Test is one of ways to know the competence of students indirectly, in which it can be known by the student’s response to stimulate or questions given (Mardapi, 2008, p. 67). The response of the students to the question shows the competence of the students itself. It is clear that a test is going on its own purpose or objective. It can be achieved if only the test items are well-made (Kusdinar, 2004). Actually, the purpose of the test should be considered carefully before constructing the test itself. The important thing in a test is that the knowledge of how far those purpose have been achieved. Talking about the test means talking about the evaluation whether the instrument is good or not. So, we have to know the quality of test that we give. The test will be able to full-fill the purpose of the test if the test has the characteristics of good test. A test is considered as a good test if the test has validity, reliability, objective, practically, and economy (Arikunto, 2013, p. 72). However, a good test must have two fundamental Terms validity and reliability because test which has low validity and reliability cannot be called as a good test (Arifin, 2011, p. 246).

Some previous researchers which are have the similar purpose are such as conducted by Dewi (2012) she took Quality Analysis the English test Try out test 2011/2012 for third year students student of SMP 3 Pasuruan. The conclusion is the test has high content validity because 100% items are appropriate with the SKL (standart Kompetensi Kelulusan), reliability of the test are had high, item difficulty test is not good enough, because the sum of the difficult, moderate and easy numbers is not same, and discriminating power are categorized as poor. Another research was made by Aminin (2012) he took English the final test in the second year semester at SMP N 1 Gondang Wetan. Overall, the test is good content validity that has been given because 92%
material includes in syllabus. The reliability of the test is high with the result of the coefficient in 0.70 the items difficulty is not too difficult and not too easy. It because 45 items has difficulty level 0.25 until 0.75 (enough) and 5 items has low difficulty level >0.75(too easy item), it indicates that the difficulty level is high, the discriminating power is poor, it because from 50 item 43 item are poor, 3 item are satisfactory, and 4 item are good.

Another research was made by Minawati (2013) investigated The English Final Test in The ninth grade of MTS Mftahul Huda Kalirejo. The conclusion is the test has good content validity because 100% items are appropriate with the syllabus, There are 30% in the first semester and 70% in the second semester. Reliability of the test is very high because it has 0.9 as the coefficient of reliability. Item difficulty test is not good to be given to the students because it has low level in difficulty so the test cannot motivate the students. And discriminating power of the test is poor.

Based on the explanation above, the researcher wants to know the quality of the English proficiency test for junior high school at the 24th anniversary of English student’s Association at STKIP PGRI Pasuruan.

METHOD

In this study, the researcher qualitative research focuses on content analysis and descriptive quantitative. Content analysis is a research method for making replicable and valid inferences from data to their context, with the purpose of providing knowledge, new insights, a representation of facts and a practical guide to action (Krippendorff:1980). The aim is to attain a condensed and broad description of the phenomenon, and the outcome of the analysis is concepts or categories describing the phenomenon. Usually the purpose of those concepts or categories is to build up a model, conceptual system, conceptual map or categories. The researcher makes a choice between the terms ‘concept’ and ‘category’ and uses one or the other (Kyngaa’s & Vanhanen: 1999). Descriptive quantitative was used to analyze the reliability.

This research is conducted in STKIP PGRI Pasuruan which is located on Ki Hajar Dewantara street number 27-29 Pasuruan. The researcher chooses English proficiency contest for junior high school at the 24th Anniversary of English Students’ Association at STKIP PGRI Pasuruan. The researcher uses two instruments in her research those are human instrument and documentation. In qualitative research, Creswell (2012:223) stated that Documents consist of public and private records that qualitative researchers obtain about a site or participants in a study, and they can include newspapers, minutes of meetings, personal journals, and letters. Sugiono (2010:296) stated that he is believed that the researcher influence the study findings through their instruction with the study participants, and that the human as instruments is the best capable of grasping and evaluation the meaning of the interaction.

