



SNAP TO READ

NEGOTIATING MEANING AND GRADES IN A SOCIOCULTURAL INVESTIGATION OF FORMATIVE CLASSROOM ASSESSMENT PRACTICES IN AN INDONESIAN ENGLISH LANGUAGE EDUCATION DEPARTMENT

I Putu Andre Suhardiana

*Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar, Indonesia
putuandresuhardiana@gmail.com*

Ida Bagus Gde Phalaguna

*Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar, Indonesia
ibphalaguna@gmail.com*

Astuti Wijayanti

*Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar, Indonesia
wijyanthi91@gmail.com*

Ida Ayu Oka Purnami

*Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar, Indonesia
idaayuokapurnami@gmail.com*

I Gusti Ayu Indah Triana Juliari

*Universitas Dwijendra, Indonesia
indah.triana@gmail.com*

First received: February 16, 2026

Final proof received: June 10, 2026

ABSTRACT

Research on formative assessments in the classroom has primarily been based on lecturer's assessment literacies in the form of questionnaires, and have ignored the elaborative, real time, assessment meaning negotiations that occur between lecturer and students. This qualitative case study examined the negotiations of the formative assessment criteria and feedback in the English as a Foreign Language

speaking classroom in Indonesia. This study contributes to the social assessment dialogic process research. This study was part of a 16-week longitudinal qualitative research study conducted in the Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar, and involved one lecturer and 25 students, while 7 focal students were chosen to participate in interviews. Data were obtained through Classroom Observational Research, video recorded interactions, Semi-Structured interviews, Stimulated Recall exercises, Reflective Journals, and Document analysis. Thematic analysis yielded four patterns of criteria negotiations; the (1) explicit/informal feedback rubric criterion conflict; (2) student-initiated clarification through questioning and requesting feedback exemplar; (3) lecturer-initiated clarifications as a response to feedback observed/perceived difficulties; and (4) the rare co-determination of rubric criteria through interactions, especially during collaborative rubric construction for the panel discussion tasks. The diverse assessment interactions also shaped the lecturer-student interactions relationships and changed students' fear based learning identities to positive confident learning identities. The study concludes that formative assessment is essentially interpretive and relational, and that it necessitates pedagogical intentionality to guarantee that all learners have equitable engagement.

Keywords:

Formative assessment, sociocultural theory, classroom discourse, assessment negotiation, Indonesian EFL, speaking assessment, learner identity

INTRODUCTION

From a pedagogical perspective, language assessment can be viewed as substantial learning feedback, beyond ascertaining student data, as the assessment has the power to influence the teacher's approach to delivery, as well as the learner's perception of learning the outcome (Black & Wiliam, 1998; Shohamy, 2014). For the purposes of this study, classroom assessment has been considered as a continual and collaborative approach to formative assessment, unlike summative assessment, which is time bound and conveys little to no information to the learner (Black & Wiliam, 2009; Carless, 2012). It is ongoing because feedback loops function across many instructional

episodes not only at the end of a unit. It is collaborative because meaning-making regarding criteria of performance arises through lecturer-student interaction, not through unilateral instructor judgement. Hattie and Timperley (2007) have recognized the potential of formative assessment to foster student ownership of learning. However, as they caution, this potential is contingent upon educational systems and cultural contexts and this contingency requires additional empirical exploration.

Meanwhile, the literature has acknowledged the potential for formative assessment to encourage student ownership of learning; however, diverse educational systems and cultures require further analysis (Hattie & Timperley, 2007). The context of this study is in English language education, which has been recognized in a number of studies as a component of the globalisation of education and employment opportunities, and the need for English proficiency for graduates to participate in globalised work structures (Mirizon, 2021; Sulistiyo et al., 2020). More recently, Lengkanawati (2016) and Widodo (2016) have stated that research shows the pressure Indonesian universities face has also resulted in the implementation of English language requirements in all fields of study and that the English language education departments have to prepare future teachers to be teachers with high levels of communicative competence.

Within these departments, they view the evaluation of speaking skills as paramount, as speaking skills serve as the most demonstrative example of communicative competence. Moreover, speaking skills are the most popular target and prioritized by both teachers and students as the aspect of language proficiency (Lambert, 2010; Sato, 2025). In the assessment of various language skills, speaking is the most complex areas of pedagogy (May, 2009; Nakatsuhara et al., 2020) because, unlike the assessment of grammar and vocabulary, which can be assessed by objective tests, speaking assessment entails real time interaction, and subjective evaluation, thus making it the most complex domain of pedagogy.

There is a growing body of research over the past 20 years that has

focused on Language Assessment Literacy (LAL), in particular the research field of assessment literacy (AL), with the focus on what teachers know concerning the scope and breadth that has been applied in various contexts of the classroom. Fulcher (2012) and Malone (2013) pioneered research that has shown the field of language assessment literacy is far more complex than the technical aspects of test construction and statistic, but also encompasses the social and ethical dimensions of assessment, the impact of assessment on learners, and the classroom environment.

In relation to the Indonesian context, researchers have modified these frameworks to study assessment literacy of EFL teachers finding that while a number of teachers have an adequate understanding of the principles of assessment, the majority have difficulty of sustaining the practice of formative assessment because of the existence of big class size, poor training, and institutional demands that promote the adoption of summative evaluation (Izyani & Mohamed Amin, 2016; Saputra et al., 2019). Most studies in this line of inquiry have used surveys and questionnaires which while useful in providing though brief, illustrations of teachers' beliefs and their self-reported practices, do not offer sufficient understanding of how assessment takes place in real time during the dynamics of a classroom (Azis, 2014; Istiarto Djiwandono & Zein, 2025). Alongside this trend in research, studies on formative assessments in the EFL contexts focus on teacher feedback, use of rubrics, and student involvement in assessments.

Research by Carless & Boud (2018) and Hattie & Timperley (2007) shows that feedback is effective if it is timely, specific and actionable. However, studies in Asian EFL classrooms, including Indonesia, show that students fail to respond to feedback due to the culturally influenced face-saving and deference to teacher authority (Lee, 2017; Ansas & Sukyadi, 2019). In the Indonesian higher education context, studies on the use of rubrics show that while rubrics are provided to students, they are often seen as mere completion documents and are rarely used as discussion aids for the criteria of assessment. Often, students are not provided the opportunity to,

negotiate, or even discuss the criteria of the rubric (Siregar & Zulkifli, 2025; Friyandini, 2025). The relevant literature in this area concerns classroom interaction and discourse studies, especially studies that focus on the feedback dialogue between teachers and students. Using the frameworks of conversation and discourse analytics, some researchers have illustrated how teachers avoid potential criticism by employing hedges, praise sequences, and indirectness, while students demonstrate an understanding, lack of understanding and resistance through different verbal and non-verbal signals (Seedhouse, 2010; Walsh, 2011).

