Item Analysis Of English Final Test Of Junior High School

Muhammad Rifqi Pratama Putera, Abdul Muth'im, and Fahmi Hidayat
Universitas lambung mangkurat
a1b215234@mhs.ulm.ac.id

Abstract

This study aimed to examine the quality of the English final test items of seventh grade in a public school located in Banjarmasin. The research used the descriptive method with a quantitative approach because it tries to describe the test items' quality. It was analyzed by using *ITEMAN 3.0*. The sample of this research was the English final test for seventh-grade students at SMPN 24 Banjarmasin. The research findings show that: (1) based on the content validity, all items are in harmony with the learning objective in the syllabus. Moreover, based on item validity, there are 41 items (82%) considered valid items; (2) the reliability of the English final test is categorized as high reliability with *Cronbach* alpha 0.800; (3) there were 16% of items considered as difficult items, 44% items were considered as moderate items, and 40% items were considered as easy items; (4) 2% of the items have very poor discrimination power, 16% of the items have poor discrimination power, 22% of the items have satisfactory discrimination power, 50% of the items have good discrimination power, and 10% of the items have very good discrimination power; (5) 9% of the items have very bad distractor efficiency, 16% of the items have bad distractor efficiency, 24% of the items have good distractor efficiency, and 54% of the items have very good distractor efficiency; (6) based on the quality of the item, 64% of the items need to be revised or discarded. In conclusion, the test still needs some revisions and improvements.

Keywords: test item analysis, final semester test.

INTRODUCTION

Brown (2004) defines a test as a method to measure the student's knowledge and ability. The functions of tests are; as the measurement of how successful the teaching-learning process is, as the instrument to see the students' achievement, and as an instrument to make the educational decision (Sugianto, 2017). Considering the importance of the test as an instrument the writer analyzed a test made by the teacher, particularly the English final test. The scope of this study is the first semester of the seventh-grade students of SMPN 24 Banjarmasin. SMPN 24 Banjarmasin is a junior high school in Banjarmasin, Indonesia. The analysis would show the quality of the English final test items in this case based on the validity, reliability, level of difficulty, discrimination power, and distractor efficiency.

The objective of the research is to determine the quality of the English final test made by the English teacher of SMPN 24 Banjarmasin. The quality of the test items was viewed from the validity, reliability, level of difficulty, discrimination power, and distractor efficiency. The writer chose SMPN 24 Banjarmasin because the teacher has never analyzed test items, so the writer wants to analyze the English final test to determine the quality of the test items.

Definition of Test

A test is essential in getting an info of students' knowledge and ability in the teaching learning process, as Brown (2004:384) mentioned that the test is a method to measure the students' knowledge and ability in a given domain. Arifin (2012:18) also defined test as a technique or method used in order to carry out measurement activities, in which there are questions, statements, or a series of task that must be done or answered to measure the students' behavior.

Two types of tests usually occur in the teaching-learning process, as Brown (2004) mentioned: formative and summative tests. The formative test aims to evaluate the students forming their competencies skills while still helping them continue developing their knowledge; one example is the

middle test exam. On the other hand, the summative test aims to measure what students have grasped by the end of the course; one example of a summative test is the final test.

Item Analysis

Item analysis is essential in knowing test item quality through a systematic procedure to get specific information about the items from the test (Arikunto, 2013). Sudjana (2011) also defined item analysis as the study of the test to get the material to make a good question in the test. In the other word, test item analysis is a process of collecting, summarizing, and getting the information from the students' answers in order to decide their grades and also the quality of the test itself.

Gronlund (1977) mentioned some benefits of doing test item analysis are; first, it contributes helpful information for class discussion of the test. Second, it provides helpful input to encourage students improving their learning outcome. In addition, item analysis is essential in developing the quality of the test items that will be reused, revised, or discarded for the later test.

Validity

Validity of a test is the extent of measurement which is measure what is supposed to measure (Heaton, 1988:155). The validity of the test must be considered in measurement due to the test must measure what supposed to measure. A test must be relevant; a teacher must examine the consistency and the relevancy of an instruments before it used as measuring instruments

Reliability

Reliability refers to consistency of the test outcome. Bachman (2004) stated reliability is the consistency of measurement across different condition. Singarimbun and Soffian (2008) stated that reliability is an index that points how far the measurement can be trusted. It means that a test can be said as reliable if the test gives consistent result when the test is re-testing to the test taker.

