

# DEVELOPING ASSESSMENT FOR SPEAKING

Yenny Rahmawati & Ertin  
Syarif Hidayatullah State Islamic University of Jakarta  
yenny\_uin@yahoo.com

## ABSTRACT

Recently there have been debates on assessing students' performances on speaking since the cultural and subjective issues embedded in bringing awareness on how teachers construct their speaking assessment. The main focus of this paper is a way to design assessment for speaking suitable for the Indonesian context at a university level. This paper stresses the criteria of effective assessment proposed by Brown and Abeywicrama which consists of a specific criterion, an appropriate task, a maximum output and practical and a reliable scoring procedure. It is recommended that teachers develop their speaking assessment which is appropriate and contextual.

**Key Words:** assessment, effective assessment, speaking

## ABSTRAK

*Terdapat banyak perdebatan dalam hal menguji kemampuan speaking siswa, yang disebabkan adanya isu budaya dan subjektivitas dalam menyusun tes kemampuan speaking. Fokus utama dari paper ini adalah mengembangkan sebuah cara untuk mendesain tes speaking yang sesuai dengan konteks Indonesia untuk level universitas yang mengacu pada teori tentang kriteria penilaian yang dikembangkan oleh Brown dan Abeywicrama, yaitu: a specific criterion, an appropriate task, a maximum output and practical dan a reliable scoring procedure. Paper ini juga merekomendasikan agar para pengajar mampu menyusun tes speaking mereka secara tepat dan kontekstual.*

**Kata Kunci:** penilaian, penilaian yang efektif, berbicara

## INTRODUCTION

The development of English as a global language has strengthened its position as a lingua franca. As a consequence, most countries in the world, especially non-English speaking countries, consider English as an important language to be learnt. This condition impacts on the educational system in many countries; some have English as a medium of instruction, while others have English as a compulsory subject at school. English occupies important position in Indonesia educational system. As a foreign language, English is learnt and tested at Indonesian schools. English teaching in Indonesia aims primarily to serve the "instrumental function" (Nababan, 1991, p.123), that is, to serve as future orientations - to obtain jobs, to gain knowledge in the fields of science and technology, and most importantly, to build an open-minded attitude toward cultural differences.

Instead of assessing four basic skills, only writing skills are assessed in the public sector of educational institutions including universities in Indonesia. The assessment of writing skills alone gives high grades and students work hard for mastery in writing excellent pieces. English Speaking Skills have rarely been assessed.

As assessment becomes very powerful, therefore careful considerations should be taken into account to build a fair and a valid assessment. Assessment is often considered as an important instructional step (Bachman, 1990). The way learners are taught and activities carried out in the classroom are greatly influenced by assessment. Further, Fulcher (2003) said that the success of a learning program is commonly determined by the result of assessment.

There are many challenges in the assessment of oral skills in a second-language including defining language proficiency, avoiding cultural biases, and attaining validity (Sánchez, 2006). Assessment of speaking skills often lags far behind the importance given to teaching those skills in the curriculum (Knight, 1992).

Several factors also contribute to the low quality of speaking assessment, as some studies show that teachers are lacking of knowledge on how to assess their students due to the poor training conducted in Indonesia. The teachers are either reluctant to test oral ability or lack of confidence in the validity of their assessments (Knight, 1992). If the teachers are lack of knowledge on how to assess their students speaking performance, their competences in teaching are also far from effective. Therefore, they need to know criteria to

assess speaking performance. This paper suggests a speaking assessment for the university level on the basis of the criteria of effective assessment proposed by Brown and Abeywicrama, which include a specific criterion, an appropriate task, a maximum output and practical and a reliable scoring procedure.

## **SPEAKING TYPES**

Before assessing speaking, we need to acknowledge five basic types of speaking. Brown and Abeywickrama (2010, p. 184-185) propose five types of speaking as explained in the following.

### **1. Imitative**

This type of speaking requires the test takers to copy a word, phrase, or a sentence. Pronunciation is the main aspect of the assessment although grammar also takes part as the scoring criteria. What needs to be highlighted in imitative speaking is that the communicative competence of the language is not essential. They need to acquire some information, and then reproduce it orally without having to add extra explanation. What comes out from them is solely the information they hear.

### **2. Intensive**

Unlike imitative, intensive speaking does not emphasize on

pronunciation or phonological aspect. Understanding meaning is needed to respond certain tasks but the interaction with the counterpart is minimal. The activity sample is reading aloud, sentence and dialogue completion.

