CHAPTER IV
RESULT AND DISCCUSION

This chapter describes the research result and discussion based on the methodology and the data obtained. In the obtaining, the data the writer used the two aspects of this research. Those were qualitative that analysis of the test appropriateness with the national curriculum and the syllabus of seventh grade students then the appropriateness with the students levels in ESL. The next aspect was the quantitative analysis included the items difficulty levels, discrimination power, validity and reliability of the test.

4.1 The Result and Discussion of the Research

The English final test that the researcher analysis was apply in south district Gresik in 2016/2017 academic years. This test consisted with fifty questions of the multiple-choice form. The test was consisting of the five aspects, which fulfill of the requirements in basic assessment in ESL. The five aspects were speaking, listening, writing, reading and structure. The questions number that included in speaking was five. Those were number 7, 8, 9, 37, 38 and 39. The next, which fulfill in the listening, was 6. They were numbers 1 until 6. The requirement in writing was numbers 16, 17, 18, 19, 20, 30, 33, 34, 35, 39, 41, 43, 48, 49, and 50 and the total was fifteen numbers. The third aspect was reading. Those were numbers 10, 11, 14, 15, 21, 22, 23, 24, 25, 26, 27, 29, 40, 42, 44, 45, 46 and 47. The total was nineteen numbers. The numbers the appropriate with structure's aspect were numbers 12, 13, 31, 32 and 36. In the test were created with using the dialog, description texts, fill in the blank, picture, tree diagram to explain the themes which included in the basic competence on the syllabus usage at seven grade students.

The English final test was usage in 43 – 50 schools (MTS) in south district Gresik. The obtained data the researcher was take from random sampling which requirements schools of each sub-district. The three samples were namely AS for the first sample, BS of second sample, and CS to third sample. The analysis of the data to gets the aim from this study. Moreover, the aspects to analyzing were qualitative and quantitative of the inquiry. The qualitative inquiries were the test appropriateness regarding to basic competence in syllabus using the curriculum checklist and the appropriateness toward students' levels in ESL with a good assessment checklist based on Brown and Heaton theories. Next inquiry was quantitative. The quantitative were analyzing of the difficulty level, discrimination power, validity and reliability test in the items test. Through this analyze, the researcher gets the evidenced about inadequacy the items to find the ways to revised them. The researcher attained the data analyze as resulting.

4.2   The Test Appropriateness Regarding to Curriculum and Syllabus

After get the paper questions test, the writer adapted the form of curriculum checklist from the syllabus and the form of checklist for selection the content and regarding in ESL from Cunningsworth (2008:58-61). The first result was curriculum checklist consists of two standard competences and eight basic competences. Those are points 3 and 4 of the standard competence and points 3.1 until 4.4 of the basic competence. The result of the fifty questions can be distinguished by the competence standard and basic competence from the syllabus in K13 that implementing in 2016/2017 academic years in South District Gresik. To summarize more the table 3.5.1 the curriculum checklist result on the appendix 1.

The table 3.5.1 the curriculum checklist, there was six columns. The first column is the standard competence. The standard basic was explains to understanding of the matter based on students' felling of the knowledge, technology, art, culture phenomenon and the real fact in point 3 and to trying, managing, and present about the concert data to calculating, reading, draw with the lesson in the school or the others resource. Then, the column of basic competence was about some points included on the matter. The third column was about the themes or materials that included the basic competence. Then, column of items test that appropriated with basic competences. The fourth column of the total number of the items test, and the last column about the percentage of the total numbers of particular items represent the elated basic competence. To summarize more on the table result as follows:

Table 7 - Table 4. 2 The Result of Test Appropriateness Regarding To Basic Competence

Competence Standard	Basic Competence	Themes / Materials	Items Test that Appropriate with the Basic Competence	Total number of items test (Σ)	Percentage of Total Numbers of Particular Items Represent the Elated Basic Competence	Checklist (√)
Memahami pengetahuan (faktual, konseptual, dan prosedural) berdasarkan rasa ingin tahunya tentang ilmu pengetahuan,	3.1 Mengidentifikasi fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi interpersonal lisan dan tulis yang melibatkan tindakan	Interaksi antara peserta didik di dalam di luar kelas yang melibatkan tindakan menyapa, berpamitan, berterimakasih, meminta maaf yang	1, 2, and 3	3	6%	√

teknologi, seni, budaya terkait fenomena dan kejadian tampak mata.		menyapa, berpamitan, mengucapkan terimakasih, dan meminta maaf, serta menanggapinya, sesuai dengan konteks penggunaannya	dapat menumbuhkan perilaku yang termuat di KI				
	3.2	Mengidentifikasi fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi transaksional lisan dan tulis yang melibatkan tindakan memberi dan meminta informasi terkait jati diri, pendek dan sederhana, sesuai dengan konteks penggunaannya. Perhatikan unsur kebahasaan dan kosa kata terkait hubungan keluarga; pronoun (subjective, objective, possessive)	Deskripsi diri sendiri sebagai bagian dari keluarga: ayah, ibu, kakak, adik,yang dapat menumbuhkan perilaku yang termuat di KI	7 & 47	2	4%	√

	3.3 Mengidentifikasi fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi transaksional lisan dan tulis yang melibatkan tindakan memberi dan meminta informasi terkait nama hari, bulan, nama waktu dalam hari, waktu dalam bentuk angka, tanggal, dan tahun, sesuai dengan konteks penggunaannya. (Perhatikan kosa kata terkait angka kardinal dan ordinal)	Waktu kejadian/peristiwa/ kegiatan terkait kehidupan di sekolah, rumah, dan lingkungan sekitar peserta didik yang dapat menumbuhkan perilaku yang termuat di KI	4, 18, 19, and 20	4	8%	√
	3.4 Mengidentifikasi fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi transaksional lisan dan tulis yang melibatkan tindakan memberi dan meminta informasi terkait nama dan jumlah binatang, benda, dan bangunan publik yang dekat dengan kehidupan peserta didik sehari-hari, sesuai dengan	Benda, binatang, dan bangunan yang biasa dijumpai dalam kehidupan nyata di rumah, sekolah, dan lingkungan sekitar peserta didik yang dapat menumbuhkan perilaku yang termuat di KI	21, 22, 23, 31, 37, 38, 39, and 48	8	16%	√

	konteks penggunaannya. (Perhatikan unsur kebahasaan dan kosa kata terkait article a dan the, plural dan singular)					
4. Mencoba, mengolah, dan menyaji dalam ranah konkret (menggunakan, mengurai, merangkai, memodifikasi, dan membuat) dan ranah abstrak (menulis, membaca, menghitung, menggambar, dan mengarang) sesuai dengan yang dipelajari di sekolah dan sumber lain yang sama dalam sudut pandang/teori.	4.1 Menyusun teks interaksi interpersonal lisan dan tulis sangat pendek dan sederhana yang melibatkan tindakan menyapa, berpamitan, mengucapkan terimakasih, dan meminta maaf, dan menanggapinya dengan memperhatikan fungsi sosial, struktur teks, dan unsur kebahasaan yang benar dan sesuai konteks.	Interaksi antara peserta didik di dalam di luar kelas yang melibatkan tindakan menyapa, berpamitan, berterimakasih, meminta maaf yang dapat menumbuhkan perilaku yang termuat di KI	5, 6, 41, and 50	4	8%	√
	4.2 Menyusun teks interaksi transaksional lisan dan tulis sangat pendek dan sederhana yang melibatkan tindakan memberi dan	Deskripsi diri sendiri sebagai bagian dari keluarga: ayah, ibu, kakak, adik,yang dapat menumbuhkan perilaku yang termuat di KI	8, 9, 10, 11, 12, 13, 24, 25, 26, 27, 28, 29, 33, 34, 35, 40, 42, and 46	18	36%	√

	meminta informasi terkait jati diri, pendek dan sederhana, dengan memperhatikan fungsi sosial, struktur teks, dan unsur kebahasaan yang benar dan sesuai konteks				
	4.3 Menyusun teks interaksi transaksional lisan dan tulis sangat pendek dan sederhana yang melibatkan tindakan memberi dan meminta informasi terkait nama hari, bulan, nama waktu dalam hari, waktu dalam bentuk angka, tanggal, dan tahun, dengan fungsi sosial, struktur teks, dan unsur kebahasaan yang benar dansesuaikonteks	Waktu kejadian/peristiwa/ kegiatan terkait kehidupan di sekolah, rumah, dan lingkungan sekitar peserta didik yang dapat menumbuhkan perilaku yang termuat di KI	14, 15, 16, 17, 30, 32, 43, 44, 45, and 49	10	20%	√
	4.4 Menyusun teks interaksi transaksional lisan dan tulis sangat pendek dan sederhana yang melibatkan tindakan memberi dan meminta informasi	Benda, binatang, dan bangunan yang biasa dijumpai dalam kehidupan nyata di rumah, sekolah, dan lingkungan sekitar peserta didik yang	36	1	2%	√

		terkait nama dan jumlah binatang, benda, dan bangunan publik yang dekat dengan kehidupan siswa sehari-hari, dengan memperhatikan fungsi sosial, struktur teks, dan unsur kebahasaan yang benar dan sesuai konteks	dapat menumbuhkan perilaku yang termuat di KI				

(Syllabus on 2013 curriculum based on revision)

The result based on the appendix 4.1 was identifying the social function, structure, and basic language in the oral and written communication of the themes in introduction, apologize, greeting, farewell with appropriate in the content. The numbers that included in 3.1 basic competences were numbers 1, 2, 3 with the percentage of 6%. The next of the point 3.2 basic competence, it was explain the identifying the social function, text structure, and linguistics of the oral and written communication to asking and giving self-information with the requirements in language which used in family of the basic linguistics. The items numbers that fulfill the requirement in the basic competence 3.2 are numbers 7 and 47 of the percentage 4%. The basic competence point 3.3 was categorizing the social function, text structure and linguistics about namely of day, month, time in numeral, date, year according to the content in the real life or the area of school, home and surrounds. The numbers of this basic competence were 4, 18, 19, and 20 with 8%. Then, the basic competence point 3.4. It was classifying the social function, text structure, and linguistics about transactional in oral or written of sharing and giving information namely of animal, thing, and the building in plural or singular word in the every days. Moreover, eight numbers in point 3.4 of the basic competence has 16%, the numbers of it was numbers 21, 22, 23, 31, 37, 38, 39 and 48.

