CHAPTER 1

This chapter presents an overview of the study. It consists of the background of the research, research questions, research objectives, research significances, research framework, and previous studies.

A. Background

In education, assessment has a very important role, especially to provide feedback in the learning and teaching processes, and enables to review and improve the whole process (Haladyna, 2002). A variety of tools and techniques is commonly used in assessing students' learning, such as assignments, tests, essays, portfolios, projects or oral examinations. In the testing area, Multiple Choice Questions (MCQs) are often chosen by teachers as a tool to assess reading comprehension ability as a result of student learning. This is due to the large number of students and heavy schedule of teachers so this MCQ is easy to rating in a short time and provides objective results.

The questions must be about good quality of an assessment instrument in order to determine the level of learning achievement of students. Purwanto (2012) states that the standard of the test must be considered during the assessment process, as poor test questions would provide an effect to the quality of education. Item (question) analysis is one method that can be used. It is a method that evaluates learner's answers to specific test items in order to evaluate the content of those items and the test as a whole (Kocdar et al., 2016). Daryanto (2012) contends that the validity, reliability, level of difficulty, item discrimination, and effectiveness of distractors can all be considered in the analysis of item quality.

Although multiple choice items are often used in reading test in vocational school in Tasikmalaya, there is no evidence that the quality of the questions in this school is good enough. Based on the results of interviews with the eleven grade English teacher at one of the vocational schools in Tasikmalaya, it was found that the teacher had never assessed the quality of the

questions before. In addition, the scores of students in these schools are mostly below average. This is due to a lack of knowledge about how to assess the quality of the question.

Several researchers have conducted research that focuses on analysis of the quality of Multiple Choice Questions (MCQs). First, Al-Ariqi's (2019) research, namely "The Analysis of MCQs in a Newly Developed Reading Comprehension Ability Test: A Study on Yemeni Undergraduate EFL Learners". Second, a research by Kocdar et al. (2016) that is analysis of the difficulty and discrimination indices of MCQs according to cognitive levels in an open and distance learning context. Third, a research was conducted by Odukoya et al. (2017) with the title "Item Analysis of university-Wide Multiple Choice Objective Examinations: The Experience of a Nigerian Private University". However, this research is different from previous research. While the previous research only focused on item facility (IF), or item discrimination (ID), or distractor efficiency (DE) or by comparing them all, this research should focus to examine all items thoroughly. This study also looked at validity and reliability to determine the overall quality of the questions. This research was conducted in one of the vocational schools in Tasikamalaya Indonesia, while the previous research was conducted at universities.

B. Research Questions

There are several research questions regarding the problem mentioned in the background:

1. How is the quality of the MCQs on the English test in terms of Validity and Reliability?

2. How is the quality of the MCQs on the English test in terms of the Item Facility (IF), Item Discrimination (ID), and Distractor Eficiency (DE)?

C. Research Objectives

To respect the research questions, this research is aimed to:

1. To identify validity and reliability of the multiple choice English test.

2. To find out the quality of MCQs on the English test in terms of the Item Facility (IF), Item Discrimination (ID), and Distractor Eficiency (DE).

D. Research Significances

Based on two points along these lines, this study is expected to have some significance:

1. Theoretical Significances

The result of this research is expected to give a useful contribution to English learning teaching especially in the assessment of English exam questions. Moreover, to make teachers are aware of the importance of quality MCQs in the English test.

2. Practical Significances

This research is aimed for teacher and educational policymakers. Teachers can use this research as an additional reference to construct good questions and to help them reach the curriculum goals. Furthermore, this is directed at educational policymakers to think about the policies they create for the development of education in Indonesia. In the area of constructing queries, consideration is required. As a result, teachers will improve their ability to construct good questions.

E. Research Framework

An assessment is one tool that can help with the teaching and learning process. Teachers will learn about their students' skills by assessing their

performance and testing their knowledge. Assessment serves a variety of functions, including providing information about students' learning and growth, as well as teaching quality and program accountability (Opre, 2015). Assessment is an important aspect of the teaching and learning process. It determines where learners are now and what level they have reached, provides feedback on their learning, diagnoses learners' developmental needs, and allows curriculum, materials, and activities to be planned (Alderson, 2005 in Sahinkarakas, 2012).