### **3. Responsive**

Authenticity in a conversation is important. Therefore, the speaker is stimulated to speak promptly. To response a short conversation, making a simple request comment is a kind of activity that belongs to this type of speaking.

### **4. Interactive**

The load and complexity of the sentences is the major different between responsive and interactive speaking. The number of the speakers also matter as sometimes it needs more than two people in the conversation.

### **5. Extensive**

Extensive speaking involves a wide range of speech production. Also, the speaker will need to interact with the counter speakers, which could be answering question, making discussion. It can be said that extensive speaking is the ultimate speaking skill that requires strong language components.

## ASSESSMENT OF SPEAKING

Assessment on speaking can be a very judgmental issue, in which people tend to relate on native/nonnative speakers on the basis of pronunciation (Luoma, 2004). Additionally, Nunan (1999) viewed that speaking requires someone to be linguistically competence in term of well articulating the sound, having sufficient vocabulary, and mastering structural or grammatical components. To speak also needs functional competence which means answering questions completely and logically. Another competence is strategic competence in which the speaker is able to use repairing strategies when conversation breaks down. And the last one is sociolinguistic/cultural competence. It demands the speakers to use the language appropriately to the context.

This theory then developed as the criteria of speaking test assessment. However, the design of speaking assessment may vary; depend on the types of speaking assessed. Then, what should to be tested? (Nunan, 1999).

### Grammar

Test takers are assessed on how to control its usage within sentences, to construct, to use it appropriately and accurately and to avoid grammatical errors in speaking.

### Vocabulary

The range, precision, and the usage of vocabulary features in a conversation used by test takers indicate the level of how proficient they are.

### Comprehension

Understanding the context of the conversation and able to give appropriate response according to the question.

### Fluency

The language fluency indicates that the production of speech in a conversation is well delivered. Have confidence in delivering the speech and able to responds specific theme without many hesitation in choosing words.

### Pronunciation

Pronunciation deals with how often errors in pronunciation occur and how the pronunciation aspect interfere the communication are the criteria of the assessment.

### Task

Task deals with finishing the command given during the speaking test.

Like all test scores, speaking scores must be dependable, fair, and above all useful for the intended purposes (Luoma, 2004). To ensure

speaking skill assessment is trustworthy, there are factors that should come into consideration (Hughes, 2003; Luoma, 2004; Nunan, 1999):

1. Practicality

The first principle of making language assessment is practicality. Before deciding a test, we need to analyze how practical the test is to be used considering the time constraint of running and interpreting the scoring of the test, budget limitation, and facilities.

2. Validity

Consistently accurate measurement must be provided to assign a valid test. It has to measure what should be measured by excluding all irrelevant variables to be tested. When speaking skill is tested, making essay is not a valid test as it fails to provide information of the test takers speech production. In result, it may not bring about the real test takers' ability.

In accordance to types of speaking, test designer should decide what kind of speaking types would be tested as it will influence the design of assessment. Using interview test for imitative speaking may lead to invalidity of assessment.

3. Reliability

The need of consistent scoring measurement is very important to

make a test reliable. In addition, clear rubric and scoring criteria is also a must. Sometimes a reliable test may not be a valid test. However, a test designer should struggle to keep a reliable test as valid as possible.

As a need of reliable scoring system is unavoidable in speaking test, before conducting speaking test, test takers must prepare a standard scoring system. The items on it should represent all aspects of what are to be assessed from the students. The weight of the score must be printed clearly on the form as well to make sure each student's ability is well presented. During the test, the standard scoring system will be used to record students' work. Without a good scoring system, it is hardly possible to have a reliable result of the test.

4. Authenticity

It refers to a contextual language or language in use. Students are asked to represent something related to their values. In that case, the language produced is authentic.

One goal of language testing is its backwash effect. It tells both teacher and learners of the effect of the learning and teaching (Hughes, 2003, p.53). As it is important, therefore, this issue should also be explored in designing a test.

## DEVELOPING ASSESSMENT FOR SPEAKING

This section describes a proposed test made by the writers. It explores the usage of assessment, the assessment instruction, scoring assessment, and oral presentation criteria.

### The Usage of Assessment

The assessment is designed to assess students' extensive speaking skill. The result of the test will decide whether test takers pass or fail the subject (speaking subject).

### Assessment Instruction

In this task, the instruction given is as follow:

- a) Students are required to perform 10 minutes oral presentation which consists of 8 minutes of presentation and 2 minutes of discussion time. Topic for presentation is free. Students can pick any themes that interest them. During discussion time, presenter has to lead the discussion to make sure it is not out of topic.
- b) Presentation is delivered by using power point or overhead projector. The media is provided but students need to prepare the materials. Scoring criteria sheet is given to acknowledge students about the skills going to be assessed.