Subsequently, the point 4.1 of the basic competence was explain to organized the interactional text in oral or written communication that included of the topics greeting, farewell, apologize with the social function, text structure, and linguistics in content. This basic competence has 8%. In addition, numbers were 5, 6, 41, and 50. Then, the basic competence 4.2 has 36% of total numbers 18. Those were numbers 8, 9, 10, 11, 12, 13, 24, 25, 26, 27, 28, 29, 33, 34, 35, 40, 42, and 46. The requirements of this basic competence was organize the interactional text in oral or written to gift and shared the self-information to introducing themselves as the member of the family that included with the text structure, content and linguistics. The point 4.3 of the basic competence description about the organization of interactional text in oral or written

communication to sharing and giving information of day, month, time, year, numeral, date with the text structure, linguistics, and content of language. Moreover, for this basic competence, there were ten numbers. Those were numbers 14, 15, 16, 17, 30, 32, 43, 44, 45, and 49. The percentage was 20% for this basic. The last basic competence was point 4.4. It was about the organized the interactional text in the oral or written communication of the information about animal, things, and building in our life. This basic had 2% for one number that includes was number 36.

4.3 The Test Appropriateness toward Students Levels in ESL

The good assessment checklist was to know what the items test appropriateness with the standard in ESL or not. The items test was good for apply in ESL. The researcher analyze from the fifteen number of the good assessments checklist, only five points do not fulfill the requirement. Moreover, 11 points in the items test made by MGMP in South District Gresik. Furthermore, the ten points make the tools of assessment can use to test the students' ability. In addition, the point's numbers 4, 8, 10, and 14 are not necessary. The detailed explanation as follow the table below:

Table 1 - Table 4.3 The Result of Test Appropriateness Toward students Levels

No.	Characteristics of Good Multiple Choice	Checklist (√)
1.	Presented out of simple context	√
2.	The Stem; the choices from which the students select their answers are referred to as option/responses/alternative	√

3.	The option in multiple choice are:
	   a. One option is correct option or key
	   b. Other option is distractors
		√

4.	Multiple choice items are all receptive, or selective, responses	X
5.	Design each item to measure a specify objective	√
6.	State both stem and options as simply and directly as possible	√
7.	Make certain that intended answer is clearly the only correct one	√
8.	Use item indices to accept, discard, or revise items	x
9.	Not need long time to answered the question or choose the correct one.	√
10.	All the multiple choice items should be at a level appropriate to proficiency level of education	x
11.	The questions or statements should be grammatically correct	√

12.	The stems should not give clues or question to other question.	√
13.	Chronological on the optional answer	√
14.	Only one feature at the time should be tasted	x
15.	Pictures, graphics, tables, and diagrams should be clear and in function	√

Based on the appendix 4.3 above, the researcher had finished the analyzing of the checklist of a good assessment in ESL. Indeed, the items test was present of the language and the themes were necessary with the matter and the levels of test-taken. The stem, the context of the question has the option / responses / alternative of choices from which the students select their answer and it has on this test-pack. The test-pack has the one correct answer and the others options of the distractor as the false answer. Then, the design for each item was to measure students' ability. Statement both the question (stem) and the options have the connection of the structure text. In addition, it has simply and clearly sentence for the students understand the meaning of the test. It not difficult for students chooses one correct option. According to the test-pack, the writer was thinking that not need must time to students answer the each questions. The grammar of each question is correctly and it has the structure content. Then, the each question not makes students thinking other options except in the option in the paper test. The picture and the chronology each questions was detail and clearly to students understand.

4.4 Difficulty Levels of the Items Test

Items difficulty level is the test to find the level of the item test-pack. The researcher analyzing the data from the three samples by the random sampling data those get from three sub-districts in South District Gresik. The formula to find the result from Sudjana (2011:137) to looks the index of the test-pack as follows:

.

$$I = B/N$$

The index percentage of students who gets the correct answered from fifty questions from three samples. The researcher divide the level of items test-pack addicted to three categories. Those were hard level of the index 0 – 0.30, medium level with the index 0.31 – 0.70, and easy level of the index 0.71-1. In details, the results of the data analysis as follow:

a. **The A Sample (AS)**

The first sample was calling the A sample (AS). The data has 28 students. The result data from fifty questions in this sample has two categories of the difficulty level. Those were hard and easy. The Items that can be classified into the difficult category that were six numbers of questions, namely the matter of numbers were 8 (0.036), 19 (0.11), 24 (0), 26 (0.036), 28 (0) and 47 (0.071). The next was the items that had easy as category. The numbers of questions that can be categorized were 44 numbers, and it was number 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 21, 23, 25, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48, 49 and 50. The difficulty level item from this sample was *0.77* that means the items test-pack from the category in this sample was easy. Then, the most index of hard level is *0*, and the index of easy level is *0.96*. Therefore, why the items test-pack this sample as call as easy categories. The explanation of computation of items difficulty level can see on *the appendix 2 of the first sample (AS) of items difficulty level*

Table 2 - Table 4.4 The Result of Difficult Level in the A Sample (AS)

NO.	QUESTIONS NUMBER	STUDENTS ANSWER (N)	STUDENTS ANSWER CORRECTLY (B)	INDEX (B/N)	LEVEL
1	Q1	28	27	0.964285714	EASY
2	Q2	28	26	0.928571429	EASY
3	Q3	28	21	0.75	EASY
4	Q4	28	21	0.75	EASY
5	Q5	28	20	0.714285714	EASY

48

6	Q6	28	26	0.928571429	EASY
7	Q7	28	25	0.892857143	EASY
8	Q8	28	1	0.035714286	HARD
9	Q9	28	24	0.857142857	EASY
10	Q10	28	23	0.821428571	EASY
11	Q11	28	26	0.928571429	EASY
12	Q12	28	25	0.892857143	EASY
13	Q13	28	25	0.892857143	EASY
14	Q14	28	24	0.857142857	EASY
15	Q15	28	25	0.892857143	EASY
16	Q16	28	26	0.928571429	EASY
17	Q17	28	27	0.964285714	EASY
18	Q18	28	27	0.964285714	EASY
19	Q19	28	3	0.107142857	HARD
20	Q20	28	22	0.785714286	EASY
21	Q21	28	27	0.964285714	EASY
22	Q22	28	26	0.928571429	EASY
23	Q23	28	26	0.928571429	EASY
24	Q24	28	0	0	HARD
25	Q25	28	23	0.821428571	EASY
26	Q26	28	1	0.035714286	HARD
27	Q27	28	27	0.964285714	EASY
28	Q28	28	0	0	HARD
29	Q29	28	20	0.714285714	EASY
30	Q30	28	21	0.75	EASY
31	Q31	28	25	0.892857143	EASY
32	Q32	28	27	0.964285714	EASY
33	Q33	28	26	0.928571429	EASY
34	Q34	28	26	0.928571429	EASY
35	Q35	28	26	0.928571429	EASY
36	Q36	28	25	0.892857143	EASY
37	Q37	28	26	0.928571429	EASY
38	Q38	28	26	0.928571429	EASY
39	Q39	28	23	0.821428571	EASY
40	Q40	28	20	0.714285714	EASY
41	Q41	28	23	0.821428571	EASY
42	Q42	28	26	0.928571429	EASY
43	Q43	28	22	0.785714286	EASY
44	Q44	28	25	0.892857143	EASY
45	Q45	28	25	0.892857143	EASY
46	Q46	28	23	0.821428571	EASY
47	Q47	28	2	0.071428571	HARD

48	Q48	28	25	0.892857143	EASY
49	Q49	28	26	0.928571429	EASY
50	Q50	28	25	0.892857143	EASY
			Index Total	38.82142857	
			Mean	0.776428571	

After analyzing the data first sample (AS) used table above, the writer get the conclusion and explain in the diagram percentage the analysis of items difficulty level as follows:
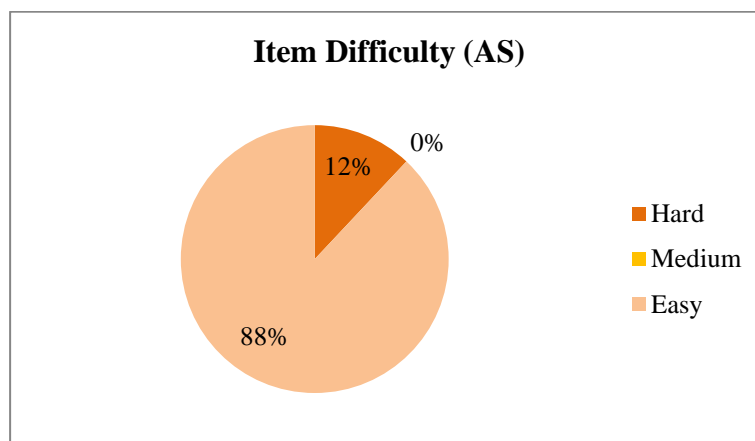
Figure 1 - Figure 4.4 the Result of Difficulty Level in the A Sample (AS)

On the figure above, the writer sum up that items test included the requirement good level of the test did not visible in this data. It where the data analysis of difficulty levels has indicated two categorize from three categorizes must have in the items test. The two categories were easy and hard categories. The percentages of each categorizes were 88% for easy and 12% for hard levels. According to it, the data has not necessary. The levels did not balance in the items test pack. The items test pack must have three categorizes that included in test items, those were medium, easy and hard levels with the ratio 3:4:3 and it means that the percentages of all level as like as 30% for easy, 40% for medium, and 30% for hard level.

