



Developing Assessment in Indonesian EFL Speaking Classroom

Oktavia Widiastuti^{1*} 

English Study Program, Fakultas Bahasa dan Sastra, Universitas PGRI Kanjuruhan Malang

E-mail addresses: oktavia.widiastuti@unikama.ac.id (Corresponding Author)

ARTICLE INFO

Article history:

Received June 04, 2025

Revised June 10, 2025

Accepted June 12, 2025

Available online June 17, 2025

Kata Kunci :

Penilaian, Penilaian Efektif,
Berbicara, Bahasa Inggris Sebagai
Bahasa Asing

Keywords:

Assessment, Effective Assessment,
Speaking, EFL

ABSTRAK

Baru-baru ini, perdebatan yang cukup besar telah muncul mengenai penilaian kinerja berbicara siswa, terutama karena pengaruh faktor budaya dan subyektif yang membentuk pendekatan dosen dalam menyusun penilaian berbicara. Artikel ini menyoroti perlunya peningkatan kesadaran di antara para pendidik tentang kompleksitas yang terlibat dalam mengevaluasi kemahiran berbicara. Tujuan utama dari artikel ini adalah untuk mengusulkan sebuah metode untuk merancang penilaian berbicara yang secara khusus disesuaikan dengan konteks pendidikan tinggi di Indonesia. Dalam melakukan hal tersebut, artikel ini menekankan pentingnya mengadopsi kriteria penilaian yang efektif seperti yang diuraikan oleh (Brown & Abeywickrama, 2020). Menurut para ahli ini, penilaian berbicara yang efektif harus didasarkan pada kriteria yang jelas, melibatkan tugas-tugas yang sesuai dan produktif secara maksimal, dan menggunakan prosedur penilaian yang praktis serta dapat diandalkan. Berdasarkan prinsip-prinsip tersebut, artikel ini merekomendasikan agar dosen mengembangkan penilaian berbicara yang tidak hanya baik secara metodologis, tetapi juga relevan secara kontekstual dan responsif terhadap kebutuhan mahasiswa. Dengan demikian, dosen dapat memastikan bahwa praktik penilaian mereka adil dan bermakna. Penerapan model penilaian yang diusulkan memungkinkan dosen melakukan evaluasi keterampilan berbicara secara adil, objektif, dan sesuai dengan konteks EFL di Indonesia. Pendekatan ini tidak hanya meningkatkan validitas dan reliabilitas penilaian, tetapi juga mendukung pengembangan kompetensi komunikasi mahasiswa serta kualitas pembelajaran Bahasa Inggris di tingkat perguruan tinggi.

ABSTRACT

Recently, considerable debate has emerged regarding the assessment of students' speaking performance, particularly due to the influence of cultural and subjective factors that shape lecturers' approaches to constructing speaking assessments. These ongoing discussions highlight the need for heightened awareness among educators about the complexities involved in evaluating oral proficiency. The primary aim of this paper is to propose a method for designing speaking assessments that are specifically tailored to the Indonesian higher education context. In doing so, the paper emphasizes the importance of adopting the criteria for effective assessment as outlined by (Brown & Abeywickrama, 2020). According to these scholars, an effective speaking assessment should be grounded in clearly defined criteria, involve tasks that are both appropriate and maximally productive, and utilize practical as well as reliable scoring procedures. Based on these principles, this paper recommends that lecturers develop speaking assessments that are not only methodologically sound but also contextually relevant and responsive to the unique needs of their students. By doing so, lecturers can ensure that their assessment practices are both fair and meaningful within the Indonesian university setting. The implementation of the proposed assessment model enables lecturers to conduct speaking evaluations that are fair, objective, and tailored to the Indonesian EFL context. This approach not only enhances the validity and reliability of speaking assessments but also supports students' communicative competence and improves the overall quality of English language instruction at the university level.

1. INTRODUCTION

The evolution of English as a worldwide language has reinforced its status as a lingua franca. As a result, the majority of nations globally, particularly those where English is not the primary language, regard English as a crucial language to acquire. This issue affects the educational systems in various countries; some utilize English as the medium of instruction, while others mandate English as a required subject in schools. English holds a significant significance within the Indonesian educational system. English is taught and assessed as a foreign language in Indonesian schools. The primary objective of English instruction in Indonesia is to fulfil an instrumental purpose (Nababan, 2019), which includes preparing students for future employment, acquiring information in science and technology, and, crucially, fostering an open-minded perspective towards cultural diversity.

In Indonesia's public educational institutions, including universities, just writing skills are evaluated, rather than the four fundamental competencies. The evaluation of writing proficiency alone yields elevated grades, prompting students to exert considerable effort to achieve mastery in producing exemplary compositions. English speaking proficiency has seldom been evaluated. As assessment gains significant influence, meticulous considerations must be undertaken to establish a fair and valid evaluation. Assessment is frequently seen as a crucial educational phase (Bachman, 2022). Assessment significantly influences the methods of instruction and classroom activities. Moreover, (Fulcher, 2023) stated that the efficacy of a learning program is typically assessed by the outcomes of evaluation.

