

CHAPTER IV

FINDING AND DISCUSSION

This chapter presents the findings of this research. It deals with the presentation of the data, data analysis, and discussion about data obtained from observation, and documentation. These data analysis and discussion are presented descriptively.

A. Research Finding

The important thing in this research is the concept of the research that was held. The concept of this research is descriptive research in the field of education. Based on the result of observations through interviews and consultation to the headmaster and English teacher, the researcher got the information a condition of school, especially in grade twelfth about English test. The researcher analyzed the English items of second exam of MAN 3 Banjarmasin. To get the data, she took the students' answer sheet, English items test and answer key to analyzed. the total numbers of test items are 40 items which consist of 40 multiple choice items. The test held on 21 April 2018 with the given time 120 minutes started 10.30 until 12.30 a.m.

This research was conducted to determine the quality of the English items in second semester exam grade XII at MAN 3 Banjarmasin in academic year 2017/2018 reviewed by the aspects of validity, reliability, level of difficulty, item discrimination, and distractor efficiency. The data is analyzed by Microsoft Excel. These are the result of documentation:

1. Validity

The validity testing in the form of multiple choice items can be done using *correlation point biserial* formula. There are 223 students of grade XII at MAN 3 Banjarmasin who take the test. But researcher only took half them, then took 27% from them where 27% is upper group and 27% is lower group. The results of the calculation were consulted to r_{table} at a significance level of 5%. So, there were 60 students of each group, 30 is upper group and 30 is lower group.

Students who take exam as many as 60 students so get the r_{table} standard at 0,254 and if $r_{pbis} \geq 0,254$ can mean the item is valid and if the result is $\leq 0,254$ means the item invalid. Items in the form of multiple choice question is said as valid totaled 31 items or 77% and items which declared invalid amounted to 9 items or 22%. Distribution of English items based on its validity standards that have been defined are as follows:

Table 3. Distribution of English items of second semester exam of grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 based on the validity index.

N	Validity Index	Item Number	Total	Percentage
	$\geq 0,254$ (Valid item)	1, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37, 39, and 40.	31	77%
	$\leq 0,254$ (Invalid item)	2, 8, 14, 22, 23, 24, 29, 30, and 38.	9	22%