50

b. The B Sample (BS)

The second sample was B sample (BS). This sample has 30 students, and the N was 30. The writer has finished analyzing of the difficulty levels. The result of analyzes there were three categorizes from fifty questions that descript in the appendix of the result B sample of difficulty levels analysis. The writer was find out of the items difficulty level has three categories, those were the items that can category as hardly were four numbers (8, 26, 28, and 47), as medium level was 22 numbers. Those were numbers 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 30, 33, 34, 26, 29, 42, 45, 48 and 49). The last the items as easily category were 24 numbers (1, 2, 3, 4, 5, 6, 9, 21, 23, 25, 29, 31, 32, 35, 37, 38, 40, 41, 42, 44, 46, and 50). From the outcome of the writer get, the items difficulty level can be classified like as the most index of hardly category in 0 (Zero), the most medium category in 0,37 (zero point thirty seven), and the last was the most index of easily category in 1 (one). Thus, by the finding data as of BS sample it can be concluded that the mean value for the BS data in 0.67 and it as call as the medium category.

Table 3 - Appendix 3 - Table 4.4 The Result of Difficulty Levels in the B Sample (BS)

NO.	QUESTIONS NUMBER	STUDENTS ANSWER (N)	STUDENTS ANSWER CORRECTLY (B)	INDEX (B/N)	LEVEL
1	Q1	30	29	0.966666667	EASY
2	Q2	30	29	0.966666667	EASY
3	Q3	30	27	0.9	EASY
4	Q4	30	22	0.733333333	EASY
5	Q5	30	27	0.9	EASY
6	Q6	30	26	0.866666667	EASY
7	Q7	30	19	0.633333333	MEDIUM
8	Q8	30	3	0.1	HARD
9	Q9	30	24	0.8	EASY
10	Q10	30	15	0.5	MEDIUM
11	Q11	30	18	0.6	MEDIUM
12	Q12	30	25	0.833333333	EASY
13	Q13	30	21	0.7	MEDIUM

14	Q14	30	20	0.666666667	MEDIUM
15	Q15	30	11	0.366666667	MEDIUM
16	Q16	30	18	0.6	MEDIUM
17	Q17	30	13	0.433333333	MEDIUM
18	Q18	30	15	0.5	MEDIUM
19	Q19	30	15	0.5	MEDIUM
20	Q20	30	21	0.7	MEDIUM
21	Q21	30	23	0.766666667	EASY
22	Q22	30	19	0.633333333	MEDIUM
23	Q23	30	28	0.933333333	EASY
24	Q24	30	16	0.533333333	MEDIUM
25	Q25	30	26	0.866666667	EASY
26	Q26	30	2	0.066666667	HARD
27	Q27	30	24	0.8	EASY
28	Q28	30	0	0	HARD
29	Q29	30	30	1	EASY
30	Q30	30	17	0.566666667	MEDIUM
31	Q31	30	22	0.733333333	EASY
32	Q32	30	22	0.733333333	EASY
33	Q33	30	18	0.6	MEDIUM
34	Q34	30	14	0.466666667	MEDIUM
35	Q35	30	22	0.733333333	EASY
36	Q36	30	15	0.5	MEDIUM
37	Q37	30	28	0.933333333	EASY
38	Q38	30	30	1	EASY
39	Q39	30	19	0.633333333	MEDIUM
40	Q40	30	29	0.966666667	EASY
41	Q41	30	24	0.8	EASY
42	Q42	30	28	0.933333333	EASY
43	Q43	30	17	0.566666667	MEDIUM
44	Q44	30	26	0.866666667	EASY
45	Q45	30	19	0.633333333	MEDIUM
46	Q46	30	25	0.833333333	EASY
47	Q47	30	1	0.033333333	HARD
48	Q48	30	18	0.6	MEDIUM
49	Q49	30	14	0.466666667	MEDIUM
50	Q50	30	26	0.866666667	EASY
			Index Total	33.33333333	
			Mean	0.666666667	

As the result of this data analyzes were obtain in the figure 2 of the diagram from the B sample (BS). Indeed, the writer gets the conclusion as follows:

Item Difficulty (BS)

Hard
Medium
Easy

8%
48%
44%
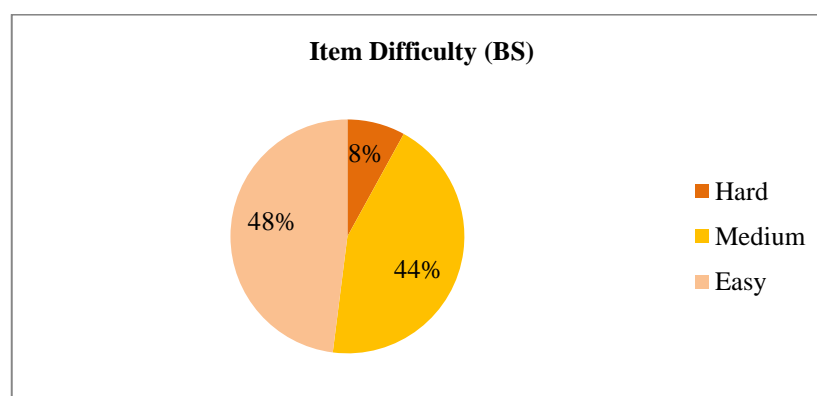
Figure 2 - Figure 4.4 The Result of Difficulty Levels in the B Sample (BS)

The second figure was description of the B sample (BS) in the items difficulty level, which have three categories in the items test. The hard categorize has 8%, medium in 44%, and easy categorize with 48%. The items test did not harmonies because the ratio of each categorize have not appropriate with the requirement of the ideal a good test pack. To sum up, the writer get which hard level was less than others levels. This analyzes from the data students' answer of fifty questions for 30 students. The diagram above concluded that the medium and easy levels were dominant in this analysis with the index of hard level in zero (0), medium in 0.37, and the last level was easy with one (1) of the index.

c. **The C Sample (CS)**

Thirty-seven students who can answer fifty questions from the CS sample, the writer identify the items difficulty levels, which has three categories. They were hard, medium and easy. The data outcomes were twelve items that can call hard category. The numbers of them were 8, 12, 13, 15, 17, 22, 26, 28, 33, 34, 47, and 49. Then the numbers of category easy were 1, 2, 5, 21, 25, 29, 31, 37,

38, 40, 41, 42, and 50. The total of the easy category were thirteen items. The last category was medium level. The items that included of the medium level had 25 items. Those were 3, 4, 6, 7, 9, 10, 11, 14, 16, 18, 19, 20, 23, 24, 30, 32, 35, 36, 39, 43, 44, 45, and 48.The writer can conclude that from this sample, the items test-pack is medium category. The mean of this data finding is 0.524. The most index of the hard category in 0.08, medium category in 0.7, and easy category in 0.97. The explanation of it can see in the appendix 4 of the third sample of difficulty levels analysis.

Table 4 - Table 4.4 The Result of Difficulty Levels in the C Sample (CS)

NO.	QUESTIONS NUMBER	STUDENTS ANSWER (N)	STUDENTS ANSWER CORRECTLY (B)	INDEX (B/N)	LEVEL
1	Q1	37	34	0.918918919	EASY
2	Q2	37	33	0.891891892	EASY
3	Q3	37	25	0.675675676	MEDIUM
4	Q4	37	13	0.351351351	MEDIUM
5	Q5	37	31	0.837837838	EASY
6	Q6	37	24	0.648648649	MEDIUM
7	Q7	37	25	0.675675676	MEDIUM
8	Q8	37	3	0.081081081	HARD
9	Q9	37	24	0.648648649	MEDIUM
10	Q10	37	17	0.459459459	MEDIUM
11	Q11	37	15	0.405405405	MEDIUM
12	Q12	37	11	0.297297297	HARD
13	Q13	37	9	0.243243243	HARD
14	Q14	37	13	0.351351351	MEDIUM
15	Q15	37	7	0.189189189	HARD
16	Q16	37	13	0.351351351	MEDIUM
17	Q17	37	4	0.108108108	HARD
18	Q18	37	14	0.378378378	MEDIUM
19	Q19	37	18	0.486486486	MEDIUM
20	Q20	37	20	0.540540541	MEDIUM
21	Q21	37	27	0.72972973	EASY
22	Q22	37	9	0.243243243	HARD
23	Q23	37	24	0.648648649	MEDIUM
24	Q24	37	23	0.621621622	MEDIUM

54

25	Q25	37	36	0.972972973	EASY
26	Q26	37	1	0.027027027	HARD
27	Q27	37	24	0.648648649	MEDIUM
28	Q28	37	1	0.027027027	HARD
29	Q29	37	35	0.945945946	EASY
30	Q30	37	23	0.621621622	MEDIUM
31	Q31	37	31	0.837837838	EASY
32	Q32	37	22	0.594594595	MEDIUM
33	Q33	37	8	0.216216216	HARD
34	Q34	37	5	0.135135135	HARD
35	Q35	37	12	0.324324324	MEDIUM
36	Q36	37	17	0.459459459	MEDIUM
37	Q37	37	33	0.891891892	EASY
38	Q38	37	27	0.72972973	EASY
39	Q39	37	15	0.405405405	MEDIUM
40	Q40	37	32	0.864864865	EASY
41	Q41	37	30	0.810810811	EASY
42	Q42	37	36	0.972972973	EASY
43	Q43	37	19	0.513513514	MEDIUM
44	Q44	37	24	0.648648649	MEDIUM
45	Q45	37	16	0.432432432	MEDIUM
46	Q46	37	26	0.702702703	MEDIUM
47	Q47	37	2	0.054054054	HARD
48	Q48	37	16	0.432432432	MEDIUM
49	Q49	37	9	0.243243243	HARD
50	Q50	37	34	0.918918919	EASY
			Index Total	26.21621622	
			Mean	0.524324324	