The assessment of speaking abilities in a second language has numerous obstacles, including the definition of language proficiency, the avoidance of cultural biases, and the achievement of validity (Sánchez, 2016). The evaluation of speaking skills frequently trails the emphasis placed on instructing those skills within the curriculum (Knight, 2022). Multiple factors contribute to the substandard quality of speech assessments, since various research indicate that lecturers lack the requisite skills to evaluate their students effectively due to inadequate training in Indonesia. The instructors are either hesitant to evaluate oral proficiency or lack trust in the reliability of their assessments (Knight, 2022). If instructors lack understanding in assessing students' speaking performance, their teaching competencies are likewise ineffective. Consequently, they must understand the criteria for evaluating speaking performance. This study proposes a university-level speaking assessment based on the effective assessment criteria established by (Brown & Abeywickrama, 2020), which encompass a specified criterion, an appropriate task, maximum output, and a practical, reliable scoring system.

Prior to evaluate speaking, it is essential to recognize five fundamental categories of speaking. (Brown & Abeywickrama, 2020, pp. 184-185) delineate five categories of speech as elucidated in table 1 below:

Table 1. Types of Speaking

No	Types of Speaking	
1.	Imitative	This form of speaking necessitates that test takers replicate a word, phrase, or sentence. Pronunciation is the primary component of the evaluation; however, grammar also contributes to the score criteria. It is important to emphasize in imitative speaking that communicative competence in the language is not crucial. They must obtain information and subsequently articulate it verbally without providing more clarification. Their output consists exclusively of the information they receive.
2.	Intensive	In contrast to imitative speaking, intensive speaking does not focus on pronunciation or phonological elements. Comprehending meaning is essential for addressing certain tasks; nevertheless, engagement with the counterpart is limited. The activity sample consists of reading aloud, as well as completing sentences and dialogues.
3.	Responsive	The authenticity of a conversation is paramount. Consequently, the speaker is compelled to articulate without delay. Responding to a brief chat by making a straightforward request is an activity characteristic of this form of communication.
4.	Interactive	The primary distinction between responsive and interactive

		speaking lies in the weight and complexity of the phrases. The number of speakers is also significant, as conversations may require more than two participants.
5.	Extensive	Extensive speaking encompasses a broad spectrum of verbal expression. The speaker must engage with the opposing speakers, which may involve answering questions and facilitating discussions. Extensive speaking is the paramount speaking skill that necessitates robust linguistic components.

Assessment of speaking can be highly subjective, sometimes leading individuals to compare native and nonnative speakers based on pronunciation (Luoma, 2024). (Nunan, 2019) posited that speaking necessitates linguistic competence, which includes the ability to articulate sounds clearly, possess an adequate vocabulary, and grasp structural or grammatical elements. Effective communication necessitates functional competence, which entails providing comprehensive and logical responses to enquiries. Another skill is strategic competence, wherein the speaker employs repair tactics to address conversational breakdowns. The final aspect is sociolinguistic and cultural competency. It requires speakers to utilize the language suitably for the context. This theory then evolved into the criterion for assessing speaking tests. Nevertheless, the design of speaking assessments may differ based on the categories of speaking assessed.

Speaking task involves completing the directive provided during the speaking examination. As with any assessment scores, speaking scores must be reliable, equitable, and, most importantly, applicable for their intended goals (Luoma, 2024). To guarantee the reliability of speaking skill assessment, several aspects must be considered (Hughes, 2023; Luoma, 2024; Nunan, 2019):

a. Practicality

The primary concept of language assessment is practicality. Prior to selecting a test, it is imperative to evaluate its practicality in relation to the constraints of time for administration and scoring interpretation, budget limitations, and available facilities.

b. Validity

Consistent and precise measurement is essential for the correct assignment of a test. It must quantify only the pertinent variables while removing any extraneous factors from consideration. Essay writing is not a valid assessment of speaking skills, as it does not accurately reflect the test takers' speech production capabilities. Consequently, it may not accurately reflect the true capabilities of the exam takers. In accordance with the types of speaking, the test designer must determine which speaking types will be assessed, as this will impact the assessment's design. The utilization of interview tests for imitative speech may result in assessment invalidity.

c. Reliability

The requirement of a consistent scoring measurement is of utmost significance to ensure the reliability of a test. In addition to that, a defined scoring rubric and standards is an absolute necessity. If a test is trustworthy, it does not necessarily mean that it is valid. On the other hand, a test designer ought to make every effort to maintain a trustworthy test that is as valid as possible. It is impossible to ignore the requirement of a dependable scoring system when it comes to speaking tests; therefore, test takers are required to develop a standard scoring system before the speaking test is administered. All the characteristics of what is to be evaluated from the pupils should be represented by the items that are included on it. The significance of the score must also be expressed in a clear manner on the form to guarantee that the capabilities of each student are accurately represented. The work of the students will be recorded using the normal scoring system while the test is being administered. It is quite difficult to obtain a solid result from the examination if there is not a trustworthy scoring system in place.

d. Authenticity

It refers to a language that is being used or a language that is contextual. The students are in charge of representing something that is connected to their core beliefs. If this is the case, then the language that is produced is genuine.