Source: Primary Data Processed

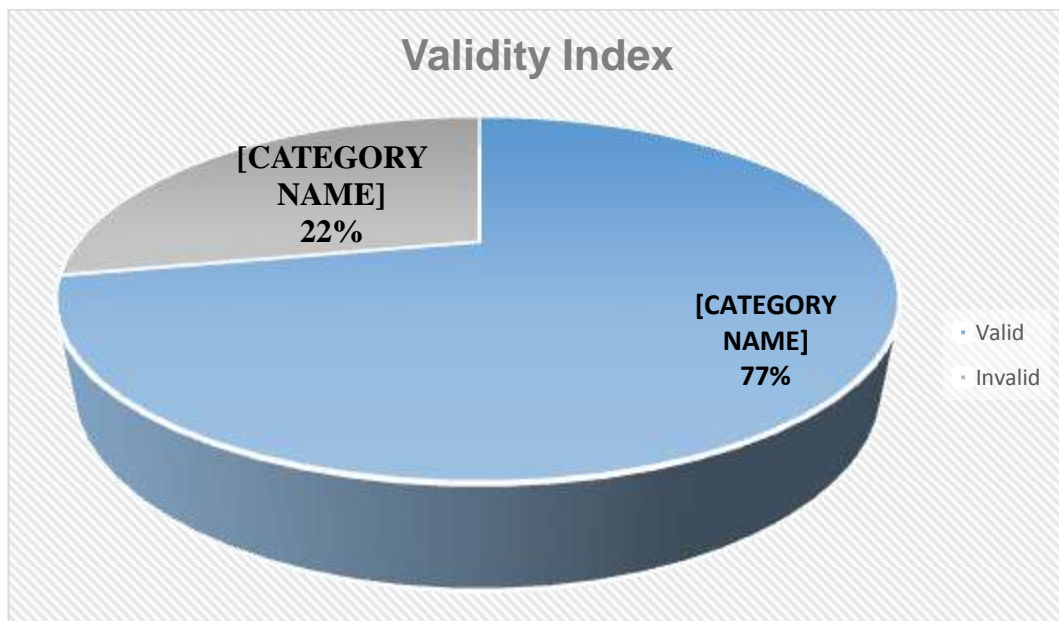


Figure 1. Distribution of English items of second semester exam of grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 based on the validity.

2. Reliability

Result of this research and analysis on the reliability of English items of second semester exam grade XII at MAN 3 Banjarmasin in academic year 2017/2018 is based on the validity index benchmark by $r_{11} \geq 0,70$, the item has a high reliability. While if $r_{11} \leq 0,70$ the item has a low reliability. From the overall calculation, known that English item of second semester exam grade XII at MAN 3 Banjarmasin 2017/2018 has high $r_{11} \geq 0,70$ or equal to 2 so that the item is said to be reliable.

3. Level of Difficulty

The classification used to interpret the calculation results of the level of difficulty is 0,00 to 0,30 including difficult category; 0,30 to 0,70 including medium

category; and 0,70 to 1 including the item categorized as easy. Based on the analysis of the result, it showed that the number of English item test categorized as difficult item category to 6 items or 15%, , medium category 21 items or 52,5%, and easy category 13 or 32,5% items. Distribution of English item based on the level of difficulty is as follows:

Table 4. Distribution of English items of second semester exam of grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 based on the level of difficulty.

No.	Level of Difficulty	Item Number	Total	Percentage
1.	0,00 – 0,30 (Difficult)	2, 6, 22, 24, 25, and 38.	6	15%
2.	0,30 – 0,70 (Medium)	3, 4, 5, 7, 10, 11, 14, 16, 17, 18, 19, 20, 21, 23, 26, 30, 32, 33, 34, 36, and 37	21	52,5%
3.	0,70 – 1,00 (Easy)	1, 8, 9, 12, 13, 15, 27, 28, 29, 31, 35, 39 and 40.	13	32,5%

