A QUALITY ANALYSIS OF ENGLISH FINAL TEST FOR THIRD GRADE STUDENT AT MTs DARUL ULUM KARANGPANDAN REJOSO PASURUAN

Ulfah Setyawati Sugih Arto Putri, Diah Anita Pusparini
English Education Departement STKIP STIT PGRI Pasuruan
ulfahsetyawatisap@gmail.com

Abstract: To analyze the contents validity, reliability, the item difficulty, the discriminating power, the item distractor of the final test for the third grade students of MTS Darul Ulum Karangpandan Rejoso Pasuruan.

The research design which is applied in this research is a qualitative focused on content analysis and descriptive quantitative. Content analysis to analyze the content validity and descriptive quantitative to analyze the reliability, the item difficulty, the discriminating power, the item distractor. The subject of the study are the third grade students of MTS Darul Ulum Karangpandan, Rejoso Pasuruan, there are 4 classes for the third grade in this school, it consist of 2 classes of man and 2 classes of woman. The instrument use to collect the data were: documentation and human instrumen. Based on the finding, there were 50 items which are related to the syllabus (KTSP) curriculum in seven grade, eight grade and nine grade. So the content validity is good. From the score of coefficient reliability by using K-R 20, the reliability of the English final test is low because the result is 0.37. The result of the item difficulty is good because most of the items is moderate. The final test has a easy level it means that the final test has a poor discriminating power. Item distractor is efficient because many item are excellent distractor. The conclusion from the analysis: content validity is good, reliability is low, the item difficulty is good, discriminating power is poor and item distractor is efficient.

Keywords: content validity, discriminating power, item difficulty, item distractor, reliability

In junior high school, senior high school and vocational school, English is one of the core subjects that has to be thought. It is hoped that English can help students for self-development in the field of science, technology and art culture (Depdikbud, 1993:1). That is why, English becomes one of the essential lessons at school. It is very important for the students to master English as well as other objects.

In teaching learning process the teacher should know the development of the students. To know the development of the students, the teacher have to give the students’ evaluation. According to Rogers and Badham (2005:2), evaluation is the process of systematically collecting and analyzing information in order to form value judgements based on firm evidence. Evaluation is an important part of every teaching and learning activities. It gives many contribution for teaching and it provides an information about the students’ progress which can be used by the teacher to manage learning task and students. Though evaluation teacher will know the students’ achievement in mastering the lesson efficiency.

Meanwhile, Heaton (2006:1) states that both testing and teaching are so closely interrelated that is virtually impossible to work in either field without being constantly concerned with the other. In other words, testing is important for both teacher and students.

According to Hughes (2006:28), making test in not easy, because the test result is used as information about the students’ level achievement in mastering certain materials and also about the successfulness of a teaching and learning program. Hughes (2006:29) also adds that making test should be clarified at the outset that if the items, which are the building blocks of a test, meet the criteria that have been introduced before, the whole test will be most likely acceptable. However, Hughes (2006:29) regrads to points that good items will necessarily produce a good test may not always come true. Therefore
Hughes concluded that a good test is needed for getting accurate information.

According to Kakajani (2007:12), the characteristic of good test are valid, reliable, objective, comprehensive, simple and scorable. Kakajani (2007:12) elaborates that a test is considered as valid when it measures what it is supposed to measure. Then she also states that validity is arguably the most important criteria for the quality of a test. In this term she also adds that the term validity refers to whether or not the test measures what it calims to measure. So, Kakajani (2007:12) regards to points that on a test with high validity the items will be closely linked to the test’s intended focus.

The validity of a test is the extent to which the test measures accurately what it is intended to measure. There are several ways to decide whether the test has sufficient validity, they are content validity, criterion-related validity, and construct validity (Hughes, 2003:26). In this case, the reasercher use content validity as the test measurement. According to Supranata (2006:51), content validity is used for educational assesment. It is also calles curriculum validity which implies a validity based on the curriculum content.

Other characteristic is reliability. According to Brown (2003:20), reliability test is consistent and deendable. It is mean that to be reliable, a test must be stable and consistent in its measurements. If the teacher gives the same test to same students or matched subjects in different occasions, the test should yield similar result. While, Arikunto (2003:87) states that a reliable measure in one that provides consistent and stable indication of the characteristic being investigated.