There is a backwash effect that is one of the goals of language assessment. It provides the lecturer as well as the students with information regarding the impact of the learning and teaching (Hughes, 2023, page 53). Considering the significance of the matter, it is imperative that this matter be investigated while developing a test. Moreover, this study fills a critical gap by providing a practical, validated, and contextually relevant model for assessing speaking in Indonesian EFL classrooms. Its novelty lies in its

systematic approach, integration of global best practices with local needs, and its potential to raise the standard and fairness of English-speaking assessment in Indonesia.

2. METHODS

This study describes a proposed test made by the writer. It investigates the application of assessment, the instruction for assessment, the criteria for scoring assessment, and the criteria for speaking presentation. The proposed test refers to the newly developed speaking assessment instrument designed specifically for the Speaking 1 course at Universitas PGRI Kanjuruhan Malang. The decision to use a proposed test as the main research instrument aligns with the study's goal of creating a valid, reliable, and contextually appropriate tool to measure students' speaking abilities.

The proposed test serves as both a product and a means of investigation to evaluate how well the newly designed assessment meets the criteria of effective language testing, including validity, reliability, practicality, and fairness. It is intended to replace or improve upon existing speaking assessments that are often informal, subjective, or insufficiently structured. It was administered as a summative assessment in the Speaking 1 course. Students prepared and delivered their presentations, which were then scored independently by trained raters using the rubric. This allowed the researcher to collect empirical data on the test's effectiveness and to identify areas for further refinement. By using a proposed test, this study not only evaluates the speaking skills of students but also provides a practical model that lecturers can adopt or adapt. This approach ensures that the test is grounded in the local educational context and responsive to the needs of both students and lecturers.

The Usage of Assessment

The assessment is designed to assess students' speaking skill. The result of the test will decide whether test takers pass or fail the subject (speaking 1 subject) in English study program, Universitas PGRI Kanjuruhan Malang.

The Steps in Developing Speaking 1 Assessment

The writer develops the speaking assessment by following these stages: identifying the instructional goal based on the course description and formulating the instructional objectives and syllabus of Speaking 1 course, developing the speaking instruction on selected topics, developing the scoring rubric, employing content and face validity for validation, administering the assessment instrument, and finally estimating the rater reliability. This process is illustrated in Figure 1.

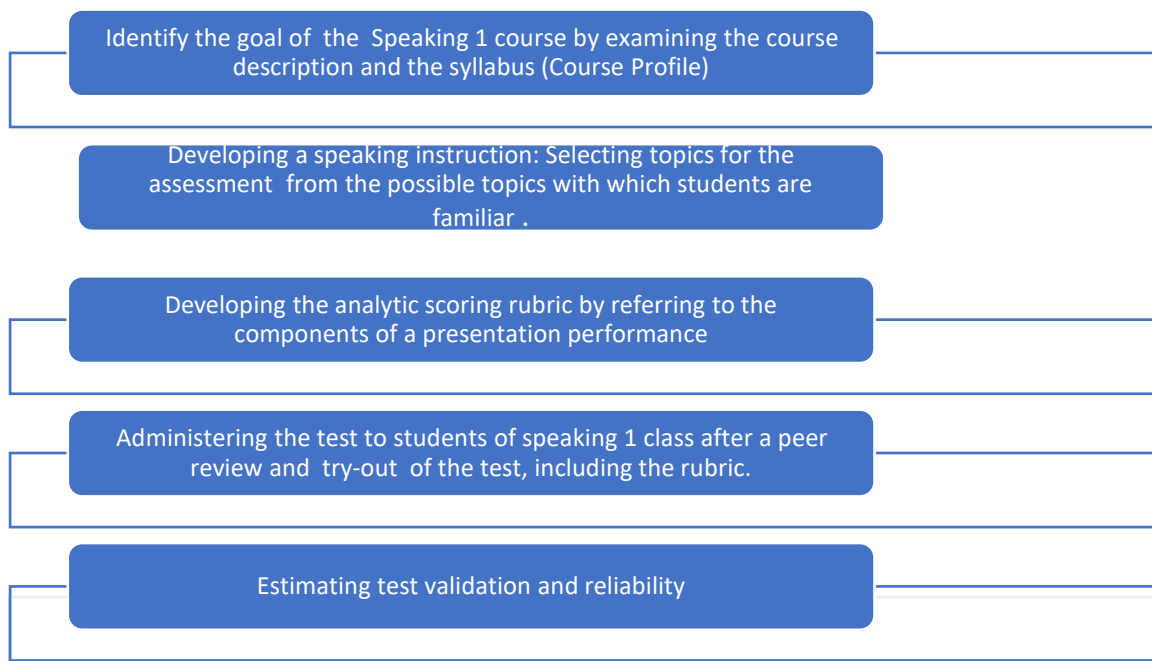


Figure 1. The Steps in Developing Speaking Assessment Instrument

Identifying the Goal of the Speaking Course

Before developing the speaking instruction and the scoring rubric, the first stage was examining the course description of Speaking 1 course (English Study Program Syllabus Universitas PGRI Kanjuruhan Malang, 2020, p. 26) as follows.

