

ADVISORS' APPROVAL SHEET

This thesis By

Name : Ema Rahmatun Nafsah

NIM : D05207072

Title : AN ANALYSIS OF ENGLISH MULTIPLE CHOICE

QUESTION (MCQ) TEST OF 7TH GRADE AT SMP

BUANA WARU SIDOARJO

has been approval by the thesis advisors for further approval by the Board of Examiners.

Surabaya, 15th Juli 2011

Advisor,

<u>Dra. Irma Soraya, M.Pd</u> NIP 196709301993032004

APPROVAL SHEET

This thesis by EMA RAHMATUN NAFSAH entitled AN ANALYSIS OF ENGLISH MULTIPLE CHOICE QUESTION (MCQ) TEST OF 7TH GRADE AT SMP BUANA WARU SIDOARJO has been approval by the thesis advisors for further approval by the Board of Examiners.

Surabaya, 19th Juli 2011

Dean.

DR H. Nur Hamim, M. Ag 196203121991031002

The Board of Examiner Chair.

Dra. Irma Soraya, M.Pd 196709301993032004

Secretary,

Rizka Safriyani, M. Pd 198409142009122005

Examiner I,

<u>Akh. Muzakki, Grad Dip SEA, M. Ag, M. Phil, Ph. D</u> 197402091998031002

Examiner II.

<u>Dra. Arbaiyyah Ys, MA</u> 196405031991032002

LEMBAR KEASLIAN

Bahwa saya yang bertanda tangan di bawah ini:

Nama: Ema Rahmatun Nafsah

NIM : D05207072

Judul: An Analysis of English Multiple Choice Questions (MCQ) Test of 7th grade

at SMP BUANA Waru Sidoarjo

menyatakan bahwa semua tulisan yang ada di dalam skripsi ini adalah murni tulisan saya, dan bukan hasil tulisan orang lain yang saya akui sebagai tulisan saya.

Apabila suatu hari ditemukan bukti bahwa tulisan ini merupakan hasil tulisan orang lain yang saya akui sebagai hasil tulisan saya, saya siap menerima sanksi dalam bentuk apapun sesuai dengan peraturan yang berlaku.

Tertanda,

EMA RAHMATUN NAFSAH NIM D05207072

ABSTRACT

Nafsah, Ema Rahmatun. 2011. An analysis of English Multiple Choice Questions (MCQ) Test of 7th grade at SMP BUANA Waru Sidoarjo. A Thesis. English Education Department, State Institute for Islamic Studies Sunan Ampel Surabaya. Advisor, Dra. Irma Soraya, M.Pd.

Key Words: validity, face validity, content validity, reliability, item analysis, index of difficulty, index of discrimination, effectiveness of distractors.

Teaching is related to the learning process and testing is related to an evaluation process. Testing is to determine he achievement and progress of instruction. Testing and teaching are so closely interrelated that it is virtually impossible to work either field without being constantly concerned with the other. It can be seen that teaching learning activities can be considered successful since testing is aimed to determine the achievement of the objective of education. Teacher as a constructor of the test should construct a good test so that the test will be valid and reliable. Test that is made by the teacher called teacher made test, it is still to be questioned whether the test is valid and reliable or not because teacher rarely tried out the test first before giving it to the students. Knowing this fact, teacher should analyze the test so that the teacher will know the quality of the test. By analyzing the test, the teacher will know which item can be used or revised.

Based on the explanation above An analysis of English Multiple Choice Questions (MCQ) Test of 7th grade at SMP BUANA Waru Sidoarjo are analyzed. It is analyzed whether it is constructed in right way, following the right principles or not. From the problem there a statement of problem. It is "What is the quality of Multiple Choice Question (MCQ) of 7th grade at SMP Buana Waru based on the standard of test?"

This study is expected to be useful for teachers in constructing a good English test, for the students in showing their real achievement in their studying, and for those who are involved in teaching learning process, I order to determine the effectiveness in teaching learning process. This research is descriptive qualitative research. The source of the data in this study is the final test items, the students' answer sheet, and the students' scores of 7th grade students in SMP BUANA especially for 7B, 7D, and 7E, as a sample study because the writer take the data randomly.