Level of Difficulty

Arikunto (2013) stated level of difficulty is the ability of a test to tell how difficult the test is. A test is said as a good test if an item from the test has an equal difficulty index (proportional). The item that categorized as easy item means to encourage non-master students and also it serves as warm-up items. However, the item that categorized as difficult items means to challenge master students (Brown, 2004:59). In this level of difficulty index, there are minimum and maximum scores. The lower index indicates the items is too difficult for the test taker, and the higher index indicates the item too easy for the test taker.

Discrimination Power

Discriminating power is the ability of a test to separate fast learner students from slow learner students (Sudijono, 2011). Zaman et al (2010) define discrimination power as the relation between students' ability to answer an item in the test with the test's total item. In other word, discrimination power of the test is the ability of the test item to distinguish the ability of fast learner students with slow learner students. fast learner students are the students who got high scores, and slow learner students are the students who got a low score on the test given.

Therefore, if the test item has a high coefficient in discriminating power, the item could differentiate the fast learner students and slow learner students. Hughes (2005:226) stated that if the discrimination power has a negative point. It means that fast learner students identify as slow learner students and vice versa. A good item is an item that can be answered correctly by fast learner students and can not be answered by slow learner students.

Distractor Efficiency

There are alternative options in multiple choice question. It purpose is to distract the students from the correct answer. Good item in test is if the distractor distributes equally by the test taker who has chosen the wrong answer. However, it is bad if the distractor distributes unequally by test taker who have chosen the wrong answer. Mardapi (2008) demonstrated that a good distractor is if the option is chosen by least 5% of test-takers. If the distractor does not meet the requirement above, it is suggested to change to another distractor that can persuade the test taker to choose the option.

RESEARCH METHOD

The design of this research is descriptive research with quantitative approach. This research used descriptive method due to it aims to describe English final test quality for seventh-grade students in SMPN 24 Banjarmasin academic year 2019/2020.

Moreover, this research uses a quantitative approach to analyze the data. It is called quantitative because the data are obtained from a document of the test items, and students answer sheet needs to be calculated; thus, the research presented in the numeral. This research concerns the components in the final test, which covers test item validity, reliability, level of difficulty, discrimination power, and distractor efficiency.

Subject of the research

The subject of this study is the English final test for the first semester of seventh-grade students of SMPN 24 Banjarmasin in the academic year 2019/2020. The English final test consisted of 50 items. So the data source was the question sheet that consist of 50 items and 150 students' answer sheet of seventh grade students at SMPN 24 Banjarmasin.

Instruments

In collecting the data, the researcher used the documentation method. The researcher got the document from the English teacher of SMPN 24 Banjarmasin. The documentation consisted of questions on the test paper, students' answer sheets, lattice of the English final test for seventh grade students, and answer key.

Research Procedure

Based on the research objective, which is to determine test item quality based on its validity, reliability, level of difficulty, discrimination power, and distractor efficiency. There are few steps that researcher do. First, collecting the documentation of the research which is the English final test question, students' answer sheet, answer key, and lattice of the English final test from SMPN 24 Banjarmasin. Second, the researcher analyzes the content validity by matching the lattice of the English final test with the syllabus. Third, the researcher analyzes the item validity, reliability, level of difficulty, discrimination power, and distractor efficiency using ITEMAN 3.0.

Data Analysis

In analyzing the data, the researcher uses a quantitative approach. Quantitative approach is done by analyzing the quality of the test item based on the item validity, reliability, level of difficulty, discrimination power, and distribution of the distractor. The researcher used ITEMAN 3.0 to help the researcher in analyzing the data. The content validity is done by matching the lattice of the English final test with the syllabus

FINDING

Validity

To determine the content validity is by matching the question framework whether the test items are in harmony with the indicators. The distribution of the items based on the content validity displayed in Table 1:

Table 1. Content validity analysis of the test items

Basic Competence	Indicators	Material	Skills	Item Number
3.1 Understanding the social function, the structure of the text and linguistic element in expression of greeting, farewell, gratitude, asking forgiveness as well as the response.	1. Provided incomplete dialog, students will be able determined the appropriate response of greeting in English correctly	Greeting, Gratitude, Congratulation, Asking Forgiveness	Reading	8,34
	2. Given short text, students will be able to determined the communicative goal correctly		Reading	36
	3. Given short dialog of congratulating someone, students will be able to determined certain information correctly		Reading	5,6
3,2 Understanding the social function, the structure of the text and linguistic element in the introduction of self and response, according to the context of its used	1. Provided short monolog text about introducing oneself, students determined the implied information correctly	Self Introduction	Reading	2
	2. Provided short monolog text about introducing oneself, students will be able to determined certain information in the text correctly		Reading	3, 4
3.3 Understanding the social function, the structure of the text and linguistic elements of the text to state and asked the name of day, the month, the year, the name of the times in a day, and the time as number, date, and year	1. Given a picture of a clock, students will be able to determined the time in English correctly	Time, Day, Month, Year, Daily activity, Cardinal Number, and Ordinal Number	Reading	1
	2. Provided an incomplete dialog about time, students will be able to determined the appropriate response of time correctly		Reading	9, 10, 49
	3. Given a short dialog of asking floor, students determined the ordinal number in English correctly		Reading	20
	4. Provide picture of calendar, students will be able to determined the time in form of days, month, and year in English correctly		Reading	32
	5. Provided a jumbled words about daily activities, students can arrange these word into a correct sentence		Writing	24
	6. Presented an incomplete descriptive text about everyday activity, students can complete the text with correct nouns and verbs		Writing	37,39
	7. Presented an incomplete descriptive text about everyday activity, students can complete the text with appropriate verbs		Writing	38, 40
3.4 Understanding the social function, the structure of the text, and the text elements of the linguistic identity exposure, according to the context of its used	1. Provided short information of identity exposure, students determined certain information in the text correctly	People around me	Reading	17,18,19
	2. Given a family tree of someone, students determined certain information in the family tree correctly		Reading	21,22,23
	3. Provide a short text of someone, students determined the occupation of the person in English correctly		Reading	44
3.5 Understanding the social function , the structure of the text and	1. Provided a short text of public building, students determined certain information correctly	Things around us, Public building, and Animal	Reading	16,27,48

linguistic element in the text to express and ask the number of animals, things, and public building close to the students ' every day life	2. Given a short dialog and picture of things, students determined certain information correctly		Reading	7, 50
	3. Given descriptive text of animal, students determined the referent correctly		Reading	33
	4. Provided an short dialogue of asking number of things, students determined the appropriate to asking the number of things in English correctly		Reading	26
	5. Given a picture of animal, students determined the number of animal in English correctly		Reading	25
	6. Given a descriptive text of someone garden, Students determined the certain information correctly		Reading	41,43
	7. Given a descriptive text of someone garden, students determined the reference correctly		Reading	42
	8. Given a picture of public building, students determined certain information correctly		Reading	46,47,48
	9. Presented a few sentences of school garden , students can arrange these sentences into coherent descriptive text.		Writing	35
	10. Given a list of the thing around house, students determined certain information correctly		Reading	11,12,13,14,15
3.6 Understanding the social function, the structure of the text, and the linguistic elements of the text label name, and a list of items, according to the context of its used	1. Given birthday shopping list, students determined the communicative goal correctly	Label and Shopping list	Reading	28,29,31
	2. Given birthday shopping list, students determined the word meaning (synonym) correctly		Reading	30
	3. Given a label, students determined certain information correctly		Reading	45

The content validity analysis show that the English final test for seventh-grade students at SMPN 24 Banjarmasin has good content validity due to all of the items in the test are in harmony with the competencies in the syllabus.

The validity of the test item obtains by using biserial point correlation formula. Total samples of the research are 150 students; hence the r_{table} is equal to 0.160. The test item validity analysis of English final test for 7th grade students at SMPN 24 Banjarmasin based on standard Ypbi $\geq$ 0.160 means the item is valid, and if Ypbi < 0.160 means the item is invalid. The results of the item validity analysis of the English final test for seventh-grade students at SMPN 24 Banjarmasin can be seen in Table 2:

Table 2. Analysis of validity of test item

Number	Validity Index	Number of Items	Total Items	Percentage
1	Invalid (< 0.160)	1,6,11,14,17,29,37,38,43	9	18%

2	Valid (≥ 0.160)	2,3,4,5,7,8,9,10,12,13,15,16,18,19,20,2 2,23,24,25,26,27,28,30,31,32,33,34,35, 36,39,40,41,42,44,45,46,47,48,49,50	41	82%