This course is intended to give students capability of specific speaking skill such as:

- a. (Basic knowledge of communication) describing person, expressing like dislike, describing personal change, stating advantages and disadvantages, talking about rule stating fact, making recommendation, describing process and strategy, giving reason and purpose, talking about hope and dream, describing and giving information about place, stating preference, presenting contrasting information, categorizing and evaluating issue.
- b. (General knowledge of communication) describing daily routine and habit, giving advice, explaining reason and condition, giving interpretation of meaning, comparing custom and habit, making generalization, presenting concern and offering solution, describing mental process, describing types, talking about appropriate behavior, reporting what someone else said.
- c. (Specific knowledge of communication) presenting information in chronological order, expressing regret about the past, evaluating traits of character, describing people attitude and believes, and doing monologue; reporting news. Besides the above skill, the comprehensibility, fluency, accuracy and pronunciation are also treated well to support the student's speaking skill.

By referring to various theories of instructional design (Brown, 2015; Finney, 2022; Marzano & Kendall, 2017; Nation & Macalister, 2020), a course profile, also known as a syllabus, was constructed. This was done based on the description of the course. Next, it was followed by the process of matching the instructional objectives with the course description, which entails making certain that the writing test is valid in terms of its written content. This course's ultimate objective is to help students improve their ability to communicate effectively through speaking. In the following manner, the goal and its instructional objectives were developed in accordance with the criteria of successful assessment that were proposed by (Brown & Abeywickrama, 2020). These criteria include a particular criterion, a suitable task, a maximum output, and a practical and a reliable scoring system.

Goal

"The students are able to have the ability of general knowledge of communication by presenting concern and offering solution orally to an issue effectively, correctly and appropriately."

Instructional Objectives

- a. Use varied information, of which all is accurate, relevant, and at the right level of detail.
- b. Demonstrate ideas in a polished, well-organized manner, introduction captures audience interests and established the purpose of the presentation, use evidence and examples to support main ideas in a fluent manner, and conclusion reinforces the main ideas in the presentation.
- c. Choose good quality of visual aids that are appropriate for the presentation setting.
- d. Show effective use of time.
- e. Present good capability of speaking fluency and coherence (pronunciation, intonation, syllable-stress, pauses, with no hesitation).
- f. Establish good aptitude of speaking accuracy (grammar).

The instructional objectives as the indicators of students' speaking ability in the assessment instruction were used as the components, or features, of a speaking scoring rubric.

Developing the Speaking Instruction

In this assessment, the instruction given is as follow:

SPEAKING 1
Midterm Test

Direction:

1. You are required to perform 15 minutes oral presentation.
2. Topic for presentation is “Schools Problems during Covid-19 Pandemic”.
3. Presentation is delivered by using interesting visual aids.
4. Your presentation performance will be evaluated on the basis of:
 - a) Varied information, of which all is accurate, relevant, and at the right level of detail
 - b) Ideas, well organized manner, introduction captures audience interests and established the purpose of the presentation, use evidence and examples to support main ideas in a fluent manner, and conclusion reinforces the main ideas in the presentation
 - c) Quality of visual aids that are appropriate for the presentation setting
 - d) Use of time
 - e) Capability of speaking fluency and coherence (pronunciation, intonation, syllable-stress, pauses, with no hesitation)
 - f) Aptitude of speaking accuracy (grammar)

Developing the Scoring Rubric

Based on the instructional objectives, the scoring rubric was developed by including the features of speaking presentation performance and the instructional objectives as the components.

The Scoring Rubric

The student's achievement was evaluated using a scoring rubric that served as a measuring tool. This was done to grade the student. Conducting an evaluation of pupils' oral performance is equivalent to evaluating their speaking ability. As a result, the rubric was meticulously designed to ensure that the results of the assessment would provide data that is both valid and reliable regarding the learning outcomes of the students. The purpose and the instructional objectives that were specified in the Course Profile were analyzed prior to the development of the scoring rubric. This was done since the instructional objectives would act as performance indicators that would be utilized in the process of establishing the scoring criteria. When the goal and instructional objectives were taken into consideration, the components of the scoring rubric, also known as the criteria and indicators of performance, were developed in accordance with the blueprint presented in Table 2.

Table 2. The Blueprint of the Speaking Assessment Adapted from Brown and Abeywickrama

The purpose of the assessment:

The assessment is directed to measure the students' ability of general knowledge of communication by presenting concern and offering solution orally to an issue effectively, correctly and appropriately.