Source: Primary Data Processed

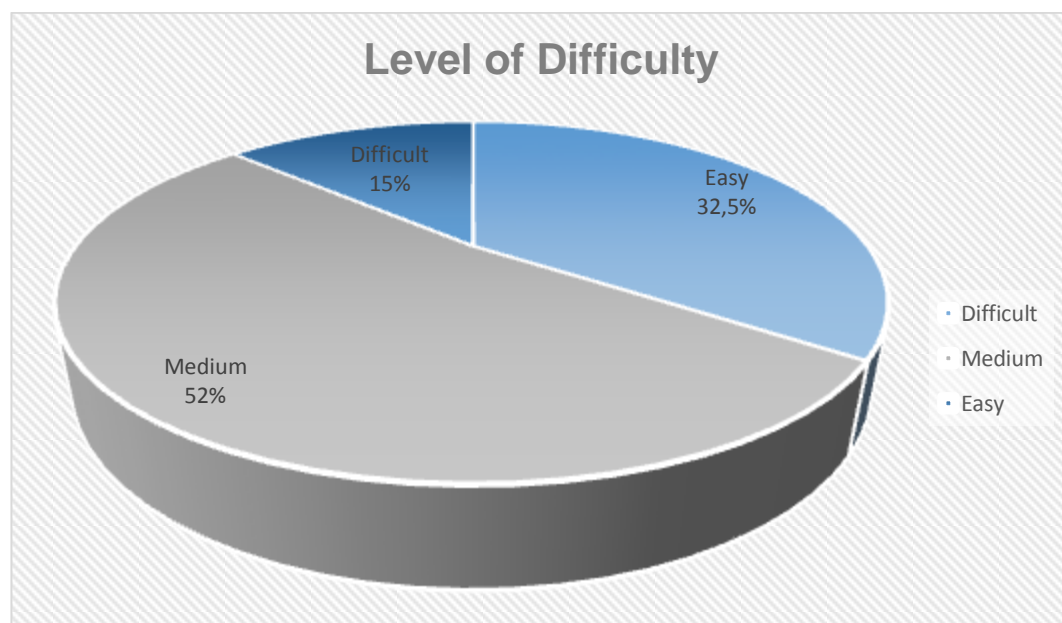


Figure 2. Distribution of English items of second semester exam of grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 based on the level of difficulty.

Based on the analysis above, the English item of second semester exam grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 can be said as good because the items were not too difficult and not too easy domination. Even though there were 6 six difficult items and 13 easy items.

4. Item Discrimination

Classification used to interpret the calculation results of discrimination power is 1,00 to 0,70 are included in the category of excellent; 0,70 to 0,40 are included in the category of good; 0,40 to 0,20 are included in the category of satisfactory; and 0,20 to 0,00 are included in the category of poor. Based on the analysis result, the number of English item which have excellent categories are 3 items or 7,5%; good categories are 15 items 37,5%; satisfactory categories are 13 items or 32,5%; poor categories are 7 items or 17,5%; and worst categories are 2 items or 5%. Distribution of English item based on item discrimination is follows:

Table 5. Distribution of English items of second semester exam of grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 based on item discrimination.

No.	Item Discrimination	Item Number	Total	Percentage
1.	1,00 – 0,70 (Excellent)	3, 5, and 37.	3	7,5%
2.	0,70 – 0,40 (Good)	4, 7, 8, 9, 12, 13, 14, 15, 17, 28, 30, 31, 33, 34, and 36.	14	37,5%
3.	0,40 – 0,20 (Satisfactory)	1, 6, 10, 11, 16, 18, 19, 20, 21, 25, 26, 32, and 35.	13	32,5%
4.	0,20 – 0,00 (Poor)	2, 22, 23, 27, 29, 39, and 40	7	17,5%
5.	$\leq 0,00$ (Worst)	24 and 38	2	5%

Source: Primary Data Processed

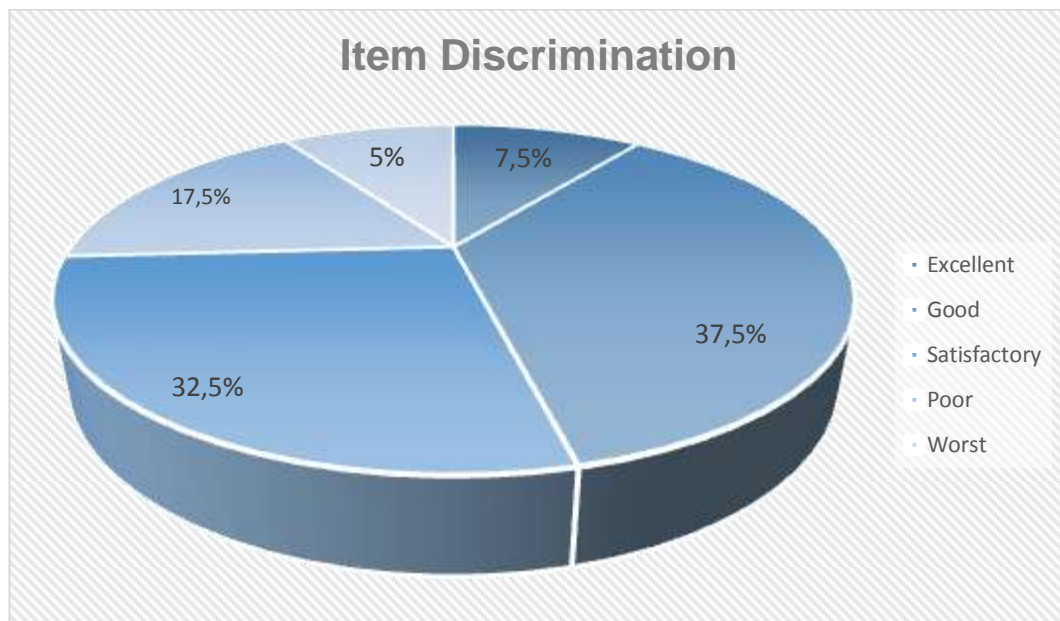


Figure 3. Distribution of English items of second semester exam of grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 based on item discrimination.

