

**AN ANALYSIS OF MARKING SYSTEM
USED BY SPEAKING LECTURERS OF STAIN CURUP
IN TESTING STUDENTS' SPEAKING ABILITY**

Leffi Noviyenty, M. Pd.
STAIN Curup-Bengkulu
iffel_me@yahoo.co.id

Abstract

It is important for English Speaking lecturers to refer theories in scoring their students' speaking ability in order to increase the objectivity and minimize the subjectivity. The purpose of teaching speaking itself also needs to be considered. This study is a case study which investigates the classification of marking, scoring scheme used by speaking lecturers and the reason in selecting the classification. English speaking lecturers are the subjects of this research. Observation and interview are the techniques of collecting data by using a checklist and interview guidance. The findings show that English speaking lecturers of STAIN Curup has already guided theories of testing speaking in scoring their students' speaking ability. The classification of marking are fluency, grammatical accuracy, comprehension/content and pronunciation. Unfortunately the score for each element is still not clear. The main reasons the lecturers use in deciding the scoring scheme and classification are efficiency and effectivity in giving the test.

Keywords: Marking system, criteria of marking dan scoring scheme

INTRODUCTION

Communicative testing must be devoted not only to what the learner knows about the foreign language and about how to use it (competence) but also to what extent the learner is able to actually demonstrate this knowledge in a meaningful communicative situation.

Nunan states that the measurement of student performance is the key to program evaluation (Nunan, 1992). The researcher who

uses assessment data as the key element in a evaluation has to give careful consideration to three factors: these are: 1. The nature of the evidence to be used, 2. The relationship between evaluation and the program goals, and 3. The appropriate measurements to be used (Nunan, 1992, p.186). A test of discrete grammatical items constructed for this purpose might be found to correlate highly with an external criterion, for instance another established test concurrently administered or a measure taken at a later date, such as final academic grades.

Related to this argument, the researcher tries to describe the relationship among evaluation, measurement and test as in following diagram:

Evaluation can be defined as the systematic gathering of information for the purpose of making decisions (Bachman, 1990). Evaluation does not necessarily entail testing, while tests are often used for pedagogical purposes, either as a means of motivating students to study, or as reviewing material taught. Test may also be used for purely descriptive purposes only when the results of tests are used as a basis for making a decision that evaluation is involved. Test is a measurement instrument designed to elicit a specific sample of an individual's behavior.

Nowadays a goal of testing English skills is not only to the competence of English language that is the knowledge of language but also to the performance of those skills. This term is familiar with communicative competence which can be applied for all English skills, reading, speaking, writing, and listening. In related to this goal, it is important to carefully design test for testing English skills. There are variety of test formats offered by some English experts which is suitable for each skill, such as multiple choice, essay, short answer question for testing reading, role play for testing speaking, summary for testing writing and many others. The variety of test format need to introduce to the students in order to elicit their knowledge not only the competence but more to the use of the knowledge in communication.

Other aspect of communicative language testing is validity and reliability (Weir, 1993). Weir includes the point of validity and reliability as the general principles for test construction. To the extent that tests can have a beneficial influence on the teaching that precedes them, there can be a positive wash back effect from tests

on teaching. It is important therefore that test sample. As widely as possible relevant, criteria and communicative items from the syllabus or from the future target situation where this can be specified. The more representative the sample of tasks from their domain, the better the washback effect. The purpose of the test must be clear to all students taking it and teachers preparing candidates for it. The more it enhances the achievement of desirable language objectives the greater its contribution to successful teaching and the more all concerned will see the value of testing in the curriculum. If a test is unreliable, it cannot be valid. For a test to be valid, it must be reliable. However, just because a test is reliable does not mean it will be valid. Reliability is necessary but not sufficient condition for validity.

In STAIN Curup, the evaluation system is given to the lecturers independently. The institution only writes the marking guidelines and for the final achievement test. Lecturers' knowledge and understanding about how to design communicative language test is not yet evaluated and supervised. In other side, the goal of teaching the four basic English skills is to develop students' communicative competence. Moreover, the role of *Dosen payung* who act as senior lecturers is also not yet maximal since their credits are over limites and almost have no spare time to discuss the evaluation, particularly the marking system, for each English skill.

