TEACHER-MADE SPEAKING ACHIEVEMENT TEST

Khairuddin[1]
[1]Department of Applied Science of Health,
Politeknik Kesehatan Kemenkes Malang
Jl. Besar Ijen No.77C, Oro-Oro Dowo, Klojen, Kota Malang, Jawa Timur 65119
Email: fadilkhairuddin@gmail.com

*Abstract***:** *One of the main jobs of an English teacher teaching speaking skill is to develop and administer a speaking test in order to measure how much or little the students have achieved the instructional objectives and materials upon completion of a language program. Through the test results, the teacher is enabled to make decisions such as deciding the effectiveness of the language program and passing or advancing the students to the next level of the program.To qualify herself in developing a valid and realiable speaking test, both conceptual and practical knowledges are needed. The article presents the needed knowledge about speaking test covering the concepts of speaking skill construct, criterion-referenced test, analytic approach, criteria setup, and scoring and grading. The article also reports the implementation of the teaher-made speaking test to testing the speaking skills achievement of the students in Politeknik Kesehatan Kemenkes Malang.*

Key words: Speaking test, Achievement test, Criterion-referenced test, Analytic approach, Scoring, Grading

INTRODUCTION

Normally, an English teaching teaching speaking skill is required to test the students' speaking achievement upon the course completion. Typically, in the test, the students are given a task that requires them to use the language communicatively so the examiner-teacher can get samples of their speaking performance to be scored. To pass judgment on the students' speaking performance, subjective scoring method should be in place, where correctness or acceptability of answers is a matter of degrees. Two approaches to subjective scoring have been known to commonly take place: holistic and analytic approaches. Analytic scoring of a speaking test which separates speaking skill into its features or sub skills to be scored separately is more appropriate for classroom purpose than holistic approach where scoring is just based on the rater's overall impression. The current article reports the development and

implementation of a speaking test at Medical Record and Health Information department of Politeknik Kesehatan Kemenkes Malang (Malang State Health Polytechnic), which covers the issues of criterion-referenced test (CRT), speaking-skill construct, analytic scoring , setting-up of the criteria and descriptors, and grading guide.

Criterion-referenced testing

CRT is an important household tool in the big family of teaching profession. This instrument functions as a test which measures a student's performance according to a particular standard or criterion which has been agreed upon even before classroom instruction is started (Richards, Platt, and Weber, 1985; Cohen, 1994; Djiwandono, 2008), which is the objective of the instruction. CRT is produced to measure well-defined and fairly specific instructional objectives (J.D. Brown 2005). Often these objectives are specific to a particular course, program, school district, or state. An example of a very strict instructional objective would be the following: *"By the end of the course the students will be able to underline the sentence containing the main idea of an academic paragraph of 200-250 words at the eleventh grade readability level*

with 60 per cent accuracy." However, objectives come in many forms. Other objectives might be defined in terms of tasks we would expect the students to be able to perform by the end of the term, or experiences we would expect them to go through. For example, *"By the end of the term students will be able to watch at least five English language movies with no subtitles."*

Douglass Brown (2004) suggests that criterion-referenced tests be also designed to give test-takers feedback, usually in the form of grades, on specific course or lesson objectives. Classroom tests involve the students in one class, and is connected to a curriculum, hence the result of the tests are expected to be useful for the pursuit of teaching effectiveness in the class and the curriculum repair efforts, or what Oller (1979, in D.H. Brown 2004) called "instructional value." In a criterion-referenced test, the distribution of students' scores across a continuum may be of little concern as long as the instrument assesses the objectives. From the results of CRT, several decision-makings like classroom-level achievement decisions and classroom level-diagnostic decisions can be based on.

In terms of interpreting the test scores, the interpretation of scores on a CRT is

considered absolute. Each student's score is meaningful without reference to the other student's scores, as in norm-referenced testing. In other words, a student's score in a particular objective indicates the percentage of the knowledge or skill in that objective that the student has learned. Moreover, the distribution of scores on a CRT need not necessarily be normal. If all students reach 100% of the objectives, then they all should receive the same score with no variation at all. Therefore, on a CRT final examination, students who have learned all the course material should all be able to score 100 per cent on the final examination. Thus, very homogeneous scores can occur on a CRT. In other words, very similar scores among students on a CRT may be perfectly logical, acceptable, and even desirable if the test is administered at the end of a course. In this situation, a normal distribution of scores may not appear. In fact, a normal distribution on CRT scores may even be a sign that something is wrong with the test, with the curriculum, or with the teaching (J.D. Brown 2005).