Figure 1. Pie diagram of validity of test item
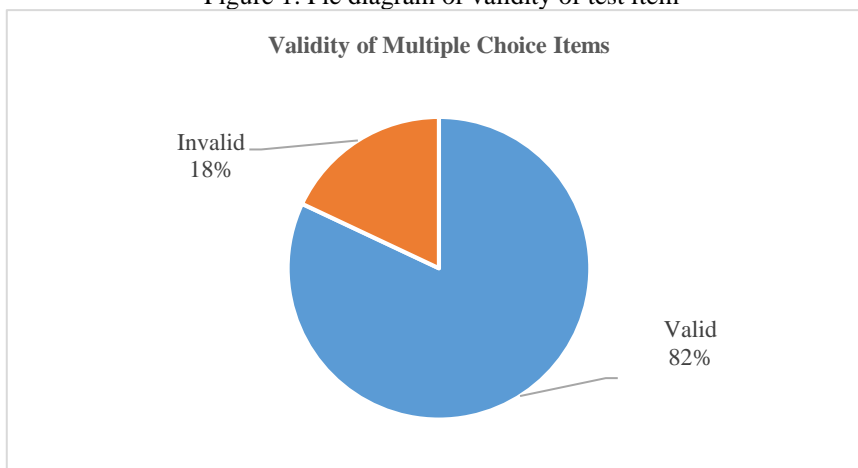
Based on table 2 and figure 1, The total of valid items of the English final test is 41 items (82%), while the total of invalid items is 9 items (18%). It means the English final test items for seventh grade students at SMPN 24 Banjarmasin have good quality in terms of validity because most of the items are considered valid.

Reliability
A test item called reliable if the test gives consistent result (Bachman, 2004). To measure the reliability of the test, the researcher used *alpha Cronbach* Coefficient. The result of reliability analysis of the English final test for seventh grade students at SMPN 24 Banjarmasin can be seen in Table 3:

Table 3. Analysis of reliability of test item

Total Item	Total Students	Alpha Cronbach Coefficient	Reliability
50	150	0.800	High Reliability

Table 3 shown the result of reliability analysis is 0.800. It means that the reliability of English final test for seventh-grade students at SMPN 24 Banjarmasin categorized as high reliability.
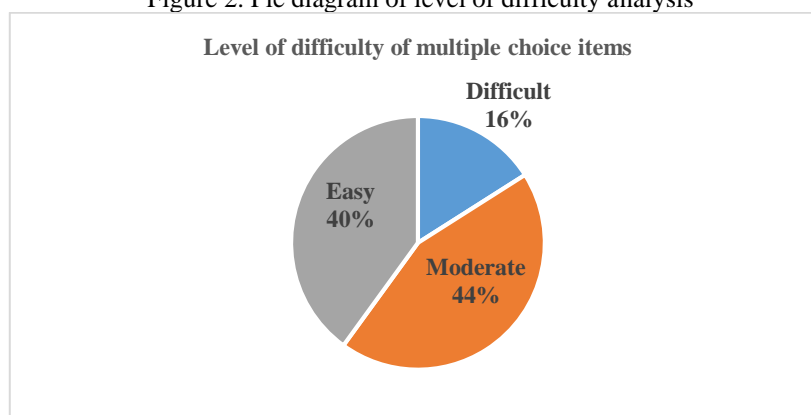
Level of difficulty
Arikunto (2013) stated a good test if a test had proportional level of difficulty. If the item is too easy for the students, the students who got the correct answer will be high and vice versa. Thus, the difficulty index of an item should be spread evenly. The level of difficulty analysis of the English final test item for seventh grade students at SMPN 24 Banjarmasin can be seen in Table 4:

Table 4. Analysis of level of difficulty of test item

Number	Category	Difficulty Index range	Number of item	Total Item	Percentage
1	Difficult	0.000-0.299	1,13,14,17,29,38,40,42,	8	16%
2	Moderate	0.300-0.700	2,3,6,7,10,11,12,15,16,20,21, 22,26,35,36,37,39,41,43,45,47,50	22	44%
3	Easy	0.701-1.000	4,5,8,9,18,19,23,24,25,27,28, 30,31,32,33,34,44,46,48,49	20	40%

Figure 2. Pie diagram of level of difficulty analysis

Based on table 4 and figure 2, the level of difficulty of an English final test item for seventh-grade students at SMPN 24 Banjarmasin mostly moderate. Of the 50 items, there are 44% moderate items, 40% are easy items, and 16% are difficulty items.

