

The Advantages of Using an Analytic Scoring Procedure in Speaking Assessment

Nur Mukminatien
Universitas Negeri Malang

Abstract: This article describes the advantages of using analytic procedure in speaking assessment. An analytic scoring guide, as compared to the impressionistic one, has a double function: as an instrument to measure the learner's speaking proficiency and as a diagnostic procedure for remedial teaching. Thus, it provides reliable sources of information in the form of scores of the speaking components and can be used as feed-back for the teacher and learner to identify which component needs improvement.

Key words: speaking, assessment, analytic procedure

A language proficiency test of a productive level, such as a speaking test, has got little attention from language teachers and even from experts on language testing. It is not uncommon that speaking assessment for class-room purposes has been done without any theoretical basis. In the English Department of Universitas Negeri Malang, even though the teachers of speaking have a routine meeting to discuss the lesson plan including the materials and classroom activities, they do not pay a great attention to the evaluation procedure. So far, they use their own technique in assessing the learner's achievement by making a kind of field notes on the basis of their direct observation on the learner's performance in the classroom activities. The so-called on-going evaluation is quite practical and easy to do, as the

teacher does not need to make a formal speaking test with a suitable scoring guide. In this case he/she uses his/her feel judgment to score the learner's performance. By doing so, the teacher determines the scores on the basis of his/her general impression without necessarily making an analysis of the aspects or features of speaking ability.

This article is directed to show that using an analytic procedure in speaking assessment, in which the scoring procedure is based on the analysis of features of speaking ability, is more advantageous than using an impressionistic one (sometimes called a holistic approach) for two major reasons. First, an analytic scoring procedure gives the rater a clear description of the components of speaking ability being assessed. Second, it functions as an instrument to measure the learner's performance and as a diagnostic procedure for remedial teaching. It can be used as reliable information on the learner's level of proficiency from many different perspectives. Thus, this analytic procedure is a suitable approach for classroom purposes as part of the teaching and learning process.

To show the advantages of using an analytic procedure in speaking assessment, the following discussion covers a definition of a speaking test, types of speaking tests, and its scoring procedures. Finally, it describes the advantages of using an analytic scoring procedure, accompanied by a sample of a scoring guide and a scoring sheet for classroom assessment.

SPEAKING TEST

A speaking test is a procedure to measure speaking ability that requires the learner to speak, or to produce utterances and he/she is assessed on the basis of his/her utterances (Underhill, 1987). In an elementary level, a speaking test may be focused on assessing one or two aspects of speaking skills such as the pronunciation, intonation, and stress. In advanced level, on the other hand, the assessment is focused on the language function. In other words, it does not merely measure the language components, which have been mastered by the learner, but it covers all the communicative ability of the learner covering fluency, grammatical accuracy, vocabulary choice, and interactive communication. Harmer (1993) states that learning

a foreign language for communicative purposes requires the learner to master the pronunciation, grammar, vocabulary, discourse, and other relevant skills. Therefore, a functional speaking test should be one which measures the learner's speaking ability covering all the language components and his/ her interactive skills.

In general, a speaking test consists of two parts: (1) the speaking prompt (SP), or an instrument to elicit the learner's utterances and (2) the scoring sheet (SS) which contains scoring scales with descriptors or scoring criteria, accompanied by a scoring guide for the rater on which he/she bases his/ her judgment. A speaking prompt varies, and it might be in the form of a picture, a photograph, a reading passage, or a topic introduced by the examiner.

Unfortunately, so far as speaking assessment is concerned, not many experts in evaluation provide a description on how to develop a scoring guide or a scoring sheet in detail. Experts in language testing such as Oller (1979), Heaton (1975), Vallette (1977), and many others have discussed types and techniques of language testing without giving information on how to score the test. Ebel and Frisbie (1986) discuss statistical procedure to estimate the reliability of a test including an essay type one, but they do not provide information on how to make a scoring guide in assessing the productive level of learner's language proficiency especially in speaking. Discussions on the productive level of tests have been dominated by the measurements of written communicative competence, or writing tests with the approaches, and scoring procedures (Lloyd-Jones, 1977; Jacobs et.al, 1981; White, 1985; Spandel and Stiggins, 1990).