- c) Due to a large number of students, the test will be held in two times meetings. Students may choose to deliver their presentation on the first or second meeting. The turn is not based on alphabetical order but student's willingness.

### Scoring Assessment

Brown and Abeywickrama (2010) contend that to provide effective assessment, there are four rules that need to establish: specify criterion, give appropriate tasks, present maximum output, and set practical and reliable scoring procedures. For this assessment, the table on oral presentation criteria below is going to be used to evaluate students' performance. Each criterion is designed to ease teacher to score students' presentation. It is also practical as teacher only needs to put ticks on the appropriate score presented in.

The criteria used to evaluate students' performance are based on those developed by Brown (2007). He suggests there are at least are six criteria to assess speaking skill: pronunciation, fluency, grammar, vocabulary, discourse feature, and task accomplishment. In addition, presentation skill checklist will be added to oral presentation assessment criteria. However, the point is not more than twenty percent of overall score to

maintain the validity of assessment which focuses on speaking skill (Table 1).

Each rating criteria is worth some points. The table below shows the numbers.

**Table 2. Rating Points**

Initial	Criteria	Score
E	excellent	5 points
VG	Very good	4 points
G	Good	3 points
S	satisfactorily	2 points
P	poor	1 point

**Table 1. Oral Presentation Assessment Criteria by Brown & Abeywickrama (2010)**

Criteria	E	VG	G	S	P	Comment
<b>SPEAKING SKILLS</b> <i>Fluency and Coherence</i> <i>speaks fluently with only rare repetition or self-correction; Speaks coherently and develops topics fully and appropriately</i>  <i>Lexical resource and range</i> <i>Express with some flexibility and appropriateness, giving effective descriptions and expressing viewpoints on a variety of topics.</i>  <i>Grammatical range and accuracy</i> <i>Complex sentence use and minor grammatical occurrence.</i>  <i>Pronunciation</i> <i>Pronounce words correctly, articulate clearly, intonate appropriately</i>  <i>Interaction (Listen and respond)</i> <i>Good contribution to other</i> <i>Active in conversation development</i>  <i>Task accomplishment</i>						
<b>PRESENTATION SKILLS</b> <i>Presentation was organized, information was logical and presented in well sequencing.</i> <i>The assignment was completed according to instruction provided.</i> <i>Presentation done within time allocation</i>						
<b>Total</b>						

Comments: \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

## ASSESSMENT DISCUSSION

### **To What Extent is Your Assessment Practical?**

Brown and Abeywickrama (2010) state that practical tests are not expensive, within time constraints, easy to conduct, and procedure of scoring is specific and efficient in time. Based on the factors above, it is unquestionable that the assessment designed for speaking fulfills Brown and Abeywickrama's requirements. Firstly, it does not need a lot of money to run the oral presentation. Students are given freedom to choose their own topics. Thus, the workload and the cost are up to students' ability. Test takers can choose the right media to deliver their presentation. Extra proctor will not be needed as the teacher alone can handle the assessment.

Secondly, each student is assigned to have 10 minutes presentation and 2 minutes discussion time inclusive. With 35 students in class, the time needed to finish the test is 350 minutes or 4 hours and 10 minutes. Thus, the test will be conducted in two meetings but still within the allotted time. Thirdly, to conduct the test, it does not need complicated technique or media.

The last reason, direct assessment scoring is used in the test. Teacher does not have to listen to students' recording which is very time-consuming as grade is given on the spot. Moreover, criteria of scoring are clearly provided.

### **To What Extent is Your Assessment Reliable?**

Four components to assure test reliability are student-related reliability, rater ability, test administration reliability and test reliability (Brown and Abeywickrama). The assessment is done within two time meetings and test takers' presentation turn is not based on alphabetical order name. Two-meeting assessment also benefits test takers. If they miss the first meeting due to sickness, the test takers still have another opportunity to be tested. The turn arrangement should also increase student-related reliability. At this point, test takers are given opportunity to choose the right timing to undergo their assessment which helps them overcoming anxiousness. In conclusion, factors that might influence student-related reliability are anticipated

To minimize the risk of unreliable test due to the rater factor in oral presentation test, clear and precise criteria of scoring are provided. Rater is avoided from complicated technique scoring during the test which can lead

to inconsistency and confusion marking. Teacher or rater is only required to tick appropriate rating points.