In conclusion, the difficulty index of the English final test for seventh-grade students at SMPN 24 Banjarmasin is not good because the total number of items in easy or difficult level are 28 items (56%). The item that is in easy or difficult level should be revised.
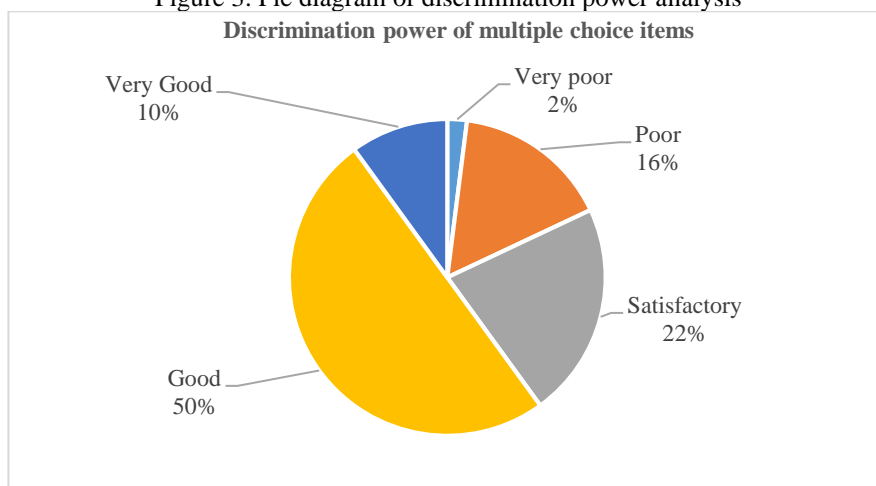
Discrimination Power

Sudijono (2011) mentioned discrimination power as the ability of the test item to separate master students from non-master students. Discrimination power with high value means that the test item can discriminate the ability of the master students and non-master students very well and vice versa. The discrimination power analysis of the English final test item for seventh-grade students at SMPN 24 can be seen in the table 5:

Table 5. Analysis of discrimination power of test item

Category	Range	Item Number	Total	Percentage
Very Poor	D = Negative	14	1	2%
Poor	0.000-0.199	1,6,11,17,29,37,38,43	8	16%

Satisfactory	0.200-0.399	3,5,7,8,21,22,34,35,36,40,42	11	22%
Good	0.400-0.699	2,4,9,10,12,13,16,20,23,24,25,26,27,28, 30,31,32,33,39, 41,44,45,47,49,50	25	50%
Very Good	0.700-1.000	15,18,19,46,48	5	10%

Figure 3. Pie diagram of discrimination power analysis

Based on table 5 and figure 3, the discrimination power of the final test for seventh-grade students at SMPN 24 Banjarmasin tends to be an acceptable level. Based on 50 items that the researcher has identified, 11 items (20%) of the test items are satisfactory because the index of discriminating power is in the range of 0.200-0.399. There are 25 items (50%) of test items categorized as good items because the discriminating power ranges from 0.400-0.699. There are 5 items (10%) of the items are categorized as very good because the discrimination power ranges from 0.700-1.000. There are 8 items (16%) of the test item categorized as poor items because their discriminating power ranges from 0.000-0.199. There is 1 item (2%) of test items categorized as very poor because the item has negative discrimination power.

In conclusion, the discrimination power of the English final test for seventh-grade students at SMPN 24 Banjarmasin is good with totals 41 items (82%) can distinguish the ability of the master students with non-master students.
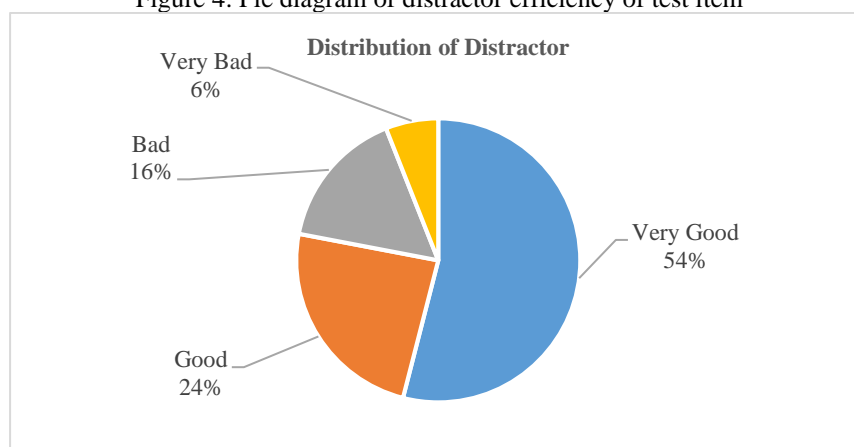
Distractor Efficiency

