

# Developing a Practical Rating Rubric of Speaking Test for University Students of English in Parepare, Indonesia

Ammang Latifa<sup>1</sup>, Asfah Rahman<sup>2</sup>, Arifuddin Hamra<sup>2</sup>, Baso Jabu<sup>2</sup> & Rafi'ah Nur<sup>1</sup>

<sup>1</sup> Muhammadiyah University of Parepare, Parepare, Indonesia

<sup>2</sup> State University of Makassar, Makassar, Indonesia

Correspondence: Ammang Latifa, Fakultas Keguruan dan Ilmu Pendidikan (FKIP), Universitas Muhammadiyah Parepare, Parepare, Sulawesi selatan, Indonesia. Tel: 62-4212-2757. E-mail: latifa\_ammang@yahoo.com

Received: February 6, 2015 Accepted: April 3, 2015 Online Published: May 28, 2015

doi:10.5539/elt.v8n6p166 URL: <http://dx.doi.org/10.5539/elt.v8n6p166>

## Abstract

This study aims to develop a practical rating rubric of speaking ability in the classroom setting. This research study involves the English speaking lecturers at a number of higher education institutions in Parepare, Indonesia. The product is designed based on Research and Development (R&D) approach, which is adopted from Gall, Gall, and Borg (2003). This Approach is implemented in six steps: problem identification, need analysis, designing product, revising product, field test, and implementation. The data is collected through interview for need analysis issue and product development. The other data collection is from tryout process that aims to evaluate the new product. Both of interviewed data and tried out data are analyzed quantitatively and qualitatively. This research study finds a new design and scoring mechanism of analytical rubric of speaking test. This rubric calls Practical Rating Rubric of Speaking Test (P2RST). It enables the rater to perform scoring mechanism more shortly than the existing analytical rubric. It is also supported by some components such as a) evaluation criteria of communicative competences; grammar, vocabulary, pronunciation, discourse, and strategic competence, b) band score; scale 0, 1, 2, 3, 4, c) descriptors.

**Keywords:** practical rating rubric of speaking test

## 1. Introduction

Many teachers face difficulties in conducting speaking test in the classroom instruction. These difficulties frequently existed in testing speaking implementation (Fulcher, 2003; Luoma, 2004). One example of the difficulties is the excessive use of time in scoring the students' test result. This problem leads lecturers become bored and frustrated to conduct a test. Thus, it brings low frequency to practice the test of speaking competence, especially in the classroom interaction.

The fewest frequency of conducting a test of speaking can be commonly found in the countries where English has enrolled as a compulsory subject in the curriculum, for example, in Indonesia especially at **Parepare** on the teaching English speaking universities. As described by the researchers, some problems may frequently appear in testing speaking by using the scoring mechanism of the existing rubric either holistic or analytical rating rubric. The most problem that frequently appear in the test both holistic and analytical rubric system is excessive use of time of scoring the students' test result. The mechanism of scoring that has implemented in a number analytical rubric revealed into four components of speaking, namely; content, accuracy, comprehension, and fluency. In accumulating the score, the components; content, accuracy, comprehension, and fluency must firstly accommodate the analyzed data from grammar, vocabulary, pronunciation, and discourse competence, and then to be continued to the step of the scoring process.

In the dynamic assessment of speaking test development, the typical existing rubric does not help much the raters to assess the students' achievement in speaking efficiently and effectively. This rubric enforces rater to consume much time and energy to converse and accumulate the score in the rating the result of the test; it must be processed more than one step. It characterized where the rater finishes working to analyze the data of each communicative competences. Then, it conducted again for the analysis of content, accuracy, comprehension, fluency of each communicative competence.

The teachers' responsibility for the language testing and assessment has already phenomenally risen up in the first decade of 21<sup>st</sup> century. This phenomenon results the increasing of profession needs such teacher to create the new pedagogic material design which is relevant to the language testing and assessment (Fulcher, 2012). In university, the assessment plays role to decide whether a nonnative learners of English provide language competence to enter university. This is very important to help the student in their academic interaction, for example, a student should have ability to listen the length lecture in English presentation and catches the meaning of the main idea (Tannenbaum & Cho, 2014). Huang (2012) and Huxham (2010) stated that the assessment of language testing plays also an important role to encourage the students in improving their ability especially in orallanguage. For these reasons, many academic institutions's assign language testing and assessment asa requirement to candidate students to enter university.

A rating rubric of a test is a scoring that consists of specific basic features. It is used to evaluate the students' work in assessment (McNamara, 2000). In specific function, the rubric is used to identify the target of selected criteria that is unused at all (Rinto, 2014). McNamara and Rinto agree that rubric plays important role to work in assessing the peoples' work by using thecriteria to select the expected target in the assessment.

The rating rubric of speaking consists of two types, namely, holistic and analytic rubric. The holistic rubric leads the rater to evaluate or score the overall components of communicative competences without considering another component of language production separately. The other one is analytic rubric. This rubric requires the rater to evaluate or score the components of language production separately (Moskal, 2000; Nitko, 2001).

Holistic rubric is formally used when high errors possibly transpire in some parts of the production process (Chase, 1999). Nitko (2001) further declares that the use of a holistic rubric is probably more appropriate when the task requires participants to create various responses. Principally, the holistic rating rubric reports the overall quality, proficiency or understanding of the content and skills (Mertler, 2001). This typical rubric only provides limited feedback to the students' language development in their profile of score. It potentially brings a curiosity to the students' scores report and it frequently reflect a discourage appearance to the students to attend the next test. The scoring mechanism of the holistic rubric can be illustrated in the following Figure 1.