In the processing of the data analysis, the researcher takes some steps. The researcher will analyze the content validity refers to curriculum (KTSP)’. The researcher will used Kuder-Richardson method to analyze the reliability of the multiple choice test items. It will be calculated by using the formula given by Djiwandono. The steps to estimate the reliability of proficiency test will apply by the following instruction: the first, Making tabulation of the testee’s scores for each test items. Second, measuring the mean of the testee’s scores. It will be gotten by dividing the sum of separate scores with the total number of testee’s. Third Measuring the standard of deviation by using the formula djiwandono, 2011. The fourth, Measuring the coefficient of the reliability by using the formula proposed by Kuder-Richardson in djiwandono, 2011. Checking the result of the reliability with the criteria to determine whether or not the test is reliability by using criteria best in Djiwandono (1995).

The index of the difficult (or the facility value) of an item simply shows whether the test easy or difficult for the particular item provide in the test, Arifin (2011:266). arrange the student’s score into upper group and lower group. Computing the item difficulty by using the formula proposed by brown in (Djiwandono: 2011). Checking the result of item difficulty by using criteria of Arikunto (2013). Analyzing the discriminating power, the researcher uses the formal proposed by Heaton. The steps to determine the index of discriminating power. The first Arrange the script in rank order to rank total score and divide into two groups of equal size. The second Count of the number of those candidates in upper group answering the first item correctly; and account the number of lower group answering the item correctly. The third Subtract the number of correct answer the lower group from the number of the correct answer in the upper group; i.e. find the difference in the proportion passing in the upper group and the proportion passing in the lower group. Count use formula by Djiwandono,
2011. Checking the result of discriminating power by using criteria of Sudijono, 1995.

FINDINGS

The researcher found that there are 89 students’ answer sheet. From 100 items test the researcher found 82% items test appropriate with the syllabus junior high school and just 18% are not appropriate with the syllabus because the materials include in Senior high school syllabus. It shows that validity of the test is very high. The reliability of the test, after the researcher calculate by using formula as previous explained, the researcher found that the result of reliability is 0.9. It shows that the reliability is very high. Index of difficulty. Based on the interpreting item difficulty, There are 16 difficult items because the index difficulty level is less than 0, 25. There are 71 moderate items because the index difficulty level is 0,25–0,75. There are 13 easy items because the index difficulty level is more than 0,75. Discriminating power of the test, there are 44 items test are poor, 38 are moderate, 17 are good, and just one items test is excellent. It means that the discriminating power of the English proficiency test is poor indiscriminating power.

DISCUSSION

Content validity

Based on the finding, it is found the item test which is given to the student is appropriate with syllabus (KTSP 2006). Where 3% item test are related to 5.2, 19% items test are related to 1.2, 2% item test are related to 8.1, 2% items test are related to 10.1, 4% items test are related to 10.2, 7% items test are related to 11.3, 4% items test are related to 1.1, 1% items test are related to 3.2, 15% items test are related to 5.1, 22% items test are related to 6.1, 3% items test are related to 9.1, and 18% items test that are not related to syllabus. As previous (Arikunto: 2013) have been explained that the characteristic of a good test possess two qualities: validity, reliability. According to Djiwandono (2011:168) there are fifth criteria of validity and the degree, the first criteria 0,80–100 is very high, the second criteria 0,60–0,79 is high, the third criteria 0,40 - 0,58 is moderate, the fourth criteria 0,20–0,39 is low, and the fifth criteria 0,00–0,19 is very low. Based on the criteria, the English proficiency test has high content validity because 82% or can be said 0,82 tests cover the syllabus of junior high school. If we want to increase the content validity a test, we must know what measure the test after we know the objective the test, we can fix the test with the curriculum. As previous chapter Kubiszyn and Borich (2003:300) stated that contend validity is dealing with the test items tested whether those items reflect on the content of curriculum measured or not.