5. Distractor Efficiency

The distractor efficiency obtained by calculate the number of students who choose answer a, b, c, d, or e. From the distractor efficiency can be determined distractors may malfunction or not to work. Distractors are functioning properly at least chosen by 5% of all test participants. The result showed that 11 items or 27,5% had distractors that works excellent, 15 items or 37,5% had good function distractors, 9 items or 22,5% had distractors that works not bad, 4 items or 10% had less good distractors, and 1 item or 2,5% had bad distractor. Distribution of 40 items based on distractors efficiency are as follows:

Table 6. Distribution of English items of second semester exam of grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 distractor efficiency.

No.	Distractor Efficiency	Item Number	Total	Percentage
1.	Excellent (0)	2, 3, 5, 6, 19, 23, 32, 33, 36, 37, and 38.	11	27,5%

2.	Good (1)	4, 7, 10, 12, 13, 17, 18, 20, 21, 22, 25, 26, 30, 34, and 39.	15	37,5%
3.	Not Bad (2)	8, 14, 15, 16, 24, 28, 31, 35, and 40.	9	22,5%
4.	Less Good (3)	1, 9, 11, and 29.	4	10%
5.	Bad (4)	27	1	2,5%

Source: Primary Data Processed

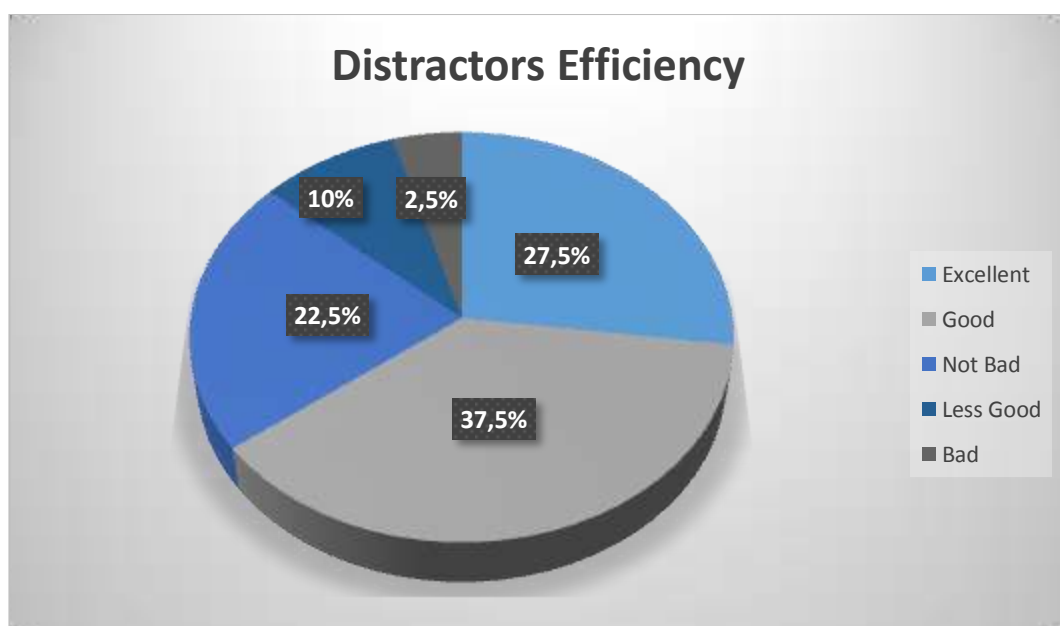


Figure 4. Distribution of English items of second semester exam of grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 based on distractor efficiency.