In the Indonesian context, Astrid et al. (2019) focused on classroom discourse patterns and identified a continuum of teacher dominant talk where all students are only able to respond to teacher initiations with minimal extended contributions, a situation that severely limits the classroom discourse needed for dialogic negotiation of formative assessments. Recent research have revealed that this pattern of teacher-dominated speaking is still the standard in Indonesian EFL classrooms. For example, Nursehag and Amalia (2024) discovered that teacher speaking made up 74.4% of the classroom interaction, with lecturing being the most common activity while student engagement mostly consisted of replying to instructor questions.

Likewise, Ghozali, Marshalina, and Hanafi (2024) reported that teacher talk was consisted of 69.4% directive speech acts, which made students the recipients of instructions rather than active interlocutors. The results suggest the persistent difficulty of creating dialogic assessment interactions in Indonesian classrooms. Mcnamara (2001) and Poehner (2008) present some of the first ethnographically informed research articles on assessment and draw attention to the importance of a sociocultural approach, one that does not treat assessment as a purely technical or neutral process, but as an inherently social and cultural practice that is tied to social relationships, identities, and institutional histories. This foundation was built upon by Poehner (2008) and Poehner & Lantolf (2005) to develop the concept of dynamic assessment in Vygotskian sociocultural theory to formulate the concept of dynamic

assessment, and emphasis has been placed on the relational construction of the self and the identity work that is a result of learner evaluation (Norton & Kramsch, 2013; Block, 2009), especially in the context of high-stakes speaking assessments as performance is seen to be closely linked to the participants' self-concept.

In an Indonesian context, Widodo (2016) and Lamb (2012) have argued that learners' interest in learning English is tied to the imagined identities of becoming global citizens. More recently, researchers have called for greater attention to how classroom assessment practices shape learner identity (Lin & Curle, 2025; Al Awlaqi & Ghozali, 2023), as the quality of interaction and feedback directly influences students' willingness to communicate and their self-perception as language learners. However, little has been done to explore the extent to which classroom assessment practices facilitate or hinder learners' identity construction.

Regarding the empirical studies and the developments cited above, the lack of longitudinal studies investigating how formative assessment is co-constructed through lecturer-student dialogue in Indonesian EFL speaking classes, which discursive techniques focus on the giving/receiving feedback, how the co-construction affects the participants' relationships and identities, and how the co-construction is affected by the participants' relationships and identities is noteworthy. In order to contribute to the research gap, the current study presents an Indonesian university case study of assessment-in-action using sociocultural theory. Since there is an impressive volume of literature focusing on language assessment literacy and formative assessment practices, there is still a gap when taking a critical look at the literature.

Although there is an amount of evidence regarding what teachers understand about assessment and what types of formative practices they claim to use, there is very little understanding about the real-time classroom interactions with teachers and students, and the collaborative processes of negotiating assessment meaning and outcomes. Research has shown that the emphasis has been placed on the teacher as the main party implementing

assessment and placing the students in a position that is only passive (where students only receive evaluation feedback). This position neglects the role of students as active participants in the assessment process where their own interpretations, responses, and negotiations can impact the assessment process and its results (Leung & Lewkowicz, 2013; Poehner & Lantolf, 2005). This teacher-oriented focus is most apparent in the assessment literacy field that has almost exclusively focused on pinpointing the ‘what’ of the knowledge, skills and beliefs of the teacher in relation to assessment and through what modalities capture what the teacher claims to do rather than what actually is the case in the context of the teacher-learner complex interaction (Fulcher, 2012; Malone, 2013; Scarino, 2013).

Although some studies of formative assessment in certain Asian EFL environments are needed for identifying problems regarding students’ reluctance to respond to the feedback provided (critically), these studies have relied on previous studies and are without a proper analysis of the feedback, as it has been delivered, perceived, interpreted, and opposed (Lee, 2017; Zheng & Yu, 2018). The literature lacks an analysis of the issues that are essential for the context of assessment as a lived experience, and more plainly ‘how’ the phenomenon is lived. The issue remains ‘How do lecturers and students negotiate the criteria and meanings of performance during formative assessment interactions in EFL speaking classrooms? This study looks at speaking assessment, which Mcnamara (2001), Poehner (2008), and Shohamy (2014) encourage by calling for more specific studies that focus on assessment as a social, relational, and dialogic process rather than a technical process.

METHOD

Design

Contextualized within a sociocultural framework, this study employed qualitative case study design (Yin, 2018). The rationale for this design is that the research question, ‘How do lecturers and students negotiate assessment

criteria?’ needs a detailed and contextualized analysis of a current occurrence in its real classroom context (Tisdell et al., 2025). A case study design, unlike surveys / experiments, preserves the comprehensive quality of formative assessment negotiations. A case study design is ideal when the purpose of the research is to analyze a phenomenon within its real setting in depth (Tisdell et al., 2025; Stake, 2006).

Participants

For the case study, participants were purposefully sampled. The first participant was an EFL lecturer teaching the Public and Academic Speaking course at the third semester level of the English Language Education Department, Faculty of Dharma Acarya, Universitas Hindu Negeri I Gusti Bagus Sugriwa Denpasar. This university was chosen because the researcher had institutional access and ethical consent and the English Language Education Department is a typical Indonesian venue to investigate formative classroom assessment techniques. The main reason for the selection of this lecturer was that this subject involved formative assessment, that is, providing students with feedback on their presentations and discussions on an ongoing basis throughout the semester.

Although other lecturers in the department also used formative assessment to varying extents, this particular lecturer was chosen because (a) the Public and Academic Speaking course involved multiple low-stakes speaking tasks throughout the semester, thus providing opportunities to observe assessment negotiations, and (b) the lecturer was willing to participate in stimulated recall interviews and allow video recording of all classroom sessions. In case study research, purposeful sampling is advised to select information-rich cases that elucidate the research issue (Tisdell et al., 2025).