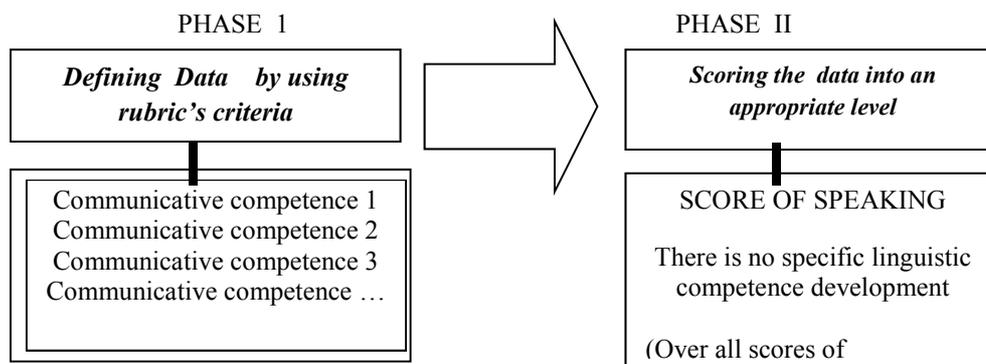


Figure 1. Illustration of the scoring mechanism of holistic rating rubric model

The other type analytic rubric is usually preferred when attention fairly focuses on the response to an interaction (Nitko, 2001). Analytic rubric results initially in many scores, followed by a summary total score representing an assessment on a multidimensional level (Mertler, 2001). The scoring process of analytic rubrics is substantially slower performance that the holistic rubric as the consequence of the assessment process for several different skills. This rubric model requires raters to do scoring process in several times. Principally, both holistic and analytic rubric sometimes do not bring comfort performance in conducting speaking assessment. The general rule of the analytic rubric of scoring is an individual's work must be assessed in a separate time for each performance of the task (Mertler, 2001). However, the beneficial use of the analytic rubric is substantial enough to influence the test takers in order to be more aware of their test (Liao & Hsu, 2014). As stated that the use of the analytic rubric may reflect the scoring process to be much slower. It requires the rater to review the performance production several times. For clear information, the Figure 2 illustrates the scoring mechanism of the analytic rubric in the following capture.

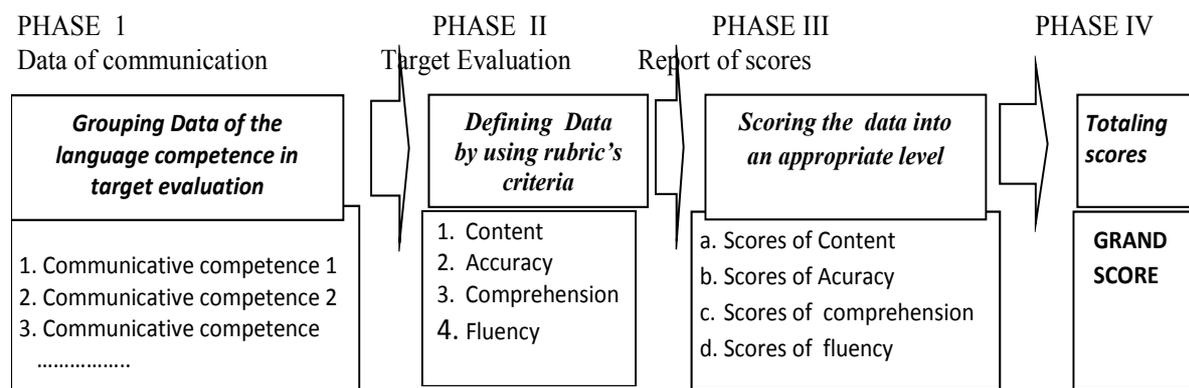


Figure 2. Illustration of scoring mechanism of analytic rating rubric model

The two types of rubrics have strengths and weaknesses in their operation. The holistic rubric has beneficial use of scoring mechanism, where the rater does not have to waste their time to review every aspect on determining the students' score in their test. However, this rubric does not provide detailed profiles of students' language development in their test. The other one is analytic rubric which can provide the information of every aspect of students' language development in the last scoring report. This rubric has a weakness; it expects the rater to review of students' performance in several times.

Nowadays, in language assessment, the rater of language assessment needs a practical and an effective rubric than can work faster and provide a detailed profile of the students' language development. This expectation became intentional project in this research which proposed a development of a practical rubric of speaking test for university students at Parepare.

The word "practical" is the adjective form of "practice". It is a designable action to provide an effective solution to problem (Cambridge Advanced Learners Dictionary, 2008), while rubric of a test is a guided scoring scheme which has a function to measure the performance of the student in the assessment.

In the case of an effective guided scoring, the design of rating rubric development aims to obtain the appropriate need of raters the rubric operation itself. In other words, the main intention of rubric development is to develop a practical scoring mechanism.

The criteria for a practical guided scoring may require some aspects, for instance, possibility of providing the profile about the students' strengths and weaknesses, 2) not time consuming in the scoring process. Bahman & Palmer (1996, p. 36) stated that the criteria can be seen in the needs that already exist in the product design and development. Further, they stated that the resources are also available in the product's activities. The rubric is expected to be a practical instrument in operation of an assessment. The other view is from Douglas (1980, p. 211) who stated that practicality of a product must be considered by financial limitation, efficient time of use, and easy administration.