Speaking Components: (The components were developed on the basis of the instructional objectives)	Description: (the criteria for qualified presentation performance)
1. Content	<input type="checkbox"/> Used varied information, of which all was accurate, relevant, and at the right level of detail.
2. Organization	<input type="checkbox"/> Presented ideas in a polished, well organized manner. <input type="checkbox"/> Introduction captured audience interests and established the purpose of the presentation. <input type="checkbox"/> Used evidence and examples to support main ideas in a fluent manner. <input type="checkbox"/> Conclusion reinforced the main ideas in the presentation.
3. Visual Aids	<input type="checkbox"/> Visual aids are professional looking. <input type="checkbox"/> Followed text size/amount guidelines. <input type="checkbox"/> Graphics and pictures are attractive, creative and precise so as to enhance the presentation
4. Time Management	<input type="checkbox"/> Carefully timed so that it fit into the time allowed. <input type="checkbox"/> Spent appropriate amount of time on topics. <input type="checkbox"/> Allowed time for questions, and answered questions effectively
5. Fluency and Coherence	<input type="checkbox"/> Speaks fluently with only rare repetition or self-correction using good pronunciation, intonation, syllable-stress, pauses, with no hesitation <input type="checkbox"/> Speaks coherently and develops topics fully and appropriately using good pronunciation, intonation, syllable-stress, pauses, with no hesitation
6. Accuracy	<input type="checkbox"/> Complex sentence use and minor grammatical occurrence

In their article from 2020, Brown and Abeywickrama argue that to deliver an effective assessment, there are four guidelines that need to be established: defining the criteria, assigning appropriate activities, presenting the maximum output, and establishing scoring processes that are both practical and reliable. The performance of the students is going to be evaluated based on the criteria listed in the table that pertains to oral presentations for this assessment. Each criterion is intended to make it simpler for the lecturer to evaluate the presentation of the pupils. Additionally, it is practical because the professor merely needs to mark the proper score that is supplied in to complete the task.

Those criteria that were defined by (Brown, 2017) are the basis for the criteria that are used to evaluate the performance of pupils. He suggests that there are at least six criteria that should be considered when evaluating speaking ability. These criteria are pronunciation, fluency, grammar, vocabulary, discourse feature, and assignment completion. As an additional component, a presenting competence checklist was included in the criteria for evaluating oral presentations.

Administering the Instrument

Prior to the speaking instrument being used to evaluate the students' performance, the draft of the instrument was subjected to peer review by two members of the team who were not included in the participants of the test. After that, it was modified and submitted for evaluation by undergraduate students from a separate group who were enrolled in the English Study Program and were taking the Speaking 1 course during the second semester of the academic year 2024/2025. It was intended for the purpose of evaluation, and the tryout was directed in terms of three aspects: appropriateness, comprehensibility of the direction, and the amount of time required to complete the test. Based on the comments received from the peer review, a few minor adjustments were made, including a change to the amount of time allotted for the test and the selection of themes and topics.

Employing Test Validation

For an evaluation to be considered valid, it must accurately measure the linguistic ability that is being evaluated. The exam takers should be given a speaking test rather than a writing test to evaluate their speaking abilities. According to (Brown & Abeywickrama, 2020), there are five different types of validity: face validity, content-related evidence, criterion-related evidence, construct-related evidence, and consequential validity.

Evidence that is related to the content also refers to the validity of the content, whereas test content should assess what needs to be measured. In this instance, the exam that was devised is used to evaluate the speaking abilities of students. To do so, they are required to give oral presentations that involve demanding and comprehensive tasks. For the purpose of evaluating the speaking ability, oral presentation is selected.

Estimating the Rater Reliability

According to (Brown & Abeywickrama, 2020), there are four components that contribute to the assurance of test reliability. These components are student-related dependability, rater ability, test administration reliability, and test reliability. During the two-time meetings, the evaluation is carried out, and the presenting turn of the test takers is not determined by the order of their names in alphabetical order. In addition, test takers can profit from the two-meeting assessment. Even if they are unable to attend the initial meeting due to illness, those who are taking the test will still have another chance to be evaluated. Additionally, the turn arrangement ought to improve the dependability of student-related matters. When they reach this step, test takers are provided with the ability to select the most appropriate time to conduct their evaluation, which assists them in overcoming their anxiety.

3. RESULT AND DISCUSSION

This study resulted in the development of a speaking assessment specifically tailored for first-year university students in the English Study Program at Universitas PGRI Kanjuruhan Malang. The process followed a systematic approach, beginning with the identification of instructional goals and culminating in the creation of a validated assessment instrument and scoring rubric. The assessment was designed to measure students' ability to present concerns and offer solutions orally, in accordance with effective assessment criteria as outlined by (Brown & Abeywickrama, 2020).

Stages of Assessment Development

- a. The instructional objectives were derived from the course syllabus, ensuring alignment between the assessment and the learning outcomes. These objectives included the use of accurate and relevant information, organized presentation structure, effective use of visual aids, time management, fluency and coherence, and grammatical accuracy.
- b. Based on these objectives, assessment instructions and a detailed scoring rubric were developed. The rubric provided clear descriptors for each performance criterion, supporting both validity and reliability in scoring.
- c. Validation was conducted through content and face validity checks, involving expert review and piloting with a sample group of students. This ensured that the assessment tasks were contextually relevant and that the rubric accurately reflected the intended competencies.