The second set of the participants comprised the whole class of 25 students enrolled in this Public and Academic Speaking course. The participants were therefore the full class, as formative assessment negotiations happen not only between lecturer and individual students but also through

whole-class feedback sessions, small group interactions and peer listening during presentations. Any student not seen would have meant an incomplete picture of the patterns of talk in the naturalistic classroom. The researcher was able to observe all 25 student assessment interactions (who opens the negotiation, who is silent, and how the lecturer changes feedback for different students) before purposively selecting a smaller focus group for in-depth interviews. From this bigger class, a smaller focus group of seven students was purposefully chosen for several interviews. The seven students were chosen based on two criteria: they should stand for a range of speaking proficiency (high, medium and low), as per the lecturer's first assessment, and they should show different degrees of interaction during the first several class observations, such as those who responded to the feedback a great deal and others who appeared to be more passive and resistant. A sample of seven was selected because this quantity enabled representation from each competence level (at least two students per level) and was still doable for in-depth semi-structured interviews and stimulated recollection sessions within the 16-week semester. This method of selecting a subsample from a broader cohort for further follow-up is congruent with qualitative case study procedures (Creswell & Poth, 2016).

Instruments

A range of instruments was utilized to ensure data triangulation (Denzin & Lincoln, 2011; Flick, 2022) and to showcase the complexity of the various dimensions of classroom assessment. The main instrument was the researcher, who served as a non-participant observer and used classroom observation protocols as well as audio and video recordings of all classroom interactions to capture the intricacies of the verbal and non-verbal discursive strategies employed by the lecturer and students during formative assessment interactions. Classroom observational research was conducted utilizing an adapted observation technique that focused on identifying instances of criterion negotiation, feedback delivery and student responses. All classroom lessons were video-taped for multiple viewings and micro-analysis of verbal

and non-verbal exchanges. Semi-structured interview guides (Brinkmann & Kvale, 2018) were created for the lecturer and each of the seven focus-group students. These guides were developed around open-ended questions that focused on the negotiation of feedback and grades. The interview guide for the lecturer and students are presented in the table 1.

Table 1. The interview guide

Participant	Category	Question
Lecturers	Philosophy of Assessment and Rubric Creation	1. How do you choose what to include in your speaking rubric?
		2. What do you do when a student has met the rubric, but you still feel they performed poorly?
		3. How often do you revise your rubric, and why?
		4. Do you ever add criteria during the semester based on student performance?
	Feedback Judgments	5. Tell me about a time when you changed your feedback based on a student's reaction.
		6. How do you determine whether to give feedback immediately or wait until later?
		7. What do you do if a student looks confused by your feedback?
		8. Do you give different feedback to high- and low-proficiency students?
		9. How do you provide negative feedback without discouraging the student?
	Student Negotiation and Identity Transformation	10. Have you seen students become more confident because of your feedback?
		11. How do you deal with a student who does not agree with your assessment?
		12. Do you think your feedback affects how students see themselves as English speakers?

Students	Understanding of Assessment Criteria	1. What is your understanding of "content organization" in your speaking tasks?
		2. Do you know what a rubric is? Which parts of the rubric are confusing to you?
		3. Have you ever read the rubric and still not understood what you were expected to do?
	Experiences of Receiving and Responding to Feedback	4. Tell me about a time when you improved as a result of feedback.
		5. What do you normally do after receiving feedback from your lecturer?
		6. Do you prefer written or spoken feedback? Why?
	Negotiation of Grades or Assessment Criteria	7. Have you ever asked your lecturer to explain a grade? What happened?
		8. Would you ever tell your lecturer that you disagree with a grade? Why or why not?
	Changes in Confidence and Learning Identity	9. Do you feel more confident speaking English now than at the beginning of the semester?
		10. Has the feedback you received changed how you see yourself as an English speaker or as a learner?

A different guide for stimulated recall sessions (Gass & Mackey, 2017; Lyle, 2003) with the lecturer was constructed using classroom assessment videos to reflect on the lecturer's moment-to-moment assessment decisions and the rationale for their strategies. In the stimulated recall activities the lecturer saw video clips of feedback occurrences in the classroom and was asked to explain her thinking at each point. Furthermore, a protocol for document analysis (Bowen, 2009) was employed to scrutinize relevant classroom artifacts such as syllabi, assignment instructions, grading rubrics, samples of the lecturer's written feedback on student work, and other course materials that situated the assessment framework within which formative interactions took place. Moreover, reflective journals were obtained from the seven students in the focus group on a weekly basis. Students were asked

to write on their grasp of assessment criteria, their responses to feedback received and any changes in their confidence or learning identity.

Data Collection

The collection of data was carried out longitudinally over the course of one academic semester of 16 weeks, in order to gauge the shifts in assessment mechanisms and the relationships involved. The researcher, in the capacity of a non-participant observer, attended all classes of the Public and Academic Speaking course, and provided detailed field notes and captured video and audio recordings of all dialogues pertaining to formative assessments. For the duration of the semester, the researcher conducted a number of semi-structured interviews, and interviewed the lecturer three times, including one interview to capture their assessment philosophy and plans, one interview to capture their reflective assessment practices, and one interview to capture their summative reflective assessment of the semester.

The researcher also conducted 2-3 stimulated recall interviews with the lecturer in order to address the lecturer's rationale in response to the specific classroom videos that the lecturer viewed. Stimulated recall is an established method for accessing teachers' decision-making processes (Gass & Mackey, 2017). Seven target students were also interviewed at the start and the end of the semester to capture their initial assessments of the process and their perceptions following the semester, in addition to gauging the evolution of their identities as learners as a result of the formative feedback.

Then, all interviews were recorded and transcribed, and all pertinent course materials were gathered throughout the semester (including syllabus and grading policy materials collected at the beginning, assignment sheets and rubrics, and student work where the lecturer's written formative feedback was present). Classroom observation entailed the researcher sitting at the back of the room during all 16 sessions writing field notes of assessment related talk. All 16 sessions were video recorded from a single camera in the back corner. Semi-structured interviews with lecturers and students (45-60 minutes) were audio-recorded and transcribed. Stimulated recall

was conducted in three sessions of 45 minutes each in which the lecturer saw 5-7 video clips of feedback moments and described her thoughts. The seven students responded to queries about criteria, criticism, and confidence in weekly reflective journals. Document analysis coded 23 documents for criteria, feedback language and negotiation chances.