Based on the two operational criteria of a practical rubric of speaking test, it can hardly be found any typical criteria of rubric model employed in speaking testing in most universities in Indonesia especially in Parepare. This reason motivates the researcher to have initiation for developing a new design of a practical rubric of speaking test.

The focus of developing a rubric is always based on the tasks of oral communication (O' Loughlin, 2001). The tasks are important because the experience of the test is based on the given tasks. The score validity depends on the rating criteria and the relationship between the criterion and the tasks. McNamara and Carsten (2006) stated that a risk condition (a mismatch recording) may happen if the criteria is in a long process of coverage after the tasks presentation. It may reflect to a loss of information about the quality of the performance. Therefore, the components and implementation of the rating rubric must be practical and economical use. These practical criteria tend to keep the examiners' psychology in order to be relaxed to interpret the test taker's data to be an appropriate band score. For this consideration, the rating developers do effort to gain the objective of the rating rubric of test. Some respected ways are very helpful to construct a rubric, for example, communicative performance, scale or band score, descriptors, and test validation.

The assessment construction determines how the criteria play a role in the rubric. It is important to know because the students need to have good preparation of what is the assessment required in the test (Sarwar et al., 2014). For this reason, the rubric designer must be aware of all criteria of the rating tasks, validate all criteria on some tasks and components of the rubric. This treatment will bring a wise decision to clarify the meaning of the proposed rating criteria of the rating rubric. It may also help the rubric developer further to modify his or her concept of the criteria of the rubric itself. The definition of the constructed test is very important to determine the decisions of the rating process, for example, the tape-based and highly structured live tests. It often contains short responses, which then, scored one by one (Luoma, 2004, p. 171).

The development of rubric always based on the institutional requirement. This effort consists of three phases. The first, the rubric designer, needs to get agreement on the condition about what performance or behavior want typical rubric to be elicited in the observation as data for scoring, or it is a typical administration (McNamara, 2000, pp. 36-37). The second, the communicative performance that need to be a critical measurement in the speaking test is formally various, for instance, grammar, vocabulary, pronunciation, discourse, and strategic competence (Scarcella, 1992). The third, the rater of test allocates the grade based on the elicited performance (McNamara, 2000, pp. 36-37).

Another important thing of rubric development is administration of collecting spoken data. This administration plays an important role to administer the data collection of the test taker. For this reason, administration must be considered as a preliminary requirement for the criteria of speaking performance where this administration must be in audio-visual recording.

The audiovisual recording has function to collect spoken data that vary in some communicative performances for example, grammar, vocabulary, pronunciation, discourse, strategic competence. These communicative competences are intentional observation in the language production. It is an essential tool for the speaker to communicate ideas or receipt messages in language interaction (Milanovic, 1989; Canale & Swain, 1980; Scarcella, 1992). In addition, Davis (2011) stated that grammar knowledge as one of communicative competence that is considered in standard language operation. Therefore, audio-visual recording is an appropriated way to collect the typical competences such as grammar, vocabulary, pronunciation, discourse, and strategic competence.

## **2. Research Method**

Some steps are popular to develop a product, namely, 1) identifying problems, 2) need analysis, 3) designing product, 4) revising product and, 5) field test and 6) implementation. These steps also become as prime tendency project in a rubric development.

### *2.1 Identification of Problem*

Problem identification aims to know the gap between reality and expectation in conducting the scoring of the speaking test result at the university students of English at Parepare. This identification involves the lecturers of English speaking. In the session of problem identification (gap validation), the identification result qualitatively showed, most of the respondents felt discourage and bored to conduct a speaking test. It is caused by the unpractical rating rubric test that already used.

### *2.2 Need Analysis*

The second step of this design is need analysis that plays a role to get brief and clear information about what rater need to do, to have and to pursue in the new rubric of speaking test. For this reason, this rubric needs an appropriate analysis that is applicable for scoring the test of speaking. For important consideration, the need analysis must have a valid instrument in order the rubric developer can develop the rubric with an appropriate need for speaking test at certain level. In analyzing the needs of new rubric, the researcher adopted the approach of Hutchinson and Waters (1987). They analyzed the students' need in material design while, in the need analysis of P2RST rubric, the developer explores the needs of the new rubric by using some questions in relevant material for speaking test.

### *2.3 Designing Practical Rubric of Speaking Test*

This research study performs an overview and a summary of the existing theories and rubric models to design the new rubric model. According to Frantz, Bailey, Starr, and Perea (2014) a developer of assessment should share their knowledge through an overview and summary to prepare a standard English Language Instruction and assessment. Regarding with the knowledge of rubric development, Namara (2000, pp. 36-37) offers three aspects of rating rubric development. The first aspect, the rubric developer should meet the condition of the test taker (speaker). The test taker condition refers with the elicitations of test taker's behaviors when producing language

in the test. It probably appears in audio or audiovisual performance. The second aspect, the rubric developer should determine a particular communicative performance that become target assessment. The third aspect, the developer determines the criteria of competence by allocating grade or score. These aspects may be extended into four efforts; determining the way of data recording, determine the typical speaking tasks (speaking activity), involving the communicative competencies as the criteria of the speaking rubric, designing the scales and descriptors.