Speaking and speaking test

To speak is to express ideas orally, to make a person's mind or thought known to and understood by others as the addressees.

To make his oral expression known and understood well a speaker need to attend to factors bringing to oral message delivery success: (1) definite message, problem or topic to be delivered, (2) well organized message delivery, (3) clarity of message which can be established through implementing the right words choice and use, accurate grammatical rules, and intelligible pronunciation and appropriate fluency (Djiwandono 2008). In addition, Harmer (1993) posits that speaking ability should cover interactive skill. Those components of speaking skill are decisive to successful communication, hence, should be taken into consideration when developing a speaking ability test.

Speaking test approach

Two approaches to productive-skills testing, analytic approach and holistic approach, are commonly used in testing practices for educational purposes. They are used for setting up criteria for measuring up the test-takers' performance in productive language skills. Holistic approach uses a single general scale to give a single global rating for each test-takers language production (Djiwandono, 2008). In this approach, the rater judges the learner, say, speaking ability on the basis of his/her

overall impression on the learner's oral performance without separating the speaking ability into its features. In the procedure of holistic scoring approach, the rater directly comes to a single score which measures the aspects of speaking ability like content, fluency, pronunciation, vocabulary, grammar, etc., as one whole for concluding the test-takers oral performance. A high expertness of rating skill is truly required to be able to apply this scoring approach responsibly and rightly for speaking ability judgment (Djiwandono, 2008), otherwise validity and reliability of the scoring results are at risk here. Although holistic approach can be applied more easily and practically for scoring criteria, it is not easy to get explicit and specific feedbacks that can tell information about the areas of student's strengths and weaknesses in their speaking achievement.

Analytic approach, on other hand, refers to a procedure of scoring the learner's speaking ability by separating the features of speaking skill into sub skills. In this procedure, the rater scores each feature and then sums up the sub scores into a final score (Underhill 1987). Perhaps some people would think that putting or designing the skill into more separated analytic aspects or components would be hard to implement,

and the scoring would be too complicated to do for the rater. However, with a more discretely separated feature of the speaking skill would allow the rater to do the scoring in a more confident way? The rater would not experience problematic decision making as happening when a single score should be assigned or given to assess or evaluate two or three components or sub skills which are treated and scored as a single unit. By doing so, the scoring can be done more objectively, hence, with higher validity and reliability estimate.

Scoring and weighting

In speaking test, a judgment is called for in the part of the scorer, thus the scoring is said to be subjective. In second and foreign language teaching, subjective marking is usually required for scoring writing and speaking tests (Henning 1987; Underhill, 1987; Hughes, 2003). In subjective scoring, "examiners are required to make judgments which are more complicated than the right or wrong decision…their job is to assess how well a candidate completes a given tasks" (Alderson, Clapham, & Wall, 2002:p. 107).

Related to scoring is weighting. Weighting refers to the values that are placed on certain test items within the test.

Test items may be weighted differently. One item for example is weighted higher as compared to other items based on the consideration that the mastery of the objective tested with the first item reflects more learning than mastery of another objective measured with the latter test item (Cohen 1990).

Speaking is basically an activity of transferring or conveying messages/ideas between persons involved in a communication. This ability of delivering understandable message content and understanding delivered message content represent the communicative competence of the communicator. To be able to deliver an understandable content of a message and to understand an addressed content of a message a person needs to make use of not only his linguistic competence, but also his discourse, sociolinguistic, and strategic competence, which together integrated they make up a person's whole communicative competence (Savignon 1983). By understanding this nature of communication, consequently, content factor should be the main consideration when judging or measuring a person's speaking ability. Thus, in developing a speaking test, content should be considered the most important criteria of speaking ability, which therefore be given

the most weighting than the other components of speaking ability.

Grammatical or linguistic competence is the smallest construct of communicative competence as opposed to the other constructs of communicative competence, i.e., discourse competence, sociolinguistic competence and strategic competence (Savignon, 1983:35-42). Agreeing with this theoretical stance, grammar aspect should be considered less important criteria of measuring speaking ability. Thus, grammar criterion should be given lesser degree of weighting in a speaking test development.