Data Analysis

Data analysis took place repeatedly as there were themes to be processed to fit the sociocultural theoretical framework. The investigation was conducted in five stages. First, all transcripts were read several times to find all situations in which assessment criteria were debated, questioned, and clarified. Second, the researcher categorized each instance for who initiated the negotiation (lecturer / student) and for what provoked the negotiation (confusing rubric, student query, lecturer observation of difficulty). Third, sociocultural ideas of mediation and scaffolding were used to think about how each case included language, rubrics, and peer-to-peer interaction as instruments of shaping understanding. Fourth, patterns across occurrences were categorized into themes such as explicit versus implicit criteria / student-initiated versus lecturer-initiated bargaining. Fifth, the data were evaluated for signs of identity transformations, specifically in terms of statements in which students indicated changes in confidence, fear, or self-perception as speakers. This happened simultaneously as data collection, with each of the data phases providing information to the phases of the data collection process.

Data were analyzed through NVivo in terms of field notes, relevant documents, and transcribed verbatim audio/video of the interviews and classroom videos that were grouped and organized accordingly. Thematic analysis was conducted following the procedures outlined by Braun and Clarke (2021). Analysis was done in several steps beginning with open coding as the researcher read through the transcripts to find concepts relating to negotiations, discursive strategies, and relationships. After open coding was accomplished, axial coding took place in each of the transcripts. More classroom interaction data were thematically analyzed with particular

attention to the presence and the roles of the strategies and constituent elements (questioning, hedging, direct correction, student negotiated, and other linguistic and paralinguistic) that mediated the interaction and the assessment process. The development of themes that responded directly to the research question occurred through the iterative process of coding and cross-case comparison. Examples of themes that were developed through the negotiation of criteria came from instances that emerged when students asked questions for clarifications about a rubric, themes of resistance or accommodation were built from patterns of student response to feedback, while themes of identity shaping were constructed from the interview data of students who, depending on the feedback they received over a semester, felt more or less confident as speakers. To ensure the credibility of the findings, a number of strategies were included for this research, such as the triangulation of data sources from observations, interviews, and documents, member checking through presenting the initial findings to the lecturer and focus-group students for their approval, peer debriefing by discussing the coding and the themes that have emerged with a fellow researcher to question and adjust interpretations as well as ease the biases, and the prolonged engagement in the field through a full semester of observation to establish trust and acquire a great deal of information about the environment.

FINDINGS AND DISCUSSION

FINDINGS

Tension between Explicit and Implicit Criteria (Theme 1)

Every student was given a course syllabus at the start of the semester, which included a speaking rubric. Four speaking rubric categories were included: fluency, accuracy, pronunciation, and organization. Each category was described in further detail in a grading rubric that was also given to the students which described the performance level of each category as 'excellent' to 'needs improvement'. As an educator, A1 (All names are coded) makes an effort to include what she describes as 'foundational' criteria in

her syllabus.

'I always provide students with a rubric from the first class, because it's something that's needed. It's extensive, students know what I'm looking for. If they don't know what I'm judging them on, how can they meet them? The rubric is my contract with them.'

By examining the syllabus and the rubrics. It can be assumed that A1 had a focus on how criticism and detailed elements were given to the students. What does the rubric say? How does that define each of the values? In practice, a lot of the criteria did not seem to be defined and explained in class. In an operational use case, however, feedback from the instructor and her comments on student speaking performances were used to imply what she meant by successful speaking.

In Week 3, after a student named B1 presented on Balinese traditional ceremonies, A1 commented on B1's presentation.

A1 said, 'B1, your content was very interesting, I learned a lot about the ceremony. But you said 'ceremony' as 'seremony' multiple times. In academic speaking, pronunciation is very important because if you mispronounce words, the audience will not understand you.'

In this case, the criteria of 'pronunciation' was exemplified by the particular case of the word 'ceremony' and audience comprehension was explained as a justification for the importance of the pronunciation.

In the same way, in Week 5, when B2, struggled to answer a question in a discussion task, A1's comment focused on fluency.

A1 told B2, 'I noticed you paused for a long time when asked about your opinion. In academic speaking, it is okay to pause to think, but you can use fillers like 'let me think about that' or 'that's an interesting question' to maintain the flow. Fluency is about managing the conversation smoothly.'

Once more, the criterion of 'fluency' was exemplified by the problem that B2 was facing. Lecturer feedback featured an array of comments and illustrated the various elements of criticism, especially the criticism of implicit and explicit elements of assignment criteria. Some comments refer to criteria that are not mentioned on the assignment rubric whereas the rubric

document rater comments and feedback on criteria that are not mentioned in the rubric.

Meanwhile, one student received the comment.

'Your arguments are logical, but consider your audience, are they familiar with this term? Audience awareness is part of a good content organization.'

This feedback introduced audience awareness, an implicit criterion that the rubric did not mention but A1 held in high esteem. When A1 was prompted with this feedback during the stimulated recall interview, she said, *'The rubric cannot capture everything.'* She further commented on the feedback by saying,

'Content organization on paper is one thing, but when you are actually speaking, you have to think about who is listening. This is something students learn through practice and feedback, not through reading the rubric.'

This contradiction in the rubric and assignment feedback was reflected on by students in the focus group. This is illustrated with B3, a student with mid-range proficiency who commented in the end-of-semester interview,

'At the beginning, I studied the rubric very carefully. I thought if I followed the rubric, I would get a good grade.' After a number of feedback sessions, *I understood that A1 values criteria outside the rubric, such as audience engagement and handling questions that weren't prepared. So, I needed to understand what she wanted beyond the paper.'*

This narrative suggests that the criteria negotiation started with the formal documents, but also developed into a continuous process of interpretation as students learned to look beyond the rubric and respond to the unarticulated values in the lecturer's comments. Theme 1 shows that written rubrics do not adequately represent formative evaluation criteria and that students need to learn how to comprehend implicit values included in lecturer feedback.

Negotiating Criteria Initiated by Students (Theme 2)

The majority of student participants negotiating criteria involved asking questions about the criteria, especially before speaking activities. During Week 4, before the first graded presentation, a student named B4 raised his hand to ask the question,

'Bu, I am still confused about the content organization. In the rubric, there is something that states that there must be a 'logical sequence of ideas.' What does logical mean? Is it introduction, body, conclusion, or is it something else?'

