

CHAPTER I

INTRODUCTION

A. Background of The Research

To measure learners' ability or performances in learning is using a test. It is usually conducted in certain period of time or in the class after getting the input from the teacher. The result of the test is to reflect learners' capacity, ability, and performance for further learning. According to Brown (2010), the test is subset of assessment, a genre of assessment techniques. The test is prepared by administrative procedures that occur in identifiable time in the curriculum. In language test, test has to qualify the criteria of good and useful test. A good test has to require the principles of language assessment to be considered as the most important aspect in testing and assessment. Brown (2010) explains that there are five major principles of language assessment. They are validity, reliability, practicality, authenticity, and backwash.

The first principle is validity. Validity means a test is to measure the objective and purpose of the material. There are five types validity. They are content validity, construct validity, face validity, criterion validity, and consequential validity. In content validity, the test is constructed to contain a representative sample of course or relate to what materials have been taught in class or according to the syllabus of the course. Construct-related validity or construct validity relates to any theory, hypothesis, or model that explains observed phenomena in our universe of perceptions towards language. Face validity refers to what the test item looks like familiar with the test takers, well-designed, and good format. Therefore, the test is appropriately finished and there is no misunderstanding to answer it. The criterion validity refers to any criteria that the test has actually been reached. The consequential validity emphasizes the consequences of the test before, during, and after the test conducted for teachers or learners.

Reliability refers to the consistency and the dependency of the test. If the examiner gives the same test to the student or matched students on two different conditions, the test is to show similar results. For instance, the research is conducted by Shahivand, Paziresh, and Raezadeh (2014). The study is an attempt to investigate whether different test formats had any significant effect on the performances of Iranian language students. As it is mentioned that one or two different test formats cannot be assumed as reliable measurement of the learners' language knowledge. Basically, the reliability of a test is enhanced when the learners' language knowledge be

measured regarding to different test formats, thus a fine judgment of their language abilities can be made.

Meanwhile, practicality refers to the budgetary limits. The test is practical if it uses efficient money to make the test. For example, the test by using a computer on the daily quiz needs a lot money rather than paper based. Therefore, the test of a computer used is less practical in budget than paper used. In addition, practicality refers to the administrative issues involved in making, giving, and scoring and assessment instrument or testing. For example, the TOEFL test is more practical to administer rather than the IELTS test. TOEFL is all in multiple-choice format while IELTS has many other test items such as gap-filling and rearranging task. Thus, IELTS takes more time to check.

The authenticity reflects how the test is to be natural to the real world. For example, the test of reading needs to replicate the real world task. The reading text for elementary can be taken from the shopping list, label, etc. Then, for higher level, the source of reading text can be taken from newspaper, magazine, etc. Thus, those sources have authenticity.

Backwash in language assessment is to find out the impact of test for the learner and teacher for further learning and teaching development. The teacher can give feedback after the test. By doing so, the students know their weaknesses and improve their learning. Besides, for the teachers, the backwash is important to evaluate their teaching method to be better.

There are two similar studies focusing in principle of language assessment. The first previous study is by Razi (2012) from Çanakkale Onsekiz Mart University, Turkey that focuses on establishing validity and reliability along with readability and vocabulary frequency of a 32-item reading test which was conducted at Dokuz Eylül University and Çanakkale Onsekiz Mart University with three colleagues and 100 undergraduate students. In validity, the reading test is valid in terms of its content, face, and criterion-related test. Then, for the reliability, all the items in the reading test, except from the items 25 and 29 are appropriate to be used in the test. Therefore, the 30-question reading test is acceptably reliable.

The second previous study is by Sims (2015). The study focuses on presenting the procedures a university followed to create a language proficiency exam with an appropriate validity, high reliability, and strong correlations to establish standardized exams in Taiwan. The researcher analyzes the exam named the New English Placement Exam (NEPE) which is conducted by English Language Center, Tunghai University, Taiwan. In short, the correlation results imply that the NEPE has strong concurrent validity similar to the Test of English for International Communication (TOEIC) and a strong to moderate concurrent validity to Intermediate level of General English Proficiency Test (GEPT).

Those previous studies mostly take part in analyzing the tests to find out the validity and reliability of those tests. The present study finds out the validity and the authenticity of the English final tests X and XI grade *SMK*

DKI Jakarta academic year 2015/2016. In the English final tests *SMK* DKI Jakarta, there are two sections which are listening and reading section. This study only takes reading section. The reading section of the English final tests X and XI grade *SMK* DKI Jakarta consists of rearrangement item, dialog completion, grammar, and reading comprehension. The research of test evaluation is important. For the teachers, it is used to evaluate the teaching method and materials for further learning. For the students, a better quality of test becomes a better measurement. Thus, the result of students' learning using a good test will be more objective.

B. Statement of Problems

This study focuses on the analysis of the English final tests X and XI grade *SMK* DKI Jakarta academic year 2015/2016. The statements of the problem area:

1. How is content validity of the English final tests X and XI grade *SMK* DKI Jakarta academic year 2015/2016?
2. How is construct validity of the English final tests X and XI grade *SMK* DKI Jakarta academic year 2015/2016?
3. How is authenticity of the English final tests X and XI grade *SMK* DKI Jakarta academic year 2015/2016?

C. Research Objectives

1. To analyze the English final tests X and XI grade *SMK* DKI Jakarta academic year 2015/2016 whether it is valid or not based on content validity
2. To analyze the English final tests X and XI grade *SMK* DKI Jakarta academic year 2015/2016 whether it is valid or not based on construct validity
3. To analyze the authenticity of the English final tests X and XI grade *SMK* DKI Jakarta academic year 2015/2016

D. Scope and Limitation

The study analyzes the text of the English final tests X and XI grade *SMK* DKI Jakarta academic year 2015/2016. There are two sections of the English final tests X and XI grade *SMK* DKI Jakarta academic year 2015/2016. They are listening and reading section. This study only takes reading section. It is because the listening test requires tape script or listening material that the school does not allow the outsiders to use them. There are 40 questions consisting of rearrangement, dialog completion, and grammar. Then, 30 questions are from 10 reading texts. The writer analyzes two types of validity which are content validity and construct validity, and the authenticity.

E. Research Methodology

1. Method