### 2.3.1 Determining the Way of Data Recording

The way of data collection depends on the typical performances that are used to measure the spoken production of the speaker. The typical performances were proposed by the raters through the need analysis for example, vocabulary, grammar, discourse, pronunciation, strategic competence. These performances are available for observation in audio-visual recording. Vocabulary, grammar, pronunciation, and discourse competence are observable performances while, the strategic competence could not to be observed through audio recording. It must have a visual recording because the rater must make sure whether the speaker uses or not nonverbal (NV) language to maintain his or her goal of communication, even though the speaker sounds hesitated to produce utterance. For this reason, the typical data such as vocabulary, grammar, discourse, pronunciation, and strategic competence requires audiovisual recording process in data collection.

### 2.3.2 Determine the Typical Speaking Tasks (Speaking Activity)

There are some typical speaking task such as pair conversation, interview, and telling story. These speaking activities get a prior portion from the respondents in the need analysis exploration because these activities are comfortable and easy for the rater to observe and elicit data from the test takers.

### 2.4 Involving the Communicative Competencies as the Criteria of the Speaking Rubric

The rater must focus on some speaker's (test takers') competencies in communicating oral language. Kangsumin in Richard (2002), Canale and Swain (1980) and Scarcella and Oxford (1992) agree that there are some competencies must have considered attention for examples, linguistic competences which consist of grammar rule, vocabulary, pronunciation (accent), discourse competence, and non-linguistic competence (strategic competence). Communicative competences are used by the speakers to deliver their goals in spoken interaction systematically. These competences integrated one another to set up a meaningful way of communication.

Relating to the need analysis of the new rubric design, there are some competences proposed by the respondent (raters) to be measured critically, they are grammar, vocabulary, pronunciation, discourse, and strategic competence. As Canale and Swain (1980) stated that grammar, discourse, sociolinguistic is the competences of communication which reflected the use of the linguistic system in communication. In addition, Scarcella and Oxford (1992, p. 141) stated that grammar competences are syntax, morphology, vocabulary, and basic sound. These competences enable then speaker to use the English language unhesitatingly and structurally. Another expert is Gibson (2009) who issued that the discourse presented remarkable similarities to the ways in which face to face conversation has been seen to operate in terms of the organization of conversational turns, the application of specific interactional rights, the lineal development of topics of conversation and the structural use of question-answer turn pairs.

### 2.5 Designing the Scales and Descriptors

Luoma (2004, pp. 82-86) issued that there are three methods probably used on developing speaking rating scale in a rubric, i.e, intuitive method, qualitative method, and quantitative method. In this research project, the researcher felt comfort to employ the intuitive method as an approach to develop the scale and descriptor in the new rubric. The reasons of choosing intuitive method are the available experiences, the comparison of the existing scale of rubric, the existing curriculum, experts' helps, and coach and trial report. These reasons follow the view of Namara (Mc2000) who stated that intuitive methods of scale development are not based on data collection but principle interpretation of experience in teaching and assessment at relevant ability levels. In addition, the rubric developer must also consult the existing scales, curriculum documents, teaching material, other material, and the information into draft descriptors at an agree number of levels. Even in a small context, the developer also should make revision on scales and descriptors twice. In broader effort, the committee of rubric project development take meeting in several times and consulted other experts, and tried out of the scale to achieve a usable formulation of the scale Clark and Clifford (1988) scales and descriptors rubric need to be consulted to the existing scales and descriptors. He further described that the small committee may play role to develop scales of rubric.

### *2.6 Exploring and Comparing the Existed Rubric Scale*

Based on the direction of intuitive method, the researcher tried to explore and compare some previous existed scales and descriptors which are already introduced by developer of rating scale in the speaking test area, Luoma (2004.) At the end of exploration and comparison of document, curriculum, empirical study, and previous model of speaking rating scale, the researcher adopted some information from the previous scales and descriptors. Some experts and institutions have introduced rubric of speaking test, for example, North (1994), the American Council for the Teaching of Foreign Language (ACTFL, 1999), The Common European Framework of Reference (CEF) (Council of Europe, 2001) EFL/ESL Native English Teacher Speaking Tests in a Korean Public High School Kimci Icecream (2010). In this step, the researcher compared some previous existed scales and descriptors with their characteristics in terms of 'how easy' to use in defining and to convert the raw data into a certain level of scale. Empirically, the researcher analyzed and formulated the descriptors in every scale for each communicative competence.

### *2.7 Revision*

The new product was revised based on the experts' validation. This revision was conducted until twice. The first revision was performed after validating the new product, then; the second revision is performed after obtaining the field test.

### *2.8 Field Test and Implementation*

The new product was examined by the rater in a limited and macro scope of the institution which has already developed English program. In this step, the user (Lecturer of English) applied the new product of rubric. In this test, the impact of the rubric was evaluated by analyzing the intra-rater and inters rater's performance in a different period. Furthermore, this test was carried out until twice in two weeks range, and it was performed by the same rater. This aim is to confirm about consistency of the raters' comprehensible input on the use of the rubric. This last process of this design is the product was ready to be used or implemented in the real application, in the curriculum of evaluation for oral language production.

This research employs a qualitative and qualitative method to analyze the data of this research. Some steps operated in the Research and Development namely, Problem identification, collecting information, designing product, expert validation, revising product design, pre-field testing, revising product, main field testing, last revising for product, produce or design and perform the product (Sugyono, 2007, p. 298). These steps attempted to design a new rating rubric speaking achievement. These steps were processed in three phases namely, need analysis, develop a product, and Evaluation product. To see the practicality of the rubric, this study operated questionnaire to explore the rubric practicality in the mechanical scoring of time, cost of operation, and the accumulation of the rubric scoring process. The respondents are the lecturers of English in speaking subject at Universities in Parepare, Indonesia. The respondents were chosen purposively.