Implementation and Findings

- a. The speaking assessment was administered as a summative evaluation for the Speaking 1 course. Students were tasked with presenting a concern and offering a solution, using supporting evidence and appropriate visual aids.
- b. The assessment results revealed that students generally demonstrated strength in organizing their presentations and providing relevant information. The majority were able to engage the audience with clear introductions and conclusions, and most utilized visual aids effectively.
- c. However, challenges were observed in areas of fluency and grammatical accuracy. Some students hesitated during delivery or made frequent grammatical errors, indicating the need for further instructional support in these areas.
- d. The scoring rubric facilitated objective and consistent evaluation by multiple raters. Inter-rater reliability was estimated using standard procedures, confirming that the rubric supported consistent scoring across different evaluators.

Student and Lecturer Feedback

- a. Qualitative feedback from students indicated that they appreciated the transparency and clarity of the assessment criteria. They found the rubric helpful for understanding expectations and preparing their presentations.
- b. Lecturers reported that the assessment instrument was practical and aligned well with course objectives. The detailed rubric enabled them to provide targeted feedback and support student learning more effectively.

Summary of Key Outcomes

- a. The research demonstrated that a carefully designed, contextually relevant speaking assessment can provide valid and reliable measures of oral proficiency in an Indonesian EFL context.
- b. The use of clear instructional objectives, a structured scoring rubric, and validation procedures contributed to the fairness and meaningfulness of the assessment process.
- c. The findings suggest that ongoing training and support for both students and lecturers are necessary to address persistent challenges in fluency and grammatical accuracy.

Based on the result of this study and the principles outlined by (Brown & Abeywickrama, 2020), here is an example of a speaking assessment suitable for an Indonesian EFL (English as a Foreign Language) university classroom, along with a sample scoring rubric.

Example of Speaking Assessment

Title: Presenting a Concern and Offering a Solution

Instruction for Students:

You are required to deliver an oral presentation in English on a real-life concern or problem relevant to your community, school, or daily life. Your task is to:

- a. Clearly describe the concern or problem.
- b. Explain why it is important or relevant.
- c. Propose a practical solution to address the concern.
- d. Support your ideas with evidence, examples, or data.
- e. Use appropriate visual aids (e.g., slides, posters) to enhance your presentation.
- f. Speak for 3–5 minutes.

Additional Guidelines:

- a. Prepare your presentation in advance
- b. Practice to ensure fluency and confidence.
- c. Visual aids should be clear, relevant, and easy to understand.
- d. You may use note cards, but do not read your entire presentation.
- e. Be ready to answer one or two questions from the audience or lecturer after your presentation.

Table 3. Example of Scoring Rubric

Criteria	Excellent (4)	Good (3)	Fair (2)	Needs Improvement (1)
Content Accuracy & Relevance	All information is accurate, relevant, detailed	Mostly accurate and relevant, minor errors	Some inaccuracies or irrelevant info	Many inaccuracies, lacks relevance
Organization & Structure	Clear intro, logical flow, strong conclusion	Mostly clear, some minor lapses in flow	Lacks clear structure, weak conclusion	Disorganized, hard to follow
Use of Visual Aids	Visuals are clear, well-integrated, enhance talk	Visuals are appropriate, somewhat helpful	Visuals are present but not effective	No visuals or visuals are distracting
Time Management	3–5 minutes, well-paced	Slightly over/under time, mostly well-paced	Significantly over/under time, rushed/slow	Far from required time, poor pacing
Fluency & Coherence	Smooth delivery, clear	Minor hesitations, mostly clear	Frequent pauses or unclear	Many hesitations, difficult to

	pronunciation, no pauses	pronunciation	pronunciation	understand
Grammatical Accuracy	Few or no errors, complex structures used	Some errors, do not impede understanding	Frequent errors, sometimes impede meaning	Many errors, hard to understand

Scoring:

Each criterion is scored 1–4,

Maximum total score: 24,

Passing score: (Set by lecturer, e.g., 15/24).

Sample Assessment Prompt for Students:

“Think about a problem you have noticed in your campus environment (for example: waste management, lack of green spaces, or student stress). Prepare a 3–5-minute presentation in English where you describe the problem and propose a realistic solution. Use at least one visual aid to support your ideas.”

How to Use This Assessment:

Preparation: Give students the prompt and rubric in advance.

Delivery: Students present individually or in small groups.

Evaluation: Lecturer and/or peer assessors use the rubric during each presentation.

Feedback: Provide specific comments based on rubric criteria.

In line with the result of this study, it can be concluded that the development and implementation of a speaking assessment instrument tailored for the Indonesian EFL university context resulted in a practical, valid, and reliable tool for evaluating students’ oral proficiency. The systematic process beginning with the identification of instructional goals, formulation of clear objectives, design of assessment instructions, and construction of a detailed scoring rubric, ensured that the assessment was aligned with both theoretical principles and the specific needs of the local context. Validation and reliability checks confirmed the instrument’s effectiveness, while the use of transparent criteria and structured tasks supported fair and meaningful evaluation. Overall, the results demonstrate that adopting contextually relevant and methodologically sound assessment practices can significantly enhance the quality and fairness of speaking evaluations in Indonesian higher education EFL classrooms.