In collecting the data, the writer makes some steps. First is finding the test items. Second is finding the students' answer sheet. Third is finding the scores of the

students from the students answer sheet which have been scored by the teacher. Fourth is arranging the students' score and then classifying the students into two groups they are upper and lower group based on their score. And the last step is doing analysis to answer the statements of problem.

The result of the study leads to the conclusion that An analysis of English Multiple Choice Questions (MCQ) Test that has been constructed by an English teacher of 7th grade in SMP BUANA Waru Sidoarjo has good test based on the characteristics of a good test, good face validity and high content validity, high reliability, good index of difficulty but poor index of discrimination.

CONTENTS

Cover	i
Advisor's Approval Sheet	ii
Approval Sheet	ii
Motto i	iv
Dedication Sheet	v
Abstract	⁄i
Acknowledgment vi	iii
Contents	ix
List of Tablesx	ii
List of Diagramsxi	iii
List of Appendicesxi	iv
Chapter I Introduction	
1.1 Background of the Study	
1.2 Statement of Problem	5
1.3 Objectives of the Study6	5
1.4 Significances of the Study	7
1.5 The Scope of the Limitation	7
1.6 Definition of Key Terms	8
Chapter II Review of Related Literature	
2.1 Testing and Teaching	1
2.2 The Kinds of Test and Testing Based on the Purposes	
a. Placement Test	13
b. Achievement Test	l 4
c. Diagnostic Test1	6
d. Placement Test	17
e. Direct Versus Indirect Test	18

f. Objective Testing Versus Subjective Testing	20
2.3 Evaluation	21
2.4 Evaluation in Education	
2.5 Evaluation in Teaching English Process	22
2.6 Evaluation in Teaching English Process in Secondary School	23
2.7 Definition of MCQ test	
a. MCQ test good for	24
b. Multiple Choice Items	25
c. The Characteristics of a Good MCQ Test	26
2.8 Validity	
a. Face Validity	27
b. Content Validity	28
2.9 Reliability	29
2.10 Item Analysis	
a. Item Difficulty	31
b. Item Discrimination	32
c. Effectiveness of Distractors	32
2.11 Previous Studies	33
Chapter III Research Methodology	
3.1 Research Design	37
3.2 Subject and Object of Research	38
3.3 Research Instruments	38
3.4 Data and Source of Data	40
3.5 Data Collection Techniques	40
3.6 Data Analysis	41
Chapter IV Data Analysis	
4.1 Analyzing of a Good MCQ Test	47
4.2 Analyzing of Validity	

a. Analyzing of Face Validity	48
b. Analyzing of Content Validity	49
4.3 Analyzing of Reliability	51
4.4 Analyzing of Item Test	
a. Analyzing index of Difficulty	52
b. Analyzing index of Discrimination	61
c. Analyzing Effectiveness of Distractors	71
Chapter V Conclusion	
5.1 Conclusion	75
5.2 Suggestion	78
Bibliography	80
Annendices	92

LIST OF TABLES

Table		Page
3.6	Table of Analyzing the Characteristic of a Good Test	41
3.6	Table of Analyzing the Content Validity	43
3.6	Table of Analyzing Index of Difficulty and Index Of Discrimination	46
3.6	Table of Analyzing the Effectiveness of Distractors	46

LISH OF DIAGRAMS

Diagrams		Page
4.4	Diagram of Analyzing Index of Difficulty for 7B	55
4.4	Diagram of Analyzing Index of Difficulty for 7D	58
4.4	Diagram of Analyzing Index of Difficulty for 7E	60
4.4	Diagram of Analyzing Index of Discrimination for 7B	64
4.4	Diagram of Analyzing Index of Discrimination for 7D	67
4.4	Diagram of Analyzing Index of Discrimination for 7E	70