Pronunciation is phonological knowledge (Canale & Swainin Brown 2002). In an analytic scoring approach, pronunciation may be isolated and scored analytically or separately from the grammar criteria. By treating it so, the examiner will be able to get practical benefit for measuring fairly and confidently the two components separately. Pronunciation may be given similar weighting as grammar criteria because they both represent linguistic competence.

Test development

State Health Polytechnic of Malang is a vocational higher education institution specifying in health education. Medical

Record and Health Information (MRHI) is one study program in the institution which prepares the students to be effective English users. Thus, two series of English subjects of English 1 and English 2 are provided in the first and second semester. In both classes the students are taught and trained on speaking skills. The class used for this project is English 1.

Since the test to be developed is aimed at testing speaking achievement of MRHI students, the MRHI curriculum is to be considered. The general objective of English 1 related to speaking skill is to make the students competent in talking about issues related to health in general, such as "health and the environment,""food and health,""the importance and development of medical record," "mental health," "health and lifestyles," "diseases and illnesses," etc.

Test topics

When a test is meant for measuring speaking ability, supposedly the speaking test measures the students' ability in speaking about the topics which they once discussed to speak about during instructional activity. Therefore, four topics are selected from the aforementioned list of activity topics. But yet, the students should not know they are going to talk about the topics they

once used during class activities. To add to the degree of spontaneity of their speaking performance, the students may be asked spontaneous questions prepared by the examiners about the chosen topic. Thus, the four topics selected for the topics of the test include: (1) The effect of environment to Health; (2) The Habit of Cigarette Smoking; (3) Drugs Abuse; (4) The Importance and Development of Medical record; The four topics are selected because they are assumed as the topics with relatively equal level of difficulty.

Test tasks

The test tasks to be selected in the speaking test should be those relevant with the students' ability level and those which students are already familiar with. The tasks used to obtain information or data about the students' speaking performance in the current speaking test model are recount and question-and-answer.

To recount is to tell somebody about something (Oxford Dictionary 2000). Recount is used as a test task in the speaking test because the activities of telling and sharing story, experience, opinion, knowledge, etc., are used as the main instructional activities during learning process. Thus, in the recount session of the

speaking test, the students tell about their story, knowledge, experience, opinion, examples, etc., about the topic they have chosen. The students do individually the recount test task face to face with the tester.

Another task that can be used in combination with recount is question-and-answer task. After telling about the given topic, the students should answer questions from the tester about the topics they have told. Question-and-answer task is implemented because the task is believed to be effective for measuring spontaneous speaking ability. This task is also used since question and answer is a typical activity during group discussion which is considered main activity in the learning process in English 1 Class for MRHI students. The tester should have supplies of questions available for each of the four topics to be asked to the students. The questions developed are wh-questions that supposedly require elaborate explanations for the students to answer.

Set-up of criteria and indicators

When the speaking construct is classified into discrete components of content relevance, content completeness, grammar and pronunciation, these components become the criteria for referencing the students' speaking performance in the speaking test. Then, the test developer gives descriptions or indications each criterion covers. The following table shows the component criteria and their each criteria descriptors or indicators.

Table 1. *Criteria and criteria description of speaking test*

COMPONENT CRITERIA	DESCRIPTION/INDICATION
RELEVANCE OF CONTENT	The speech content is relevant with the topic to be spoken about. The speech content of the answer is relevant with the question asked.
COMPLETENESS OF CONTENT	The speech content is supported with good knowledge, details, examples, facts, and other supporting arguments relevant with the topic to be spoken about and the question to be answered.
GRAMMAR ACCURACY	The oral language use applies accurate grammar that can deliver the clearest and the most understandable meaning of the content.
PRONUNCIATION (Intelligibility/fluency)	The pronunciation is intelligible. The speed speech or fluency is natural.

Each component is then graded into degrees or levels of mastery in a range of "very good," "good," "adequate" or "fair", "inadequate" and "poor", where very good category is given score 5, good 4, adequate 3, bad 2, and poor 1. Thus, if the student could perform best in the tasks when the performance are referred the three components of content, grammar, and pronunciation, then he/she would be

assigned the highest mark or score 5 for each component criteria. And, if the student perform so poorly in the speaking test when referred to the three components, then he/she would get the lowest score 1 for each component. The setting-up of descriptors or indicators for each level of mastery in each component criteria and the marking or assigning of marks to each component mastery level is shown in Table 3.