Moreover, as one of the four basic skills, testing speaking is likely to be more subjective test than the others. As a productive and spoken test, it also has some aspects to be tested as the marking criteria and needs spesific attention on scoring scheme. It is important to design a marking system that may represent the real ability of students. The students will bw able to evaluate themselves by recognising their weaknesses and developing their strengths. By offering some criteria of marking, the teacher would also be proffesional in helping the students to develop their communicative competence. Furthermore, the marking system which has detail supported criteria and clear scoring scheme colud become a valid source to describe the level of testeed, the students.

Based on some theories and facts above, the researcher is intended to investigate the marking system used by English lecturers of STAIN Curup in testing students' speaking ability.

There are many subjects in English *Tadris* Study Program and each subject perform test as the evaluation, including the marking system. Since each test has its own criteria of marking and scoring scheme, this research only discuss the marking system in Testing Students' Speaking Ability. There are so many aspects of analysis that could be researched related to English Tests, especially in testing speaking, this study only cover two aspects of analysis: the criteria of marking and the scoring scheme of Speaking Test. These two aspects are needed to be investigated as the first step of Test Analysis in English *Tadris* Study Program since they are the general principles and basic guidelines in constructing language test.

The problem of this research will be about the marking system use by speaking Lecturers of STAIN Curup in Testing Students' Speaking ability. The objectives of this research are to investigate:

1. The criteria of Marking used by speaking lecturers of STAIN Curup in testing students' speaking ability.
2. The Scoring Scheme used by speaking lecturers' of STAIN Curup in testing students' speaking ability.
3. The reason for speaking lecturers of STAIN Curup in designing the marking system to test students' speaking ability.

THEORITICAL FRAMEWORK

Testing Speaking

Speaking ability involves many aspects which can be analyzed into the elements of the speaking skills and the overall speaking proficiency (speaking for functional purposes). At the element level of speaking (primary level), the speaking might involve pronunciation, intonation, stres and other suprasegmental features. At this stage, the speaking also requires the correct use (structure), and the correct idiomatic use (vocabulary) of the target language. At the functional level, speaking involves the integration of the elements of the language and the function of using language either for transaction or for interaction. On the basis of its function language can be used for social relationship (interaction function) and for giving information (transactional function). In testing, the interactive speaking can be in the form of interview, role play, discussion and the like, while the trans-active speaking may take

the form of storytelling. Oral report, describing object/person/thing, addressing speech, and so on.

The two levels of assessment in speaking test cause problems in choosing criteria in assessing students' ability. The problems relate to the decision to determine the aspects to be looked for: Do the examiners focus on the elements of speaking skills or the overall speaking proficiency (speaking for functional purposes). The test designers, therefore, should determine the purpose of conducting tests, which can be derived from the objectives of language learning. From the purpose and objectives of the test, they can employ the appropriate types and approaches of testing procedures whether to employ discrete-point, integrative or pragmatic test. A discrete-point test refers to a test that attempts to assess a particular element of language at a time such as pronunciation, stress, intonation, structure, and vocabulary. An integrative test attempts to assess learners' ability to use many bits of their skill at a time. A pragmatic test refers to a procedure or task that requires learners' to process sequences of elements in a language that conforms to the normal contextual constraints of that language and to relate sequences of linguistic elements to extra linguistic contexts in a meaningful way.

In speaking test, is not always easy to get students to speak. Sometimes the tasks we expect to be capable of motivating students to speak do not work as expected. To overcome this situation, in addition to the careful design of the speaking tasks to fulfill students' level and to meet speaking aspects to be assessed, the examiner can function himself as partner in stimulating the students to speak.