After gets data analyzes from the table above, the writer has the percentage of this third sample (CS) with the three levels of difficulty from the 37 students as the test-taken. It was as follows:
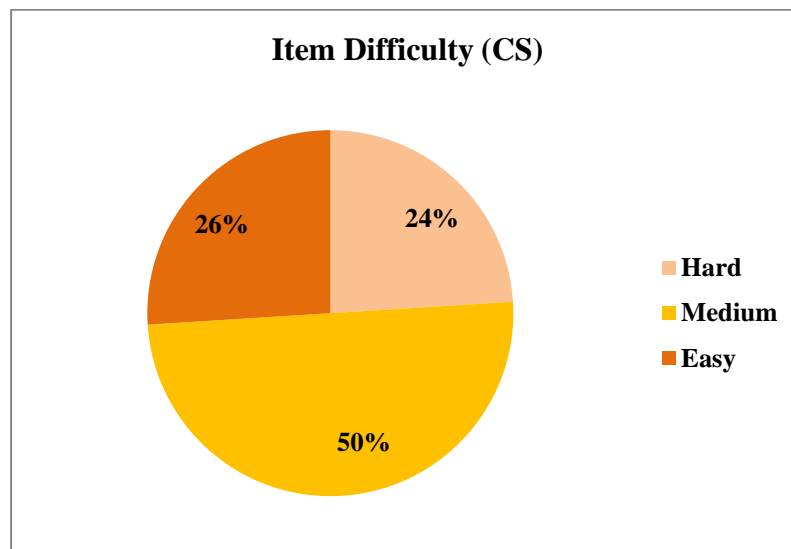
Figure 3 - Figure 4.4.3 The Result of Difficulty Levels in the C Sample (CS)

Based on the diagram above, the researcher can conclude that the items test pack from the sample CS in medium categorize with the percentage of 50%. The easy categorize more than hard categorize on the percentage of 26% > 24% from the fifty questions. Indeed, this items test based on this data is equal with the three categorize that must be had of the medium, easy and hard categorized. The writer get the results which the items test have not the appropriateness of the qualification in the items difficulty levels, because the items tests do not the balance in the categories of the ration 3:4:3 that meaning the hard and easy categories same and the hard categorize was the high value more than others.

4.5 Discrimination Power of the Items Test

The discrimination power for knowing the quality of the items test-pack, and which the test-pack has the distinction. The writer used the formula from Sudjana and found to discrimination power using the table of Ross & Stanley. The formula of discrimination power as follows:

$$DP = SR - ST$$

The formula to know the levels of the distinction in the test-pack which fulfill that included the requirement in discrimination power at each items. It was being of the three samples and the results described in the below:

a. The A Sample (AS)

The writer had finished analyzing data from the first sample (AS). This data has 28 students as N. From this data, there were three categories as the students' group. The three categorizes were students in the upper, intermediate, and the lower group. The students' group was creating from the students' score that following the test-taken. It was explain in the table-9 below:

Table 5 - Table 4.5 The Result Students' Group of the First Sample (AS)

NO.	NAME OF STUDENTS	SCORE	RANK
1	AS10	88	1
2	AS5	84	2
3	AS6	84	2
4	AS16	84	2
5	AS18	84	2
6	AS17	82	6
7	AS1	80	7
8	AS15	80	7
9	AS19	80	7
10	AS21	80	7
11	AS4	78	11
12	AS9	78	11
13	AS23	78	11
14	AS24	78	11
15	AS26	78	11
16	AS27	78	11
17	AS28	78	11

18	AS2	76	16
19	AS3	76	16
20	AS7	76	16
21	AS22	74	19
22	AS8	72	20
23	AS12	72	20
24	AS13	72	20
25	AS14	72	20
26	AS20	72	20
27	AS11	70	25
28	AS25	70	25

Based on the table above, the writer classified the students from the three groups and it according to table of the criteria students' group by Ross & Stanley (Sudjana, 2011:142). This data has n = 8 and it has four option in the form of the test-pack. From it, the result of students' group was three categorizes. The first categorize was the eight students in upper-group. There were AS10, AS5, AS6, AS18 – AS17, AS1, and AS15. Those were students had classified of the first rank until seventh (student who gets seventh rank point two as like eighth rank / 7.2 = 8) with the higher score was 88 and the lower score was 80.

The second categorizes was intermediate's group. The students from this group were twelve students. They were AS19, AS21, AS4, AS9, AS23, AS26 – AS28, AS2, AS3, and AS7. The classified of this group were students from the seventh rank (students who gets rank of seventh point three until sixteenth point three / 7.3 = 9 until 16.3 = 20). The last was lower group with eight students. There were AS22, AS8, AS12 – AS14, AS20, AS11, and AS25. The rank of this group were categorized as the students rank of nineteenth until twenty-fifth (students who get twenty-ones until twenty-eighth rank). After the writer classified the students group, next was the analysis of discrimination power. It from students who has the wrong answered in the test-packs. It is seeing on the appendix of the table 10 about the analysis discrimination power of the first sample below:

Table 13 - Table 4.5 The Result Discrimination Power of the A Sample (AS)

Questions Number	Total number of students who answered incorrectly (SR)	Total number of students who answered incorrectly (ST)	SR - ST	INFORMATION
1	0	0	0	NO
2	1	1	0	NO
3	3	0	3	NO
4	2	0	2	NO
5	3	1	2	NO
6	1	0	1	NO
7	3	0	3	NO
8	7	8	-1	NO
9	2	0	2	NO
10	1	1	0	NO
11	1	0	1	NO
12	3	0	3	NO
13	2	0	2	NO
14	2	1	1	NO
15	0	1	-1	NO
16	0	1	-1	NO
17	1	0	1	NO
18	0	0	0	NO
19	6	7	-1	NO
20	4	0	4	NO
21	0	1	-1	NO
22	1	1	0	NO
23	1	1	0	NO
24	8	8	0	NO
25	3	0	3	NO
26	8	8	0	NO
27	0	0	0	NO
28	8	8	0	NO
29	1	1	0	NO
30	3	1	2	NO
31	1	0	1	NO
32	0	0	0	NO
33	1	0	1	NO
34	0	1	-1	NO
35	0	0	0	NO
36	2	0	2	NO

37	1	0	1	NO
38	0	1	-1	NO
39	**5**	**0**	**5**	**YES**
40	5	2	3	NO
41	2	1	1	NO
42	1	0	1	NO
43	2	0	2	NO
44	2	0	2	NO
45	1	0	1	NO
46	2	2	0	NO
47	7	8	-1	NO
48	3	0	3	NO
49	1	0	1	NO
50	0	2	-2	NO

Indeed, the writes was summarize of the table above that from fifty question which following of sixteen students from the upper and lower group who had the wrong answered. The discrimination power from it only number 39. The number 39 has 5 of the r calculation as like as the r table and it has one of items number test-pack. It explained that this first sample did not have a discrimination power on the items test.

b. The Second sample (BS)

The second sample is BS schools. This sample has 30 students as N. From this N, there were three categories as the students' group. The three categorizes were students in the upper, intermediate, and the lower group. The students' group was creating from the students' score that following the test-taken. It was describe as below:

Table 14 - Table 4.5.3 the Result Students' Group of B Sample (BS)

NO.	NAME OF STUDENTS	SCORE	RANK
1	BS19	92	1
2	BS17	84	2
3	BS24	84	2
4	BS25	84	2
5	BS2	78	5
6	BS16	78	5
7	BS14	76	7
8	BS18	76	7
9	BS23	72	9
10	BS26	72	9
11	BS21	70	11
12	BS27	70	11
13	BS30	68	13
14	BS6	66	14
15	BS29	66	14
16	BS7	64	16
17	BS9	62	17
18	BS10	62	17
19	BS13	62	17
20	BS15	62	17
21	BS5	60	21
22	BS8	60	21
23	BS11	60	21
24	BS4	58	24
25	BS20	56	25
26	BS28	56	25
27	BS1	54	27
28	BS3	54	27
29	BS22	48	29
30	BS12	46	30

The result of this sample was three categorizes from the N30. The writer has finished analyzing that it had upper, intermediate, and lower group. The N30 has 8 of the criteria to categorize the students' group. From the upper group, there were BS19, BS17, BS24, BS25, BS2, BS16, BS14, and BS18. The classified in this group was the rank one until seventh (students who get the rank 1 until 8) with the high and low score were 92 and 76. Then, the intermediate's group was fourteen students. Those were BS23, BS26, BS21, BS27, BS30, BS6, BS29, BS7, BS9, BS20, BS15, BS5, and BS8. The ranks of this group were ninth until twenty-two.

The last group was lower group. The students who had the low rank were BS11, BS4, BS20, BS28, BS1, BS3, BS 22, and BS12. The rank categorize were twenty-three until thirtieth. The high and lower score from this group were 60 and 46. Then, the researcher analyze the discrimination power on each items test based on the upper and lower group who had the wrong answered of the items test. It was details in the table 12 below:

Table 15 - Table 4.5.4 The Result of Discrimination Power of B Sample (BS)

Questions Number	Total number of students who answered incorrectly (SR)	Total number of students who answered incorrectly (ST)	SR - ST	INFORMATION
1	1	0	1	NO
2	0	0	0	NO
3	1	0	1	NO
4	5	0	5	YES
5	2	0	2	NO
6	4	0	4	NO
7	4	0	4	NO
8	8	8	0	NO
9	4	0	4	NO
10	7	2	5	YES
11	7	1	6	YES
12	4	0	4	NO
13	1	2	-1	NO

14	5	1	4	NO
15	5	4	1	NO
16	4	2	2	NO
17	8	1	7	YES
18	7	1	6	YES
19	1	6	-5	NO
20	3	2	1	NO
21	3	1	2	NO
22	4	1	3	NO
23	2	0	2	NO
24	3	7	-4	NO
25	0	2	-2	NO
26	7	7	0	NO
27	3	0	3	NO
28	8	8	0	NO
29	0	0	0	NO
30	3	1	2	NO
31	4	2	2	NO
32	3	0	3	NO
33	4	0	4	NO
34	4	2	2	NO
35	4	0	4	NO
36	3	0	3	NO
37	1	0	1	NO
38	0	0	0	NO
39	6	0	6	YES
40	0	0	0	NO
41	2	3	-1	NO
42	1	0	1	NO
43	5	0	5	YES
44	3	0	3	NO
45	4	0	4	NO
46	4	0	4	NO
47	8	7	1	NO
48	6	0	6	YES
49	7	1	6	YES
50	1	2	-1	NO