This led A1 to respond to B4 by talking about a lot of things. She explained a lot, showed examples of the previous students' presentations, and asked the other students to explain. B4, the student, wrote in his reflection journal,

'I was nervous to ask the question, I didn't want my friends to see me and think I was a fool. But I really did not understand. I thought that if I didn't ask, I would make a mistake in my presentation, and I was really grateful to A1 for explaining it. I was really surprised that other students looked confused, I thought I was the only one.'

It means that *'perhaps they were too shy to ask,'* which shows how students had to navigate the cultural barrier of *'sungkan,'* and also demonstrates the esteem that students placed on understanding the criteria prior to any performance.

Other forms of self-initiated negotiation included asking for examples and models of performances. During Week 7, following feedback from the first graded presentations, some students went to A1 after class. As field notes document, B5 asked:

'Bu, we heard your feedback about fluency and pronunciation, but we are not sure what 'good' fluency sounds like, could you show us an example of a presentation you think is very good? Or a video maybe?'

Ayu, in her end-of-semester interview, provided more context about the situation:

'I learn more by imitating. The rubric is clear, but the rubric varies from person to person. If I can listen or watch what the good presentation is, I can visualize the goal. I understand the standard

better.'

Ayu, along with many others in the focus group, described negotiation as a means of going beyond the criteria. Less frequently, students questioned or challenged criteria in feedback when negotiating with students. B6 received feedback on his presentation and was told it was 'too informal' for an academic speaking.

The next class, B6 had this exchange:

'Bu, you said my presentation was too informal. I get that for academic speaking it needs to be formal, but I don't know what the line is. In my presentation, I used some examples from daily life to make it interesting. Is that not allowed?'

This was one of the few times a student sought a response that illustrates the effort to think critically about the criteria. While analyzing the scenario, A1 explained,

'B6 was not being disrespectful. He wanted to understand the issue. His question made me think. I hadn't explained that including a daily life example is fine as long as the language used is formal. His question helped me.'

Therefore, student-initiated negotiation helps lecturers improve how they express criteria.

Nonetheless, the student-initiated negotiation was not evenly spread out. With the non-verbal communication checklists and observation analysis, it was learned that the seven students from the interview, especially those with higher proficiency and higher interaction frequency, initiated negotiation more than other students. Students with less proficiency and more passive interaction patterns did not ask about the criteria in questions in the entire class. A less proficient student, B7, described this:

'I had a lot of questions, but I was afraid that my questions were not smart enough. Also, if I ask, everyone will look at me, and I feel embarrassed. I wait to see if my friends ask the same question and if they do, I just listen to the answer.'

Therefore, student-initiated negotiation was present, but it also required more from the students, that is confidence, proficiency, and the ability to

speak out in public. Therefore, it created a situation in which some students had a higher chance than others to negotiate more criteria than others. Thus, Theme 2 suggests that the benefit of student-initiated negotiating is awarded to the confident, high-proficiency student and the less confident student is left behind, posing equity problems.

Lecturer-Initiated Clarification of Criteria (Theme 3)

Most lecturer-initiated clarifications of criteria have happened after the lecturer picked up some students' patterns of struggles. In week 6, for the first round graded presentations, A1 spent an entire 30-minute session clarifying criteria based on what she saw. In the video, she stated:

'I have finished grading all of your presentations, and I want to talk about some patterns I noticed. Many of you lost points on pronunciation, but not on all words, specifically on words with sounds that do not exist in Bahasa Indonesia and Balinese, like 'th' in 'think' and 'this.' So, we are going to practice these sounds today because pronunciation is not just about being understood, it is also about sounding good as an academic speaker.'

Here, the lecturer converted the criteria in the rubric (general criteria), which was '*pronunciation*,' into something more constructive based on the students' needs. While responding to the mid-semester interview, A1 said,

'If I just give grades and move on, students will make the same mistakes in the next assignment.' My job is to support their development, which means I have to illustrate what the criteria entail and provide opportunities for them to practice. The rubric outlines what is needed, while my lessons and feedback provide the detail.

Another example of lecturer-initiated clarification concerned previewing criteria for major assessments. In Week 14, A1 explained the criteria for the final group presentation, giving students a simplified rubric and providing examples for each. Field notes state:

'For the final presentation, I will be paying special attention to how you organize your group's ideas. Content organization is about how your parts connect to each other. Do you have transitions between speakers? Do you refer back to what your friend said? This is something we have not emphasized before, but for the final project,

it is needed.'

A1 is adjusting the criteria to a new level of expectation by introducing the idea of '*transitions between speakers*' to students as a new consideration of the organization of content for the group task. In her stimulated recall interview and in response to the question regarding this moment, she mentioned:

'I could have just said 'content organization' again, but that would not have been fair because group presentations are different from individual presentations. Students need to know that the criteria shift depending on the task, even if the rubric uses the same wording.'

Lecturer-initiated clarification also took place regarding student affect, including situations when students appeared anxious and discouraged. During Week 9, after a difficult feedback session whereby several students did not meet their expectations, A1 stopped her planned lesson and said,

'I can see that some of you are disappointed with your feedback. But I want you to understand something needed. The criteria on the rubric are designed to reflect an ideal. They represent where I want you to be by the end of the semester. And if you are not there yet, that is fine. The purpose of feedback is to help you to get to the ideal.' Do not view a lower grade as a failure, view it as feedback on what to improve for next time.'

The lecturer here, is not saying what she means when she talks about the criteria, she is explaining the interpretive structure within which she means criteria, developing targets as opposed to zero sum evaluations. One student, B8, noted this in his interview at the end of the semester:

'When A1 said that, it changed how I looked at my grade. Before, I thought if I got a B, I was a B student. But she said the criteria are a destination and feedback is the road. That made me less afraid to fail.'

This shows that lecturer-initiated clarifications work at several levels. They clarify what the criteria are, how they function in relation to individual tasks and how students interpret and relate to the criteria over time. Hence, Theme 3 suggests that lecturer-initiated clarification accounts for criteria and influences students' emotional attachment to evaluation.

Criteria Construction via Dialogue (Theme 4)

The instance of co-construction observed was in Week 12, when A1 presented a new assignment, an academic panel discussion. Instead of giving the students a rubric, she encouraged students to devise their own means of evaluating their performance. The following is a transcript of a 25 minute engagement that shows a degree of collaboration in the construction of meaning.