Besides of stating the criteria of a good test, Kakajani (2007:15), also mentions that a tesy which has poor validity will not measure the job-related content and competencies it ought to. In this case, there is no justification for using the test result for their intended purpose. And Kakajani (2007:15) stated that a test with poor reliability, on the other hand, might result in very different scores for the examiner across the two test administrations. If a test yields inconsistent scores, it may be unethical to take any substantive actions on the basis of the test.

Based on the fact that had already mentioned above and realizing the importance of a test, the reasercher tries to analyze the English final test at MTS Darul Ulum Karangpandan Rejoso Pasuruan of third grade students, because the researcher want to know the quality of final test and that will determine whether the students eligible to graduate or not, so the test must be valid and reliable. The reasons of why researcher chooses this school because of the school have designed the own English final test and it has never done a research yet there and for the effectiveness of doing the research.

Based on the reason above, the researcher conducts this research focusing on the content validity, realibility, item difficulty, descriminating power and item distractor of English try out test of the final test for the third grade students of MTS Darul Ulum Karangpandan Rejoso Pasuruan.

REVIEWOF RELATED LITERATURE

Definition of Test

According to Danuri (2012), test a method to determine a student’s ability to complete tasks or demonstrate mastery of a skill or knowledge of content. Those are several definitions of the test by the experts although, those definitions are created in different sentence, but it shows the same point. In other words test is one of instrument which can be used for collecting some information about the strengths and weaknesses of students and measuring the students’ achievement in learning process.

The Purpose of Language Test

According to Sudjiono (2008) in general test has two function:

a. As a tool to measure the student’s capability. The test serves to measure the level of development or progress that has been achieved by learners after they go through the process of teaching and learning within a certain period.

b. As a measure the success of teaching program, because through the test will be known how far the program teaching that has been determined, it have been achieved.

Type of Test

According to Hughes (2003), there are many types of test used to measure students’ achievement. There are four types of tests: (a) proficiency test; (b) placement test; (c) diagnostic test; and (d) achievement test.

a. Proficiency Test

According to Hughes (2003), proficiency tests are designed to measure people’s ability in a language, regardless of any training they may have had in that language. According to Brown (2001), the content of a proficiency test is not based on the content or objectives of language courses that people taking but rather it is based...
on a specification of what candidates have to be able to do in the language in order to be considered proficient.

b. Placement Test
According to Brown (2004:56), placement test is administered to assign students to the classes at different level. So, this test designed to determine the students’ performance at the beginning of instruction and used to know the ability level of the students.

c. Diagnostic Test
Diagnostic tests measure students’ understanding of a subject area or skill base. According to Hughes (2003), diagnostic tests are used to identify learners’ strengths and weakness. According to Brown (2004:47), diagnostic test is designed to diagnose specified aspects of a language.

d. Achievement Test
The purpose of achievement testing is to measure some aspect of the intellectual competence of the students, what the student has learned to know or to do. Teachers use achievement tests to measure the attainments if their students. According to Hughes (2003), achievement tests are directly related to language courses, their purpose being to establish how successful individual students, group of students, or the courses themselves have been in achieving objectives.

Type of Test Based on Test Maker
Test based on the test maker can be divided into standardized test and teacher made test.

1. Standardized Test
Standardized test is a test that is administered, scored, and interpreted in the same way for all test-takers. According to Brown (2004:78), test that presupposes certain standard objectives, or certain that are held constant accross one of the test to another. There are some types of assessment used in school system and commonly used standardized tests. Such as:

A. Aptitude tests
An aptitude test is any type of assessment that evaluates the talent/ability/potential to perform a certain task, with no prior knowledge and/or training. According to Hatfield (2016) an aptitude test measure the ability of a student to acquire a set of skills or training by measuring the student’s natural talents.

B. Inteligency test
Intelligence tests are used to determine a student’s mental ability. According to Reynolds (2009), a standarized test used to establish an intelligence level rating by measuring a subject’s ability to form concepts, solve problems, acquire information, reason, and perform other intellectual operations.

C. Achievement test
Achievement tests assess what students should have learned throughout the school year based on their grade. According to Hughes (2003), achievement tests are directly related to language courses, their purpose being to establish how successful individual students, group of students, or the courses themselves have been in achieving objectives.

2. Teacher made test
Simply teacher made test is the test that made by the teacher of classroom itself. According to Arikunto in Minawati (2003), the purpose of the test to measure the degree of students “success in achieving instructional objectives of particular lesson or course are taught in the classroom”.

