CHAPTER I

INTRODUCTION

1.1 Research Background

The development of science and technology has brought big changes in human life and brought people to global competition. The challenges and development of education in Indonesia at present and in the future will be even bigger and more complex. This is due to changes in community demands for the quality and quantity of education itself. Education is one of the important sectors in development in every country. According to Constitution Number 20 of 2003, education is a conscious and planned effort to develop all the potential that students have through the learning process (Amalia & Widayati, 2012: 2).

Mathematics is also developing and continuing to support the development of science, technology, business dan goverment (Minarni & Napitupulu, 2020 : 3). Mathematics as a field of research taught in formal educational institutions is an important part of efforts to improve the quality of education. Mathematics is a subject that deals with many concepts. Concept is an abstract idea by which we can group objects into examples or not examples. Concepts in mathematics are related to one another. The interrelationship between one material concept and another is evidence of the importance of understanding mathematical concepts. Therefore, students cannot understand material if they do not understand the previous material or prerequisite material from the material to be studied (Novitasari, 2016: 8).

The success of learning process can be reflected in the results obtained by the participants which can be seen from the results of the evaluation carried out by the teacher. Learning evaluation plays a very important role in providing an overview of the success of the learning carried out. The form of evaluation is very crucial to pay attention to because the evaluation results are well and badly influenced by the form of evaluation used. Evaluations that meet various rules and requirements in evaluation can provide a picture of the actual success of learning (Siswanto, 2006: 60).

The process towards a predetermined goal or success is strongly influenced by several factors, including teacher factors towards students in the teaching and learning process in class, students who have different characteristics, students' interest in a lesson, learning activities, strategies the use of learning methods, available facilities, an atmosphere of evaluation as well as evaluation tools and evaluation materials are good so that students get good and poor learning outcomes. The teacher plays a role in preparing the evaluation in the form of tests. So every teacher is required to have the responsibility in planning and implementing evaluations (Supandi &, Farikhah 2016: 71-72).

Changes in students are known from the evaluation (assessment) of the teaching and learning process. The results of tests or evaluations are measuring tools commonly used to determine students' understanding of the material that has been delivered. In addition, from the questions used, it can be seen whether the questions can measure the curriculum objectives that have been set or not, so that the results can be used as a benchmark for the implementation of learning objectives (Hamimi, et al., 2020: 58).

Evaluation is very important and must be considered in the learning process. However, many teachers have not been able to choose a good evaluation tool. In general, the teacher-made test in designing questions, did not pay attention or did not analyze the test items so that most of them could not identify the good, mediocre, and bad questions. Apart from analyzing the questions, the tests that are arranged must also meet the requirements or characteristics of good test quality (Supandi & Farikhah, 2016: 72).

In general, assessment tools (instruments) can be categorized into two forms, namely: 1) Test; and 2) Not a test (non-test). Measurement tools included in the non-test category are: a) Questionnaires; b) Interview; c) Match List (check list); d) Observation; e) Assignment; f) Portfolio; g) Journal; h) Inventory; i) Self-assessment; j) Peer assessment. Whereas a test is a number of questions that must be answered, or questions that must be selected or responded to, tasks that must be done by the person being tested at a certain time. The test is a number of questions that have right or wrong answers, questions that require answers or be given a

response to measure a person's level of ability in certain aspects (Wening as cited in Kholis, 2017: 307-308).

The test is a measuring tool most often used by teachers to measure student learning outcomes. The teacher can find out the extent to which the predetermined goals can be achieved from the test results that have been obtained. Therefore, in order for a test to measure the results correctly, it must be properly developed. A new test will be meaningful if it consists of items that test important objectives and represent all the materials being tested (Purwanti, 2014: 82).

The essential concern of test designers when building a test is the nature and quality of test things and how examinees react to these things. The legitimacy and the unwavering quality of any test depend eventually on the characteristics of its things. These characteristics are item difficulty and item discrimination. Test theories empower the forecast of results of tests by distinguishing parameters of item difficulty, item discrimination, and the capacity of test-takers (Awopeju & Afolabi, 2016:265).

According to Safitri et al. (2019: 93), the objectives of the item analysis are: 1) to research and examine each item in order to obtain a quality question before the question is used, 2) to help improve the test through revision or to remove ineffective questions, 3) to know information diagnostic for students, whether they have or not the material that has been taught. But in fact, according to Muhammad Nasir et al (as cited in Nasir, 2015 : 338), there are many teachers in schools do not analyze the items for evaluation because: (1) lack of time to do the analysis, (2) analysis of the items is still difficult because it involves formulas that complete and tiring, and (3) the absence of tools that make it easier to perform item analysis.

Thus, an effort to find out whether the questions made by the teacher are classified as appropriate and good, and provide maximum results in measuring and increasing the level of student understanding, analysis can be carried out on each item (Sudjana as cited in Rahayu et al., 2014:40). Items analysis that carried out will be able to improve the quality of the questions through elements of validity, reliability, difficulty level, discrimination power, and effectiveness of

distractor (Salmina & Adyansyah, 2017:38). Analysis of validity and reliability can be used to determine the quality of the items as a whole, while the analysis of difficulty level, discrimination power and effectiveness of distractor are used to determine the quality of the items. Analysis of difficulty level and discrimination power can be used to measure the quality of objective items and descriptions (Rahayu & Djazari, 2016:86). This is guided by two most frequently used item analysis theories, namely Classical Test Theory (CTT) and Item Response Theory (IRT) as presented by Siri & Freddano (2011:189) and Frey (2017:1).