The type of the English final test for seventh-grade students at SMPN 24 Banjarmasin is multiple choice. Every item has four options, which only has one key answer. Every option besides the key answer is a distractor. A good item should have a distractor that is working correctly. A distractor can be called working if it chooses by at least 5% of the total test taker (Mardapi, 2008). The distractor analysis of the English final tests item for 7th grade students displays in Table 6:

Table 6. Analysis of distractor efficiency of test item

Category	Item Number	Total	Percentage

Very Good (All distractor serves as whole)	2,7,9,10,11,12,13,14,15,16,20,21,24,26,29,30,32,35,36,38,39,40,41,42,43,47,50	27	54%
Good (One malfunction distractor)	1,3,4,6,8,17,22,23,33,45,47,49	12	24%
Bad (Two malfunction distractor)	19,25,27,28,31,34,44,46,48	8	16%
Very Bad (All malfunction distractor)	5,18,46	3	6%

Figure 4. Pie diagram of distractor efficiency of test item

Based on table 6 and figure 4, the items that have good distractor efficiency total to 27 items (54%), the items that have good distractor efficiency total 12 items (24%), the items that have bad distractor efficiency total to 8 items (16%), and the items that have very bad distractor efficiency total to 3 items (6%). The result of item distractors analysis shows that some of the distractors are too flashy, misleading, and do not interest test-takers who do not understand the materials. It means that English final test of SMPN 24 Banjarmasin still needed revision, especially to the 11 items (22%) that have malfunction distractor.
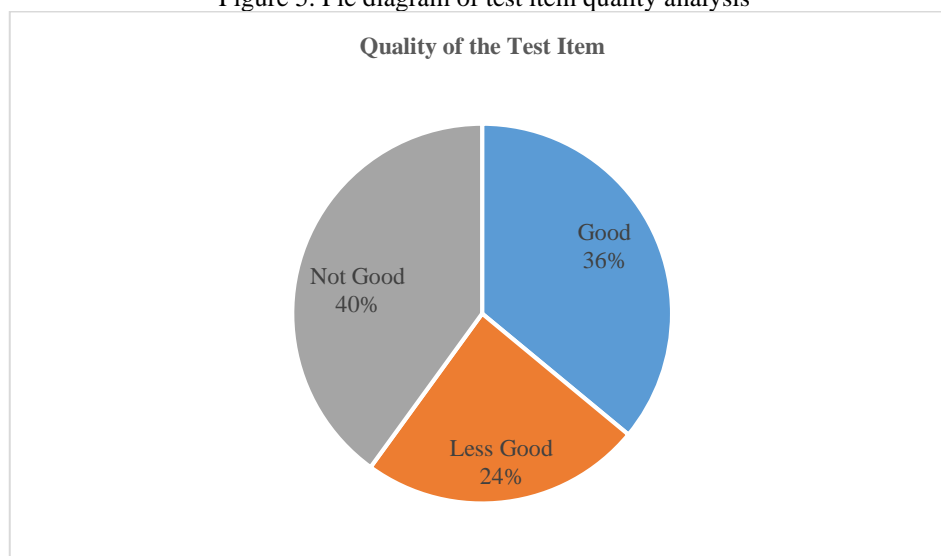
Quality of Test Item

A good test item must have all of the four criteria; the item must be valid, the difficulty index should be around 0.3 to 0.7, discrimination power > 0.2, and have very good or good distractor efficiency. Test items with bad characteristics based on their validity, difficulty index, discrimination power, and distractor efficiency should be chosen selectively. The results of the quality of the test items are displayed in Table 7:

Table 7. The quality of English final test item

Criteria	Number of item	Total item	Percentage (%)
Good (All criteria)	2,3,7,10,12,15,16,20,21,22,26,35,36,39,41,45,47,50	18	36%

Less good (3 Criteria)	4,8,9,13,23,24,30,32, 33,40,42,49	12	24%
Not good (≤ 2 Criteria)	1,5,6,11,14,17,18,19,25,27,28,29,31,34,37, 38,43,44,46,48	20	40%

Figure 5. Pie diagram of test item quality analysis

The purpose of defining the quality of the test items is to find out the item that does not meet the criteria of a good item. The item that does not meet the criteria also reduces the reliability of the test. Based on the table 7 and figure 5, the result of the quality of the test item based on the validity, difficulty index, discrimination power, and distractor effectivity show that 12 items (24%) of the item do not met 1 criterion of a good item, 20 items (40%) does not meet 2 or more criteria of good items, and 18 items (36%) of test item met the criteria of good item. The totals of the items that do not meet one or more criteria are 32 items. Deleting or revising 32 items means that there are no items that have problems with validity, level of difficulty, discrimination power, and distractor efficiency.