In line to the opinion above, there are some other reasons why it is difficult to assess speaking ability, which makes the test be avoided in practice. Those reason are (1) oral testing is very time-consuming. The neglect of the implementation of speaking test in Indonesian educational context is due to this reason. The average class size in SMA/SMK/SMP is 40-45 students in a class and a teacher should teach parallel classes of 4-5. How long do the teachers have to spend to conduct the test? As a result, a paper-pencil communicative test -- an indirect way of testing communication—is used to replace the indirect way of testing oral proficiency/achievement; (2) it is difficult to get students to say

anything interesting; although, it does not mean to expect them to entertain the examiner with brilliant conversation or witty anecdote, but it, at least, fulfils one of such criteria as: (a) the students must have a chance to show that he can use the language for a variety of purposes (describing, narrating, apologizing, etc); (b) he must have a chance to show that he can take a part in spontaneous conversation, responding appropriately to what is said to him and making relevant contribution; and (c) he must have a chance to show that he can perform linguistically in a variety of situations, adopting different roles and talking about different topics. (3) The other reason relates to the issue of assessing. What sort of criteria can we use to assess students' performance? Is there any standard guideline to be used in setting up the criteria?

Criteria of Marking in Testing Speaking

It is possible to use one method as a check on the other. An example of this in oral testing is the American FSI (Foreign Service Institute) interview procedure, which requires two testers concerned in each interview both to assign candidates to a level holistically and to rate them on a six-point scale for each of the following: accent, grammar, vocabulary, fluency, comprehension. These ratings are then weighted and totaled. The resultant score is then looked up in a table which converts scores into the holistically described levels. The converted score should give the same level as the one to which the candidate was first assigned. If not, the testers will have reconsidered whether their first assignments were correct. The weighting and the conversion tables are based on research which revealed a very high level of agreement between holistic and analytic scoring.

The criteria offered to be considered in testing oral ability according to FSI are as in the following table (Hughes, 1989).

Table1. The Criteria Levels of Testing Oral Ability based on American FSI (Foreign Service Institute)

No	Criteria of Marking	Indicator	Score
1	Accent	Pronunciation unintelligible	frequently 1
		Frequent gross errors and a very heavy accent make understanding difficult, require frequent	2

		repetition.	
		“foreign accent” requires concentrated listening, and mispronunciations lead to occasional misunderstanding and apparent errors in grammar or vocabulary.	3
		Marked “foreign Accent” and occasional mispronunciations which do not interfere with understanding.	4
		No conspicuous mispronunciations, but would not be taken for a native speaker.	5
		Native pronunciation, with no trace of “foreign accent”	6
2	Grammar	Grammar almost entirely inaccurate phrases	1
		Constant errors showing control of very few major pattern and frequently preventing communication	2
		Frequent errors showing some major patterns uncontrolled and causing occasional irritation and misunderstanding	3
		Occasional errors showing imperfect control of some patterns but no weakness that causes misunderstanding	4
		Few errors, with no patterns of failure.	5
		No more than two errors during the interview.	6
3	Vocabulary	Vocabulary inadequate for even the simplest conversation.	1
		Vocabulary limited to basic personal and survival areas (time, food, transportation, etc).	2

		Choice of words sometimes inaccurate, limitations of vocabulary prevent discussion of some common professional and social topics.	3
		Professional vocabulary adequate to discuss special interest; general vocabulary permits discussion of any non-technical subject with some circumlocutions.	4
		Professional vocabulary broad and precise; general vocabulary adequate to cope with complex practical problems and varied social situations.	5
		Vocabulary apparently as accurate and extensive as that of an adequate native speaker.	6
4	Fluency	Speech is so halting and fragmentary that conversation is virtually impossible.	1
		Speech is very slow and uneven except for short or routine sentence.	2
		Speech is frequently hesitant and jerky; sentence may be left uncompleted.	3
		Speech is occasionally hesitant, with some unevenness caused by rephrasing and grouping for words.	4
		Speech is effortless and smooth, but perceptibly non-native in speech and evenness.	5
		Speech on all professional and general topics as effortless and smooth as a native speaker's	6
5	Comprehension	Understand to little for the simple type of conversation.	1

Understand only slow, very simple speech on common social and touristic topics; requires constant repetition and rephrasing	2
Understands careful somewhat simplified speech when engaged in a dialogue, but may require considerable repetition and rephrasing.	3
Understands quit well normal educated speech when engaged in a dialogue, but requires occasional repetition of rephrasing	4
Understand everything in normal educated conversation except for very colloquial or low-frequency items, or exceptionally rapid or slurred speech.	5
Understands everything in both normal and colloquial speech to be expected of an educated native speaker	6

