

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter orderly presents theoretical foundation: definition of rating scale, writing, and assessment. Then, it is continued by the explanation of rating scale in writing assessment, explanation of EBB scale, and EBB scale as a rating scale descriptor in writing assessment. Next is about previous studies, which have the relation of the research.

A. Theoretical Background

1. Rating Scale

In language testing, Namara states that a rating scale is a series of ascending descriptions of salient features of performance at each language level. A language performance can be assessed by examining either the whole impression of the performance or the performance according to different criteria. In this regard, there are two types of rating scales: holistic scales which describe learners' performances as a whole (e.g. the American Council for the Teaching of Foreign Languages scale); and analytic scales which consist of a number of criteria referring to particular aspects of performance such as grammar, fluency and content (e.g. the Test of Spoken English).

2. Writing

Writing is a fundamental aspect of academic literacy and communicative competence in the current educated world.¹ According to Hammad, writing is a form of language outcomes as a real form of language input. In writing, to produce an essay, the authors should have an idea which written with a correct grammatical, mechanical writing, content and communicative language.

Furthermore, writing as a thinking process which involves generating ideas, composing these ideas in sentences and paragraphs, and finally revising the ideas and paragraphs composed. Good writing also requires knowledge of grammatical rules, lexical devices, and logical ties.² Moreover, White and Arndt define 'writing' as "a form of problem-solving which involves such process as generating ideas, discovering a voice with which to write, planning, goal setting, monitoring and evaluating what is going to be written, and searching with language with which to express exact meanings".

3. Assessment

Traditionally, the word “assessment” has referred to the way teachers assign letter grades on tests and quizzes. Assessment has also been used as a way to discuss teaching effectiveness. However, assessment is now

¹ Nadia Behizadeh and George Engelhard, “Historical view of the influences of measurement and writing theories on the practice of writing assessment in the United States”, *Assessing Writing*, vol. 16, no. 3 (2011), p. 190, accessed 27 Jul 2016.

² Hammad Enas Abdullah, "Palestinian University Students' Problems with EFL Essay Writing in an Instructional Settin", *Journal of Second and Multiple Language Acquisition – JSMULA*, vol. 2, no. 1-21 (2014), p. 2, accessed 31 May 2016.

Assessment used by the teacher to measure student ability in final lesson. The teacher gives assessment to student as a product for the student's input. The assessments are appropriate with students' creativity in applying the lesson. Teacher's goals visible with assessment criteria gives to student, it must success or not student utilize their knowledge of the lesson to get the perfect assessment.

The conceptualization of rating scale as part of the test construct has opened a new horizon on examining different aspects of rating scales functioning in performance assessment. Writing assessment as a kind of performance assessment is affected by the quality of the rating scale used. To a large extent, the common thread of arguments on rating scales converges on the important issue of construct validity. Weigle summarizes McNamara, identifying the importance of rating scale in the quality of

⁴ Nicole Williams, *Reflective journal writing as an alternative assessment* (2008), p. 2, <http://205.133.226.154/Files/pdf/Education/JTIR/VolumeIII/williams.pdf>, accessed 31 May 2016.

1. First, there are also some researches related to using rating scale for assessment in productive skills. Harai and Koizumi were done the study. They were analyzing validity and reliability between two rating scale, those are; EBB scale that compared with an analytic scale. The research was conducted to know the rating scale more valid and reliable to assessing productive skills. This research using Story Retelling Speaking Test (SRST) to verify two rating scales. The finding shows that EBB scale more valid and reliable for SRST, whereas analytic scale more practically. The EBB scale was slightly superior in reliability and validity, whereas the analytic scale excelled in practicality. However, the results helped us find points for revision in the scales. First, the descriptors of EBB Grammar & Vocabulary criterion should be modified. Second, the Communicative Efficiency and Content criteria of the EBB scale can be combined to enhance its practicality. Third, the current EBB binary format might be changed into one similar to the analytic scale. Since both scales have strengths, combining the good aspects of these scales may enable us to create a better scale in the future.¹¹
2. Second is the research was done by Mu-shuan Chou, entitled “Teacher Interpretation of Test Scores and Feedback to Students in EFL Classrooms: A Comparison of Two Rating Methods”. The research was conducted to know interpreted teacher in scoring speaking and student’s benefit from feedback of two rating scale descriptor. This research using