Based on level of difficulty and item discrimination of the test, there are 11 items that did not meet criteria both in level of difficulty and item discrimination. They are items number 2, 6, 22, 23, 24, 25, 27, 29, 38, 39, and 40.

B. Discussion

Quality of English item of second semester exams grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 can be seen through some aspects. Those

are: validity, reliability, level of difficulty, item discrimination, and distractor efficiency. The following discussion of each of the aspects as follow:

1. Validity

A test can be said to be valid if the test can be measure the object that should be measured. In this research, the validity of the items was calculated using the formula correlation point biserial r_{pbi} obtained from the calculation which further consulted to r_{table} at a significance level of 5%. According Gay (1081, p. 110) states that the validity is the degree to which a test measures what is supposed to measure and consequently, permits appropriate interpretation score. Result of this research in MAN 3 Banjarmasin to English item second semester exams grade XII shows that students amounted 60 participants so that $n= 60$, r_{table} showed 0,254. If r_{pbi} more than or same with r_{table} then the item was declared valid. Conversely, if r_{pbi} less than r_{table} the item was declared invalid.

The result showed that the items were declared valid are 31 items, while the item are invalid aggregate 9 items. The result of this study when compared with Dwi Rusmiana (2015) which indicates a valid items total to 16 items or 40% and the category is invalid 24 items 60%, the study result is not the same. The total of English item valid less than the total of items that were invalid. Based on description above, it can be concluded that the English item of second semester exams grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 including invalid items should be improved by increasing technical mastery of the preparation of the test items. The items were valid because its construction is good and includes material truly

represent the target of measurement. Teacher or item maker could ask the opinion of expert to establish the validity of the items that have been made. Thus, the validity is a characteristic that are important tests. The invalid items need to be eliminated or revised and the activity should be truly conducted by the teacher in order to be suitable with normal validity index of a high-quality test. This information should let test constructor to master the items which work on the ability of those items to measure what are supposed to measure.

2. Reliability

Reliability is the level or degree of stability or consistency of a test to measure so that it can be trusted. Referring to the result of data elaboration, the result of reliability of these test items calculated by using the *split half* formula. Interpretation of reliability coefficient r_{11} . When $r_{11} \geq 0,70$ the item being tested has a high reliability and can be said as reliable. But if $r_{11} \leq 0,70$ the item being tested has a low reliability and can be said as unreliable.

The result showed that reliability of English item second exams grade XII at MAN 3 Banjarmasin was 2. Based on the result of these calculations can be concluded that English item second semester grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 included as the test that is high reliability or $r_{11} \geq 0,70$. Basically, it is the degree to which a test consistently measures whatever it is measuring. It is completely in same assumption with Heaton's point of view (1988, p. 162) that reliability is the extent to which the same marks or grades are awarded if the same test papers are marked by two or more different examiners or the same

examiner on different occasion. Shortly, to be reliable, a test must be consistent in its measurement.

3. Level of Difficulty

Level of difficulty of items is the proportion of the total students who answered an item correctly by the total number of students who take the test. Items can be expressed as a good item if tested not too difficult or too easy, but that is the good items will be categorized as medium. Too easy items will not stimulate students to solve that items. Conversely items that is too hard will make students feel desperate and do not have the power to try again because beyond the reach of the students. Items that fall into the category of medium has an index of 0,31 – 0,70.

The result showed that the items in the form of multiple choice items which categorized as difficult 6 items or 15%, medium category 21 items or 52,5%, and easy category 13 items or 32,5%. Items in the category of medium can be included into question bank for reuse as an evaluation tool in the future. Items that easy or difficult can be re-examined to determine the cause of the question to be easy or difficult, so can revised and tested on the next test.