DISCUSSION
The quality of the English final test items for seventh-grade students at SMPN 24 Banjarmasin Academic Year 2019/2020 can be seen through some indicators of validity, reliability, level of difficulty, discrimination power, and item distractor. The item analysis of the validity, reliability, level of difficulty, discrimination power, and distractor efficiency are discussed below.

First, the validity of the test items; a test can be valid if the test can measure the object measured. Validity reflects how far the precision and accuracy of test instruments can function as a measurement tool for learning outcomes. The content validity can be viewed from the question framework which shows the item of English final test for seventh-grade students has good content validity. The distribution of the items was in line with the curriculum. 50 items in the test represented the indicator and the basic competence, meaning that the test had 100% agreement with the curriculum. However, this test only measures reading and writing skills as informed by the teacher. The test missed the listening and speaking skills.

The other validity of the test items is item validity. To calculate the item validity, the researcher used point biserial correlation formula, which was further consulted to r_{table} at a significance level of 5% (0.05). The result of item validity analysis of the English final test for seventh-grade students at SMPN 24 Banjarmasin shows that the total samples of the research are 150 persons so that N = 150,

r_{table} shows the number 0.160. If point biserial correlation is more or the same as the value of r_{table} means that the item is considered valid. However, if it is less than the value r_{table}, the item is considered invalid items; the results showed that the total of valid items is 41 items, while the invalid item is 9 items. The invalid items should be revised or discarded, while the valid items can be reused for the later test.

Second, the reliability of the test items. The result of the reliability of the English final test for seventh grade students at SMPN 24 Banjarmasin academic year 2019/2020 based on the standard that if $r \geq 0.700$ means that the test has high reliability. However, if $r < 0.700$ means that the test has low reliability. To calculate the reliability, the researcher used the alpha *Cronbach* formula. The calculation of overall reliability from the items of the English final test for seventh-grade students at SMPN 24 Banjarmasin academic year 2019/2020 has the reliability index of 0,800. It means that the test has high reliability. It means that when the test is re-testing to the same group without reflecting with the previous test, they will get the same result.

Third, the level of difficulty of the test items. The level of difficulty of items is the proportion of students who answered an item correctly over the total candidates. Items can be considered as good items if it is not too difficult or not too easy. Items that are too easy will cause students not to increase their effort to solve the problems. However, if the items are too difficult, it will cause the students to become unmotivated. The item that belongs to the moderate category has a difficulty index from 0.300-0.700. Based on the analysis, the level of difficulty of every item in the English final tests for seventh-grade students at SMPN 24 Banjarmasin academic year 2019/2020 shows that 8 items (16%) are difficult, 22 items (44%) are moderate, and 20 items (40%) are easy. The results of the analysis show that most of the items are considered moderates. It means that the level of difficulty of the English final test is good.   The items that categorized as easy or difficult level should be revised for future test.

Fourth, the discrimination power of the test items. Discrimination power is the item's ability where the item's scores can distinguish a group of master students from non-master students. The result of the analysis showed that 1 (2%) items considered as very poor items, 8 items (16%) considered as poor items, 11 items (22%) considered as satisfactory items, 25 items (50%) considered as good items, and 5 items (10%) considered as very good items. The research results showed that most of the items were categorized as very good, good, and enough to be retained. In contrast, items with poor and very poor discrimination power should be discarded or revised.

Fifth, the distractor of the test items. Item distractor obtained with less count the number of testee who choose answer A, B, C, D or not choose any answer at all. The distractor called working if chosen by at least 5% of the students. There are 150 students' answer sheets, so the distractor is called working when selected at least 8 students. The result showed that 27 items (54%) have very good distractor efficiency, 12 items (24%) have good distractor efficiency, 8 items (16 %) have bad distractor efficiency, and 3 items (6%) have very bad distractor. The percentage of items showed that 11 items (22%) have malfunction distractors. Items that have malfunction distractor should be discarded or revised with the other distractor that interest the students.