METHOD

To begin the test, students are called one-by-one to do the speaking test face-to-face with the tester. After a short greeting "good morning" and "how are you" to make the student feel comfortable or to ease anxiety in the test session, the student called is asked to choose randomly one of the four faced-down cards containing one topic each (Appendix 1) that they are going to recount or tell about to the tester. After choosing the card, the student is asked to read loudly the topic stated in the card, for example "The Importance of Medical record" and also the short instructions: "tell as much as possible for 2.5 minutes about the topic". Then the examiners welcome the student to begin telling about the topic "Now please tell about your opinions about the importance of medical records" and inform him that "The

examiner would give notification when the time allotted for recount is up." When the students are doing the recount the examiner should be listening attentively and not doing anything that could interrupt or annoyed the students doing the task. Then the test is continued for the next 2.5 minutes with question-answer session. The tester could use the list of ready-to-use questions already made available. The examiners should ask the questions to the students in well-articulated manners.

Table 2. *The list of questions to be used during question-answer task*

a. The habit of cigarette smoking
• Who should be blamed when the number of smokers gets higher?
• Why do you think many school-aged boys and girls begin to smoke?
b. The habit of cigarette smoking
• Who should be blamed when the number of smokers gets higher?
• Why do you think many school-aged boys and girls begin to smoke?
• What (and why) is your attitude toward a smoker smoking in public places like in canteen & bus?
• What (and why) do you say to smokers who think they can't stop smoking ever?
• What (and why) would you say to passive smokers?
c. Drugs abuse
• Why do you think people start using drugs in the first place?
• What (and why) would you do if you see a drug transaction before your eyes?
• Who (and why) should be blamed on the widespread abuse of drugs among young people?
• Who (and why) should be most responsible for drugs abuse combat?

What would you say to drug addict and why?
- What (and why) would you suggest the government do concerning drug addict?
- What (and why) would you say if a friend of yours is trapped in drugs abuse?

d. The importance and development of medical record
- Can you describe the job of a medical recorder?
- How have medical records documentation developed?
- What are the pros and cons of electronic health and medical records?

RESULT AND DISCUSSION

The tester rates the students' speaking performance trough analytic approach by using and referring to the score sheet he/she should have prepared. The score sheet contains the cells for each of the components scored: content, grammar, and pronunciation. Each component criteria is divided into five mastery levels. Each level of mastery of each component is described with defined indicators/descriptors. The tester check"√"in the relevant cell under each descriptor's cell to measure each student's speaking performance in each component.

The score sheet can be seen in the following table:

Table 3. *The score sheet for speaking test*

Component	Marks/Descriptors				
	5	4	3	2	1

CONTENT RELEVANCE	Highly relevant with the topic	Finely relevant with the topic	Adequately relevant with the topic	Inadequately relevant with the topic	Poorly relevant with the topic
	√				
CONTENT COMPLETENESS	Very well supported with good knowledge, details, example, facts, other sup-porting argument	Well supported with knowledge, details, examples, facts, and other supporting argument	Adequately supported with knowledge, detail, example, fact, and other supporting argument	Inadequately supported with knowledge, details, examples, facts, other supporting argument	Poorly supported with knowledge, detail, example, fact, and other supporting argument
		√			
GRAMMAR ACCURACY	Highly appropriate grammar use, and error-free	Very few grammatical error but meaning not harmed	Some errors in grammar use which somewhat disturb meaning	Frequent grammar errors, many times meaning is difficult to understand	Almost every sentence contains error, meaning is hardly understood
	√				

PRONUNCIATION (Intelligibility and Fluency)	Perfectly intelligible, very natural speech spee	Satisfactorily intelligible, natural speech speed, very few mispronunciat	Slightly slow speed, mispronunciation is occasional causing occasi	Very slow speech speed, so many hesitation, frequent mispr	speech speed too slow and full of incompetent
		√			

Each component of content, grammar and pronunciation is graded in a range of "very good," "good," "adequate," "inadequate," and "poor," with the scores of 5, 4, 3, 2, and 1, respectively. Thus, if the student could perform best in the tasks for the three component criteria, then he/she would be assigned the highest mark or score 5 for each component criteria. And, if the student perform so poorly in the speaking test when referred to the three components, then he/she would get the lowest score 1 for each component. The raw scores or marks each student gets for each component criteria is multiplied with the weighting percentage each component bears: content relevance 30%, content completeness 30%, grammar 20%, and pronunciation 20%. Consequently, their marks on content relevance criteria and content completeness criteria are multiplied with 3, while the marks they get in grammar and pronunciation criteria are multiplied with 2,

to get the students' scores for each component criteria. Each student's scores in each component criteria is added up to generate the student's total score and thus the final grade. The marking, weighting, and scoring is shown by the following Table 4.