It has been observed that many teachers provide only a speaking prompt without any explicit scoring guide in their evaluation. They base their judgment on pre-determined implicit criteria without necessarily making more explicit ones that show what aspects of speaking ability are assessed. In other words, the speaking performance is assessed on the bases of the rater's general impression of the learner's performance. That is why this kind of procedure is called an impressionistic scoring procedure.

TYPES OF SPEAKING TESTS

A functional speaking test is basically a test to measure the learner's communicative competence. Such a test on productive level can be done in two ways in the forms of direct and indirect measurement. A direct measurement is done by asking the learner to speak, and he/she is assessed on the basis of his/her utterances. Thus, while the learner is speaking, the rater scores his/her performance. This is what is actually called a speaking test. In contrast to the direct test, the indirect test does not necessarily require the learner to speak but it requires him/her to make a written response on a given communicative event, such as completing a dialogue, or choosing the best response for a given utterance. This kind of test, called a paper-and pencil test (Brown, 1987) is not a good test of communicative competence, as it does not ask the learner to speak in a 'real communication'. In other words, the test does not measure the actual production of the learner's communicative competence. It is not suitable for assessing classroom achievement; but it is practical for a large-scale test involving hundreds of examinees.

Speaking tests vary according to the language elements being assessed and the objectives of the test. Based on the language elements, a language test in general, can be classified into discrete-point test, integrative test, and pragmatic test (Oiler, 1979). A discrete-point test is a test, which measures one of the language components, such as pronunciation, intonation, grammar, vocabulary. A discrete-point speaking test, then, is a speaking test which measures one of the language elements mastered by the learner, while an integrative test is a test which measures all the language components at a time. The last type, the pragmatic test, is an integrative test which can be considered as a functional speaking test because it measures the learner's ability in using the target language for communicative purposes in a given context.

According to the purpose of the evaluation, a speaking test can be classified into two, that is to measure the learner's ability in using the target language (functional speaking test), similar to Oiler's pragmatic test, and to measure one or more language components, or a discrete-point test. A

functional speaking test can be done in an interactive communication or transactive one. Brown and Yule (1983) proposed two classifications for language function: interactive function and transactive function. The first refers to the function of language to maintain social interaction while the second refers to a type of communication that is focused on conveying the message, not on the interaction. On the basis of the classification, an inter-active speaking test might be conducted in the form of interview or role play, while a transactive speaking test might be in the forms of story telling, giving a speech, reading an announcement, presenting a report, and many others. For classroom purposes, a speaking test should be made by determining the objectives and the types of test and the approach of the scoring procedure. Then, in reference to the objectives, the types, and approach, the scoring guide and scoring sheet are developed.

THE SCORING PROCEDURE OF SPEAKING TESTS

The scoring procedure of a discrete-point test and a functional speaking test is different. In a discrete-point test, the scoring procedure is simpler because it measures only one or two language components and the rater scores it by referring to the pre-determined criteria. A functional speaking test, on the contrary, is more complicated as it requires the rater to pay attention to the overall components of speaking skill at one time. Consequently, he/she has to concentrate not only on the language components being assessed but also on the criteria on the scoring sheet. Therefore, due to the nature of the complicated scoring procedure, it is not easy for the examiner to obtain a high reliability of this test.

The speaking test discussed further in this article refers to the functional speaking test that measures the learner's ability to use the target language for communicative purposes. Theoretically speaking, there are two approaches in the scoring procedures of a speaking test: the analytic approach and the impressionistic approach (Underhill, 1987). An analytic scoring procedure is a technique of scoring the learner's speaking ability by separating the components of speaking skill into sub skills, and the rater scores each component, and then sums the sub scores into a final score. In

contrast to the analytic approach, in the impressionistic approach the rater judges the learner's speaking ability on the basis of his/her general impression on the learner's performance without necessarily separating the speaking components. Thus, the rater directly comes to a single score without totaling the sub scores such that in the analytic approach. The term impressionistic approach is similar to the term holistic approach for writing assessment (Lloyd-Jones, 1977; White, 1985; Spandel and Stiggins, 1990). They state if the procedure of scoring is based on the analysis of features, it is called analytic, and if the scoring procedure is based on the judgment of the general impression as a whole without any separable features, it is called holistic.