A good test is a test which is not too easy or too difficult to students. It should give optional answer that can be chosen by students and not too far by the key answer. Very easy items are to build in some affective feelings of “success” among lower ability students and to serve as warm up items. In the other hand, very difficult items can provide a challenge to the highest-ability students (Brown, 2004, p. 59). It makes students know and record the characteristics of teacher’s test if the test given

always comes to them too easy and difficult. Thus, the test should be standard and fulfill the characteristics of a good test. The number that shows the level of difficulty of a test can be said as difficulty index (Arikunto, 2006, p. 2017).

a. Item Analysis based on easy and difficult

Table 5. An error analysis of English items of second semester exam of grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018

No	Item	Analysis
2.	Whale is huge. Tuna is Tiny. What is the best combination of the two sentences. a. However whale is huge, tuna is tiny. b. Whilst whale is huge, tuna is tiny. c. Nevertheless whale is huge, tuna is tiny. d. Whale is huge even if tuna is tiny. e. Whake is huge if tuna is tiny.	All distractors have a similar sound.
6.	Snow leopards are solitary in nature. Gray wolves live in group. What is the best combination of the two sentences? a. Even though snow leopards are solitary in nature, gray wolves live in a group. b. Nonetheless snow leopards are solitary in nature, gray wolves live in group. c. Snow leopards are solitary in nature unless gray wolves live in group. d. Snow leopards are solitary in nature; however, gray wolves live in group. e. Snow leopards are solitary in nature if gray wolves live in group.	Complex vocabulary which not all of students understand the meaning.
22.	He decided to go..... a. In spite of I begged him not to. b. Nevertheless I begged him not to. c. But I begged him not to anyway. d. Even though I begged him not to. e. Despite I begged him not to.	Less a verb in every options
24.	They're coming next week,.....I don't know which day. a. Although b. Despite c. In spite d. Much e. Whatever	Not parallel options
25., she is taking a German language course. a. Despite the fact that she has to cope with two small daughters. b. But she has to cope with two small daughters.	Unfamiliar word (low frequency word)

	<ul style="list-style-type: none"> c. She has to cope with two small daughters. d. In spite of she has to cope with two small daughters. e. However she has to cope with two small daughters. 	
27.	<p>“Oscar soon comes to realize that his claim may have serious consequences.” (Paragraph 2) The synonym of the underlined word is....</p>	Not clear instruction. No underlined word in this item.
38.	<p>I can now see how if it would have upset you to return after the holiday weekend and find your belongings moved for no <u>apparent</u> reason. The synonym of the underlined word is.....</p> <ul style="list-style-type: none"> a. Good b. Clear c. Suitable d. Specious e. Acceptable 	Unfamiliar word
39.	<p>The text give us.....</p> <ul style="list-style-type: none"> a. Explanation how to tie dye old T-shirt b. Instruction how to tie dye old T-shirt c. Information about tie dyeing d. Description about tie dyeing e. Reusing an old T-shirt 	Instruction not clear. Not parallel options.

Based on the table above, it can be seen that 11 items that did not meet criteria both in level of difficulty and item discrimination. Item number 24, option D was too short compared to the other options. Options D was only one word with one syllable. Item number 39, option E was not parallel with the other options. Option E contained verb, while the other options contained noun.

4. Item Discrimination

Discrimination index the ability of items where the scores can distinguish participants from the highest to the lowest group. Interpretation of item discrimination from 0,00 to 0,19 are included in the category of poor; 0,20 to 0,39 are included in the category satisfactory; 0,40 to 0,69 are included in good categories; and 0,70 to 1,00 are included in the excellent category. The result showed that English items second semester exams grade XII at MAN 3 Banjarmasin which

have excellent categories are 3 items or 7,5%; good categories are 15 items 37,5%; satisfactory categories are 13 items or 32,5%; poor categories are 7 items or 17,5%; and worst categories are 2 items or 5%.