The writer summarizes of analyze based on the table 12 there were nine numbers which fulfill on the requirements of the discrimination power. Those were number 4, 10, 11, 17, 18, 39, 43, 48, and 49. The upper r-value was 7, it has of the number 17. The resume in this data has the discrimination power with the N30 and from fifty questions of the test-pack.

c. *The Third Sample (CS)*

The third sample was CS school. There were 37 students and the criterion of students' group was 10. There were three categories as the students' group. The three categorizes were students in the upper, intermediate, and the lower group. The students' group was creating from the students' score that following the test-taken. It was explain in the table-13 below:

Table 16 - Table 4.5.5 The Result Students' Group of The C Sample (CS)

NO.	NAME OF STUDENTS	SCORE	RANK
1	CS4	76	1
2	CS20	72	2
3	CS19	64	3
4	CS26	64	3
5	CS10	62	5
6	CS6	60	6
7	CS7	60	6
8	CS15	60	6
9	CS3	58	9
10	CS11	58	9
11	CS18	58	9
12	CS24	58	9
13	CS13	56	13
14	CS14	56	13
15	CS22	56	13
16	CS23	54	16
17	CS5	52	17
18	CS8	52	17
19	CS9	52	17
20	CS12	52	17

21	CS29	52	17
22	CS33	52	17
23	CS35	52	17
24	CS16	50	24
25	CS37	50	24
26	CS2	48	26
27	CS1	46	27
28	CS25	46	27
29	CS32	46	27
30	CS34	46	27
31	CS28	44	31
32	CS30	44	31
33	CS31	44	31
34	CS27	42	34
35	CS21	40	35
36	CS36	40	35
37	CS17	18	37

According to the appendix of the students' group in the third sample, the writer has finished analyzing of the divided students in the three groups. Those were upper, intermediate, and lower group. The first was explaining upper group. There were CS4, CS20, CS19, CS26, CS10, CS6, CS7, CS15, CS3, and CS11. The upper score was CS4 with 76, and the lower score of this group was CS11 with 58. The ranks in this group were one until tenth. The next was intermediate group. The ranks in this group start from eleventh until seventeenth. In addition, the students who classified in the intermediate group were CS18, CS24, CS13, CS14, CS22, CS23, CS5, CS8, CS9, CS12, CS29, CS33, CS35, CS16, CS37, CS2, and CS1. The high score in this group was 58 and the low score was 46.

The third group was lower. These groups namely were CS25, CS32, CS34, CS28, CS30, CS31, CS27, CS21, CS36, and CS17. The upper score get CS25 with 46, and lower score get CS17 with 18. The ranks of the intermediate group start from 28 until 37. After the writer has analyzed of the students group, then would like to analyze the discrimination power based on the appendix of table 14, as follows:

Table 17 - Table 4.5.6 The Result Discrimination Power of The C Sample (CS)

Questions Number	Total number of students who answered incorrectly (SR)	Total number of students who answered incorrectly (ST)	SR - ST	INFORMATION
1	3	0	3	NO
2	3	0	3	NO
3	6	1	5	YES
4	0	5	-5	NO
5	6	0	6	YES
6	6	1	5	YES
7	2	4	-2	NO
8	8	10	-2	NO
9	5	2	3	NO
10	7	4	3	NO
11	9	6	3	NO
12	10	3	7	YES
13	10	5	5	YES
14	8	6	2	NO
15	10	7	3	NO
16	8	7	1	NO
17	10	8	2	NO
18	6	8	-2	NO
19	8	5	3	NO
20	6	4	2	NO
21	3	1	2	NO
22	9	6	3	NO
23	6	1	5	YES
24	8	1	7	YES
25	2	0	2	NO

26	10	10	0	NO
27	6	1	5	YES
28	9	10	-1	NO
29	1	0	1	NO
30	5	4	1	NO
31	2	2	0	NO
32	7	2	5	YES
33	9	6	3	NO
34	10	7	3	NO
35	10	3	7	YES
36	4	7	-3	NO
37	3	0	3	NO
38	4	0	4	NO
39	8	3	5	YES
40	5	0	5	YES
41	4	0	4	NO
42	2	0	2	NO
43	6	5	1	NO
44	6	0	6	YES
45	5	6	-1	NO
46	5	1	4	NO
47	10	10	0	NO
48	6	4	2	NO
49	9	7	2	NO
50	1	0	1	NO

Based on the appendix above, the writer analyze there has numbers of the appropriate of the discrimination power of the items test pack. Ultimately, the writer takes 13 items of the test-pack that have discrimination power. The items numbers are 3, 6, 13, 23, 27, 32, and 39 with the r-value 5. The next items numbers are 5 and 44 = 6. Then, 12, 24 and 35 has r value = 7.

As the summary, the discrimination power of the items test based on the random data, there was one number has on the each sampling. It was number 39, but the indicator of the rank in the upper group each sampling was different. In the AS only has number 39 with the value 5, the BS has nine numbers with

different indicators for each numbers, and in the number 39 the indicator is 6. The CS has the higher total numbers of the discrimination power with 13 numbers, and one of the item number was 39 in the indicator 5. It is the indicator same as in AS. Indeed, the item test pack has the discrimination power. The instrument of the test pack was enough to say it. The test packs can usage repeatedly in the next test in south district Gresik.

4.6 Validity of the Test

The validity test using the SPSS 16.0 for analyzes the valid data. This test based on Priyanto theory (2008). The data writer gets to validity test was from the students paper sheets. It was about fifty questions (50), than it from three random samples, those are AS with 28 Students, BS with 30 students, and CS with 37 students. The validity test is valid if the value of the r calculation more than the r table (**rc > rt**). From the validity calculation, the writer got the result as follows:

$$r_{i(x-i)} = \frac{n \sum i(x-i) - \left(\sum i\right)\left(\sum (x-i)\right)}{\sqrt{\left[n \sum i^2 - \left(\sum i\right)^2\right]\left[n \sum (x-i)^2 - \left(\sum (x-i)\right)^2\right]}}$$

(Priyanto. 2008:22)

Where:

|	
i	= Item score (correlation)
(x-i)	= Total score item minus item score
n	= Total of subject

a. The A Sample (AS)

These sample has N = 28 (students) and the questions of test was 50. The significant of alpha was 0.05. The outcome of this data can looks from the corrected Item − Total Correlation based on *the appendix of the table 15*. It the value of correlation that the researcher was gets. The r calculation as compared to the r table, the r table found to significant of the alpha 0.05 with the two test and total of data (n) = 28, which the r table is 0.3739. The detailed of the r table can see on the appendix of the table r (koefisien korelasi sederhana) according to Junaidi (2010:1). For the detailed computation can see in this below:

Table 6 - Table 4.6 The Result of Validity Test in The A Sample (CS)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q1	37.8571	5.683	-.094	.143
Q2	37.8929	5.507	.047	.121
Q3	38.0714	5.180	.129	.089
Q4	38.0714	5.550	-.053	.151
Q5	38.1071	5.358	.030	.123
Q6	37.8929	5.507	.047	.121
Q7	37.9286	5.106	.301	.057
Q8	38.7857	5.952	-.384	.183
Q9	37.9643	5.443	.038	.122
Q10	38.0000	5.778	-.158	.178
Q11	37.8929	5.433	.108	.109
Q12	37.9286	5.180	.247	.070
Q13	37.9286	5.402	.090	.109
Q14	37.9643	5.295	.129	.096
Q15	37.9286	5.921	-.252	.189
Q16	37.8929	5.877	-.245	.178
Q17	37.8571	5.386	.241	.095
Q18	37.8571	5.683	-.094	.143
Q19	38.7143	5.915	-.249	.189
Q20	38.0357	4.925	.288	.037
Q21	37.8571	5.905	-.334	.176
Q22	37.8929	5.729	-.131	.156
Q23	37.8929	5.729	-.131	.156
Q24	38.8214	5.634	.000	.129

Q25	38.0000	4.963	.298	.040
Q26	38.7857	5.582	.018	.127
Q27	37.8571	5.608	-.012	.131
Q28	38.8214	5.634	.000	.129
Q29	38.1071	5.877	-.204	.203
Q30	38.0714	5.402	.018	.127
Q31	37.9286	5.328	.142	.097
Q32	37.8571	5.683	-.094	.143
Q33	37.8929	5.507	.047	.121
Q34	37.8929	5.803	-.188	.167
Q35	37.8929	5.284	.233	.083
Q36	37.9286	5.106	.301	.057
Q37	37.8929	5.358	.170	.096
Q38	37.8929	5.877	-.245	.178
Q39	38.0000	4.815	.389	.010
Q40	38.1071	5.433	-.005	.136
Q41	38.0000	5.556	-.040	.144
Q42	37.8929	5.433	.108	.109
Q43	38.0357	5.369	.046	.118
Q44	37.9286	5.180	.247	.070
Q45	37.9286	5.624	-.060	.145
Q46	38.0000	5.481	.000	.133
Q47	38.7500	5.898	-.262	.181
Q48	37.9286	5.106	.301	.057
Q49	37.8929	5.358	.170	.096
Q50	37.9286	5.995	-.298	.200

The result of this analyze can find that item number $1 - 38, 40 - 50$ the value of r calculation more less than 0.3739 (r table), and the indeed was the items test was invalid. Then, on the item number 39 the r calculation = 0.389 was more than r table = 0.3739. In addition, for this data, the researcher only was doing one validity test without retesting again. Indeed, the percentage of this analysis was detail in the figure 4, as below:

The Validity Test of the First Sample (AS)