Discussion**To What Extent is the Assessment Practical?**

According to (Brown & Abeywickrama, 2020), practical tests are not expensive, they are easy to administer, they are within the time constraints limited, and the technique of scoring is particular and efficient in terms of time that is available. In light of the aforementioned considerations, it is indisputable that the evaluation that was developed for speaking satisfies the standards that Brown and Abeywickrama have outlined. First of all, the oral presentation does not require a significant amount of money to be carried out. Students are given the opportunity to select their own themes independently. Both the amount of work and the expense are dependent on the capabilities of the students. Examinees have the ability to select the appropriate medium in which to deliver their presentation. The lecturer is capable of handling the assessment on their own, thus there will be no need for an additional proctor.

The second assignment is for each student to give a presentation that lasts for fifteen minutes. In a classroom with twenty students, the amount of time required to complete the test is three hundred minutes. As a result, the examination will be carried out in two separate meetings, but it will still be completed within the allocated time. Thirdly, in order to carry out the test, it is not necessary to make use of any difficult methods or media (Knight, 2022).

In conclusion, the test makes use of direct assessment scoring for the final reason. As the mark is awarded on the spot through a zoom meeting session, the lecturer does not have to listen to the recordings that the students have made, which would be a very time-consuming process. In addition, the criteria for scoring are presented in a transparent manner.

To What Extent is the Assessment Reliable?

In this study, elements that can influence student-related reliability are expected. In order to minimize the possibility of performing an unreliable exam due to the rater factor in the oral presentation test, clear and exact rules for scoring are supplied. During the test, the rater is protected from the practice

of scoring using intricate techniques, which can result in inconsistency and confusion in the marking process. The only thing that is required of the lecturer or rater is to cross off pertinent rating criteria. The environment in which the exam is conducted is a significant factor in determining the reliability of the test. The lecturer or rater of the test should make sure that all the material and the zoom meeting application are ready to use before the test begins. This will ensure that the dependability of the test administration does not contribute to an unreliable factor (Sanchez, 2016).

By providing very specific guidance and instructions in advance, it is possible to avoid unreliability in the test. Also contributing to the inaccuracy of the test is the amount of time required to complete it. The fact that students are only given fifteen minutes to show their work, on the other hand, means that time will not be an issue. In addition, they have been provided with information regarding the test a significant amount of time before the deadline, which means that they have a lot of time to be ready for it. These two factors ought to eliminate the possibility of the test being less reliable than it otherwise would have been (Fulcher, 2023).

To What Extent is the Assessment Valid?

It is guaranteed that the content validity of the test will be met. For the purpose of evaluating speaking abilities, the criteria for scoring the exam have been constructed in accordance with the criteria that were developed by (Brown's, 2007) evaluation performance principles.

As a result of taking this examination, individuals who take the examination will either pass or fail the topic. Those individuals who are unable to pass the examination are required to review the material once more. As a result, the goals of consequence validity are very obvious. The term "face validity" refers to the extent to which individuals who take a test are aware that the test is designed to evaluate their specific abilities. It is essential that the rater or lecturer provides clear guidance and direction in order to enhance the level of awareness among test takers on the skill against which they will be evaluated. Based on the fact that this test provides both guidance and instruction, it appears that face validity has been satisfied (Nababan, 2019). The information regarding the examination is presented in the clearest possible manner. In addition, students will be provided with grading criteria, which will allow them to be aware of the specific linguistic components that will be evaluated.

To What Extent is the Assessment Authentic?

There are a large variety of authentic aspects that are included in this oral presenting test. In the first place, themes are selected according to current issues. It means that they are able to take any resources from the real world and use them in their research. While the themes are being presented, test takers are required to demonstrate a wide range of language abilities and components, including speaking, listening, writing, reading, structure, pronunciation, vocabulary, and many more. It is at this moment that the presentation to other people is integrated with the language abilities and language components that have been acquired. They learn to use language in its entirety, rather than in isolated parts (Bachman, 2022).

In real-world situations, the ability to communicate information orally is required. When students combine their oral presenting skills with their speaking skills, it is unquestionably easier for them to practice the skills that they will require in the future (Douglas, 2024). The participants in the exam and the audience are exposed to real-life communication during the discussion time. The questions and responses that are asked during this time are not based on any specific events. It is accurate to assert that the test possesses a very high level of language authenticity because the facts that are presented are supported.

Will the Assessment Create Positive Wash Back? How? Why?

One of the advantages of having detailed grading standards is that it allows students or anyone taking tests to have a better understanding of their own potential and limitations. As a result, kids are aware of which aspects or talents related to language have been learnt and which ones require further development. For the purpose of providing students with a detailed score of the skills that were evaluated during the test, the grading sheet for this test is meant to provide clear information about the performance of the students.

The remark section of the grading sheet is supplied; hence, raters are able to write feedback that is both generous and particular. This can provide students with an innate interest in the subject matter, which in turn increases the likelihood of good wash back (Brown, 2020).