Types of Test Items
According to Arikunto (2008), there are two types of test items. Both of them are based on the scoring criteria while the type of test items are subjective and objective test.

a. Subjective Test
Subjective test is the test that permite the student’s to organize and present an original answer. According to Crocker (2009), subjective tests are more commonly called constructed response items.

b. Objective Test
According to Jhonson (2015) objective, which require students to select to correct response from several alternatives or to supply a word or short phrase to answer a question or complete a statement. For example: multiple choice, true-false, matching, completion.

According to Heaton (2010:12), there are several kinds of objective test which are usually used:

A. Multiple Choice Test Item
According to Brame (2016), multiple choice test items have several potential advantages: (1)
Versatility. Multiple choice test items can be written to assess various levels of learning outcomes, from basic recall to application, analysis and evaluation. Because students are choosing from a set of potential answers, however, there are obvious limits on what can be tested with multiple choice items. (2) Reliability, reliability is defined as the degree to which a test consistently measures a learning outcome. Multiple choice test items are less susceptible to guessing than True/False question, making them a more reliable means of assessment. (3) Validity is the degree to which a test measure the learning outcomes it purports to measure.

B. True-False Items
Alternative choice in which the student indicates whether each of several statements is true or false. According to Sudweeks (2011 in Albanese, 2013:13), a true-false test is a test which asks the students to judge the accuracy of a statement. He also states that this test can be used in any level of education. This is because students will know how to answer each question so that they are going to learn about their mistakes.

C. Matching Test Items
According to Sudweeks (2011 in Albanese, 2013:14) states that matching test items are consisting of series of question and series of answers. He adds that each question has answer in series of answers. The students look for the answer and place it which is match with the question.

D. Short Answer Items
Tuckman (2007:13), defines that short answer tests shows the empty space which provided for the tasted to write the answers according to the direction.

Example: Complete the sentence by writing an appropriate form of the verb that is given in parentheses.
She...................................(go) to market by motorcycle.

Criteria of Good Test
Based of Brown (2004), there are five criteria for testing a test:

1. Practically, an effective test is practical. This means that it is no excessively expensive, stays within appropriate time constraints, relatively easy to administer and has a scoring/evaluation procedure that is specific and time-efficient.

2. Reliability, a reliability test is consistent and dependable.

3. Validity, the extent to which inferences made from assessment result are appropriate, meaningful and useful in terms of the purpose of the assessment.

4. Authenticity, the degree of correspondence of the characteristic of a given language test task to the feature of a target language task.

5. Washback, the effect of testing on teaching and learning.

Validity
According to Hughes (2003), Validity refers to whether test measures accurately what it is intended to measure. There are several ways to decide whether the test has sufficient validity or not, they are:

1. Content Validity
According to Hughes (2003:27), a test is said to have content validity if it represents all materials taught in the class, it means the relationships between the test items and the course objectives are apparent.

2. Criterion-Related Validity
According to Phelan and Wren (2005), criterion-related validity is used to predict future or current performance, it correlates test results with another criterion of interest.

3. Construct Validity
According to Pope (2009), construct validity refers to whether how well an assessment, or topics within an assessment, measure the educational/psychological constructs that the assessment was designed to measure. For example, if the construct to be measured is “sales knowledge and skills” construct.

Reliability
According to Brogan (2009), reliability is the consistency of a measure. In educational testing, reliability refers to the confidence that the test score will be the same across repeated administrations of the test.

According to Djiwandoro in Tantowi (2015) there are several ways in order to know the test reliable or not, namely 1) Test retest, 2) Split-half method, 3) Cronbach Alpha Method

Item Difficulty
In addition, Brown (2004:59), stated that the formula that can be used to measure the discrimination power of multiple-choice test items is:
Discriminating Power

Discriminating power is to distinguish the test taker into 2 groups. According to Hughes (2003), it is the extent to which an item differentiates between high and low-ability take-takers. The formula that can be used to measure the discrimination power of multiple-choice test items is:

Item Distractor

The step to measure item distractor:
1. Making tabulation of students’ answer choice
2. Analyzing the item distractor

The formula to measure item distractor:

RESEARCH METHOD

The research design which is applied in this research is a qualitative focused on content analysis and descriptive quantitative. Content analysis to analyze the content validity and descriptive quantitative to analyze the reliability, the item difficulty, the discriminating power, the item distractor. Content analysis is one of several qualitative methods currently available for analyzing data and interpreting its meaning (Schreier, 2012). The object of content analyzing can basically be any kind of recorded communication, i.e. transcripts of interviews / discourses, protocols of observation, video tapes, written documents in general etc. (Mayring, 2000b, 468-469). Descriptive quantitative studies are aimed at finding out “what is,” so observational and survey methods are frequently used to collect descriptive data (Borg & Gall, 1989). Studies of this type might describe the current state of multimedia usage in schools or patterns of activity resulting from group work at the computer. It can involve collections of quantitative information that can be tabulated along a continuum in numerical form, such as scores on a test or the number of times a person chooses to use a certain feature of a multimedia program, or it can describe categories of information such as gender or patterns of interaction when using technology in a group situation.