Result of this research are in accordance with the theoretical review of Anas Sudjono (2011) that item discrimination is the ability of an item of achievement test to be able to distinguish between students who have a high capability with the low capability. Hence, the item discrimination of English items second semester exams grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 said to be good because of items are in category of satisfactory, good, even excellent are dominant. Bad items in poor and worst categories should be discarded and which has yet to reach good category should be revised and looking for the cause.

5. Distractor Efficiency

Any distractor that has been selected by less than 5% of the students is considered to be a non-functioning distractor (NF-D). Total of items in English item of second semester exam grade XII at MAN 3 Banjarmasin 40 items. Meanwhile, there are 5 distractors alternative. Those are a, b, c, d, and e. So, there are 200 distractors all of items and 40 is right distractors.

Then, by used *Likert Scale* it is also interpret: Excellent category if all distractors are functioning properly. Good category if item just has 1 distractor malfunction because chosen by less than 5% of all students. Satisfactory category if item has 2 distractors malfunction because chosen by less than 5% of all students. And item included in poor category if it has 3 distractors malfunction. Items are categorized as not good or worst category if an item of the whole distractors can not

function (malfunction) properly. Based on the description that has been described, it can be concluded that the English item second semester exam grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 based on distribution distractor efficiency included in a pretty good item. The purpose of the investigation was to discover the relative rank order to of efficiency of eight distractor categories in multiple-choice. Brown and Priyanvada (2010, p. 72) said that the distractors are responses in a multiple choice item used to divert the test-taker from the correct response.

6. Item Analysis

The overall results of test English item analysis of the second semester exams grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 based on validity, level of difficulty, item discrimination, and distractor efficiency. These are the overall results of English item second semester exams grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 as follow:

Table 7. The overall results of English item second semester exams grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018 based on validity, level of difficulty, item discrimination, and distractor efficiency.

No.	Criteria	Item Number	Total	Percentage
1.	Good (4 Categories)	3, 4, 5, 7, 10, 15, 16, 17, 18, 19, 20, 21, 33, 34, and 36.	15	37,5%
2.	Less Good (3 Categories)	1, 6, 11, 12, 13, 14, 23, 25, 26, 28, 30, 31, 32, 37, 39, and 40.	16	40%
3.	Not Good (≤ 2 Categories)	2, 8, 9, 22, 24, 27, 29, 35, and 38.	9	22,5%