## **3. Findings and Discussions**

The finding of this study begins with the development of design model of the rubric which then was analyzed the practicality of the PRRST rubric. The finding product was tried out to the rater and then, it was reflected to the raters. This is intended to see the raters' perception about the practicality of the rubric toward the operational of time, accumulation of scoring, and cost of operation.

### *3.1 The Design Model of PRRST Rubric*

The formulated descriptors and other established components were arranged into a set of a new rubric here after called as Practical Rating Rubric of Speaking Test (PRRST). So, PSSRT was then, validated by the experts who were available in speaking assessment. The information from the expert was then brought to the second couch to get a similar understanding with the user of rubric, and then, sustained by revision. The result of the revision was examined in the second trial user of the rubric. The descriptor revision was showed in the revision of The PRRST Rubric.

Table 1. The last revision of PRRST rubric

| Scales | Descriptors:  |
|--------|---|
| 0      | The speaker is <b>not able</b> to use the rule of (vocabulary competence, syntactic competence, pronunciation, discourse competence, strategic competence) on his or her utterances <b>at all</b> .   |
| 1      | The speaker <b>is able to</b> perform (vocabulary competence, syntactic competence, pronunciation, discourse competence, strategic competence)but<br>The speaker still makes <b>many (approximately 57% to 75 %)</b> ungrammatical production<br>The speaker Only produced the related vocabulary (approximately <b>25% to 49%</b> )<br>The speaker still makes mispronunciation approximately <b>66-86 %</b><br>Speaker disconnected discourse (approximately <b>67 to 83.3%</b> disconnected meaning)hesitation in utterance  |
| 2      | The speaker <b>is able to</b> perform (vocabulary competence, syntactic competence, pronunciation, discourse competence, strategic competence) even he or she still makes <b>some</b> ungrammatical production/ unrelated vocabulary/mispronunciation/ disconnected discourse/ hesitation in utterance<br>Even, he or she still makes some lacks of formality (approximately 50 % to 60%)<br>The speaker only produced 38% to 50% words still employ a number of times to recall the words.<br>Still makes some mispronunciation (approximately 50% to 62%)<br>The speaker still produces some unconnected utterances (approximately 50% to 66%)<br>The speaker still has some problems to communicate his/her meaning (approximately 50%).   |
| 3      | This level describes that the speaker is <b>able to</b> perform (vocabulary competence, syntactic competence, pronunciation, discourse competence, strategic competence)in (he or she still makes <b>few</b> ungrammatical production/unrelated vocabulary/mispronunciation/ disconnected discourse/hesitation in utterance<br>The speaker still makes few lacks of formality (approximately 10-30%)<br>The speaker is able to produce approximately 67-90% the related words but still employ few times to recall the words.<br>The speaker still makes few mispronunciations (approximately 10-33%)<br>The speaker still making few unconnected meaning in communication (approximately 33%)<br>The speaker still has problem to communicate his/ her meaning (approximately 16.7%) |
| 4      | This level describes that the speaker <b>is able</b> to perform (vocabulary competence; syntactic competence, pronunciation, discourse competence, strategic competence) <b>no mistake</b> .<br>Seems there is no mistake in grammar<br>Very fluent & seems there is no difficulty to produce words.<br>Utter the words, phrases, sentences close to native ‘ pronunciation<br>The speaker is fluently uttered the language which seems there is no problem in communication.   |

### 3.2 The Scoring Mechanism of P2RST

We already compared the phase of mechanical scoring of holistic and analytic rubric in the Figure 1 and 2. The strength of holistic rubric was providing a short phase mechanical scoring while the analytic rubric provided detailed information of the tester. This beneficial mechanism inspired the mechanical scoring development of P2RST as a practical rubric. On the other word the P2RST rubric provided a short phase and detailed information of the tester. In the following illustration, the scheme about the process of scoring of P2RST only provided two phases and information of test taker language development in the test in term of the score of communicative competences. This feature can be described in the following illustration of mechanical scoring of PRRST rubric.

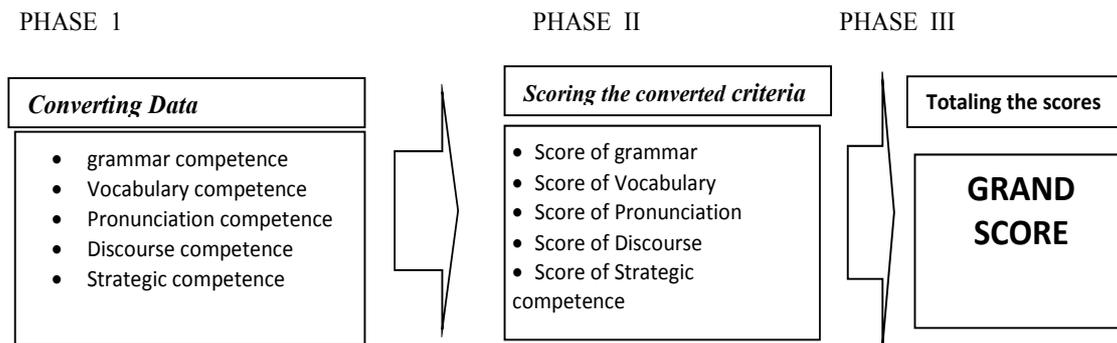


Figure 3. The illustration of mechanical scoring or P2RST

This illustration revealed that the mechanical scoring only employs three phases in the scoring process. As addition, the report of scores provides detailed information on the language development of the test taker. The P2RST rubric mechanical scoring slightly combined the mechanical scoring of holistic and analytic rubric.