This study offers a systematic and contextually relevant model for developing speaking assessments tailored to Indonesian EFL classrooms. The proposed assessment instrument and rubric are grounded in effective assessment principles, ensuring clarity, fairness, and reliability. By providing clear instructional objectives and a validated scoring rubric, the study empowers lecturers to conduct more objective and

transparent evaluations, reducing subjectivity and bias in speaking assessments. Additionally, the model supports student development by clarifying expectations and promoting comprehensive communicative competence.

The findings provide English lecturers in Indonesian higher education with practical tools and guidelines for designing and implementing fair and reliable speaking assessments. The adoption of this model can enhance the validity and consistency of speaking evaluations, ultimately improving the quality of English language instruction and supporting students' readiness for real-world communication demands. Furthermore, the model can be adapted and replicated in other institutions, contributing to broader improvements in EFL assessment practices across Indonesia.

Despite its strengths, the study has several limitations. The assessment instrument was developed and validated within a single university context, which may limit its generalizability to other institutions with different student populations or resources. The study also relied on a relatively small sample for pilot testing and rater reliability estimation, which could affect the robustness of the findings. Additionally, the research primarily focused on summative assessment and did not explore the integration of formative assessment practices or the impact of ongoing feedback on student learning.

Future research should consider expanding the sample size and involving multiple universities to enhance the generalizability of the findings. Investigating the effectiveness of the assessment model in diverse educational contexts and with different proficiency levels would provide deeper insights. Further studies could also explore the integration of formative assessment strategies, peer assessment, and the use of technology to facilitate speaking evaluations. Longitudinal research examining the impact of the proposed assessment model on students' speaking development over time would also be valuable.

4. CONCLUSION

When it comes to constructing a speaking assessment, significant considerations need to be taken into account because speaking assessment falls under the category of being subjective. It is possible to apply the criteria that were produced by (Brown & Abeywickrama, 2020) as guidelines to construct an assessment for speaking. These criteria include a specified criterion, an acceptable task, a maximum output, and a scoring process that is both practical and reliable. The lecturers are the ones who are responsible for carrying out this kind of evaluation, thus it is necessary for them to be familiar with the issues of the practicality, validity, reliability, authenticity, and wash-back impact. Additionally, because there may be a large number of lecturers who have a limited understanding of speaking evaluation, it is recommended that the relevant institutions and the government contribute to the provision of supports for lecturers to develop their professionalism. These supports could take the form of trainings, workshops, or seminars, for instance.

The practical implication of this study is that it provides English lecturers in Indonesian higher education with a systematic, validated, and contextually relevant model for assessing students' speaking skills, enabling them to implement fair, transparent, and reliable speaking assessments that align with effective assessment criteria. By adopting the proposed assessment method and rubric, lecturers can minimize subjectivity and cultural bias, offer more meaningful feedback, and better support students' communicative competence, ultimately improving the overall quality of English language instruction and assessment in Indonesian universities.

5. REFERENCES

- Alderson, J. C., & Banerjee, J. (2022). Language testing and assessment (Part 2). *Language Teaching*, 35, 79-113.
- Bachman, L. F. (2022). *Language assessment and measurement: Issues and practice*. Routledge.
- Brown, H. D. (2017). *Teaching by principles: an Interactive approach to language pedagogy* (4th ed.). Pearson Education.
- Brown, H. D. & Abeywickrama, P. (2020). *Language assessment: principles and classroom practice* (3rd ed.). New York: Pearson Education Inc.
- Douglas, D. (2024). Quantity and quality in speaking test performance. *Language Testing*, 11, 125-44.
- Finney, D. (2022). *Curriculum development in language teaching*. Cambridge University Press.
- Fulcher, G. (2023). *Practical language testing* (3rd ed.). Routledge.
- Hughes, A. (2023). *Testing for language teachers* (3rd ed.). Cambridge University Press.
- Fulcher, G. (2017). Tests of oral performance: The need for data- based criteria. *English Language Teaching Journal*, 41, 287-91.
- Knight, B. (2022). Assessing speaking skills: a workshop for teacher development. *ELT Journal*, 46(3), 294-

302.

- Luoma, Sari. (2024). *Assessing speaking*. Cambridge: Cambridge University Press.
- Marzano, R. J., & Kendall, J. S. (2017). *The new taxonomy of educational objectives* (3rd ed.). Corwin Press.
- Nababan, P. W. J. (2019). English in the Indonesian education system: Its role and function. *Journal of Language and Education*, 5(2), 12–21.
- Nation, I. S. P., & Macalister, J. (2020). *Language curriculum design* (2nd ed.). Routledge.
- Nunan, D. (2019). *Teaching English to speakers of other languages: An introduction*. Routledge.
- Sánchez, A. (2016). Validity and reliability in language assessment: Issues and challenges. *Language Testing in Asia*, 6(1), 1–10.
- Upshur, J., & Turner, C. E. (2015). Constructing rating scales for second language tests. *English Language Teaching Journal*, 49, 3-12.
- Upshur, J., & Turner, C. E. (2019). Systematic effects in the rating of second language speaking ability: Test method and learner discourse. *Language Testing*, 16, 82-111.