LIST OF APPENDICES

Appendices		Page
1.	Paper Test	82
2.	Analyzing of The Characteristics of MCQ Test	86
3.	Analyzing of Content Validity	87
4.	The Score of Upper and Lower Group of 7B	96
5.	The Score of Upper and Lower Group of 7D	97
6.	The Score of Upper and Lower Group of 7E	98
7.	The Result of Analyzing Index of Difficulty and	
	Discrimination of 7B	99
8.	The Result of Analyzing Index of Difficulty and	
	Discrimination of 7D	101
9.	The Result of Analyzing Index of Difficulty and	
	Discrimination of 7E	103
10.	The Result of Analyzing the Effectiveness of	
	Distractors of 7B	105
11.	The Result of Analyzing the Effectiveness of	
	Distractors of 7D	113
12.	The Result of Analyzing the Effectiveness of	
	Distractors of 7E	121

CHAPTER I

INTRODUCTION

1.1 Background of the study

Evaluation is a process collecting of the data to measure how far the goals of learning activities can be reached. ¹ Cronbanch and Stufflebeam said that evaluation not only use to measure how far the goals of learning activities can be reached but also it uses to make a decision for student. ² For student the evaluation can be used to measure their ability to catch the materials in the class and for teacher it used to know is the method that they have used in the class during learning activities can be mastered by students.

Based on the definition above, as-a teacher we need an instrument to measure how far the goals of learning activities can be reached by teacher and students. One of instrument that often used by teacher in evaluation process is a test. By doing the test we can measure the success of student in learning activities in the class.

The kind of test that considered as the easiest in evaluation process is Multiple Choice Question (MCQ) test. MCQ test is chosen to avoid the teacher's subjectivity for some students. J.B. Heaton in his book writing English Language Tests said that the chief criticism of the multiple choice item, however, is that frequently it does not to lend itself to the testing of language as communication. The process involved in

² *Ibid*, p. 3

¹ Suharsimi Arikunto, Dasar-Dasar Evaluasi Pendidikan, (Jakarta: Bumi Aksara, 2003) p. 3

the actual selection of one out of four or five options bears little relation to the way language is used in most real-life situations. Appropriate responses to various stimuli in everyday situations are *produced* rather than chosen from several options.³

Nevertheless, multiple choice items can provide a useful means of teaching ad testing in various learning situations (particularly at the lower levels) provide that it is always recognized that such items test knowledge of grammar, vocabulary, etc. rather than the ability to use language. Although they rarely measure communication as such, they can prove useful in measuring students' ability to recognize correct grammatical forms, etc. and to make importance discriminations in the target language. Moreover, the multiple choice items can help both student and teacher to identify areas of difficulty. Furthermore, multiple choice items offer a useful introduction to the construction of objective tests.⁴

A test can be categorized as a good instrument if it has some good characteristics, such as: validity, reliability, and objectivity. In constructing process of MCQ test, the teacher must have good skill, so the test that they have made based on the material that they have taught in the class and based on the characteristics of constructing MCQ test. But, in reality, analysis in writing tests seldom has done. This case makes the material, the arrangement of test, validity, reliability, and item

⁴ *Ibid*, p. 27

³ J.B. Heaton, writing English Language Testing, (United State of America: Longman, Inc, 1989) p.27

analysis consist of distractors, item discrimination, item difficulty and discrimination always be said as poor test.⁵

In this research, the writer chooses SMP BUANA Waru as a subject of her observation because in this place there are few teachers who do analysis in test that they have made, so the validity, reliability, the weakness and the strength of the test can be known. Before the writer does this research, the writer has asked to one of the teacher there about their system in giving a test to their student. They said that after they made a test, they gave it to their student directly and after that they do the correction and gave the mark without doing the analysis. This reason makes the writer interest to do this research there.

The writer also chose SMP BUANA because it the writer graduated from. So, she wants to help this school better than ever. This school is processing to be RSBN School, so the writer wants to increase the quality especially in constructing the English Multiple Choice Question test. So, the writer did this observation to measure the quality of MCQ test which has been constructed by English teacher in SMP Buana Waru Sidoarjo, can it be categorized as a good test or yet?

The writer chose Multiple Choice Question (MCQ) test because MCQ test is easier to be analyzed, and until right now MCQ test still used by some teachers as an instrument of their evaluation to measure their student's ability during learning activities in class. MCQ test also be used by the government in Final Examination

⁵ Suharsimi Arikunto, Op Cit, p. 57

(UNAS). So, analyzing MCQ test is importance to do by the teacher to know the quality of their test.