A1: 'For this panel discussion, I want to try something a little different. Instead of me giving you the rubric, let's create it together. What do you think is important when you participate in a group discussion in an academic speaking?'

B1: 'It's important that we talk about the topic and not just go off on a line.'

A1: 'Alright, knowledge of the topic. What other criteria do we have?'

B2: 'We should listen to each other. In group discussions, there are people that just sit there and wait until it's their turn to speak, and then they talk without really listening.'

A1: 'That's the idea. So, we want to make sure that people are listening and responding to each other. In what ways would we be able to tell that someone was doing that?'

B5: 'They can talk about what the previous speaker said. Like 'I agree with what B1 said about the topic' and 'Building on what B2 said to listen to each other''

A1: 'Great. We established two criteria: showing knowledge about the topic, and listening actively and referencing others. What other criteria do we have?'

B6: 'What about language? In the presentations, do we have to be assessed on fluency and accuracy?'

A1: 'That is a good question. What do you think? In a discussion, is it more important to speak fluently or to get your point across, even if you have some errors?'

Other students participated in this exchange by suggesting criteria

and evaluating the importance of each of them, while A1 only facilitated the discussion to summarize and provide the end statements. At the end of the discussion, the class had collaboratively created a set of criteria for the panel discussion which included; knowledge of the content, ability to engage others, and the ability to provide clear answers (imperfections in fluency are acceptable). In her stimulated recall interview about this session, A1 said:

'I was hesitant to do this because I was not confident that students would engage and they would formulate criteria that made sense, but they proved me wrong. They were more engaged than in any of my other classes and the criteria they came up with was more sophisticated than whatever I would have given them.' Because they owned it, they understood it better. In her reflective journal following the session, B5 wrote: *'Today was different. We made the rubric together. I feel like A1 really values our opinions. Also, now I understand exactly what I need to do because I helped decide what matters. I will remember these criteria more than if she just gave them to us.'*

B3 was equally positive claiming that all the other panel discussions he had to participate in before did not even come close to motivating him as the one in class did because he was the one who set the goals. Most of the time co-constructing other participants' understanding was dialogue that was more casual than this one, which was because of the panel discussion, the most formal and, at least in those weeks, most structured form of class. One described in week eight of the most casual end of co-constructed dialogue. A1 was able to give a refined understanding of 'delivery' in the provision of non-verbal communication and described to her a two-way-dominating discussion. These moments, although short and little, were invaluable because they framed students as knowledgeable rather than passive recipients of externally imposed standards.

Analysis showed that co-construction tended to emerge under particular circumstances. It had emerged the most when the task was novel (the panel discussion), when the lecturer directly invited student contributions, and when the students considered the classroom environment to be safe for

contributions. The non-verbal observation checklist showed that during these episodes of co-construction, student postures were more relaxed, more students made eye contact with the lecturer and with each other, and the length and tension of the silences was much shorter and less tense than during typical teacher-fronted interactions. Given that these conditions were present less often than they were during the summer semester, that is, the semester's co-construction, despite the obvious pedagogical value, was under the co-construction value. Thus, Theme 4 suggests that the co-construction of criteria has instructional value, but only under certain conditions (new tasks, explicit invitation, safe classroom setting).

DISCUSSION

The finding detailed four patterns interlinked with negotiating assessment criteria: (i) tension between criteria that is explicit and implicit; (ii) negotiation initiated by students; (iii) clarifications initiated by the lecturer; and (iv) dialogue and collaboration. This finding builds on the existing literature on language assessment literacy by going beyond the almost exclusive focus on lecturer's assessment literacy and instead inquiries into the situated evidence of the construction of the meaning of assessment (Fulcher, 2012; Scarino, 2013). The enduring tension between the rubric and the criteria in the lecturer's feedback aligns with Leung & Lewkowicz (2013) assertion that classroom-based assessment is interpretive and contextualized. This means that both lecturer and students need to construct meaning beyond what is articulated by the rubric's criteria. In the Indonesian context, where a cultural deficit model assumes that hierarchy and face-saving cultural norms constrain negotiation, the evidence of students instigating and even clarifying the 'logical sequence' of the rubric and requesting exemplars of 'good' performances illustrates that students are not merely passive recipients of feedback (Izyani & Mohamed Amin, 2016; Ansas & Sukyadi, 2019). The unequal number of students initiating negotiation, with more skilled and confident students dominating, draws attention to issues that have received

little focus in the assessment literacy research.

Poehner & Lantolf (2005), from a sociocultural perspective, suggest that assessment practices that do not create opportunities for all learners to engage in and participate in all ways, risk reproducing the existing hierarchy and not breaking through it. B7's case illustrates the need for Carless (2012) assessment for learning pedagogy. B7 remained silent and in the case of assessment, pedagogy, and learning, B7's case illustrates the need to create pedagogy in assessment for learning for the case of assessment and learning. The study shows the clarity of an engaged formative assessment through the pedagogy of intentionality. These practices are aligned with the principles of feedback by Hattie & Timperley (2007), when feedback addresses 'where am I going' and 'how am I going' and 'where to next' questions.

On the other hand, the lecturer also exemplified the essence of Black & Wiliam (2009) on formative assessment by using evidence of learning to guide instruction to address the student's needs. The fact that the lecturer modified her feedback to students exhibiting different levels of confidence and receptivity, being more explicit with confident students such as B1 and more vague with anxious students such as B7, reflects a type of assessment literacy that combines technical knowledge with that of interpersonal sensitivity (Davison, 2019; Malone, 2013). In regard to the Indonesian cultural context, this sort of assessment literacy, or interpersonal sensitivity, is of particular importance, as direct feedback can be socially disruptive and result in a loss of face (Astrid et al., 2019). In this regard, the use of praise sandwiches, hedges, and culturally appropriate metaphors is a discursive strategy to mitigate the tension between assessment as an evaluation and assessment as a means of preserving a person's positive face (Goffman, 2017; Lee, 2017). These techniques are more than merely expressions of politeness. They are essential elements of pedagogy that allows students to leave defensively and engage with feedback (or face) to act on that feedback.