The subject of the study are the third grade students of MTs Darul Ulum Karangpandan, Rejoso Pasuruan, there are 4 classes for the third grade in this school, it is consist of 2 classes of man and 2 classes of woman. The number of students around 60 students, consist 20 student every classroom. The object of the study is English final test of MTs Darul Ulum Karangpandan, Rejoso Pasuruan for the third grade student 2016/2017. The researcher borrowed test document after the students have finished the test and after the teacher corrected the students’ answer sheet. The test will be done on April 5th, 2017. The researcher collected the data of study on Mei 8th, 2017.

FINDING AND DISCUSSION

Content Validity

here were 50 items in the English final test which included the listening, speaking, reading and writing. They were related to the syllabus of seven grade, eight grade and nine grade. The researcher checked the appropriateness of the test with the syllabus based on the KTSP curriculum.

In the seven grade first semester; only in Listening skill, there were 2 base competence in this skill including 2.1 and 2.2. There were 6 related item in base competence 2.2, they are; item number 9,10,11,15,16,17.

For example: What is the purpose of the test?
A. To give announcement about a concert
B. To give announcement about a special bazar
C. To give announcement about the meeting,
D. To give announcement about the school’s anniversary.

In the eight grade second semester, the first was reading skill, there were 2 base competence in this skill including 11.1 and 11.2. There were 4 related item in base competence 11.2, they are; item number 24,25,26,27.

Second was listening, there were 2 base competence in this skill including 8.1 and 8.2. There were 2 related item in base competence 8.1, they were; item number 1 and 2.

And in the nine grade second semester; the first was Listening skill, there were 4 base competence in this skill including: 7.1, 7.2, 8.1, 8.2. There were 8 related item in base competence 8.1, they were; item number 3,4,5,6,7,8,22,23.

And there are 4 related item in base competence 8.2, they were; item number
Second was speaking skill, there were 4 base competence in this skill including: 9.1, 9.2, 10.1, 10.2. There were 2 related item in base competence 10.1, they were item number 18,19.

Third was reading skill, there were 4 base competence in this skill including: 11.1, 11.2, 11.3, 11.4. There were 7 related item in base competence 11.2, they were item number 20,21,35,36,37,47,48.

They were 15 related item in base competence 11.3, they were; item number 12, 13, 4, 28, 29, 30, 31, 32, 33, 34, 38, 39, 40, 49, 50.

The last was writing skill, there were 2 base competence in this skill including: 12.1 and 12.2. There were 2 related item in base competence 12.2, they were; item number 45, 46.
Reliability

Measuring the coefficient of the reliability by using the formula propose by Kuder-Richardson as state below:

\[ r^1 = \left( \frac{n}{n-1} \right) \left( \frac{S^2 - \Sigma pq}{S^2} \right) \]

\[ = \left( \frac{50}{50-1} \right) \left( \frac{26.9 - 9.3}{26.9} \right) \]

\[ = (1,020)(0.65) \]

\[ = 1,020 - 0.65 \]

\[ = 0.37 \]

a. Checking the result of the reliability with the criteria to determine whether or not the reliability

The criteria of reliability and the degree (Sudijono: 2008)

- 0.81-1.00 = very high
- 0.61-0.80 = high
- 0.41-0.60 = moderate
- 0.21-0.40 = low
- 0.00-0.20 = very low

From the score of coefficient reliability using KR.20. The reliability of The English final test is low, because the result is 0.37. So, the reliability of the English Final Test for the Third Grade Students at MTs. Darul Ulum Karangpandan Rejoso Pasuruan is low.

Measurement of Item Difficulty

To measure the item difficulty is used formula:

\[ P = \frac{B}{JS} \]

To interpret the result of the difficulty index Arikunto (2008:210) state that Item with difficult index 0, 00 until 0, 30 is a difficult item. Item with difficult index 0, 30 until 0, 70 is a medium item. Item with difficult index 0, 70 until 1, 00 is an easy item.