### 3.3 The Practicality of the Rubric

The last revision of the PRRST rubric was examined and reflected by the rater to see the impact whether it is practice or not on its implementation. The practicality is specified in the time use of operation, the score accumulation process, and the cost of operation.

The information on the practical operation of P2RST rubric was explored in three items of practical operation where these items are obtained from the result of the need analysis. These items tent to see time use of operation, accumulation scores process and the cost operation. The analysis of the practical use of the rubric was presented in the following table.

Table 2. The practicality perception

| Items   | Raters' score |    |    |    | Freq. |      |     |   |    | Total |
|---|---------------|----|----|----|-------|------|-----|---|----|-------|
|   | R1            | R2 | R3 | R4 | SA    | A    | N   | D | SD |       |
| 1 The PRRST rubric operates a simple accumulation of score      | 3             | 4  | 4  | 4  | 0     | 3    | 1   | 0 | 0  | 4     |
| 2 The scoring mechanism of PRRST does not operate extra time    | 4             | 5  | 5  | 4  | 2     | 2    | 0   | 0 |    | 4     |
| 3 The operation of PRRST rubric does not need a high cost       | 5             | 5  | 5  | 5  | 4     | 0    | 0   | 0 | 0  | 4     |
| Frequency of response (F)                                       |               |    |    |    | 6     | 5    | 1   | 0 | 0  | 12    |
| % of frequency  |               |    |    |    | 50    | 41.7 | 8.3 | 0 | 0  | 100   |
| Gained score of practical perception (F)x(liker scale)          |               |    |    |    | 30    | 20   | 3   | 0 | 0  | 53    |
| The critical area; (item)x( max.scale)x(participant)/3x5x4 = 60 |               |    |    |    |       |      |     |   |    |       |
| 0-12 = Very Bad (VB)  |               |    |    |    |       |      |     |   |    |       |
| 13-24 = Bad (B)   |               |    |    |    |       |      |     |   |    |       |
| 25-36 = Neutral (N)   |               |    |    |    |       |      |     |   |    |       |
| 37-48 = Good (G)  |               |    |    |    |       |      |     |   |    |       |
| 49-60 = Very good (VG)  |               |    |    |    |       |      |     |   |    |       |

The table perception of practicality confirmed that the critical area perception was dominated by the positive area; it was indicated by 50% responses which stated strongly agree with the perception of the PRRST rubric does not need time-consuming and the P2RST rubric does not need a lot cost on its operation, then followed by 41.7% responses which stated that the P2RST rubric only need simple scoring process, it does not need a lot cost,

and it does not need time consuming, it does not need a lot cost to operate while there were only 8.3% responses which stated in neutral response, and there was nothing in strongly disagree, or even in disagree response.

In the gained score which based on the liker scale divisions, the raters' perception revealed the area in a very good perception. It was indicated by the gained score number; 53 laid in the area 49 to 60 that categorized as very good perception. This data is illustrated in the following figure.

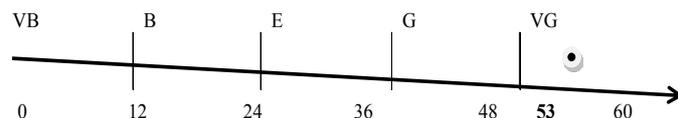


Figure 4. The critical area of practicality perception

#### 4. Discussion

As the researcher find, the model of P2RST rubric is a kind of analytic rubric of speaking which my help the rater to operate the rubric more practice rather than operating another analytical rubric of speaking. The first process of finding the model of P2RST rubric is need analysis. It aims to meet the raters' need in the rubric operation.

There are three main components employed in this rubric namely, communicative performance, scales of score, and descriptors. These components are already designed based on the raters' need to support its operation. Linguistic competence as components of the rubric is appropriately used to measure the speakers' competence in speaking as Canale and Swain (1980) stated that there are some communicative competences which utilize by the speakers to communicate their thought to others i.e., grammar, vocabulary, pronunciation, discourse, strategic competence and sociolinguistic competence. In P2RST rubric, sociolinguistic competence is not involved as part of speaker competence to be measured because its data is very difficult to elicit that potentially bring a hard problem to define the meaning for scoring. This view is also supported by Milanovic (1989), who did not also involve sociolinguistic competence in his rubric. This component is not involved to be analyzed in the rubric because they are already inclusively measured at the beginning of the scoring process; the level of score of every competence captured the accuracy of the speaker production of language.

Another completeness of P2RST rubric are scales or scores and descriptors. Scales or scores of P2RST rubric apply five scales (scale 0, scale 1, scale 3, and scale 4). These scales are simplest which is believed can enhance the raters' motivation to carry out test regularly in speaking course of the classroom interaction. The application of these scales is very easy to use where the score is only matched with the criteria of the appropriate scale. Principally the more positive ability description is the higher level of scale that the speaker has or, the less ability description, is the lower level of scale the speaker possess.

Descriptor is a pair of scale which has a function to integrate the criteria of scale level with the competence of speaking production from the speaker. In the integration process, it works to define the language production in positioning the ability to a certain level of the scale. Therefore to get an understandable definition of this process, all descriptors were designed with practical language use which used to describe the data in the appropriate level of scale.