The panel discussion task brought about the sociocultural ideal of assessment as a shared process and the co-constructed criteria of assessment

that the lecturer solicited from students as criteria she described as ‘sophisticated.’ This activity resembles what Poehner (2008) calls ‘dynamic assessment,’ where assessment and instruction simultaneously occur through a dialogue that is mediated. Students were able to use the knowledge that they had acquired from previous discussions; the need to listen, the need to speak to others and the need to refer to others, to engage in the construction of discussion criteria that, as the lecturer noted, were ‘more sophisticated’ than the ones she would have imposed. This finding shows that Poehner & Lantolf (2005) were correct in stating that the learner’s involvement in constructing the assessment criteria was a contributor to the enhancement of the learner’s metacognitive engagement and the sense of ownership towards the learning objectives.

Moreover, the motivational effect captured in the student’s saying ‘I will remember these criteria more because I helped decide what matters’ speaks to what Norton & Kramersch (2013) calls ‘investment’. Learners will invest more in achieving the outcomes of the learning process. The occurrence of this type of co-construction throughout the semester, only emerging when tasks were new and the lecturer overtly solicited participation, indicates that even in classrooms with a reflective and responsive lecturer, dialogic assessment practices appear to be the exception, and not the rule. This is consistent with Turner & Purpura (2016) who note, despite the advocacy for learning-oriented assessment, its application is still a complex and contextualized problem.

The results of this study have various implications for practice, policy and research. For practice, the four types of criteria negotiation propose that lecturers should go beyond simply presenting rubrics and instead establish structured chances for students to ask questions about criteria, see instances of outstanding performances and engage in rubric construction. The unequal nature of student-led discussion (Theme 2) suggests that lecturers cannot rely on students to ask questions themselves, but must encourage the less confident and those with lower skill to contribute. For policy, the study

suggests that teacher training programs should be designed to include assessment literacy (technical knowledge of rubrics and grading) and dialogic assessment skills, such as eliciting student questions, clarifying criteria without face-loss, and reframing feedback as developmental rather than judgmental. This study shows that for research purposes longitudinal, video-based observation captures assessment negotiations not captured by surveys and questionnaires. Future research should investigate whether structured interventions (collaborative rubric development, anonymised question systems) improve engagement in assessment negotiation, especially for silent / apprehensive students.

CONCLUSION

In this regard, the negotiation of the criteria for assessment took place in four interrelated dimensions. First, there was the tension between the explicit criteria within the documents and the criteria that remain implicit in the feedback given by the lecturer. Second, there was the student 'negotiation' by way of questions to clarify, and requests for an example. Third, there was lecturer initiated clarification as feedback to apparent challenges, in some cases, there was the provision of error-targeted pronunciation lessons, and in other cases, there was the reframing of criteria as developmental goals instead of destination judgments.

Besides, there was the case of the collaborative dialogue of the criteria, where most of the students negotiated criteria for the panel discussion task. This was an evident case of increased ownership as demonstrated by the reflection of one student that 'I will remember these criteria more because I helped decide what matters'. Despite this evidence of ownership, student negotiation was not equally distributed. More confident students 'dominated' this, while less confident students, especially those with lower language proficiency, remained mute from fear of looking stupid. In some cultures, there exists a stark compliance with the cultural norm of deferral. This situation created concerns of equity. As for the lecturer's clarifications, these

extended to all the students because most of the time, the unclear categories on the rubric were made concrete and actionable. This demonstrated that the co-construction of assessment could indeed exist where there was enough trust and an invitation. Further, criteria negotiation was a process that interprets documents and explains the process to students so that they recognize and respond to the unexpressed values in the lecturer's feedback that hinges on culturally defined values, personal self-esteem, and the lecturer's value system.

There are various limitations in this study. First, the conclusions are based on a single scenario (one lecturer, one class, one university) and cannot be applied to all Indonesian EFL classrooms. Secondly, the presence of the researcher as a non-participant observer have altered the classroom interactions, but continuous engagement over 16 weeks have diminished the effect. Third, the study focused on speaking assessments, and the results can vary for writing, listening, and reading assessments. Fourth, the selection of the seven focus students was based on proficiency and interaction patterns, but their reflective diaries and interviews have been affected by a desire to please the researcher.

Future study should address these constraints by performing similar longitudinal case studies across multiple Indonesian universities, both private and public, to examine whether the four negotiation styles are consistent across diverse contexts. Researchers would also want to look into negotiation of evaluation in other language skills (writing, listening, reading) to see if the patterns are specific to particular skills. Quantitative / mixed methods research could explore the extent to which student-initiated bargaining predicts advances in speaking proficiency / confidence. Besides, intervention studies could examine whether structured approaches (anonymous question systems, obligatory rubric critique assignments, peer-led criterion discussions) promote participation by less confident and lower-proficiency students.

REFERENCES:

- Al Awlaqi, H. A. S. M., & Ghozali, I. (2023). Is oral communication apprehension reasonable?: A study on sociocultural factors among Indonesian efl learners. *REiLA: Journal of Research and Innovation in Language*, 5(3), 252-265.
- Ansas, V. N., & Sukyadi, D. (2019). Source-based writing among undergraduate students: perspective and challenges. *International Journal of Education*, 11(2). <https://doi.org/10.17509/ije.v11i2.14757>
- Astrid, A., Marzulina, L., Erlina, D., Harto, K., Habibi, A., & Mukminin, A. (2019). Teaching writing to efl student teachers: Teachers' intervention and no teachers' intervention through peer feedback writing techniques. *Universal Journal of Educational Research*, 7(9). <https://doi.org/10.13189/ujer.2019.070908>
- Azis, A. (2014). Indonesian junior secondary school teachers' conceptions of assessment: A mixed methods study. Faculty of Education, Doctoral D(February).
- Black, P., & Wiliam, D. (1998). Inside the Black Box Raising Standards Through Classroom Assessment. *Phi Delta Kappan*, 80(2).
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1). <https://doi.org/10.1007/s11092-008-9068-5>
- Block, D. (2009). Second Language Identities. In *Second Language Identities*. <https://doi.org/10.1080/17501229.2011.576837>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, 9(2), 27-40.
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage publications.
- Brinkmann, S., & Kvale, S. (2018). *Doing interviews*. (2nd ed.). Sage publications.
- Carless, D. (2012). From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-Heritage Settings. In *From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-Heritage Settings*. <https://doi.org/10.4324/9780203128213>

- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, 43(8). <https://doi.org/10.1080/02602938.2018.1463354>
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.
- Davison, C. (2019). Using Assessment to Enhance Learning in English Language Education. In *Springer International Handbooks of Education: Vol. Part F1628*. https://doi.org/10.1007/978-3-030-02899-2_21
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The Sage handbook of qualitative research*. sage.
- Flick, U. (2022). *An introduction to qualitative research (7th ed.)*. Sage publication.
- Friyandini, R. (2025). Authentic Assessment in Indonesian Language Learning: Theoretical Review And Its Implications. *International Seminar on Humanity, Education, and Language*, 1.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2). <https://doi.org/10.1080/15434303.2011.642041>
- Gass, S. M., & Mackey, A. (2016). *Stimulated recall methodology in applied linguistics and L2 research*. Routledge.
- Ghozali, A. S., Marshalina, A., Hanafi, M., & Aylet, Y. T. (2024). Speech Act Analysis of Teacher Talk in Indonesian EFL Classroom: A Case Study at Secondary School. *Jurnal Pendidikan Humaniora*, 12(4), 1.
- Goffman, E. (2017). Interaction ritual: Essays in face-to-face behavior. In *Interaction Ritual: Essays in Face-to-Face Behavior*. <https://doi.org/10.4324/9780203788387>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1).
- Istiarto Djiwandono, P., & Zein, S. (2025). School Language Assessment and Educational Policies. In *The Routledge International Handbook of English Language Education in Indonesia*. <https://doi.org/10.4324/9781003399308-12>

- Izyani, M., & Mohamed Amin, E. (2016). Students' perception on the use of WhatsApp as a learning tool in ESL classroom. *Journal of Education and Social Sciences*, 4.
- Lamb, M. (2012). A Self System Perspective on Young Adolescents' Motivation to Learn English in Urban and Rural Settings. *Language Learning*, 62(4). <https://doi.org/10.1111/j.1467-9922.2012.00719.x>
- Lambert, C. (2010). A task-based needs analysis: Putting principles into practice. *Language Teaching Research*, 14(1). <https://doi.org/10.1177/1362168809346520>
- Lee, I. (2017). Classroom writing assessment and feedback in L2 school contexts. In *Classroom Writing Assessment and Feedback in L2 School Contexts*. <https://doi.org/10.1007/978-981-10-3924-9>
- Lengkanawati, N. S. (2016). Teacher's Beliefs about Learner Autonomy and Its Implementation in the Indonesian EFL Settings. In *Language learner autonomy: Teachers' beliefs and practices in Asian contexts*.
- Leung, C., & Lewkowicz, J. (2013). Language communication and communicative competence: A view from contemporary classrooms. *Language and Education*, 27(5). <https://doi.org/10.1080/09500782.2012.707658>
- Lin, Y., & Curle, S. (2025). Understanding willingness to communicate in english medium instruction: Cultural and pedagogical influences on Chinese university students. *The Asia-Pacific Education Researcher*, 34(5), 1593-1602.
- Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British educational research journal*, 29(6), 861-878.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3). <https://doi.org/10.1177/0265532213480129>
- May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing*, 26(3). <https://doi.org/10.1177/0265532209104668>
- Mcnamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18(4). <https://doi.org/10.1177/026553220101800402>
- Mirizon, S. (2021). Teachers of English Assessment Literacy: Unveiling

- What They Know and Do. *Indonesian Research Journal in Education [IRJE]*, 5(1). <https://doi.org/10.22437/irje.v5i1.12834>
- Nakatsuhara, F., Inoue, C., & Khabbazbashi, N. (2020). Measuring L2 Speaking. In *The Routledge Handbook of Second Language Acquisition and Language Testing*. <https://doi.org/10.4324/9781351034784-31>
- Norton, B., & Kramsch, C. (2013). Identity and Language Learning: Extending the Conversation. In *Identity and Language Learning: Extending the Conversation*.
- Nursehag, D. A., & Amalia, L. L. An Investigation of teacher talk and student talk in Indonesian EFL Classroom Interaction. *PAEDAGOGIA*, 27(2), 298-308.
- Poehner, M. E. (2008). Dynamic assessment: A vygotskian approach to understanding and promoting l2 development. In *Educational Linguistics* (Vol. 9).
- Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9(3). <https://doi.org/10.1191/1362168805lr166oa>
- Saputra, E. R., Hamied, F. A., & Suherdi, D. (2019). A Case of Authentic Assessment in Indonesian Secondary EFL Classroom Context: Teachers' struggle. <https://doi.org/10.2991/conaplin-18.2019.40>
- Sato, R. (2025). Japanese EFL Speakers' Fluctuation of Willingness to Communicate: Its Interplay and Association with Feedback. *Language Teaching Research Quarterly*, 48. <https://doi.org/10.32038/ltrq.2025.48.05>
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3). <https://doi.org/10.1177/0265532213480128>
- Seedhouse, P. (2010). The Interactional Architecture of the Language Classroom. *Bellaterra Journal of Teaching & Learning Language & Literature*, 1(1). <https://doi.org/10.5565/rev/jtl3.32>
- Shohamy, E. (2014). The Power of tests: A critical perspective on the uses of language tests. In *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. <https://doi.org/10.4324/9781315837970>
- Siregar, S. S., & Zulkifli, N. A. (2025). An Analysis Authentic Assessment

On Speaking Abilities In English Classroom. *Journal of Education and Teaching*, 6(1). <https://doi.org/10.24014/jete.v5i4.20003>

Stake, R. E. (2013). *Multiple case study analysis*. Guilford press.

Sulistiyo, U., Haryanto, E., Widodo, H. P., & Elyas, T. (2020). The portrait of primary school English in Indonesia: policy recommendations. *Education 3-13*, 48(8). <https://doi.org/10.1080/03004279.2019.1680721>

Tisdell, E. J., Merriam, S. B., & Stuckey-Peyrot, H. L. (2025). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In *Handbook of Second Language Assessment*. <https://doi.org/10.1515/9781614513827-018>

Walsh, S. (2011). *Exploring classroom discourse: Language in action*. In Taylor and Francis Ltd 5 (Vol. 9780203827826). <https://doi.org/10.4324/9780203827826>

Widodo, H. P. (2016). *Language Policy in Practice: Reframing the English Language Curriculum in the Indonesian Secondary Education Sector*. In *Language Policy(Netherlands)* (Vol. 11). https://doi.org/10.1007/978-3-319-22464-0_6

Yin, R. K. (2018). *Case study research and applications* (Vol. 6). Thousand Oaks, CA: Sage.

Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*, 37. <https://doi.org/10.1016/j.asw.2018.03.001>