

CHAPTER 1

INTRODUCTION

Chapter 1 presents introduction of the study. It contains Background of the Study, Research Question, Objective of the Study, and Pedagogical Significance.

A. Background of the Study

Evaluation plays an important role in every teaching and learning process. It contributes directly because it is integrated in the school program. Michael Russel defined evaluation as the process of judging the quality or value of the performance or a course of action¹. Lyle F. Bachman also stated that evaluation can be defined as the systematic gathering of information for the purpose of making decision.²

In summary, evaluation is one of important processes in teaching learning process which helps teacher gathering or collecting data so that they can be measured and known whether the students have reached the instructional goals which have been stated before. Doing evaluation is a must for techers, because they can be used as the systematic gathering of information for the purpose of making decisions.

There are many ways to evaluate students. One of them is by using a test. Testing is an important part of every teaching and learning experience. It measures learning in program evaluations. An improvement in test scores shows the change in skills, knowledge, or attitude attributed to the program³. H. Douglas

¹ Airasian, Peter W., *Classroom Assesments: Concepts and Applications*, (New York; McGraw-Hills companies, inc., 2012), 7th edition, p.39

² Bachman, Lyle F., *Fundamental Considerations in Language testing*, (Oxford: Oxford University Press, 1990), p.22

³ Phillips, Patricia Pulliam and Cathy A. Stawarski, *Measurement and Evaluation series 2; Data Collection, Planning For and Collecting All Types of Data*, (San Fransisco: Pfeiffer, 2008), p.39

Brown stated that a test is a method that is used to measure person's ability, knowledge, or performance in a given domain.⁴

In other word, a test is a set of question to measure person's ability or knowledge, that each of which has a correct answer and usually the test takers answer orally or writing.

To measure the student's learning progress at school a teacher commonly administers two kinds of test; formative test and summative test. The former test is held earlier than latter test which is held at the end of semester. Through both test, a teacher can measure the students' achievement level and the degree of how far the instructional objectives of learning be accomplished by them.

Brown stated that, formative evaluation is assessing students in the process of "forming" their competencies and skills in order to help them continue that growth process. Formative assessment often implies the observation of the process of learning, as opposed to the product. On the other hand, summative evaluation occurs at the end of a lesson, unit, or course and therefore attempt to measure, or summarize, what a student has grasped. It is usually administered within a relatively short time limit. They are systematic, planned sampling techniques constructed to give teacher and student an appraisal of student achievement.⁵

Here we can infer that a test can be used as a valuable teaching device. A test must be well constructed so that it can be a good test. Making a good test needs careful arrangements. Test as an instrument has to have a good quality, because the quality of the test will influence the result of the test. If the test is good, the result will provide the right information to be used by the teacher in making accurate decision to the students' achievement.

⁴ Brown, H. Douglas, *Language Assesments, Principles and Classroom Practices*, (USA: Pearson Education, Inc., 2004) p.3

⁵ Brown, H. Douglas, *Teaching by Principles: An Interactive Approach to Language Pedagogy*, (New York: A Pearson Education Company, 2001), 2nd edition, p. 402-403

H. Douglas Brown stated that a well constructed test should have five main characteristics which involve validity, reliability, practicality, authenticity and washback. While the test can be practical if it is easy to do and administer. It can be reliable if the result of the test is the same even though the test administered to the same level students in the next time and it can be valid if it can measure what is supposed to measure. Then, authenticity is defined as a concept that is a little slippery to define, especially within the art and the science of evaluating and designing tests. Meanwhile, washback is the effect of testing in teaching and learning.⁶

Selection of appropriate language items is not enough by itself to ensure a good test. Each question needs to work properly; otherwise, it can weaken the exam. Fortunately, there are some rather simple statistical ways of checking individual items. This procedure is called item analysis.