For the completeness, the product of P2RST rubric was required of a certain task in collecting the data from the speaker (test taker), for example, the speaking task in which can be performed in appropriate criteria of institutional requirement of speaking course at university, as O'Loughlin (2001) stated that the focus in designing a rating rubric is always based on the tasks of oral communication. The tasks are important because the examiners' experience of the test is based on the given tasks. For the scores validity, it depends equally much on the rating criteria and the relationship between the criterion and the tasks. McNamara and Carsten (2006) stated that if the criteria are developed long after the tasks have been finalized, as they often are, there is risk condition of a mismatch recording, which probably reflects to a loss of information about the quality of the performance.

One thing of the institutional expectation of learning to speak course at university is evaluating the students speaking competence i.e., grammar, vocabulary, discourse, pronunciation, and strategic competence. Therefore,

P2RST rubric focused on measuring the audiovisual recording data. For this criterion, the typical task of speaking activity was only selected and suitable on two activities, namely pair conversation and interview activity.

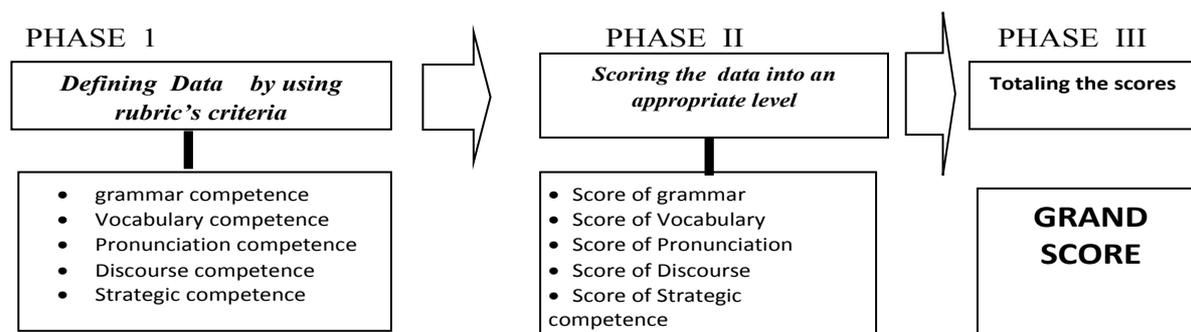
The special thing of P2RST rubric was as an a little bit of collaboration of analytic and holistic rubric of scoring mechanism. The rater does not have to waste their time to score the test result of the participant. It is like a quicker scoring process (Nitko, 2001) and it is consistent to bring up a profile the language production in the scoring report (Metler, 2001). This mechanism lets the rater scores the speaker's performance practically beside that it provides the profile of speaker's language development separately in the last report of score.

## 5. Conclusion

The model of PRRST rubric was derived from analyzing the gap between the existing rubric and the expected one. It was found that:

The existing rubrics which have been used by the lecturers at UMPAR to assess the students' speaking competence are analytical and holistic rubrics. Both of them are not practice and expected. Some existing analytical rubrics provide various components which are sometimes overlap each other, for example between content and discourse, grammar and accuracy, fluency and pronunciation. These overlap components demand the raters to measure and assess the data more than once. The holistic rubric, meanwhile, does not present any profile of the students' language development at the end of scoring process. These typical rubrics are considered to be quite complicated or lacking practical rubric. For this reason, most of English speaking lecturers feel bored to assess the students' speaking performance by using rubrics. In turn, the lecturers assess the students' performance intuitively.

Based on these phenomena, the researcher designed a new rubric model which enabled the rater to assess the students' speaking performance more practically. This new rubric is an analytical rubric because it needs the rater to make an assessment by analyzing some components of the student's performance. This rubric also adopts the criteria of the holistic rubric because it provides simple steps in scoring the components of students' performance. However, this rubric provides profile of students' language development at the end of the scoring process which does not exist in the existing holistic rubric. Simply, this analytic rubric is more practice than the existing one and more informative than existing holistic rubric.



This typical rubric is called Practical Rating Rubric of Speaking Test (P2RST). This rubric equipped four scales of band score for every component of communicative performance. The typical communicative performance that employed in P2RST rubric is namely vocabulary competence, grammar competence, pronunciation competence, Discourse competence, strategic competence.

The special thing of P2RST rubric was collaboration of analytic and holistic rubric of scoring mechanism where the rater does not waste their time to score the test result of the participant. The P2RST rubric practicality is available in the three views of perception, namely simple accumulation of scoring mechanism; the P2RST rubric does not operate extra time; the P2RST does not operate high cost.

## Acknowledgements

Our first acknowledgements go to extraordinary three people with whom we have been working closely. We are very much indebted to Prof. Dr. Haryanto, M.Pd who has generously offered his expertise, wisdom, and continuous encouragement in guiding us and mentoring us step by step through the whole research process. We are eternally and personally grateful to Dr. H. KismanSalija, M.Pd, whose profound knowledge, intelligent

guidance on every consultation meeting, and passionate love for academic work have always been an inspiration for us. We also would like to give our sincere and deep gratitude to Prof. Dr. H. BurhanuddinArafah, M.Hum for his careful correction of every session of our work, for his smart remarks in leading me to take a critical perspective on literature review and for his humorous lectures. Interpersonally, thanks are due to director of Graduate Program of Makassar State University, Prof. Dr. Jasruddin, M.Si, for his Endless strong attention support, motivation and kind help on finishing ourwork.