Table 4. *The indicators, marking, weighting and scoring system*

COMPONENTS	INDICATORS/DESCRIPTORS	GRADE	MARK	WEIGHT	SCORE
RE-LEVANCE OF CONTENT	Highly relevant with the topic	Very Good	5	X3	15
	Finely relevant with the topic	Good	4	X3	12
	Adequately relevant with the topic	Adequate	3	X3	9
	Inadequately relevant with the topic	Inadequate	2	X3	6
	Poorly relevant with the topic	Poor	1	X3	3
COMPLETENESS OF CONTENT	Very well supported with good knowledge, details, example, facts, other sup- porting argument	Very Good	5	X3	15
	Well supported with knowledge, details, examples, facts, and other supporting argument	Good	4	X3	12
	Adequately supported with knowledge, detail,	Adequate	3	X3	9

Aspect	Description	Rating	Score	Weight	Total
	example, fact, and other supporting argument				
	Inadequately supported with knowledge, details, examples, facts, other supporting argument	Inadequate	2	X3	6
	Poorly supported with knowledge, detail, example, fact, and other supporting argument	Poor	1	X3	3
GRAMMAR ACCURACY	Highly appropriate grammar use, and error-free	Very Good	5	X2	10
	Very few grammatical error but meaning not harmed	Good	4	X2	8
	Some errors in grammar use which somewhat disturb meaning	Adequate	3	X2	6
	Frequent grammar errors, many times meaning is difficult to understand	Inadequate	2	X2	4
	Almost every sentence contains error, meaning is hardly understood	Poor	1	X2	2
PRONUNCIATION (Intelligibility/ fluency)	Perfectly intelligible, very natural speech speed, free of mispronunciation	Very Good	5	X2	10
	Slightly slow	Good	4	X2	8
	s peed, mispronunciation is occasional causing occasional unintelligibility				
	Slightly slow speed, mispronunciation is occasional causing occasional unintelligibility	Adequate	3	X2	6
	Very slow speech speed, so many hesitation, frequent mispronunciation very often affect intelligibility.	Inadequate	2	X2	4
	speech speed too slow and full of incompetent paucity, hardly intelligible	Poor	1	X2	2

Since the current speaking test model is based on a criterion-reference test, the grading system commonly used in norm-referenced test is not applicable here. As known, a NRT test result is interpreted based on the concept of normal curve or normal distribution. Being so, the minimum passing grade is usually the scores which fall or scatter close around the mean score or between -1 and +2 standard deviation (SD) of the mean, usually is graded C. And the assigning of grades A, B, D and E is decided

by how many standard deviations the scores lie from the mean. For example, the scores that fall 2 SD below the mean (normal tendency) is assigned D, and the scores lying 4 SD below the mean are assigned E. On the other extreme, the scores lying 2 SD above the mean or the normal portion are assigned B, and the scores lying 4 SD above the normal portion are assigned A.

As a criterion-referenced test, the deciding of minimum passing grade and the assigning of other grades on the students' speaking performance cannot be referred to the normal curve distribution concept and its grading system of using normal portion and standard deviations, but should be referred to the set up criteria. Each student's performance on a CRT is compared to a particular criterion in absolute terms, without reference to the other students' scores (Brown 2005). So, in the current test, as mentioned earlier, every level of ability for each content component has their own specified and formulated set of criteria. The grading, i.e., from A, B, C, D, to E, therefore, should be based on or referenced to the accumulated scores of the three components. Each grade indicates detailed specified criteria containing indicators that the test takers or the students should achieve.

Thus, to generate total minimum acceptable score for the speaking test, the minimum acceptable scores of the three components should be added together. Thus, adding up the minimum acceptable content score (18) with minimum acceptable grammar score (6) and the minimum acceptable pronunciation score (6) results in the minimum acceptable speaking achievement test score (30). This score is the total minimum score for the current speaking test which represent the acceptable level and minimum criteria of performance in all the three components. The range of scores for each level of speaking ability and the assigned grades to be used for pass-or-fail and other decisions is shown in Table 5.