Weir suggests that the considerations for deciding the criteria should also come from whether the assessment will cover routine skills or improvisational skills where each of them should establish different scoring scheme assessment for routine skills will consider: (i) normal time constraints, fluency, as overall smoothness of execution of the task, would be assessed; (ii) in addition one might want to comment on the discorsal coherence. That is the internal organization of the stages of the discourse. This may especially relevant in longer turns; (iii) Appropriateness: this would include the sociocultural ability to take into account setting, topic, role relationships, formality required. Due observance of the norms of interaction in terms of silence, proximity and dealing with encoding difficulties might be looked for. If the task leads to the deployment of improvisational skills then the lecturers might also

wish to develop criteria to take account of proficiency in the use of these.

In other side, assessing the improvisational skills might involve the lecturers taking a decision on overall effectiveness in two important improvisation abilities: (i) ability to negotiate meaning in cases of comprehension or production difficulties manifested on the part of the candidate or his/her interlocutor; (ii) ability to manage interaction (agenda and turn taking) actively and flexibly. This is particularly important where speakers can be expected to be active participants. If it improvisational skills the lecturers might make detailed assessment in terms of (i) fluency: smoothness of execution. Ability to negotiate meaning would, for example, include the ability to use communication strategies with case when in difficulties; (ii) appropriateness: this could include, for example, the degree of politeness and suitability of timing in turn taking or suitability of the language used in request for clarification or disagreement. In order to measure the quality of spoken performance, we first need to establish criteria of assessment. These criteria that might be considered for assessment of the output of communicative spoken interaction tasks. These criteria can be elaborated into the following table.

Table 2. Analytic marking Scheme for Speaking

No	Criteria Of Marking	Score
1	Appropriateness	0-3
2	Adequacy of vocabulary for purposes	0-3
3	Grammatical Accurately	0-3
4	Intelligibility	0-3
5	Fluency	0-3
6	Relevance and adequacy of content	0-3

The criteria in each of the three areas need empirical validation in the particular contexts testers find themselves in. first, tester would need to specify appropriate tasks in terms of conditions and operations and decisions could be taken iteratively on the criteria that are applicable to the output generated and the levels of performance within each of these. The dimension of practicality cannot be ignored here and the criteria developed would need to

be readily deployable by teachers. It would have to be established how many criteria teachers could reliably handle. The criteria develop would need to be accessible to other teachers and the number of levels within each criterion would have to represent real distinctions in the performance of actual candidates.

Criteria of marking could describe the level of ability. The intermediate level is characterized by the speaker's ability to:

1. Create with the language by combining and recombining learned elements, though primarily in a reactive mode;
2. Initiate, minimally sustain, and close in a simple way basic communicative tasks; and
3. Ask and answer question.

Global impression (B. J. carrol, 1980) marking scheme could be arranged into assessment scale which also describe the level of speakers as in the following:

RESEARCH METHODOLOGY

The research that is carried out is called descriptive research and will be presented in qualitative way. The researcher describes the existing facts that performed on how the Speaking lecturers design their marking system which includes the criteria of marking and the scoring scheme. The analysis covers two aspects, the criteria of marking the test and the scoring scheme. Gay and Airasian (2000) stated that descriptive research which is also called survey research determines and describes the way things are. Moreover, Gay explains that descriptive research, also called survey research, is useful for investigating a variety of educational problems and issues. Mainly the interviews and observation (analysis), as in qualitative research are used as the techniques of collecting the data. This research is guided by some steps which are conscientiously executed: identify the topic or problem, select an appropriate sample of participants, collect valid and reliable data, and analyze and report conclusions.

There are three mean data taken: the Marking Sheet for Speaking Test, consists of classification of marking for Ujian Akhir semester (UAS) test for Speaking; the Speaking Lecturers who teach Speaking during the active semester, and the syllabus for speaking. The researcher also uses checklist, deep interview, field

notes, and tape recorder. The field notes are made during the interviews (analysis) in order to provide the description and understanding of each indicator performed and the lecturers' intention. Data for this research is collected in term of longitudinal. Gay states that longitudinal research collects the data at more than one time in order to measure growth or change. In detail, the data is collected through document analysis and interview.