Considering the explanation above, the writer chooses An analysis of English Multiple Choice Questions (MCQ) Test of 7th grade at SMP BUANA Waru Sidoarjo. The researcher wants to measure the quality of English Multiple Choice Question (MCQ) test that has been constructed by English teacher in SMP BUANA Waru Sidoarjo.

By doing this research, the writer hopes the teacher can construct a good MCQ test. So the result of the test can be accountable to measure the students' success during learning process in the class. Moreover, by giving MCQ test to their students, the teacher starts to prepare them to do Final Examination (UNAS), because Indonesia still uses MCQ model in UNAS.

This research ever does by some researchers before, the first is Aan Kulistyono Ujiwandi, his research was about An Analysis Of The Teacher Made English Test Items In UAS Semester 2 2005 / 2006 for the first year students of SMAN I Puri Mojokerto. His research focused on the content validity, reliability, analysis of item difficulty and index of discrimination. The differences of the writer study with this research are the writer just focused on MCQ test and the writer add face validity to be analyzed and measure the quality of the MCQ English test based on the criteria of constructing a good English MCQ test.

The second is Sigit Eko Saputro, his research was about An Analysis of The Teacher Made Of The First Year Student Of SMA 2 Magetan. His research almost same with the first research, he focused on the content validity, reliability, index of difficulty, and index of discrimination. The differences of the writer study with this research are the writer more specific in MCQ test, and the writer try to measure the validity include face validity and content validity which did not analyze by the researcher before and the writer also add an analysis of distractor in her research.

The third is Millatul Islamiyah, her research was about Content Validity and Item Analysis Of Semester II English Final Test For Tenth Grade Students Of SMAN 3 Sidoarjo. Her research was focused on content validity and item analysis only. The differences this research with the writer's research is lied on the writer focused on the English MCQ test that has been written by English teachers in SMP Buana Waru, so the result of this research can be used as an instrument to measure their quality in constructing English MCQ test, but the previous research did not mention who was the test maker. So her result did not give significant English system learning in this school especially for the teacher.

The fourth is Ria Dewi Pratiktasari, her research with the title An Analysis of Semester II English Summative Test for the Eighth Year Students of SMPN 1 Slahung Ponorogo. Her research focused on content validity and item analysis includes index of difficulty and index of discrimination. The difference of this research with the writer's research is the writer more focuses on MCQ test. The

writer analyzes about the validity includes face validity and content validity, reliability, item analysis includes index of difficulty and index of difficulty. The writer also measures the quality of English MCQ test based on the characteristic of a good MCQ test.

The fifth is Erlik Widiyanistyati, here research almost same with the other researchers before. She analyzed about the content validity, reliability, index of difficulty and index of discrimination. Her research was about An Analysis of the English Final Test for the First Semester of the Twelfth Year of SMAN 1 Surabaya was SMAN 1 Surabaya. The differences of the writer study with this research are the writer more specific in MCQ test, and the writer measure the validity include face validity and content validity which did not analyze by the researcher before and the writer also add an analysis of distractor in her research

1.2 Statement of the problem

Based on the background of the study above, the researcher would like to answer the main problem, which is formulated as follow:

1. What is the quality of Multiple Choice Question (MCQ) of7th grade at SMP Buana Waru based on the standard of test?

1.3 Objectives of the study

Derived from formulated problems above the purposes of the study are:

To measure the quality of Multiple Choice Question (MCQ) test of 7th grade at SMP Buana Waru based on the standard of test.

1.4 Significance of the study

Significance of the study that as follow:

1. For English teachers

The result of this research gives a lot of information to the teachers how the quality of Multiple Choice Question (MCQ) test which they has been constructed, and give them information how to construct a good English MCQ. And the result of the test can measure the success of teacher during learning activities in the class.

2. For the student

The students can measure their ability during learning activities in the class.