¹¹ Harai Akiyo, and Koizumi Rie, “Validation of the EBB scale: A Case of the Story Retelling Speaking Test”, p. 15.

effectiveness using local rating scale. This research was compared the assessment used traditional approach and rating scale. The research was done with some questioner given to teacher or test-taker to know how far the teacher used local rating scale. Majority of the raters in this study believed that native scales have to be appropriated in the context before application. In their ideas, unmediated application of native rating scales would surface a hidden conflict between the assumptions behind these scales on the one hand and the realities of the local context on the other hand. McNamara states that strongly questions the validity of rating scales and by tracking the origin of scale tradition in the FSI test in the 1950s shows how successive rating scales developed over the last four decades have been heavily influenced by the assumptions, and even the wording of the original work, and rare empirical validation has been done.¹³

claim for a PDT would rest upon the comprehensiveness of the description upon which it was generated, and the relevance of the assessment categories to current theories of ‘successful interaction’ within a particular context. As such, other PDTs must be developed through a careful analysis of communication in context, and a theoretical description of the constructs that underlie successful interaction, in order to generate context sensitive assessment categories.¹⁹

8. Seventh is "*Pengembangan Instrumen Penilaian Membaca Kelas VII SMP*" by Nila Maulan, Imam Agus Basuki, and Bustanul Arifin. This research was done to develop an instrument for reading assessment. The product of this research was an instrument used to assess reading dictionary, reading fast, reading ceremony text, retelling story, and giving comment on narrative text. There were two instrument for assessing reading dictionary; those are assignment of reading dictionary and rubric for assessment. Instrument for assessing fast reading were subjective text and answer clue. Instrument for reading ceremony text were reading ceremony text and rubric assessment. Instrument for assessing retelling story were retell the story and rubric assessment. And instrument for assessing giving comment for text book were give comment and rubric assessment. According to evaluator expert, practice expert and students,

¹⁹ G. Fulcher, F. Davidson, and J. Kemp, “Effective rating scale development for speaking tests: Performance decision trees”, ...p. 25.

the researcher concluded that based on validity, reliability, and practices the product appropriate to be implemented.²⁰

9. Eighth is “*Pengembangan Rubrik Penilaian Portofolio Proses Sains Siswa pada Materi Ekosistem di SMPN 1 Wedarijaksa-Pati*” by Vera Widyaningsih. This research was done to develop rubric for sains assessment. Based on the research, the writer concluded that the teacher did not use rubric for each aspect on the lesson. Teacher used rubric for assess cognitive aspect in *LKS* only. The researcher develop rubric for assessment through identify process, construct rubric, and implementation the rubric. Based on the result, students mean had not differences between three classes. Teacher and students responses were positive for rubric assessing sains. Rubric for portfolio assignment for ecosystem material to be appropriate for assessment.²¹

This research is different from all of those researches because in this research, the researcher analyze the development and implementation of empirically derived, binary choice, boundary definition scale in senior high school. The procedure to develop rating scale descriptors of empirically derived, binary choice, boundary definition scale is first give the task to a group of students drawn from the target population. Take the resulting language samples and ask the group of experts to divide them into two groups

²⁰ N. Maulani, I. Agus Basuki, and B. Arifin, "Pengembangan Instrument Penilaian Membaca Kelas VII SMP". *Universitas Negeri Malang*. 2012

²¹ V. Widyaningsih., Undergraduate thesis; “*Pengembangan Rubrik Penilaian Portofolio Proses Sains pada Materi Ekosistem di SMPN 1 Wedarijaksa-Pati*”. (Universitas Negeri Semarang, 2013)