The researcher analyzes the Marking Sheet for Speaking Test in order to search the criteria of making used and also the scoring scheme. This analysis uses theories as the guidance and will be written down on checklist. Interview is done between the researcher and the speaking lecturers. This technique is supported by Bogdan and Biklen (1982) defined an interview as a purposeful conversation, usually between two people (but sometimes involving more) that is directed by one in order to get information. The researcher asks about the lecturers' knowledge on designing Speaking Test, particularly about the criteria of marking, the scoring scheme and also the reason the lectures consider in designing the marking system to test students' speaking ability. The interview guidance is used to manage the data needed.

The instruments of this research are **(1) Checklist**, the indicator for each formal will be build based on two theories that complete each other, they are **Cyril J Weir in his book Communicative Language Testing and Arthur Hughes in his book Testing for Language Teachers**. Each criterion has a specific scoring scheme. The scoring scheme is investigated by using theories and a deep interview and also analyzing the speaking lecturers' marking sheet that consists of classification of scoring. The finding will be compared to the theories being guided; **(2) interview**, the purpose of interview is to investigate the speaking lecturers' knowledge and intention to and about marking system in testing students' speaking ability; **(3) Field notes**, the criteria and scoring scheme of speaking test which have been checked from the checklist and interview is elaborated by also including additional information taken from the field notes in order to explain the marking system and the reason of speaking lecturer in designing the system. To ensure the validity of the research, the following strategies are applied: Triangulation Data. The researcher uses more than one technique of collecting data as a

comparison, member checking, the participants are served as check throughout the analysis process and Long terms and repeated analysis.

FINDINGS

Based on the result written on the checklist, it is found that three Speaking lecturers do not use a specific theory in deciding the criteria of marking in giving their speaking test to the students; however, all them include some point of marking taken in general related to speaking skill such as in the following table:

Table 3. The criteria of Marking used by Speaking lecturers in testing students' speaking ability

No	Criteria	Speaking Lecturers	Sources
1	Accent	1	Based on FSI (Foreign Service Institute)
2	Grammar	3	
3	Vocabulary	2	
4	Fluency	2	
5	Comprehension	3	
	Grammatical Accuracy	3	Based on TEEP, CALS
	Intelligibility	3	
3	Fluency	3	
4	Relevance and Adequacy of Content	3	
1	Fluency and Coherence	3	IELTS
2	Lexical resource		
3	Grammatical Range and Accuracy	3	
4	Pronunciation	3	

From interviews, it is found that the criteria of marking used by Speaking lecturers are not decided based on a theory (FSI,

TEEP), but they tend to select the criteria based on their need in evaluating their students. Moreover, in a basic guidance based on the lecturers understanding, the criteria of marking system in testing Speaking consist of fluency, accuracy, grammar accuracy. Fortunately, some theories above include those criteria. The researcher also list definition of each criterion used by speaking lecturers as in the following:

1. Accent:
whether the students speak English highly influenced by their first language or dialect or they can suit to native speaker accent.
2. Grammar/Grammatical Accuracy
whether the students produce grammatical mistakes and errors such as tenses, from of sentences (statement, question, direct and indirect and many others)
3. Fluency:
whether the students speak English fluently or they produce many pauses and use much time to think what they want to say.
4. Comprehension
Whether the students English can be understood by others, speak suitable to the topic given, relevant vocabulary.

It is clear enough that the need of students in following the Speaking class is highly considered as the criteria of marking. Otherwise, these criteria are not specified enough in the scoring scheme. From the interview with the three speaking lecturers in order to investigate the marking scheme they use in testing students' speaking ability, it is found that they do not include the score for each criterion of marking suggested by the theories. The lecturers' marking scheme for speaking subject is based on the marking scale given by STAIN as the following:

1. 00 – 49 = E (failed)
2. 50 – 59 = D
3. 60 – 69 = C
4. 70 – 85 = B
5. 86 – 100 = A

Meanwhile, the lecturers do not design specific band or scale as the indicator of 00 – 100 that characterized speaking testing. Weir states clearly (Weir, 1993):

in oral testing, as in the assessment of written production, there is a need for explicit, comprehensive marking scheme, close moderation of tests tasks and marking scheme, and training and standardization of markers. In order to measure the quality of spoken performance, we first need to establish criteria of assessment.