The analysis of validity, reliability, level of difficulty, discrimination power, and item distractor aimed to determine the quality of the English final test for seventh grade students at SMPN 24 Banjarmasin Academic Year 2019/2020. Gronlund (1981) stated there are following considerations to determine the quality of the English final test item; first, the item is said good if the item has four criteria (validity, level of difficulty, discrimination power, and item distractor). Second, the item is said less good if the item has three criteria from all four criteria. Third, the item is said not good item if the item has two criteria or did not meet all the criteria of good items. Based on the analysis, the results of the test item quality of the English final test for seventh-grade Student at SMPN 24 Banjarmasin Academic Year 2019/2020 are 20 items (40%) have not good quality, 12 Items (24%) have less good quality, and 18 items (36%) have good quality. The items that have less good quality should be revised. However, the items that have not of good quality must be discarded.

Compared to Hanafi's (2012) research, this research obtained almost the same results. In Hanafi's case, the successful item that has met all the criteria only 20 items (40%), while the item that does not meet one or more criteria is 30 items (60%). However, the successful item from this research only 18 items (36%), while the item that does not meet one or more criteria is 32 items (64%).

In addition, the researcher realized that this research contains weaknesses. It is descriptive research related to validity, reliability, level of difficulty, discrimination power, and distractor efficiency of the English final test. This research will be more useful if the researcher helps the teacher redesign the less good and not good English final test items.

CONCLUSION

Based on the analysis of 50 items in the English final test for 7th grade students at SMPN 24 Banjarmasin academic year 2019/2020 lead to the following conclusions.

First, the validity of the English final test for seventh grade students at SMPN 24 has good content validity because all of the items are in harmonic with the learning objective in the syllabus. Moreover, the English final test items also have good item validity because the number of valid items were greater than the number of invalid items.

Second, in terms of reliability, the English final test item has high reliability, with *Cronbach* alpha is 0.800.

Third, in terms of level difficulty, the English final test item still needs some revision, with a total of easy and difficult items are 28 items (56%).

Fourth, The English final test items have good discrimination power. It was shown by 41 items (82%) that can discriminate master students with non-master students.

Fifth, in terms of distractor efficiency, some of the English final test item need revision because some of the distractor are too flashy.

Sixth, in terms of the quality of the test items, the overall quality of the English final test for 7th grade students are not good because 32 items (64%) are considered unsuccessful items.

The result of the present study indicates that some items need to be revised or discarded. Moreover, it was important for the teacher to construct an appropriate test.

REFERENCE

Arifin, Z. (2012). *Evaluasi Pembelajaran.* Bandung: PT. Remaja Rosdakarya.

Arikunto, S. (2013). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.

Brown, H. D. (2004). *Language assessment and classroom practice*. Pearson Education Limited.

Gronlund, N.E. (1977). *Constructing Achievement Test,* 2nd Edition. New York: Prentice Hall, Inc. University of Illinois.

Gronlund, N.E. (1981). *Measurement and Evaluation in Teaching (Fourth Edition)*. New York: Macmillan.

Hanafi, A. (2012). Karakteristik Butir Soal UN Bahasa Inggris SMK di Kabupaten Tabalong

Kalimantan Selatan Tahun Akademik 2010/2011. *Jurnal LingTera*. Universita Negeri Yogyakarta, Yogyakarta. Retrieved from https://journal.uny.ac.id/index.php/ljtp/article/view/2465

Heaton, J.B. (1988). *Writing English Language Tests*. New York: Longman

Hughes, A. (2005). *Testing for Language Teachers (Second Edition)*. Cambridge: Cambridge University Press.

Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendekia.

Singarimbun, M & Sofian, E. (2008). *Metode Penelitian Survey*. Jakarta: LP3ES.

Sudijono, A. (2011). *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada.

Sudjana, N. (2011). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya.

Sugianto, A. (2017). Validity and Reliability of English Summative Test for Senior High School. *Indonesian EFL Journal, 3(2), 22-38*. Retrieve from https://www.researchgate.net/publication/325425907_VALIDITY_AND_RELIABILITY_OF_ENGLISH_SUMMATIVE_TEST_FOR_SENIOR_HIGH_SCHOOL

Surapranata, S. (2004). *Panduan Penulisan Tes Tertulis* (p. 2004). Bandung: Remaja Rosdakarya.

Zaman, A., Niwaz, A., Faize, F. A., Dahar, M. A., & Alamgir. (2010). Analysis of multiple choice items and the effect of items' sequencing on difficulty level in the test of mathematics. *European Journal of Social Sciences*, *17*(1), 61–67. Retrieve from https://www.researchgate.net/publication/289572600_Analysis_of_multiple_choice_items_and_the_effect_of_items'_sequencing_on_difficulty_level_in_the_test_of_mathematics