Item Difficulty

From the table of measuring item difficulty and interpret of item difficulty, the result of difficulty level was:

- Item with difficult index 0, 00 until 0, 30 is a difficult item, there were 24 difficult item they are number: 3, 4, 6, 8, 10, 14, 16, 19, 22, 25, 26, 27, 28, 29, 34, 35, 36, 38, 41, 43, 44, 46, 47, 49. So, the percentage of item difficulty which difficult categorize is 48%.
- Item with difficult index 0, 30 until 0, 70 is a medium item, there were 26 medium item they are number: 1, 2, 5, 7, 9, 11, 12, 13, 15, 17, 18, 20, 21, 23, 24, 30, 31, 32, 33, 37, 39, 40, 42, 45, 48, 50. So, the percentage of item difficulty which medium catagorize is 52%.

The histogram above show that percentage of item difficulty, from the result there are two categorize in item difficulty they are: difficult cagorize and medium cagorize. The percentage of difficulty is 48% and the percentage of the medium cagorize is 52%. It means that the result of the difficulty index is good because according to Arikunto (2008) to make a good test, it must be neither too easy nor too difficult (10%<P<90%). From the result, the researcher conclude the item difficulty of the test is good, because this test moderate for the students.

Discriminating Power

To measure discriminating power of the test, the researcher does the following steps:

Arranging the students score from the higher score from the higher score until the lower score. Then, divide them into two group; upper and lower group. The total of the students are 60 students, so the researcher takes 33% of the total students for the upper group and 33% of the lower group. It means that 20 students for the upper group and 20 students for the lower group.(See on appendix)

After knowing the lower group and upper group, we can enter to the form bellow, it could be seen on table 4.2 measuring the discriminating power.

\[ D = \frac{BA}{JA} = \frac{BB}{JB} = \frac{PA}{PB} \]
To interpret the result the discriminating index (Arikunto, 2008:214) state that:
D = 0.00 until 0.20. It means that the item is poor
D = 0.20 until 0.40. It means that the item is satisfactory
D = 0.40 until 0.70. It means that the item is good
D = 0.70 until 1.00. It means that the item is excellent
D = negative. It means that the item is very poor and should be thrown away.

Based on the table discriminating power and interpret the result of the discriminating power, the researcher shows the result of discriminating power as follows:

Discriminating power = negative. It means that the item is very poor and should be thrown away, there are 11 items were very poor they are number: 4, 6, 8, 10, 26, 28, 35, 36, 43, 44, 46. So, the percentage of discriminating power which very poor is 22%.

Discriminating power = 0.00 until 0.20. It means that the item is poor, there are 15 items were poor they are number: 2, 12, 14, 16, 19, 25, 27, 29, 34, 37, 38, 41, 42, 47, and 49. So, the percentage of discriminating power which poor is 30%.

Discriminating power = 0.20 until 0.40. It means that the item is satisfactory, there are 11 items were satisfactory they are number: 1, 3, 7, 11, 22, 23, 24, 30, 39, 40, and 48. So, the percentage of discriminating power which satisfactory is 22%.

Discriminating power = 0.40 until 0.70. It means that the item is good, there are 10 items were good they are number: 9, 13, 15, 17, 20, 21, 31, 32, 33, and 50. So, the percentage of discriminating power which good is 20%.

Discriminating power = 0.70 until 1.00. It means that the item is excellent, there are 3 items were excellent they are number: 5, 18, and 45. So, the percentage of discriminating power which excellent is 6%.

From the histogram above, the researcher found the discriminating power of the test is categorized into 5 parts. very poor with the percentage 22%, poor with the percentage 30%, satisfactory with the percentage 22%, good with the percentage 20% and excellent with the percentage 6%. By the result analysis, the discriminating power of the test is poor because there are 15 items are poor categorized.