Weir explains that without criteria of marking, it is too subjective for the markers (lecturers) decide the score, and the quality of testing speaking performance could be not valid. The dimension of practicality cannot be ignored and the criteria developed would need to be readily deployable by lecturers. Criteria of marking could describe the level of ability. While, scoring will be valid and reliable only if, clearly recognizable and appropriate description of criteria levels are written and scores are trained to use them.

The fact that particular grammatical structures are not specified as content, and there is no reference to vocabulary or pronunciation, does not of course mean that there are no requirements with respect to these elements of oral performance. The accurate measurement of oral ability is not easy. It takes considerable time and effort to obtain valid and reliable result. Nevertheless, there backwash is an important consideration, the investment of such time and effort may be considered necessary. Speaking is probably the most difficult skill to test. it involves a combination of skills that may have no correlation with each other, and which do not lend themselves well to objective testing. There are not yet good answers to questions about the criteria for testing these skills and the weighting of these factors. a speaker can produce all the right sounds but not make any sense, or have great difficulties with phonology and grammar and yet be able to get the message across. Comprehension of spoken material depends, among other factors, on the degree to which the listener is familiar with the speaker's accent and the degree to which they share background knowledge, and so what is a problem for one listener may not be a problem for another listener. Testing speaking is also

a particular problem when it is necessary to test large numbers of students. It is necessary to test hundreds of students, and even if each student speaks for only a few minutes, this becomes a huge job.

One of the great difficulties in testing speaking is, of course, the assessment. It is necessary to develop a system of assessment that can be applied as objectively in assessment. The scale can be one general scale for overall speaking ability, or it can be divided between several aspects of the skill of speaking, such as pronunciation, grammar, organization, etc. The scale also depends on the speaking task that is used for the test. A test that uses public speaking as the task would be different from one that uses a group discussion. If possible, the speaking task should be recorded and the scoring done from the tape. In addition, the marking should be done by more than one person and their reliability checked. If the task is an interview, the interviewer should not be required to score the test at the same time as conducting the interview, if this is avoidable. Among the aspects of speaking that might be considered in the assessment scale are grammar, pronunciation, fluency, content, organization, content and vocabulary. The band descriptions for a general scale might be as follows. The number indicates the level, and it is followed by a description of the characteristics of a speaker at that level. In the classroom, during daily exercise, 2 lecturers sometimes use the following scoring scheme:

Table 4. Scoring Scheme

7	Spoken communication is fluent, appropriate, and grammatically correct, with few if any errors.
6	Communication is generally fluent and grammatically correct with only occasional errors in grammar or pronunciation.
5	Students produces numerous grammatical errors and hesitations, but these do not interfere greatly with communication. Utterances are long and connected.
4	Students produces numerous grammatical errors and hesitations, and these occasionally interfere with communication. Utterances are short and connected.
3	Student's communication is limited to short utterances and depends in part on previously memorized

conversational elements. Difficulty dealing with unpredictable elements. Many hesitations and grammatical errors. Communication only possible with sympathetic interlocuter.

2 Communication limited to short utterances, almost entirely memorized conversational elements. Unable to deal with unpredictable elements.

1 No communication possible.

Though speaking is a particularly difficult skill to assess, there are methods that can be employed to create situations that elicit speech and methods of assessing the testees' speech that are reasonably reliable. Testing speech is important for its backwash effect, even if the method of testing and of assessment are not as perfect as they might be.