According to H. Douglas Brown there are three main components of item analysis, they are: difficulty level, discriminating power and the effectiveness of the distractor. Level difficulty here is the extent to which an item is easy or difficulty for the purposed group of test-takers, and discriminating power means the extent to which an item differentiates between high and low ability test-takers, while effectiveness of distractors means one more important measure of a multiple-choice item's value in a test, and one that is related to item discrimination.⁷ According to Purwanto, a good test item should have three criteria; moderate difficulty level, high discriminating power and distractor analysis which works effectively.⁸

Here, the writer limits the problem of the study. She just focuses on the English Summative test for second grade of MA Uswatun Hasanah made by MGMP LP Ma'arif NU of Semarang district in the academic year 2013/2014, in

⁶ Sudijono, Anas, *Pengantar Evaluasi Pendidikan*, (Jakarta: PT Raja Grafindo Persada, 2008),p.93

⁷ Brown, H. Douglas, *Language Assesments, Principles and Classroom Practices*, (USA: Pearson Education, Inc., 2004) p.58-60

⁸ Purwanto, *Evaluasi Hasil Belajar*, (Yogyakarta: Pustaka Pelajar, 2009), p. 99

term of content validity, practicality, reliability and item analysis. In this research, the writer intends to analyze English summative test for whole items; multiple choice and essay test, because she found a problem especially in the second grade of MA Uswatun Hasanah Semarang. Almost all of students commented that the test is so difficult, then many students got low score. The writer tries to investigate about this problem and wants to know whether the test fulfills the criteria of a good test in term of their validity, reliability, practicality and their item analysis. Then, the writer wants to do the research under the title “AN ANALYSIS OF ENGLISH SUMMATIVE TEST FOR SECOND GRADE OF SENIOR HIGH SCHOOL MADE BY MGMP LP MA’ARIF NU OF SEMARANG DISTRICT AT ODD SEMESTER 2013-2014 (A Case Study at the Second Grade of MA Uswatun Hasanah Semarang in the Academic Year of 2013-2014).

B. Research Question

In this research, the researcher formulates six questions, those are:

1. How is the content validity of English summative test for second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma’arif NU of Semarang district at odd semester in the academic year 2013-2014?
2. How is the reliability of English summative test for second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma’arif NU of Semarang district at odd semester in the academic year 2013-2014?
3. How is the practicality of English summative test for second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma’arif NU of Semarang district at odd semester in the academic year 2013-2014?
4. How is the difficulty level of test items on English summative test for second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma’arif NU of Semarang district at odd semester in the academic year 2013-2014?

5. How is the discriminating power of test items on English summative test for second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma'arif of Semarang district at odd semester in the academic year 2013-2014?
6. How is the distractor analysis of test items on English summative test for second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma'arif of Semarang district at odd semester in the academic year 2013-2014?

C. Objective of the Study

Concerned with the problems formulated above, the objectives of this study are:

1. To know the content validity of English Summative Test for the second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma'arif NU of Semarang district at the odd semester 2013/2014.
2. To know the reliability of English Summative Test for the second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma'arif NU of Semarang district at the odd semester 2013/2014.
3. To know the practicality of English Summative Test for the second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma'arif NU of Semarang district at the odd semester 2013/2014.
4. To know the difficulty level of test items on English Summative Test for the second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma'arif NU of Semarang district at the odd semester 2013/2014.
5. To know the discriminating power of test items on English Summative Test for the second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma'arif NU of Semarang district at the odd semester 2013/2014.
6. To know the distractor analysis of test items on English Summative Test for the second grade of MA Uswatun Hasanah Semarang made by MGMP LP Ma'arif NU of Semarang district at the odd semester 2013/2014.

D. Pedagogical Significance

The purposes of this research are: Firstly, it informs for test makers or classroom teachers about the quality of multiple choice items test. Through this research, they can know the good items for the future used and the students' achievement in mastering the materials taught in order to evaluate the teacher's competence in teaching. Besides that, it is expected to give contributions or a useful input and feedback as bases for improving English Summative Test.

Secondly, it provides with the feedback to the writer especially, and the English teacher of how to analyze the test items in term of difficulty level.

Finally, the result of this study can be used as a reference for further research.