**Item Distractor**

According to Sudjana (2008:278) the step to measure item distractor:
1. Making tabulation of students’ answer choice
2. Analyzing the item distractor

According to Arifin (2009-280), to interpret the result of the index quality of distractor:
- Excellent ID = 76% - 125%
- Good ID = 51% - 75% or 126% - 150%
- Deficient ID = 26% - 50% or 151% - 175%
- Poor ID = 0% - 25% or 176 – 200%
- Very poor ID = more than 200%

From the result, Interpreting the result of the Item Distractor it could be categorized into:
- Excellent item distractor = 76% - 125%, there are 29 items they were number: 1, 4, 5, 6, 7, 10, 15, 16, 19, 20, 23, 24, 27, 28, 29, 30, 31, 32, 34, 36, 37, 39, 40, 42, 43, 45, 47, 48, 50. So, the percentage of the excellent item distractor is 25%.
- Good item distractor = 51% - 75% or 126% - 150%, there are 30 items they were number: 1, 3, 4, 5, 7, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 25, 26, 28, 29, 30, 31, 32, 33, 35, 40, 42, 44, 45, 48, 49. So, the percentage of the good item distractor is 26%.
- Deficient item distractor = 26% - 50% or 151% - 175%, there are 29 items they were number: 2, 6, 8, 10, 11, 12, 13, 14, 18, 20, 22, 24, 25, 26, 27, 31, 33, 34, 35, 36, 38, 41, 42, 43, 44, 46, 47, 48, 49. So, the percentage of the deficient item distractor is 25%.
- Poor item distractor = 0% - 25% or 176 – 200%, there are 15 items they were number: 4, 8, 9, 10, 11, 12, 17, 21, 24, 25, 35, 36, 38, 43, 47. So, the percentage of the poor item distractor is 13%.
- Very poor distractor = more than 200%, there are 11 items they were number: 2, 5, 8, 9, 13, 14, 17, 22, 41, 46, 49. So, the percentage of the very poor item distractor is 10%.

Based on the histogram above there are five categorize in item distractor: excellent item distractor with the percentage 29%, good item distractor with the percentage 28%, deficient item distractor with the percentage 26%, poor item distractor with the percentage 9% and very poor item distractor with...
the percentage 6%. So, it could be concluded that item distractor was efficient because many item were excellent distractor. The item distractor could be applied and work effectively.

CONCLUSION

In conclusion that the analysis of English Final test is fair/moderate because reliability is low and the final test had a poor discriminating power. A high criteria test should have good validity, reliability, item difficulty, discriminating power and item distractor because all the criteria can determine a test good, fair or low. Based on the analysis of English final test of MTs Darul Ulum Karangpandan Rejoso Pasuruan, it can be concluded that from 50 items of multiple choices, there were 50 items which are related to the syllabus (KTSP) curriculum in of seven grade, eight grade and nine grade. From the score of coefficient reliability by using K-R 20, the reliability of the English final test is 0,37. Based on the criteria of the degree which was used to interpret the result, the reliability of the test was low because it was in 0, 21 until 0, 40 degree. From the finding of difficult item, there were 24 difficult category.

They are number: 3, 4, 6, 8, 10, 14, 16, 19, 22, 25, 26, 27, 28, 29, 34, 35, 36, 38, 41, 43, 44, 46, 47, 49 and there were 26 medium item, they are number: 1, 2, 5, 7, 9, 11, 12, 13, 15, 17, 18, 20, 21, 23, 24, 30, 31, 32, 33, 37, 39, 40, 42, 45, 48, 50. So, the result of the item difficulty was good because most of the result was moderate. The interpretation result of the difficulty index was good, the researcher found the discriminating power of the test is categorized into 5 parts. There are 11 items are very poor, 15 items are poor, 11 items are satisfactory, 10 items are good and there are 3 items are excellent. By the result analysis, the discriminating power of the test was poor discriminating power because there were 15 items in poor category. It could be concluded that the final test had a easy level it meant that the final test had a poor discriminating power. The difficulty level of test influence the discriminating power of the test. The result of English Final test, of Item Distractor can be categorized into: excellent item distractor there are 29 items, good item distractor there are 30 items, deficient item distractor there are 29 items, poor item distractor there are 15 items, very poor distractor there are 11 items. It could be concluded that item distractor was efficient because many item were excellent distractor. The item distractor could be applied and work effectively.

The test maker should give more attention on the difficulty level and discriminating power of the test because both of them influence the score of the students. Teachers should make analysis in every test which were held before and compare the result to know the real ability of the student from the result of analysis. The items that fulfill the criteria of good items can be used for the future evaluation. For the items that do not fulfill the criteria should be revised so that they can be used for the next evaluation. The teacher should improve the effectiveness of teaching and learning process by using the result of analysis on the item. Curriculum in our country is always changing. It is also give effect to the lesson materials and test items. So, the other researcher could analyze it, or analyze another type of test item, for example subjective type of test item. He/ she can use the result of this research as the preview study of item analysis of English final test as consideration and information.

REFERENCES