In general, assessment and evaluation should be carried out by every EFL teacher. Evaluation is a process of giving meaning and value or quality about an object being evaluated (Joseph, 2015). Based on Patton (2002) the term "evaluation" refers to the method of assessing a program objectively. It entails gathering and analyzing data about the activities, characteristics, and outcomes of a program. Its aim is to make observations about a program in order to increase its efficacy and/or make decisions about information programming. Measurement and assessment activities are used in evaluation.

An MCQ consists of a stem with a question line underneath it, usually followed by three to five options. According to Cizek and O'Day (1994), one of the alternatives is the key, which is the right or acceptable response, while the others are referred to as distractors. Distractors are designed to draw students who do not know the correct answer, whereas those who do are expected to avoid them. If the incorrect choices are intended to expose common misconceptions, MCQ tests can be used to verify students' difficulties, and they can provide a more thorough sample of the subject material due to the wider coverage.

Analyzing test items is important for identifying question categories and determining the quality of each multiple choice question. Ahiri (in Badara, 2016) stated that test item analysis aims to improve test quality by revising or removing ineffective items and identifying diagnostic details on each student's understanding of the material being taught.

Item facility (IF), item discrimination (ID), and distractor efficiency (DE) are three indicators that can help with the proper selection and placement of acceptable multiple-choice items on a test (Brown, 2010). They are required for standardized norm-referenced examinations that should be administered multiple times and/or in various forms. The difficulty levels of the question for the proposed group of test-takers is known as item facility (Khairani and Shamsuddin, 2016). Its purpose is to judge whether the question is easy, moderate, or challenging. Good questions are those with moderate difficulty, not too easy and not too difficult.

Item discrimination, according to Zajda (2006), is the degree to which a test item distinguishes between high and low ability test takers. An item that scored equally well for high ability students (those who did well on the test) and low ability students (those who did not) would have weak ID since it did not differentiate between the two groups. On the other hand, items that received correct answers from the majority of the high-ability group but wrong answers from the majority of the low-ability group had good discriminatory power. Meanwhile, distractor efficiency is the effectiveness of a distractor in a certain situation. Distractor efficiency, according to Hartati and Yogi (2019), is another important measure of a multiple choice item's worth in a test and one that is linked to item discrimination. It is used to look at possibilities that are not the correct response to a query. Distractor efficiency are those that are not part of the solution key or the correct answer.

According to Sudijono (2012), when analyzing item quality, validity and reliability must also be considered. The validity of a test instrument as a tool for measuring learning outcomes demonstrates its accuracy. A test is said to be valid if it can measure the object being measured and does so in accordance with certain criteria. Meanwhile, reliability refers to an instrument's level of consistency. A test is considered reliable if it consistently produces the same results when tested on the same group at different times and on different subjects (Zainal Arifin, 2012).

F. Previous Study

There are various studies on assessing the quality of MCQs. First, a research was done by Al-Ariqi et al. (2019). This research was conducted among 134 second year Yemeni EFL students in Sana'a University, Yemen. There were 20 multiple choice questions that were analyzed to find the relationship between IF and ID, and to find out DE. The result is for 19 items categorized as "good" with percentage IF from 20% to 90%, DE $\geq$ 0.40. Overall 75% items had two non-functional distractors (NFDs), while 20% items had three functional distractors and 5% had only one functional distractor.

Second, the research of Kocdar et al. (2016) was to examine whether there was a significant difference between IF and ID on multiple choice questions according to cognitive levels. This research found that there was no substantial difference in the discrimination indices of the remembering and understanding-level questions ($p = 0.378 > 0.05$), but there was a significant difference in the discrimination indices of the remembering and applying level questions ($p = 0.003 < 0.05$) and the understanding and applying level questions ($p = .048 < 05$).

Third, research by Odukoya et al. (2017) was conducted at private universities in Nigeria with undergraduate participants. This study discussed the difficulty level (IF) of the questions and the effectiveness of the distractor (DE) with using secondary data and the ex-post facto as the research method. Almost all cases (between 65 and 97 percent of the 70 items fielded in each course) did not meet psychometric criteria in terms of difficulty and distractive indexes, and consequently needed to be moderated or deleted.

However, from all of the previous studies there were significant differences with this study. None of the studies discussed item facility (IF), item discrimination (ID), and distractor efficiency (DE) simultaneously in detail and thoroughly. The validity and reliability of the questions were also

examined in this study to determine the overall quality of the questions. So, this was used by the researcher as a gap in the research. In addition, all of the research was carried out in universities but here the researcher was conducted research in vocational high schools in Indonesia.