Test administration reliability comes from the milieu where test is administered. To make sure test administration reliability does not contribute to unreliable factor, before the test teacher/rater should ensure all media (computer, overhead projector) are ready to use. Classroom where test is held should also be prepared such as the seat arrangement. In addition, choosing classroom with minimum level of noise should be considered.

Test unreliability can be avoided by giving clear direction and instruction beforehand. Time needed to accomplish the test also triggers test unreliability. However, since students are only assigned 10 minutes presentation, time will not be a problem. In addition, information regarding to the test has been notified long before the due, thus they have plenty of time to prepare. These two considerations should eliminate greater risk of the test being unreliable.

### **To What Extent is Your Assessment Valid?**

Validity means the assessment should measure the language skill being assessed. To test speaking, test

takers should be given speaking test, not writing test. Brown and Abeywickrama (2010) divide validity into five types; content-related evidence, criterion-related evidence, construct-related evidence, consequential validity, and face validity.

Content-related evidence also refers to content validity whereas test content should measure what needs to be measured. In this case, the test designed is used to test students' extensive speaking skills, where they need to produce monologue which involves complex extensive task. Oral presentation is chosen to measure the skill as form of monologue. In conclusion, the content validity of the test is guaranteed. The scoring criteria of the test is designed based on the criteria developed by Brown's (2007) evaluation performance principles in assessing speaking skills.

The consequence of taking this test is that test takers pass or fail the subject. Meanwhile, for those who fail the test, they need to re-study the topic. Therefore, consequence validity goal is clear.

Face validity refers to test takers fully understand that a test is established to assess their particular skills. To raise test takers awareness of what skill they are going to be tested, it

is important that rater/teacher gives clear instruction and direction. In this test, face validity seems to be fulfilled as direction and instruction are given. All information of the test is delivered as clear as possible. Moreover, students will receive grading criteria so that they know exactly what language components are marked.

### **To What Extent is Your Assessment Authentic?**

This oral presentation test involves a wide range of authentic factors. First, topics are chosen based on students' interest. It means that they can take any materials from the real world perusals. In presenting the topics, test takers perform many language skills and components such as speaking, listening, writing, and reading, structure, pronunciation, vocabulary, etc. At that point, language skills and language components are combined to the presentation to other people. They learn to use language as a whole, not isolated.

Oral presentation skill is needed in real-world situation. Combining speaking skill with oral presentation definitely facilitates students to practice the skill that they need in the future. During discussion time, test takers and audience face real-life communication where questions and answers occur during the time are not based on

scenarios. To say that the test contains very high language authenticity is true as the facts given are supported.

### **Will Your Assessment Create Positive Wash Back? How? Why**

One benefit of having detail grading criteria is that students or test takers can really understand their strengths and weaknesses. Therefore, they know which language skills or components have been mastered or need improvement. The grading sheet for this test is designed to give clear information of students' performance so that students receive detail score of skills assessed in the test.

In the grading sheet, comment part is provided, therefore rater can write generous and specific feedback as this can give intrinsic interest to the students which enhance positive wash back (Brown, 2010).

### **CONCLUSION AND SUGGESTION**

As speaking assessment falls into subjective, careful considerations need to be taken into account when developing speaking assessment. The criteria developed by Brown and Abeywickrama, i.e. a specific criterion, an appropriate task, a maximum output and practical and a reliable scoring procedure, can be used as guidelines to build assessment for speaking. Since

teachers are the one who execute this type of assessment, they must be familiar with the issues on practicality, validity, reliability, authenticity and wash-back effect. In addition, for there may be many teachers who have limited knowledge on speaking assessment, it is then recommended that the related institutions and government can contribute to provide supports for teachers to develop their professionalism, for example in the form of trainings, workshops or seminars.

## REFERENCE

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Brown, H. D. (2007). *Teaching by principles: an Interactive approach to language pedagogy* (3<sup>rd</sup> ed.). New York: Pearson Education Inc.
- Brown, H. D. & Abeywickrama, P. (2010). *Language assessment: principles and classroom practice* (2<sup>nd</sup> ed.). New York: Pearson Education Inc.
- Fulcher, G. (2003). *Testing second language speaking*. London: Longman/Pearson Education.
- Knight, B. (1992). Assessing speaking skills: a workshop for teacher development. *ELT Journal*, 46(3), 294-302.
- Luoma, Sari. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Nababan, P. W. J. (1991). Language in education: the case of Indonesia. *International Review of Education*, 37(1), 115-131.
- Nunan, D. (1999). *Second language teaching and learning*. USA: Heinle.

