachers' perceptions. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, *5*(2), 265-284. https://doi.org/10.21093/ijeltal.v5i2.709

Fitriyah, I., Masitoh, F., & Widiati, U. (2022). Classroom-based language assessment literacy and professional development need between novice and experienced EFL teachers. *Indonesian Journal of Applied Linguistics*, *12*(1), 124-134. https://doi.org/10.17509/ijal.v12i1.46539

Ghaffar, M. A., Khairallah, M., & Salloum, S. (2020). Co-constructed rubrics and assessment for learning: The impact on middle school students' attitudes and writing skills. *Assessing Writing*, *45*, 100468. https://doi.org/10.1016/j.asw.2020.100468

Ghanbari, N., & Nowroozi, S. (2021). The practice of online assessment in an EFL context amidst COVID-19 pandemic: Views from teachers. *Language Testing in Asia*, *11*(1), 1-18. https://doi.org/10.1186/s40468-021-00143-4

Greenberg, K. P. (2015). Rubric use in formative assessment: A detailed behavioral rubric helps students improve their scientific writing skills. *Teaching of Psychology, 42*(3), 211–217. https://doi.org/10.1177/0098628315587618

Hentasmaka, D., & Cahyono, B. Y. (2021). Peer feedback uptakes and outcomes across EFL students' proficiency levels: A study at tertiary education in Indonesia. *International Journal of Instruction, 14*(3), 271-286. https://doi.org/10.29333/iji.2021.14316a

Jiang, L., & Yu, S. (2021). Understanding changes in EFL teachers' feedback practice during COVID-19: Implications for teacher feedback literacy at a time of crisis. *The Asia-Pacific Education Researcher, 30*(6), 509-518. https://doi.org/10.1007/s40299-021-00583-9

Kozma, R. B. (2003). Technology, innovation, and educational change: A global perspective: A report of the second information technology in education study: module 2.

International society for technology in education (ISTE). https://cir.nii.ac.jp/crid/1130000795455791232

Lee, I. (2007). Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *The Canadian Modern Language Review, 64*(1), 199–213. https://doi.org/10.3138/cmlr.64.1.199

Lee, I. (2016). Putting students at the centre of classroom L2 writing assessment. *The Canadian Modern Language Review*, *72*(2), 258–280. https://doi.org/10.3138/cmlr.2802

Lee, I. (2017) *Classroom Writing Assessment and Feedback in L2 School Contexts*. Springer Singapore. https://doi.org/10.1007/978-981-10-3924-9_8

Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing, 22*(1), 34–50. https://doi.org/10.1016/j.jslw.2012.11.003

Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). *Qualitative Data Analysis: A Methods Sourcebook.* Sage publications.

Nurhayati, A. (2020). The implementation of formative assessment in EFL Writing: A case study at a secondary school in Indonesia. *Pedagogy: Journal of English Language Teaching, 8*(2), 126-137. https://doi.org/10.32332/pedagogy.v8i2.2263

Panadero, E., Fraile, J., Pinedo, L., Rodríguez-Hernández, C., & Díez, F. (2022). Changes in classroom assessment practices during emergency remote teaching due to COVID-19. *Assessment in Education: Principles, Policy & Practice, 29*(3), 361–382. https://doi.org/10.1080/0969594X.2022.2067123

Payant, C., & Zuniga, M. (2022). Learners' flow experience during peer revision in a virtual writing course during the global pandemic. *System,* 105, 102715. https://doi.org/10.1016/j.system.2021.102715

Şenel, S., & Şenel, H. C. (2021). Remote assessment in higher education during COVID-19 pandemic. *International Journal of Assessment Tools in Education*, *8*(2), 181-199. https://doi.org/10.21449/ijate.820140

Vuogan, A., & Li, S. (2022). Examining the effectiveness of peer feedback in second language writing: A meta-analysis. *TESOL Quarterly*. https://doi.org/10.1002/tesq.3178

Wang, L., Lee, I., & Park, M. (2020). Chinese university EFL teachers' beliefs and practices of classroom writing assessment. *Studies in Educational Evaluation*, *66*, 100890. https://doi.org/10.1016/j.stueduc.2020.100890

Wang, L., & Lee, I. (2021). L2 learners' agentic engagement in an assessment as learning-focused writing classroom. *Assessing Writing*, *50*, 100571. https://doi.org/10.1016/j.asw.2021.100571

Widiastuti, I. A. M. S., Mukminatien, N., Prayogo, J. A., & Irawati, E. (2020). Dissonances between teachers' beliefs and practices of formative assessment in EFL classes. *International Journal of Instruction, 13*(1), 71-84. https://doi.org/10.29333/iji.2020.1315a

Winarti., Cahyono, B. Y., Mukminatien, N., & El Khoiri, N. (2021). Collaborative writing using process writing approach: The effect of group size and personality types. *International Journal of Instruction, 14*(4), 391-410. https://doi.org/10.29333/iji.2021.14423a

Wu, X. (Molly), Zhang, L. J., & Dixon, H. R. (2021). Implementing assessment for learning (AfL) in Chinese university EFL classes: Teachers' values and practices. *System, 101*, 102589. https://doi.org/10.1016/j.system.2021.102589