THE ADVANTAGES OF USING AN ANALYTIC PROCEDURE

Even though the analytic procedure has some drawbacks in terms of practicality and reliability, it provides two advantages, especially for class-room purposes. The advantages lie on its ability to provide information on the features or components of speaking skills achieved by the learner. This information can be used as feedback for the learner to improve his/her performance, that is, which language component needs improvement. The second advantage is beneficial for the teacher as it functions as reliable information for remedial teaching. On the basis of the scores and the accompanying field notes or comments on each component, the teacher can further make a decision on which speaking skill needs more practice in the remedial teaching. This cannot be done in an impressionistic procedure because the learner cannot receive explicit feedback and the teacher is unable to identify the learner's weaknesses as the score does not tell anything about which aspects of speaking are achieved nor about the level of the learner's proficiency. Take for example, a very simple impressionistic scoring procedure taken from Underhill (1987:60). There are only three categories in the scoring scale as described below.

- 0: inappropriate or seriously incorrect
- 1: relevant but not entirely acceptable
- 2: appropriate and correct

This kind of scoring procedure is not informative in giving feedback either for the learner or for the teacher, as there is no detail description on the features of speaking skills being assessed. Therefore, this kind of scoring procedure cannot serve as a diagnostic procedure in the teaching and learning process.

A good example of analytic scoring procedure of speaking assessment is the one used by The Cambridge University as part of the First Certificate in English Examination. It is used to select foreign students who are going to study in England (Falvey, 1989). This test uses three kinds of speaking prompt (a picture, a reading passage, and a topic to discuss), and the speaking test is done in three steps. In the first step a picture is given to the examinee, and he/she is asked to describe the picture. The second step re-quires the examinee to read him/herself a passage for a few minutes then the examiner will give some questions. In the third step, two examinees are assigned to discuss a topic so that they can interact each other. There are six components to be assessed in this test: fluency, grammatical accuracy, pronunciation of sentences, and pronunciation of words and sounds, inter-active communication, and vocabulary resources.

For classroom purposes in the English Department at Universitas Negeri Malang, a type of analytic scoring procedure shows significant evidence in terms of both validity and reliability. This procedure describes more clearly and explicitly about what is going to be assessed in the speaking achievement test in that it reflects the teaching objectives of a speaking class. In other words, it meets the requirements of content validity as it measures the expected ability stated in the instructional objectives. The major concern about speaking test reliability is the rater's consistency since the scores come from his/her subjective judgment. Widiati (1998) proves a sufficient inter-rater reliability coefficient in an experimental study: 0.72 for the pre-test and 0.88 for the posttest. Thus, it is evident that the analytic procedure shows good reliable scores. The speaking test and the scoring procedure were made by a team of speaking teachers by referring to the Cambridge First Certificate with some modifications as shown in the following scoring guide.

THE SCORING GUIDE OF SPEAKING ASSESSMENT

1. Description

Aspects to be Evaluated	Description
1. Pronunciation	1. Pronunciation of individual sounds and words
2. Grammatical Accuracy	Accurate use of structure, or how the learner gets
3. Vocabulary	The learner's ability in choosing appropriate words and how to solve the problems
4. Fluency	1. The ability to keep the conversation going
5. Interactive Communication	The ability to get the meaning across the listener

2. Scale Criteria

Scale	Proficiency	Category	Description of Criteria
0	10%-39%	Very Poor	Pron: Many wrong pronunciations GA: No mastery of sentence construction Vo: Little knowledge of
	40%-50%	Poor	Pro: Frequent incorrect pronunciations GA: Major problems in structure

2.	60%-70%	Average	Pro: Occasional errors in pronunciation GA: Several Errors in structure Vo: Occasional errors in word choice
3.	75%-80%	Good	Pro: Some errors in pronunciation GA: Minor problems in structure
4.	85%-100%	Very Good	Pro: No errors/minor errors GA: Demonstrate mastery of structure (few errors) Vo: Effective/appropriate word choice

The scoring guide consists of two parts: (1) a description of the aspects of speaking ability to be assessed, and (2) the scale criteria based on which the rater makes a judgment. Before examination, all raters should understand the scoring guide. Training or coaching session is held where necessary especially for inexperienced raters so that they can come to a close agreement in making judgment. When everything is understood, the examination can be done and two raters are assigned using a separate scoring sheet. In this case, each learner gets two different scores from two different raters. The two sets of scores are correlated to find out the inter-rater reliability. A high reliability coefficient shows that the two raters come to a close agreement in making judgment. The following is an example of the scoring sheet.