Source: Primary Data Processed

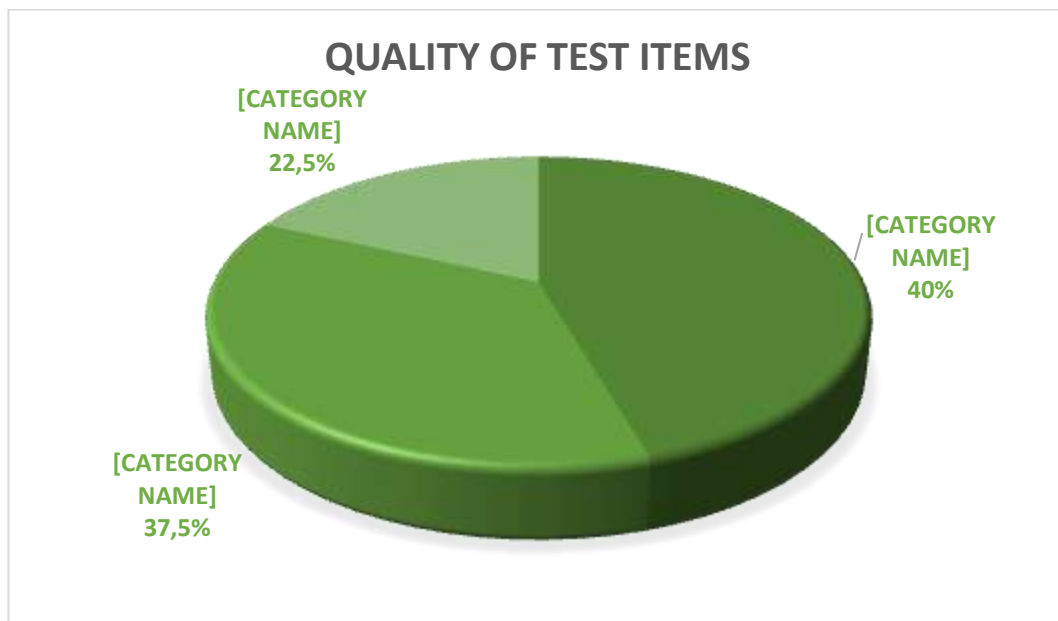


Figure 5. The overall results of English item quality analysis of second semester exams grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018.

Based on the table can be seen that the English items quality of second semester exams grade XII at MAN 3 Banjarmasin in academic year 2017/2018 in form of multiple choice questions there are 15 items or 37,5% in the good category, 16 items or 40% in the less good category, and 9 items or 22,5% in the not good category. Good items will be put into question bank and can be used in the next test, while for items because it were not fulfill one or more predetermined quality standards. Failure items that became less good and not good can be seen in the following table:

Table 8. The cause of the English items failure in second semester exams grade XII at MAN 3 Banjarmasin in Academic Year 2017/2018.

No.	Criteria	Item Number	Total	Percentage
1.	Validity (Invalid)	2, 8, 14, 22, 23, 24, 29, 30, and 38.	9	22,5%
2.	Level of Difficulty	1, 2, 6, 8, 9, 12, 13, 15, 22, 24, 25,		

	(Easy and Difficult)	27, 28, 29, 31, 35, 38, 39, and 40	19	47,5%
3.	Item Discrimination (Poor and Worst)	2, 22, 23, 24, 27, 29, 38, 39, and 40	9	22,5%
4.	Distractor Efficiency (Less Good and Not Good)	1, 9, 11, 27, and 29.	5	12,5%

Source: Primary Data Processed

It is concluded that the main cause of failure of English item mostly in the level of difficulty which mean that the items are used still too easy or difficult. The second cause of failure is in the item discrimination and validity level, which means that the item is invalid and does not have suitability or alignment direction to the total score. Less good items need to be re-investigated on every category and tested again also be re-evaluated after being repaired. The good item will be stored in the question bank and can use in the next test, but if there is no change it should discarded, while not good items should deleted and not reused in the next exam.

Item analysis uses statistics and expert judgement to evaluate tests based on the quality of individual items, item sets, and entire sets of items, as well as the relationship of each item to other items. According to Thompson and Levitov (1985, p. 163) said that it investigates the performance of items considered individually either in relation some external criterion or in relation to the remaining items of the test. It uses this information to improve item and test quality. Item analysis concepts are similar for norm-referenced and criterion-referenced tests, but they differ in specific, significant ways.

With the criterion-referenced tests, use norm-referenced statistic for pretest data and criterion-referenced statistics for posttest data. There are five categories to know how well the quality of test items. The most critical dimension is validity. Simply stated, (Anastasi, 1954, p. 203) validity is what a test measures and how well it does this. Validity is a crucial consideration in evaluating test. Second, reliability is a test score that includes a true score and random error. It is which the best single measure of a test accuracy, is the extent to which test results are consistent, stable, and free of error variance. Third, item difficulty is a characteristic of the item and the sample that take the test. It is a percentage of people who answer an item correctly. It is the relative frequency which students choose the correct response (Thorndike, Cunningham, 1991, p. 75). Then, item discrimination compares the number of high scores and low scores who answer an item correctly. It is the extent to which items discriminate among students in the high and low groups. The total test and each item should measure the same thing. The last is distractor efficiency, which is one component of multiple-choice. A discrimination power should be obtained for each option to determine each distractor's usefulness. Test developers either drop or revise items which have negative point-biserial correlations (Osterlund, 1998).