2%

■ Valid
■ Invalid

98%
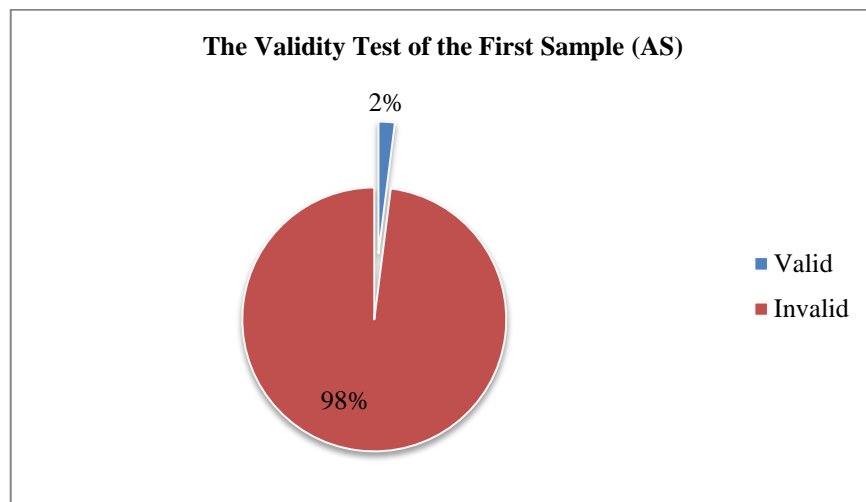
Figure 4 - Figure 4.6 The Result of Validity Test in the A Sample (AS)

As summary, the researcher had finished the analyzing of the valid data in the first sample (AS). From the table of analysis of validity test and the figure 4 above, the valid item in this test-pack only one number. The number it has the r calculation more than r table (0.389 > 0.3739). Indeed, the percentage of this analysis is 2% that item number can call as a valid data and 98% the items numbers, which do not fulfill the requirement of invalid data. The first sample was not enough with the appropriate of the valid items test, if the item was valid only one number. All the components was say invalid, because the r calculation less than r table.
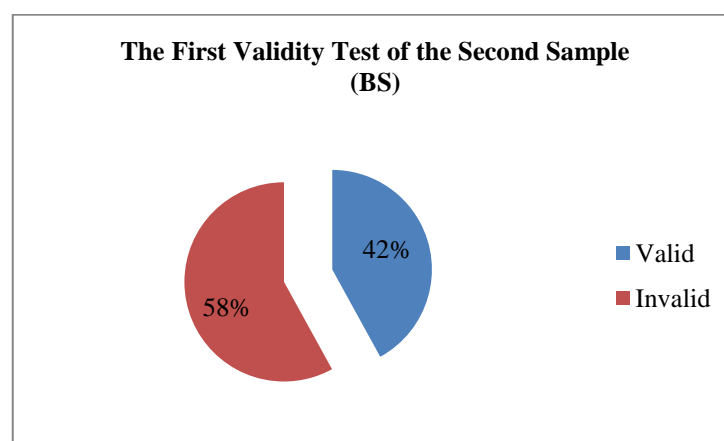
b. The B Sample (BS)

The sample data BS with the N = 30 and the questions is 50 with the significant level 0.05. The r table from the **N = 30 are 0.3610**, it is from N30 = 30 − 2 = **28**. The writer got the valid data after doing more than one analyze. The detailed computation of the first validity test as follow:

Table 19 - Table 4.6 The First Result Validity Test of the B Sample (BS)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q1	32.3667	31.344	.181	.750
Q2	32.3667	31.895	-.088	.755
Q3	32.4333	31.702	-.014	.755
Q4	32.6000	29.352	.450	.737
Q5	32.4333	30.875	.230	.748
Q6	32.4667	29.775	.491	.739
Q7	32.7000	29.114	.452	.736
Q8	33.2333	31.702	-.014	.755
Q9	32.5333	29.844	.391	.741
Q10	32.8333	29.109	.434	.737
Q11	32.7333	29.168	.433	.737
Q12	32.5000	29.569	.494	.737
Q13	32.6333	32.171	-.121	.763
Q14	32.6667	29.816	.325	.743
Q15	32.9667	31.206	.055	.755
Q16	32.7333	29.651	.341	.742
Q17	32.9000	27.817	.691	.724
Q18	32.8333	28.144	.620	.727
Q19	32.8333	34.075	-.436	.778
Q20	32.6333	31.826	-.056	.760
Q21	32.5667	31.289	.057	.754
Q22	32.7000	31.045	.085	.754
Q23	32.4000	30.455	.438	.743
Q24	32.8000	35.200	-.616	.786
Q25	32.4667	32.878	-.315	.765
Q26	33.2667	31.375	.108	.751
Q27	32.5333	30.671	.202	.748
Q28	33.3333	31.747	.000	.752
Q29	32.3333	31.747	.000	.752
Q30	32.7667	30.737	.135	.752
Q31	32.6000	30.869	.135	.751
Q32	32.6000	30.110	.291	.744
Q33	32.7333	29.168	.433	.737
Q34	32.8667	30.189	.233	.747
Q35	32.6000	30.455	.219	.748
Q36	32.8333	29.178	.421	.738
Q37	32.4000	31.490	.068	.752
Q38	32.3333	31.747	.000	.752

Q39	32.7000	29.252	.425	.738
Q40	32.3667	31.964	-.121	.755
Q41	32.5333	32.189	-.131	.761
Q42	32.4000	31.559	.044	.753
Q43	32.7667	29.082	.444	.737
Q44	32.4667	30.189	.379	.742
Q45	32.7000	28.838	.507	.734
Q46	32.5000	29.914	.408	.741
Q47	33.3000	30.907	.397	.746
Q48	32.7333	29.168	.433	.737
Q49	32.8667	28.464	.559	.730
Q50	32.4667	32.189	-.143	.760

In the result of computation above, there are 21 items numbers, which fulfill the requirement of the validity. There are items number 4, 6, 7, 9, 10, 11, 12, 17, 18, 20, 23, 36, 39, 43, 44, 45, 46, 47, 48, and 49. The twenty-one numbers was having the r calculation more than the r table 0.36210. Then, the items that do not fulfill the requirement are 29 numbers. It is 1, 2, 3, 5, 8, 13, 14, 15, 16, 19, 21, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 34, 35, 37, 38, 40, 41, 42, and 50, then for it items numbers is invalid because more less 0.3610. Then after get the data analysis of the validity test, writer created the percentage of the valid items in below:

The First Validity Test of the Second Sample (BS)

42% Valid
58% Invalid

Figure 5 - Figure 4.6 The First Result Validity Test of the B Sample (BS)

The conclusion is 42% the percentage of valid items, and 58%, which do not fulfill the requirement of invalid data. The most value of the r calculation was 0.691 in the question number 17. While the data was invalid, but it did not enough to make the writer gets the result that the items test was valid. Therefore, the writer was doing the second test. In addition, the second test has the detail computation as follows below:

Table 20 - Table 4.6 The Second Result Validity Test of the B Sample (BS)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q4	12.5333	23.292	.592	.871
Q6	12.4000	23.628	.689	.870
Q7	12.6333	22.930	.617	.870
Q9	12.4667	23.913	.499	.874
Q10	12.7667	23.771	.410	.877
Q11	12.6667	24.161	.338	.880
Q12	12.4333	23.702	.601	.872
Q17	12.8333	22.764	.633	.869
Q18	12.7667	22.806	.618	.870
Q20	12.5667	26.599	-.156	.894
Q23	12.3333	24.437	.623	.874
Q33	12.6667	23.264	.531	.873
Q36	12.7667	23.357	.498	.874
Q39	12.6333	24.240	.328	.880
Q43	12.7000	22.838	.617	.870
Q44	12.4000	24.179	.519	.874
Q45	12.6333	23.068	.585	.871
Q46	12.4333	23.909	.543	.873
Q47	13.2333	25.564	.253	.880
Q48	12.6667	24.023	.367	.879
Q49	12.8000	22.993	.578	.871

74

Base on the second validity test above, the writer had finished of analyzing with seventeen numbers of questions. Those are questions numbers 4, 6, 7, 9, 10, 12, 17, 18, 23, 36, 43, 44, 45, 46, 48, and 49. It the items numbers that have a value more than r table. Moreover, the r table was 0.3610. Then, the items numbers, which do not fulfill the requirements, it were four questions of 19%. There are items number 11, 20, 39, and 47. The writer has the percentage of the second test in below:

The Second Validity Test of the Second Sample (BS)

19%

81%
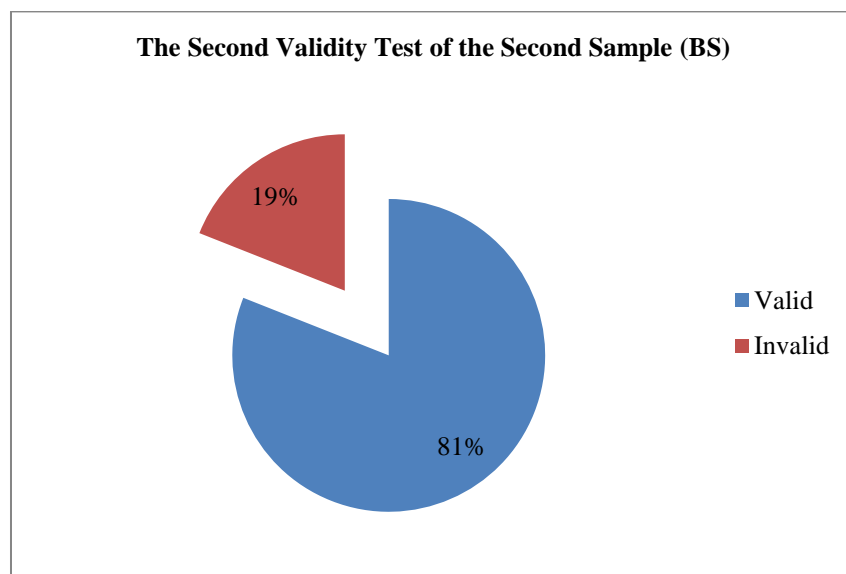
■ Valid
■ Invalid

Figure 6 - Figure 4.6 The Second Validity Test of the B Sample (BS)