Reliability

This study uses Kuder Richardson 20 method. The researcher makes tabulation of the testee’s scores for each test items, Measure the mean of the testee’s scores, measure deviation standard and calculate the form. From the theory Djiwandono (1995) have classified five criteria of reliability and the degree. the first 0, 90 up to 1.00 the criteria of reliability is very high reliability, the second 0, 70 up to 0, 89 the criteria of reliability is high , third 0, 50 up to 0, 69 the criteria of reliability is moderate, fourth 0, 30 up to 0, 49 the criteria of reliability is low, and fifth less than 0, 30 the criteria of reliability is very low. From the calculating in finding, the coefficient reliability is 0, 9. Based on the criteria of reliability of English proficiency test is very high. Bachman (1990, p. 24) mention that the validity and reliability are the two essentials to the interpretation and the use of language ability. To increase the reliability of the test, we must make a good content validity because if the test is valid can be said reliable, we can get more consistency of the test score.

Item difficulty

Item difficulty is looking for the level item test that includes in criteria difficulty index. It divides into three parts. They are less than 0, 25 (too difficult), 0, 25 - 0,75 (moderate), more than 0, 75 (too easy). From the analysis, the test has 16 items (too difficult), 71 item (moderate item) and 13 items (too easy). From the theory which has discussed in the previous chapter, According to Arifin (2011), a test has a balanced level of difficulty (proportional), it can be said that the test is good because a good test is not too difficult and too easy tide. So English proficiency test can be said good test because 71 item (moderate) but still less challenging, consider in finding that the English proficiency test fulfill level of difficulty (proportional). Arifin (2011) also stated that the difficulty level is a measurement of how much matter the level of difficulty, so we must know the criteria of the test if we want make a good test with the good standard the level of difficulty, because this is the proficiency contest if the test is too easy, it can’t motivate and challenge the students’ to be best participant in the next year.
Discriminating Power

Discriminating power is to differentiate the student’s upper and lower, from the theory that Discriminating power is used to know about learners who have mastered the competencies with students who have not mastered the competencies based on certain criteria. According to Sudijono (1995) there are 4 classification in discriminating power: first, less than 0, 20 (poor), it means the item test is very low in discriminating power and it cannot be called that the test do not have discriminating index. Second, 0,20 up to 0,40 (satisfactory), it means the item test is low in discriminating power. Third, 0,40 up to 0,70 (good), it means the item test is good in discriminating power and 0,70 up to 1,00 (excellent), it means the item test is very good in discriminating power. From the finding, the researcher find Discriminating power of the test is categorized in to 4 parts, there are 44 items from poor group including less than 0,20 (poor), 38 item test from satisfactory group including 0, 20 up to 0, 40 (satisfactory), 17 items from good group including 0, 40 up to 0,70 (good), and 1 item from excellent including 0,70 up to 1,00. By the result of analysis, the discriminating power of test is poor discriminating power. It still cannot differentiate the high achiever students and low achiever students because A good test is test which could be answered by clever students. According (Djiwandono 2011) discriminating power is capable of measuring discriminating power and it cannot be called that the test cannot motivate and compete the student in the next contest to be best participant.

Discriminating power of the test is categorized in to 4 parts, there are 44 items from poor group including, 38 item test from satisfactory group, 17 items from good group, and 1 item from excellent. By the result of analysis, the discriminating power of test is poor discriminating power.

The suggestion is expected to be considered by the test maker and the teacher in making good test. The test maker should prepare the test which really measure the student’s achievement. Besides, the test maker also pay attention to discriminating power and make good distrusters because both of them influence the students score. Secondly, the English teacher should make analysis in every test are held like research before. Because it can measure the ability of every student by knowing the result of analysis. And finally, other researcher can read this result of study as a reference.

REFERENCES


Benedetti, Kate De. 2006. Validity And Reliability In Foreign Language Tests. Mextesol jurnal. 16: 25-40.