Based on this explanation, it can be said that the success of learning can be seen through the evaluation of learning (especially the evaluation of student learning outcomes). One way to find out student learning outcomes is to carry out a test. Therefore, the test that is made must be objective and of quality so that the teacher can provide appropriate follow-up based on the student's learning outcomes. So a test analysis (item analysis) is needed to determine the level of objectivity and quality.

A more in-depth research is needed regarding the role of Classical Test Theory and Item Response Theory in item analysis. It aims to improve the quality of learning evaluation. Where Systematic Literature Review (SLR) is one method that can be used to examine more deeply about the topic. Therefore, a more in-depth study is needed of the results of previous studies regarding the role of Classical Test Theory and Item Response Theory in item analysis to determine the quality of mathematics tests. Thus, this systematic literature review research can complement the knowledge and address gaps in the literature. This research will present a configuration and conceptual framework for item analysis to determine the quality of mathematics tests related to the role of Classical Test Theory and Item Response Theory for further researches.

Systematic reviews will be very useful in synthesizing or cumulative various relevant research results, so that the facts presented to policy makers become more comprehensive and balanced (Siswanto as cited in Hadi et al., 2020: 7-8). This research will also be able to find out what kind of research should be done in the future.

Systematic review's very nature means that they tend to be of higher quality, more comprehensive, and less biased than other types of literature review, making them more likely to be published than other types of literature review. If done well, a systematic review is clearly a novel and important substantive contribution to knowledge in its own right (Siddaway at al., 2019:9-10). The point is systematic review is very useful.

It is in line with the opinion of Kitchenham (2004:1-2) about the common reasons for conducting a systematic literature review, namely : (1) to summarise the existing proof concerning a treatment or technology, (2) to recognise any gaps in current research in arrange to recomment area for further investigation, and (3) to provide a framework/background in arrange to position new research activities appropriately.

Based on this background, the researcher will conduct a systematic literature review on test item analysis to determine the quality of mathematics tests which focused on the role of Classical Test Theory and Item Response Theory. This systematic literature review focuses on national journals accredited by SINTA. Later this systematic literature review can be used as a source of information (reference) about previous researches on test item analysis to find out mathematics test quality then provide ideas for future researches.

1.2 Problem Identification

Based on the research background, several problems can be identified as follows.

1. Many teachers have not been able to choose a good evaluation tool.
2. The teachers did not pay attention or did not analyze the test items when they made the test so that most of them could not identify the good, mediocre, and bad questions.
3. The need for a more in-depth research of the role of Classical Test Theory and Item Response Theory in item analysis.
4. The need for a more in-depth research of the previous research results regarding the role of Classical Test Theory and Item Response Theory in item analysis to determine the quality of mathematics tests.

1.3 Scope of Problems

Based on the problem identification, this research is limited to these following problem:

1. The need for a more in-depth research of the role of Classical Test Theory and Item Response Theory in item analysis.
2. The need for a more in-depth research of the previous research results regarding the role of Classical Test Theory and Item Response Theory in item analysis to determine the quality of mathematics tests.

1.4 Research Questions

Based on the research identification and scope of problems, the research question in this research can be made as follows.

1. How is the role of Classical Test Theory and Item Response Theory in item analysis?
2. How is the role of Classical Test Theory and Item Response Theory on the previous research results on item analysis to determine the quality of mathematics tests?

1.5 Research Objectives

Based on the formulation of the problem, the research objectives to be achieved are as follows.

1. To find out the role of Classical Test Theory and Item Response Theory in item analysis.
2. To find out the role of Classical Test Theory and Item Response Theory on the previous research results regarding item analysis to determine the quality of mathematics tests.

1.6 Research Benefits

The implementation of this research is expected to provide the following benefits.

1. For Teachers

    This research can make teachers aware of the importance of learning evaluation, especially item analysis. Therefore, mathematics teachers will pay more attention to the importance of item test analysis.

2. For Students

    By creating the quality tests through the analysis of these items, students will know more about their actual abilities and get appropriate follow-up from the teacher.

3. For Schools and Educational Agency

    This research can be used as a reference for evaluating student learning in schools so that learning strategies can be set that can meet educational standards. In the end, students who are competent and able to compete at school and outside of school will be formed.

4. For the Researcher

    This research is useful for researchers as prospective mathematics teachers to examine more deeply about the evaluation of mathematics learning through item analysis. This can be a motivation for researchers to further develop the concept of item analysis that can be implemented in present and future learning.

5. For Next Researcher

    The results of this research are expected to be a source of information that summarizes previous researches on test item analysis to determine mathematics test quality which can support further researches.

1.7 Operational Definitions

The operational definitions of this research are as follows.

1. Item analysis is an activity that must be carried out to determine the quality of test items based on certain aspects, namely: validity, reliability, difficulty level, discrimination power and effectiveness of distractor. There are two most frequently used item analysis theories, namely Classical Test Theory (CTT) and Item Response Theory (IRT).

2. Systematic literature review also known as systematic review is a systematic method of collecting, critically evaluating, integrating and presenting findings from various researches on a particular topic.