As the result according to the figure 6, the second validity test the writer had analyze the seventeen items with the percentages of 81%, and the invalid items with 19%. Then, the items numbers, which do not fulfill the requirement, are four questions of 19%. There were items number 11, 20, 39, and 47. The r calculation for the four questions were Q11 = 0.338, Q20 = (- 0.156), Q39 = 0.328, and Q47 = 0.258. After analyzing the second test, the writer was doing the third test to make sure the valid items. The third validity test as the computation detailed was follows:

Table 21 - Table 4.6 The Third Result Validity Test of the B Sample (BS)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q4	10.5667	20.116	.590	.893
Q6	10.4333	20.461	.678	.892
Q7	10.6667	19.471	.691	.889
Q9	10.5000	20.672	.503	.896
Q10	10.8000	20.579	.404	.900
Q12	10.4667	20.533	.589	.894
Q17	10.8667	19.568	.645	.891
Q18	10.8000	19.683	.611	.892
Q23	10.3667	21.206	.612	.895
Q33	10.7000	20.010	.548	.895
Q36	10.8000	19.959	.546	.895
Q43	10.7333	19.513	.659	.890
Q44	10.4333	20.944	.517	.896
Q45	10.6667	19.540	.674	.890
Q46	10.4667	20.671	.547	.895
Q48	10.7000	21.321	.246	.905
Q49	10.8333	19.730	.602	.893

On the third test, the researcher had the outcome the seventeen 17 questions number has valid items from second test, and know in the third test she got 16 questions that fulfill the requirements of the validity that r calculation > r table 0.3610. They were 4, 6, 7, 9, 10, 12, 17, 18, 23, 36, 43, 44, 45, 46, and 49. Then, the item number that did not have the appropriate was question number 48. It has the r calculation of 0.246 more than r table of 0.3610. And the detailed of percentage the calculating based on the third validity test as follows:

The Third Avalidity Test of the Third Sample (BS)

6%

94%
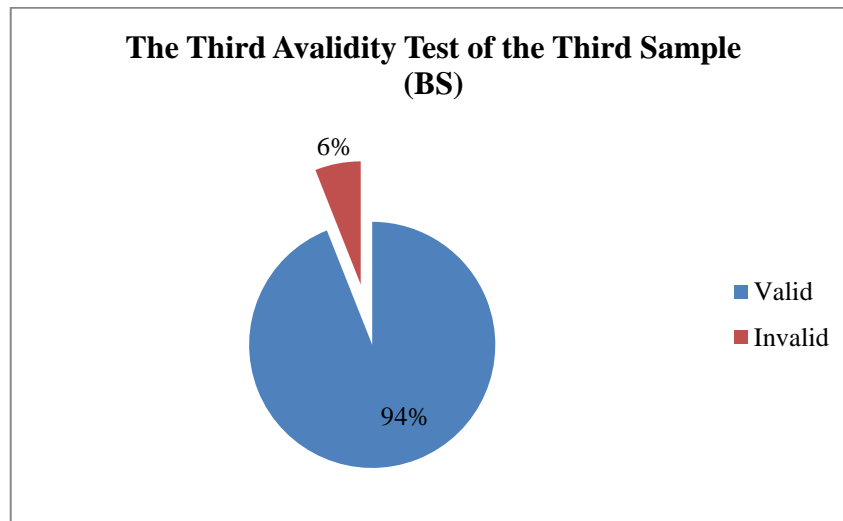
- Valid
- Invalid

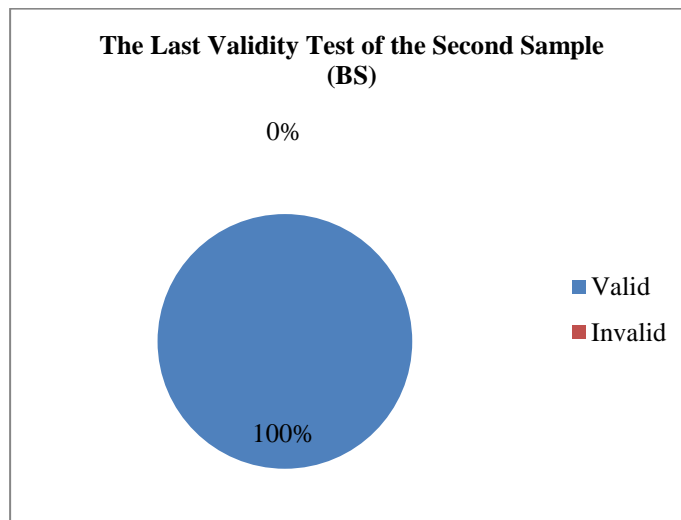**Figure 7 - Figure 4.6 The Third Validity Test of the B Sample (BS)**

The writer had the conclusion was sixteen questions number with 94% of the valid percentage and 6% of the invalid percentage. The item number that did not have the requirement of valid items was question number 48. It has the r calculation of 0.246 more than r table of 0.3610. To make sure the valid items, the writer was doing the last validity test. The last test was to know the all items was valid. Then, the detailed of the computation the last validity test can see in below:

Table 22 - Table 4.6 The Last Result Validity Test of the B Sample (BS)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q4	9.9667	18.792	.597	.899
Q6	9.8333	19.178	.668	.898
Q7	10.0667	18.133	.706	.895
Q9	9.9000	19.334	.509	.902
Q10	10.2000	19.338	.385	.907
Q12	9.8667	19.361	.544	.901
Q17	10.2667	18.202	.666	.896
Q18	10.2000	18.372	.617	.898
Q23	9.7667	19.909	.595	.901
Q33	10.1000	18.714	.547	.901

77

Q36	10.2000	18.717	.533	.901
Q43	10.1333	18.051	.704	.895
Q44	9.8333	19.661	.502	.902
Q45	10.0667	18.064	.724	.894
Q46	9.8667	19.361	.544	.901
Q49	10.2333	18.530	.580	.900

The last validity test as affirmation and assured for the researcher had analyzing that this data said valid. The result was according to the last test, where the valid data were 16 items numbers of questions. The r table is 0.3610, and the r calculation for 16 items numbers were Q4 = 0.597, Q6 = 0.668, Q7 = 0.706, Q9 = 0.509, Q10 = 0.385, Q12 = 0.544, Q17 = 0.666, Q18 = 0.617, Q23 = 0.595, Q33 = 0.547, Q36 = 0.533, Q43 = 0.704, Q44 = 0.502, Q45 = 0.724, Q46 = 0.544, and A49 = 0.580. There did not have the invalid item on the detailed computation above. The last described of this data, the researcher was creating the diagram below:

Figure 8 - Figure 4.6 The Last Result Validity Test of the B Sample (BS)

According to the figure above, the researcher has the result of the second sample (BS) analyzes which sixteen questions numbers of the items test pack was valid. The valid items gets from the fourth test that the researcher was do. The question numbers 16 items was 100%. It the value of r calculation more than r table. As summary, the valid test from the test until the last test of validity was make sure that the items test pack was valid.

c. The Third Sample (CS)

The CS data has N = 37 (students) and the questions were 50 items numbers. From this data, the writer analysis to determine the validity of the items test with the significant of the alpha 0.05 and N37 is 0.3246 (r table). The analyzing has followed the computation below:

Table 23 - Table 4.6 The First Result Validity Test of the C Sample (CS)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
p1	25.2973	24.270	.365	.647
p2	25.3243	24.225	.328	.648
p3	25.5405	23.589	.331	.643
p4	25.8649	23.731	.292	.646
p5	25.3784	23.408	.496	.636
p6	25.5676	23.197	.410	.637
p7	25.5405	25.366	-.052	.670
p8	26.1351	25.953	-.244	.672
p9	25.5676	24.474	.132	.657
p10	25.7568	24.800	.057	.663
p11	25.8108	24.991	.021	.665
p12	25.9189	23.021	.473	.633
p13	25.9730	24.249	.211	.652
p14	25.8649	24.009	.231	.650
p15	26.0270	25.305	-.030	.666
p16	25.8649	24.231	.184	.654
p17	26.1081	24.488	.242	.652
p18	25.8378	26.584	-.293	.686

p19	25.7297	24.592	.098	.660
p20	25.6757	24.614	.094	.660
p21	25.4865	25.701	-.123	.673
p22	25.9730	24.860	.067	.661
p23	25.5676	23.197	.410	.637
p24	25.5946	24.359	.153	.656
p25	25.2432	25.078	.143	.657
p26	26.1892	25.380	-.040	.662
p27	25.5676	24.474	.132	.657
p28	26.1892	25.713	-.240	.666
p29	25.2703	25.480	-.083	.664
p30	25.5946	26.137	-.206	.681
p31	25.3784	26.131	-.243	.677
p32	25.6216	23.853	.255	.648
p33	26.0000	23.889	.313	.646
p34	26.0811	24.410	.237	.651
p35	25.8919	23.710	.304	.645
p36	25.7568	25.078	.001	.667
p37	25.3243	23.781	.476	.641
p38	25.4865	23.646	.341	.643
p39	25.8108	24.269	.168	.655
p40	25.3514	23.290	.577	.634
p41	25.4054	23.303	.490	.636
p42	25.2432	25.078	.143	.657
p43	25.7027	23.992	.220	.651
p44	25.5676	23.030	.447	.634
p45	25.7838	24.619	.094	.660
p46	25.5135	23.312	.405	.638
p47	26.1622	25.473	-.080	.664
p48	25.7838	25.341	-.050	.670
p49	25.9730	24.583	.132	.657
p50	25.2973	25.048	.078	.659