The scoring scheme used by speaking lecturers seems not yet measure the quality of students' speaking ability since the score obtained comes from several clarification such as: students' participation, assignments, mid term and final examination. Each of those classifications does not define clear score related to speaking ability. For example, A until E does not explain the quality of students' speaking performance. IELTS, the International language Testing System, which is designed to assess the language ability of candidates who need to study or work where English is used as the language of communication for universities and em[ployers in many countries is one of trusted standars of English Test explains that the score should represents the level of candidates' speaking performance. IELTS includes some criteria of marking system to test speaking performance; they are fluency and coherence, lexical resource, grammatical ranmge and accuracy and pronunciation. IELTS also define the band for each criterion clearly and the scale of score such as in the following:

Table 5. IELTS Scoring Scheme for Speaking Performance

No	Scoring Scheme		Level of Speaking performance
	IELTS	PTE for Academic	
1	8 – 9	85 +	Level 5
2	6,5 – 7,5	76 – 84	Level 4
3	5 – 6	59 – 75	Level 3

4	4 - 4,5	43 - 58	Level 2
		30 - 42	Level 1

Based on some facts above, in summary, the finding of this research shows that the marking criteria of testing students' speaking ability used by speaking lecturers of STAIN Curup is designed by taking the point of speaking skill itself rather than guide theories of marking criteria for speaking performance. However, those criteria included in some theories. Each criterion used by the lecturers also does not define the score. Moreover, the scoring scheme is not clear and does not directly relate to the quality of students' speaking performance. Some reasons about practicality become crucial considerations for the lecturers in designing the marking criteria and scoring scheme.

CONCLUSION

Speaking lecturers of STAIN Curup has already decided some criteria of marking system to test the students speaking ability, they are: Fluency, Grammatical accuracy, Comprehension/content, and pronunciation. These criteria are suggested by some theories of testing speaking performance. The scoring scheme use by speaking lectures of STAIN Curup is not clear enough to measure the students' speaking ability since there are some other considerations, such as practicalities aspects and STAIN classification of scoring. It is important to guide theories in deciding the criteria of marking and the scoring scheme to test speaking performance. Each criterion of marking should be defined as possible in order to draw a suitable score as the measurement of speaking performance and to minimize the markers' subjectivity. Eventhough the classification of scoring is provided by STAIN, the speaking lecturers should still consider the purpose and the importance of teaching speaking. The speaking ability could be evaluated first before they come to the final score for speaking subject. This suggestion could bring the valid data the real condition or ability of students' speaking performance.

SHORT BIOGRAPHY

The writer is An English Lecturer in STAIN Curup, graduated her S1 from Bengkulu University in 1999 and her Master Degree of English Language Education from Padang university in 2006.

REFERENCES

- (2000). *Principles of Language Learning and Teaching*. New York: Wesley Longman, Inch.
- Ary, D. Jacobs, L.C, dan Razavieh, A. (1982). *Pengantar Penelitian Pendidikan*. Terjemahan oleh Arief Fuchan. Surabaya: Usaha Nasional.
- Ary, Donald.(1985). *Introduction to Research Education*. New York: CBS College Publishing
- Bachman F, Lyle. (1990). *Fundamental Considerations in Language Testing*.. New York: Oxford University Press.
- Bloomfield, Leonard. (1995).*Language*. Jakarta:PT Gramedia
- Bogdan, R & Biklen, S. K. (1982). *Qualitative Research for Education: An Introduction to Theory and Method*. Needham Heights: Allyn & Bacon,.
- Brown, D. H. (1987).*Principles of Language Learning and Teaching*. 2 edition. Englewood Cliffs, N.J: Prentice Hall, Inc.
- Brown, Gillian and George Yule. (1982). *Teaching the spoken Language: An Approach based on the analysis of confersational English*. Cambridge: Cambridge University Press.
- Gay, L.R.and Peter Airasian. (2000). *Educational Resarch: Competencies for Analysis and Application*. New Jersey: Prentice Hall,Inc.
- Heaton, JB. (1990). *Writing English Language Test*.. USA: Longman
- Hughes, Arthur. (1989). *Testing for language Teachers*. UK: Cambridge University press..
- Janice, C (Ed). No Year. *Communcative Competence for Individuals who Use Augmentative and Alternative Communication (AAC): from Research to Effective Prectice*, (online).
- Nunan, David. (1992).*Task-based Language Teaching* New York: Cambridge University.

Weir, Cyril J. (1993). *Understanding and Developing Language Tests*.
UK: Prentice Hall International.

Weir, Cyril J. (1998). *Communicative Language Testing*. UK:
Prentice Hall.