## References

- ACTFL. (1999). *The ACTFL Proficiency Guidelines: Speaking (revised 1999)*. Yonkers, NY: ACTFL.
- Allen, M. J. (2004). *Assessing academic programs in higher education*. Bolton, MA: Anker.
- Bachman, L. F., & Palmer, A. (1996). *Language Testing in Practice*. Oxford: OUP.
- Brown, A. (1991). *Teaching English Pronunciation: A Book of Readings*. London: Routledge.
- Canale, N., & Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1, 1-47.
- Chase, C. I. (1999). *Contemporary assessment for educators*. New York: Longman.
- Clark, J. L. D., & Clifford, R. T. (1988). The FSI/ILR/ACTFL proficiency scales and testing techniques: Development, current status and needed research. *Studies in Second Language Acquisition*, 10, 129-147.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: CUP.
- Davies, A. (2011). Does Language Need the Native Speaker? *Journal of Personality Assessment*, 11(1), 27-44. <http://dx.doi.org/10.1080/1543430.2011.570827>
- Fisher. (2006). *Students in Pragmatics: Approach to Discourse*. Elsevier. The Netherland.
- Frantz, R. S., Bailey, A. L., Starr, L., & Perea, L. (2014). Measuring Academic Language Proficiency in School-Age English Language Proficiency Assessments under New College and Career Readiness Standards in the United States. *Journal of Language Assessment Quarterly*, 11(4), 432-457. <http://dx.doi.org/10.1080/15434303.2014.959123>
- Fulcher, G. (2003). *Testing second language speaking*. Essex, U.K.: Pearson Education.
- Fulcher, G. (2012). Assessment Literacy for the Language Classroom. *Journal of Language Assessment Quarterly*, 9(2), 113-132. <http://dx.doi.org/10.1080/15434303.2011.642041>
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <http://dx.doi.org/10.1080/15434303.2011.642041>
- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational Research: An Introduction (7th ed.)*. USA, Pearson Education.
- Gibson, W. (2009). Negotiating Textual Talk: Conversation Analysis, Pedagogy, and the Organization of Online Asynchronous Discourse. *British Educational Research Journal*, 35(5), 705-721. <http://dx.doi.org/10.1080/01411920802688754>
- Huang, S. (2012). Pushing Learners to Work through Tests & Marks: Motivating or Demotivating? A Case in a Taiwanese University. *Journal of Language Assessment Quarterly*, 9, 60-77. <http://dx.doi.org/10.1080/15434303.2010.510898>
- Hutchinson, & Waters. (1987). *English for Specific Purpose: A Learning Centre Approach*. Cambridge: Cambridge University Press.
- Huxham, M., Campbell, F., & Westwood, J. (2010). Oral versus written assessments: A test of student performance & attitudes. *Journal of Assessment & Evaluation in Higher Education*, 37(1), 125-136. <http://dx.doi.org/10.1080/02602938.2010.515012>
- Kimchi Icecream: The Second Serving. (2011). *EFL/ESL Native English Teacher Speaking Tests in a Korean Public High School—Planning and giving speaking tests Part I*. Co-Teaching Issues and Stories. Stories about teaching English in South Korea.
- Liao, H., & Hsu, L. (2014). Using an Analytical Rubric to Improve the Writing of EFL College Students. *The Asian TEFL Journal*, 16(1), 296.
- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.

- McNamara, T. (2000). *Language testing*. Oxford, UK: Oxford University Press.
- McNamara, T., & Carsten, R. (2006). *Language Testing: The social dimension*. Blackwell Publishing. USA.
- Mertler, C. A. (2001). *Using performance assessment in your classroom* (Unpublished manuscript). Bowling Green State University.
- Milanovic, M., & Saville, N. (1989). Introduction. In M. Milanovic, & Saville (Eds.), *Studies in Language Testing: Performance testing, Cognition, and Assessment, selected Papers from the 15 the language testing research colloquium*. Cambridge and Arnheim. Cambridge: Oxford.
- Moskal, B. M. (2000). Scoring rubrics: What, when, and how? *Practical Assessment, Research, & Evaluation*, 7(3). Retrieved from <http://pareonline.net/getvn.asp?v=7&n=3>
- Nitko, A. J. (2001). *Educational assessment of students*. Des Moines, A. A: Prentice-Hall.
- North, B. (1995). The Development of common framework scale of Descriptors of Language Proficiency based on the theory of measurement. *System*, 23(4), 445-465.
- O' Loughlin, K. J. (2001). *Studies in Language Testing*. Cambridge University Press. United Kingdom.
- Practical. (2008). *Advance Learner's Dictionary* (3rd ed.). Cambridge Univeristy Press.
- Richards, J. C., & Lockhard, C. (1994). *Reflective teaching in second language classroom*. New York: Cambridge University Press.
- Scarcella, R. C., & Oxford, R. L. (1992). *The Tepetry of Language Learning: The Individual in the Communicative Classroom*. Boston, MA: Heinle & Heinle.
- Schiffirin, D. (1987). *Discourse Marker*. Cambridge University Press.
- Shumin, K. (2002). *Developing Adult EFL Students' Speaking Abilities in Richards: Methodology in Language Teaching: An Anthology of Current Practices*. Springer Science Business Media, Cambridge University. <http://dx.doi.org/10.1007/978-0-387-09506-6>
- Sugyono. (2011). *Metode Penelitan Administrasi: Dilengkapi Metode R&D*. Penerbit Alfabeta, Bandung.
- Tannenbaum, R. J., & Cho, Y. (2014). Critical Factors to Consider in Evaluating Standard-Setting Studies to Map Language Test Scores to Frameworks of Language Proficiency. *Journal of Language Assessment Quarterly*, 11, 233-249. <http://dx.doi.org/10.1080/15434303.2013.869815>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).