In the first test, the result based on the computation above was the items numbers, which fulfill the requirement of the validity was thirteen numbers of questions. Those were Q1 = 0.365, Q2 = 0.328, Q3 = 0.331, Q5 = 0.496, Q6 = 0.410, Q12 = 0.473, Q23 = 0.410, Q37 = 0.476, Q38 = 0.341, Q40 = 0.577, Q41 = 0.490, Q44 = 0.447, and Q46 = 0.405. Those values of the r calculation more than r table = 0.3246. Then, the items numbers of invalid data were thirty-seven questions numbers. To details result, the writer has the diagram below:

80
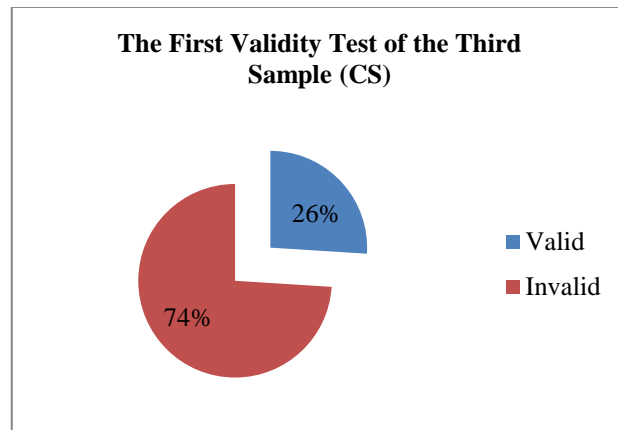
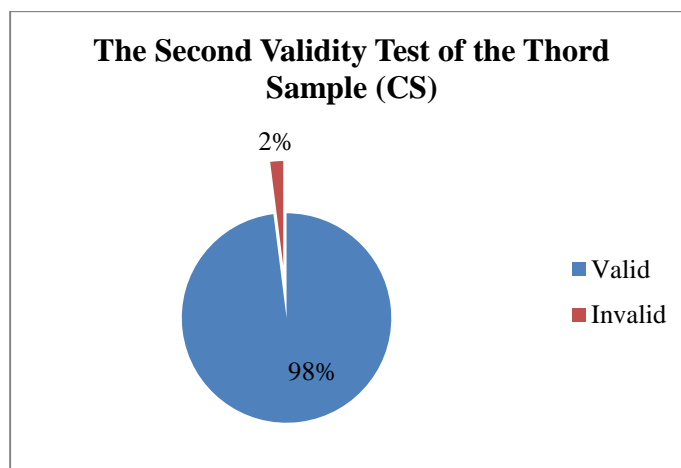**Figure 9- Figure 4.6 The First Result Validity Test of the C Sample (CS)**

In the first test, the result was the items numbers, which fulfill the requirement of the validity items was 26% and in the invalid items has 74%. The most value of the r calculation in the item number 40 with the significant of alpha 0.577. To sum up of the items test from the third sample, the writer must do the repeated test again. It to help the writer gets the result of the real validity items test in this data. Then, the second validity test as follow:

Table 24 - Table 4.6 The Second Result Validity Test of the C Sample (CS)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
p1	8.6486	8.401	.379	.808
p2	8.6757	8.447	.295	.813
p3	8.8919	7.599	.482	.800
p5	8.7297	7.980	.457	.802
p6	8.9189	7.299	.594	.789
p12	9.2703	7.814	.408	.806
p23	8.9189	7.521	.501	.798
p37	8.6757	8.281	.390	.807
p38	8.8378	7.917	.381	.809
p40	8.7027	8.048	.467	.802

p41	8.7568	7.745	.535	.796
p44	8.9189	7.465	.524	.796
p46	8.8649	7.676	.465	.801

The analyzing based on the second computation above, the writer had twelve question numbers, which had the requirements in the valid items. They were numbers Q1, Q3, Q5, Q6, Q12, Q23, Q37, Q38, Q40, Q41, Q44, and Q46. All of the items had the value of r calculation more than r table (0.3246). Then, just has one number of the second test, which has not the requirement on the valid items. It was number Q2. The Q2 has the r calculation = 0.295 < the r table = 0.3246. In addition, to detail the explanation can see on the diagram below:

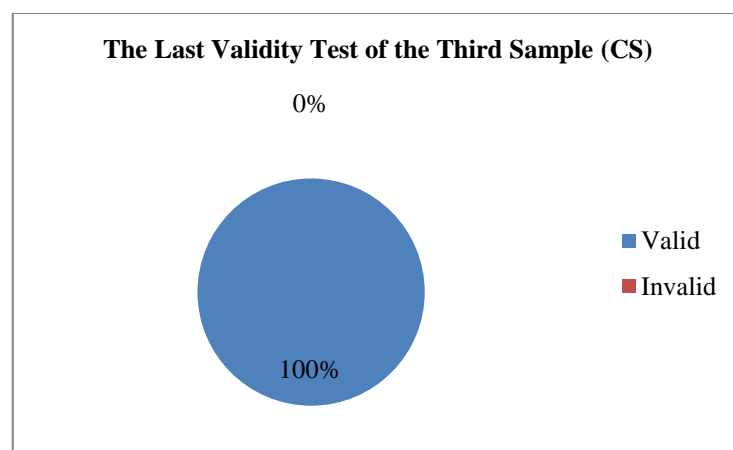The Second Validity Test of the Thord Sample (CS)

2%

98%

■ Valid
■ Invalid

Figure 10 - Figure 4.6 The Second Result Validity Test of the C Sample (CS)

The writer do re-testing to find out whether there was data still valid or there was invalid data in this items test. In the second test and the diagram above, the writer has finished analyzing that the valid items have 98% and 2% of invalid item. To further ensure of the twelve items had validity, the writer conducted the last test as follows:

Table 25 - Table 4.6 The Third Result Validity Test of the C Sample (CS)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
p1	7.7568	7.856	.332	.809
p3	8.0000	7.056	.463	.799
p5	7.8378	7.417	.437	.801
p6	8.0270	6.694	.607	.785
p12	8.3784	7.131	.445	.801
p23	8.0270	6.916	.510	.795
p37	7.7838	7.674	.387	.806
p38	7.9459	7.386	.351	.810
p40	7.8108	7.435	.472	.799
p41	7.8649	7.176	.524	.794
p44	8.0270	6.805	.558	.790
p46	7.9730	7.083	.466	.799

Based on the last test above, the all items was the evidence of the valid data. The researcher know that this instrument valid. Those were Q1 = 0.332, Q3 = 0.463, Q5 = 0.437, Q6 = 0.607, Q12 = 0.445, Q23 = 0.510, Q37 = 0.387, Q40 = 0.472, Q41 = 0.524, Q44 = 0.558, and Q46 = 0.466. Then, the writer makes the diagram to illustration the last validity test in below:

The Last Validity Test of the Third Sample (CS)

0%

100%

- Valid
- Invalid

Figure 11- Figure 4.6 The Second Result Validity Test of the C Sample (CS)

According the diagram above, the writer make a conclusion that the all items in the last validity test was valid. It proven the last computation there did not have the invalid items. The twelve items had the r calculation more than the r table of 0.3246. In addition, the most value of the significant alpha in the Q6 with 0.607.

As summary, the writer had finished analyzing of the three samples of south districts Gresik in the English final test items in academic years 2016/2017 with the items test pack was fifty questions. Then, the significant of the alpha in validity test was r calculation in two sampling more than r table. Therefore, one sample did not identify with invalid items. The sample is the first sample or AS school. Nevertheless, the total numbers of items is not balance or more low than the interpreted the researcher, but the items test-pack can be said was valid. The all the data were AS sample has one valid items, BS had 16 numbers of the valid items, and CS sample had 12 numbers of the valid items. Indeed, the entire items test was valid with the result analyzes in three sample.

4.7 Reliability of the Test

The test of reliability is to know the consistently of the assessment when the tool used repeatedly. The writer has three random sampling to reliability test. They are AS, BS, and CS. For the result of the reliability, the writer as follow:

a. The First Sample (AS)

The first sample (AS) was has 28 students as "N". The N28 has the r table of the reliability was 0.3739. Then, the first sample data can be say is reliable if the r calculation more than r table. To gets the result of the reliability test, the writer as following the result of computation below:

84

Reliability Statistics

Cronbach's Alpha	N of Items
.129	50

Based on the result above, the writer has conclusion from fifty questions test with N28 this data did have not reliable. The alpha of the data outcome was 0.129. It was less than 0.3739 of the r table. The analyzing of this instrument is not necessary if usage in the repeatedly in the next academic year.

b. *The Second Sample (BS)*

This sample has 30 students that following the test-pack. Then, with N30 the second sample has 0.3610. For detailed analysis and computation as the result, this analyzes as follows:

Reliability Statistics

Cronbach's Alpha	N of Items
.752	50

According to result of computation above, the writer concluded that the value of significant from N30 with fifty questions has the r calculation of 0.752. It is more than r table. Indeed, the test items from the sample BS has the reliable, because 0.752 > 0.3610. The items test-pack can be usage repeatedly in the next test-pack with different test-taken.

c. **The Third Sample (CS)**

The third sample has 37 students with the fifty questions test. The N was 37; the significant of N37 was 0.3246. To analyzing this data the writer has follow this computation:

Reliability Statistics

Cronbach's Alpha	N of Items
.660	50

From the computation above, the writer gets of the significant the r calculation was 0.660. Whereas, the value of the r table on the significant 0.05 with N=37 is 0.3246 (see on the appendix of the r table). Because, the value of the r calculation more than 0.3246, the writer can get the conclusion that the test items in this sample can be said was reliable items.

As summary, the writer had the result of analyzing that the three samples must have the reliability test. Form the two test-taken said reliable. Those were BS and CS sample. Then, one sample of the test-pack did not reliable. It was the first sample or AS sample. The AS have the r calculation less than r table (0.129 < 0.3739). In others wise, the BS has r calculation more than r table with the alpha 0.752 > 0.3610 and the CS is reliable too. The CS has 0.660 r calculation more than 0.3246 of the r table. Indeed, the all of the items test was reliable, because there were two samples that had the requirements of the reliable items test. And it test-pack can usage in represent in the next test-pack with replacement the sentences of the items test with the same themes or matter in south district Gresik.