Table 5. *Grading system: conversion of scores to grades*

Range of scores	Grades	
44-50	VERY GOOD	A
37-43	GOOD	B
30-36 *(MINIMUM PASSING SCORES)*	*ADEQUATE (MINIMUM PASSING CRITERIA)*	*C (MINIMUM PASSING GRADE)*
20-29	INADEQUATE/BAD	D
10-19	POOR	E

CONCLUSION

At the end of a learning program, a teacher has to evaluate how much the learning objective has been achieved by the students by developing and conducting an achievement test, for making decisions about, e.g., which students would be advanced to the next level of study, which students should graduate, or simply for grading the students. The results of an achievement test can also serve as feedbacks that can tell the change direction for improving curriculum design, staffing, facilities, materials, equipment, etc.—which factors influence the teaching and learning endeavor. Using a criterion-referenced test instrument to measure students' learning achievement can generate trustable or valid results and information about the students' achievement for CRT principally measure the students' achievement performance with reference to or based on particular standards or criteria set up in relation to the instructional objective of the course in question.

To measure speaking ability, the test instrument should measure the learner's actual performance covering the language components, functions, and interactive skill. An analytic approach to scoring which separates speaking skill construct into discrete sub skills is a very suitable and useful scoring procedure for classroom practice since it can generate explicit and specific valuable feedbacks both for the teacher and the students. In the scoring procedure, weighting may be given in different doses to the speaking components respective to the essential roles or degrees each component plays in making up the speaking construct in whole. Content component is expectedly given more weight than the others like grammar and pronunciation. Each component criteria then should be given explicit, easily understood and informative descriptors of expected mastery as a guide for rating the students' oral performance in each component.

In developing the test, firstly the general objective of the course should be identified. If the general objective is "to enable the students to participate and express their ideas spontaneously in a group discussion, storytelling, describing, and reporting," the test objective should be made accordingly, which is "to measure the students' ability in expressing their ideas spontaneously in communicative activities of group discussion, storytelling, reporting, etc." Secondly, consider the selection of topics and tasks for the speaking achievement test. The topic and task selection should consider

the objective of the test, students' level of speaking mastery and their familiarity with the topics and tasks in question. As the final words, students' achievement is what matters most in teaching profession, not the teacher's contentment. Therefore, as far as information and feedbacks for the betterment of classroom instruction is concerned, no detail is too much or too small to consider in developing and conducting a powerful speaking achievement test.

REFERENCES

Brown, Douglass. 2004. *Language Assessment, Principles and Classroom Practice.* San Francisco: Longman

Brown, J.D. 2005. *Testing in Language Program.* New York: McGraw-Hill

Brown, J.D. & Hudson, Tom. 2002. *Criterion-referenced Language Testing.* Cambridge: Cambridge University Press.

Cohen, Andrew. 1996. *Assessing Language Ability in the Classroom. Boston: Heinle & Heinle Publishers.*

Djiwandono, Soenardi. 2008. *Tes Bahasa: Pegangan Bagi Pengajar Bahasa.* Jakarta: Indeks

Foley, J.A. 2005. *New Dimensions in the Teaching of Oral Communication.* Singapore: SEAMEO RELC.

Hughes, A. 2003. *Testing for Language Teachers.* Cambridge: Cambridge University Press.

Mohtar, Tunku. 2005. *Teachers' Perceptions toward Oral Assessment and Their Implications for Teaching.* In Foley, J.A. (Ed.) New Dimensions in the Teaching of Oral Communication. Singapore: SEAMEO RELC.

Mukminatien, Nur. 2005. *Scoring Rubrics for Speaking Assessment.* In Foley, J.A. (Ed.) New Dimensions in the Teaching of Oral Communication. Singapore: SEAMEO RELC.

Harmer, Jeremy. 2015. *The Practice of English Language Teaching.* England: Pearson

Heaton, J.B. 1988. *Writing English Language Tests.* New York: Group UK Limited.

Sims, James. 2005. *A New Dimension in Assessing Oral Communication in a Foreign Language Context.* In Foley, JA. (ed), *New Dimensions in the Teaching of Oral Communication.* Singapore: SEAMEO RELC.

Tuckman, B.W. 1991. *Testing for Teachers.* Sandiego: Horcourt Brace Jovanovich

Underhill, N. 1987. *Testing Spoken Language.* Cambridge: Cambridge University Press