SCORING SHEET

Aspects to be Evaluated	Scale										
1. Pronunciation	1	2	3	4	5	6	7	7.5	8	8.5	10
2. Grammatical	1	2	3	4	5	6	7	7.5	8	8.5	10
3. Vocabulary	1	2	3	4	5	6	7	7.5	8	8.5	10
4. Fluency	1	2	3	4	5	6	7	7.5	8	8.5	10
5. Interactive	1	2	3	4	5	6	7	7.5	8	8.5	10

Total Score: $1+2+3+4+5= \dots\dots\dots$ 5 Comments:

1. Pronunciation	
2. Grammatical Accuracy	
3. Vocabulary	
4. Fluency	
5. Interactive	

The comment space is used to make notes or comments on the learner's performance and it helps the rater to make judgment and to score the learner's speaking ability. The comments are useful for planning remedial teaching as it provides information on the weaknesses of the learner's speaking proficiency. In short, the scoring procedure serves two functions: as an instrument to measure the learner's speaking ability and as a diagnostic procedure for remedial teaching.

CONCLUSION

Based on the characteristics of the analytic scoring procedure, it can be concluded that this kind of procedure is more advantageous compared to the impressionistic one. As an instrument, it meets the content validity of classroom achievement and it proves to be a reliable scoring procedure. As a diagnostic procedure, it describes clearly the aspects or components of the learner's speaking ability. Therefore, it serves as feed-back both for the learner and the teacher. The accompanying comments help the teacher identify which aspect of speaking needs improvement. To conclude, the analytic procedure of speaking assessment provides a double function: as a valid and reliable instrument and as a diagnostic procedure for remedial teaching.

REFERENCES

- Brown, H.D. 1987. *Principles of Language Learning and Teaching*. New Jersey: Prentice Hall.
- Brown, G. and Yule, G. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.
- Ebel, R.L. and Frisbie, D.A. 1986. *Essentials of Educational Measurements*. London: Prentice Hall International.
- Falvey, P. 1989. *Oxford English Video: Success At First Certificate (The Interview)*. Cambridge: Moviemaker Film and Television Production.
- Harmer, J. 1993. *The Practice of English Language Teaching*. Singapore: Longman.
- Heaton, J.B. 1975. *Writing English Language Tests*. Singapore: Longman.
- Jacobs, H.L., Zinkgraf, S.A., Wormuth D.R., Hartfiel V.F., Hughey J.B. 1981. *Testing ESL Composition: A Practical Approach*. Rowley, Massachusetts: Newbury House Publishers.
- Lloyd-Jones, R. 1977. Primary Trait Scoring. In Cooper, Charles R and Lee Ode! (Eds). *Evaluating Writing: Describing, Measuring, Judging* (pp. 33-66). New York: The National Council of Teachers of English.
- Oiler, J.W., Jr. 1979. *Language Tests at School*. London: Longman Group Ltd.
- Spandel, V. and Stiggins, R.J. 1970. *Creating Writers, Linking Assessment and Writing Instruction*. San Francisco: Jossey-Bass.

- Underhill, N. 1987. *Testing Spoken Language*. Cambridge: Cambridge University Press.
- Valette, R.M. 1977. *Modern Language Testing*. New York: Harcour Brace Jovanovich.
- White, E.M. 1985. *Teaching and Assessing Writing*. San Fransico: Jossey-Bass.
- Widiati, U. 1998. *Meningkatkan Kemahiran Berbicara Mahasiswa Jun san Bahasa Inggris dengan Menggunakan Penutur Asli dari Kaset: Kontbinasi antara Listening dan Speaking*. Malang: Laporan Penelitian Tidak Dipublikasikan.