Wulandari, I., Purwati, O., Setiawan, S., & Anam, S. (2021). Indonesian novice EFL teachers' perceptions of the use of peer assessment in speaking class. *Journal of English Teaching*, *7*(1), 67–79. https://eric.ed.gov/?id=EJ1290873

Xiang, X., Yuan, R., & Yu, B. (2022). Implementing assessment as learning in the L2 writing classroom: A Chinese case. *Assessment & Evaluation in Higher Education, 47*(5), 727–741. https://doi.org/10.1080/02602938.2021.1965539

Xu, Y., & Harfitt, G. (2019). Is assessment for learning feasible in large classes? Challenges and coping strategies from three case studies. *Asia-Pacific Journal of Teacher Education*, *47*(5), 472–486. https://doi.org/10.1080/1359866X.2018.1555790

Zhang, X. (2020). Assessment for learning in constrained contexts: How does the teacher's self-directed development play out? *Studies in Educational Evaluation, 66*, 100909. https://doi.org/10.1016/j.stueduc.2020.100909

Zhang, X. S., Zhang, L. J., Parr, J. M., & Biebricher, C. (2022). Exploring teachers' attitudes and self-efficacy beliefs for implementing student self-assessment of English as a foreign language writing. *Frontiers in Psychology, 13*, 956160. https://doi.org/10.3389/fpsyg.2022.956160

Zhao, H. (2018). Exploring tertiary English as a foreign language writing tutors' perceptions of the appropriateness of peer assessment for writing. *Assessment & Evaluation in Higher Education, 43*(7), 1133–1145. https://doi.org/10.1080/02602938.2018.1434610

Zou, M., Kong, D., & Lee, I. (2021). Teacher engagement with online formative assessment in EFL writing during COVID-19 pandemic: The case of China. *The Asia-Pacific Education Researcher*, *30*(6), 487–498. https://doi.org/10.1007/s40299-021-00593-7

Appendix A

AfL-AaLSQ practice and values and the means

Factor	Indicator	Practice	Value
Sharing learning goals with students (SLG)	I define goals of a lesson or a series of lessons based on my students' previous learning	4.02	4.33
	I communicate goals to my students in ways they understand	4.39	4.56
	I encourage my students to set their goals in learning writing	3.96	4.30
Helping students understand the standards they are working Towards (AC)	I help my students understand success criteria by providing detailed descriptions	4.13	4.43
	I use examples to help my students understand what constitutes good writing	4.44	4.63
	I give students sample texts, mini-text analysis tasks, and text improvement tasks where they attempt to apply the criteria to evaluate the quality of the texts	4.24	4.43
	I familiarize themselves with the assessment criteria that will be used to revise their text	4.04	4.39
Involving students in assessment (PSA)	I encourage my students to reflect on what they have learnt during lessons	4.00	4.46
	I provide and explain guidance to help my students assess their own work	4.11	4.48
	I teach my students how to interpret feedback from their peers	3.63	4.06
	I give my students opportunities to comment on their own work during lessons	3.81	4.20
	I give my students opportunities to comment on their peers' work	4.04	4.30
	I deliberately develop an atmosphere in which my students feel safe and comfortable about giving and receiving feedback from their peers	4.04	4.50
	I provide guidance and training and vary the demands of self- and peer assessment according to students' abilities	3.65	4.24

	I provide and explain guidance to help my students give feedback to their peers	3.80	4.20
	I encourage self-inquiry where students not only reflect on their strengths and weaknesses in writing, but also formulate their own goals and take initiatives to improve their own writing	3.70	4.31
	I Let students engage in assessment that contains different focuses	3.56	4.04
Teachers providing constructive feedback that helps (PCF)	I provide immediate feedback during lessons to my students	3.76	4.13
	the feedback I give to my students specifies how to improve to achieve their learning goals	3.94	4.35
	I guide my students how to revise their work according to my feedback	4.07	4.33
	I give my students opportunities to discuss with me about my feedback	4.24	4.50
	the feedback I give to my students specifies the areas for remediation	4.02	4.28
	The feedback I give to my students specifies their strengths	3.98	4.20
	I provide quality feedback so that students learn about their strengths and weaknesses in writing	3.98	4.46
Creating a classroom culture where mistakes are a natural part of learning and where everyone can improve (MN)	I focus on fostering motivation and emphasizing progress and achievement during the assessment	4.09	4.52
	I create secure learning atmosphere, as students feel that making mistakes is natural	4.20	4.52
	Peers and self-assessment should not be harmful for both students and peers	4.00	4.48
	I give selective error feedback to avoid insecurity	3.72	4.26
	I devise strategies and design materials that help students work on relevant aspects of their writing	3.94	4.35

Appendix B

The Interview Questions

1.   Can you describe the process of evaluating your students' writing during the online learning, from giving learning goals, providing the model and criteria, to providing feedback? Did you do self-assessment and peer-feedback?

2.   For your answer in no. 1 above, how different was your evaluation practice during the pandemic (online learning) and pre/post-pandemic (on-site)?

3.   The questionnaire result shows that in general teachers practice and value AC (giving Assessment Criteria) the most, and PSA (doing Peer and Self-Assessment) the least. In your opinion, why is this so?