1.5 The scope of the limitation

The scope limitation of this study is limited as follow:

- This study implemented to the 7th grade teacher of SMP Buana Waru Sidoarjo, furthermore it focuses on MCQ paper test.
- This study focus on the quality of MCQ test of 7th grade at SMP BUANA
 Waru based on the standard of test includes the characteristics of a good

MCQ test, Validity (face vaidity and content validity), reliability and item analysis (index of difficulty, index of discrimination and effectiveness of distractors).

 This study emphasizes in the improving of evaluation English system learning.

1.6 Definition of key terms

In order to avoid misunderstanding, the researcher defines the key terms as follow:

- 1. Multiple Choice Question (MCQ)
 - Multiple choice is a form of assessment in which respondents are asked to
- select the best possible answer (or answers) out of the choices from a list. The items of a multiple choice test are often colloquially referred to as "questions.
- 2. English Multiple Choice Question (MCQ) test

English Multiple Choice Question (MCQ) test is a one of the kind of objective test. English Multiple choice test is a form of assessment in which respondents are asked to select the best possible answer (or answers) out of the choices from a list. In this research the researcher focuses on English MCQ test that has been constructed by English teacher in 7th grade in SMP BUANA WARU SIDOARJO.

3. An analysis of English MCO test

An analysis of English MCQ test means an analysis that will be done by the researcher to know the quality of English MCQ test that has been constructed by English teacher in 7th grade in SMP BUANA WARU SIDOARJO based on validity include face validity and content validity, reliability, based on the characteristic of a good English MCQ test and based on item analysis (analysis of distractor and item difficulty).

4. The quality of test

The quality of test means how far the test that has been constructed by English teacher in 7th grade in SMP Buana Waru can be categorized as a good test or not, considering the validity (face validity and content validity), reliability, item analysis, item difficulty and discrimination and extended answer analysis based on experts. Even though, reliability and effectivities of distractor do by statistic analysis.

5. Validity

A test is said to be valid if it measures accurately what it is intended to measure. The validity here include face validity and content validity.

6. Face validity

If a test items looks right to other testers, teachers, moderators, and testees, it can be described as having least face validity.⁶

⁶ J.B Heaton, Op Cit, p. 159

7. Content validity

A test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc. with which it is meant to be concerned.⁷

8. Reliability

Reliability is a necessary characteristic of ay good test, for it to be valid at all, a test must first be reliable as a measuring instrument.⁸ The result of the test gives consistent result, if it is given to different groups of testees.

9. Item analysis

The identification of certain difficult items in the test, together with knowledge of the performance of the individual distractors in multiple choice items, can prove just a valuable in its implications for teaching as for testing. The item analysis here includes item difficulty, item discrimination and analysis of distractors.

⁹ *Ibid*, p. 178-182

⁷ Arthur Hughes, *Testing For Language Teacher*, (Cambridge University Press: Cambridge, 2003) p. 26

⁸ J.B Heaton, Op Cit, p. 162

CHAPTER II

REVIEW OF RELATED LITERATURE -

The purpose of this study is to support the writer's opinion as a basic of analysis to answer the problem in chapter one. Realizing that is important to support the writer's opinion, she would like to discuss about definition and characteristics of MCO test.

In this chapter the writer discuss about testing and teaching, the kinds of test and testing, evaluation, evaluation in education, evaluation in teaching English process, evaluation in teaching English process in secondary school, definition of MCQ test. The writer also discusses about validity (face validity and content validity), reliability, item analysis consists of distractors, item difficulty and discrimination.

2.1 Testing and Teaching

A large number of examinations in the past have encouraged a tendency to separate testing from teaching. Both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly conserved with the other. Test may be constructed primarily as a means of assessing the student's performance in the language. In the former case, the test is geared to the teaching that has taken place, whereas in the latter case the teaching is often geared

largely to the test. Standardized test and public examinations, in fact, may exert such a considerable influence on the average teacher that they are often instrumental in determining the kind of teaching that takes place before the test.

A language test which seeks to find out what candidates can do with language provides a focus for purposeful, everyday communication activities. Such a test will have a more useful effect on the learning of a particular language than a mechanical test of structure. In the past even good tests of grammar, translation or language manipulation had a negative and even harmful effect o teaching. A good communicative test of language, however, should have a much more positive effect on learning and teaching and should generally result in improved leaning habits. 10

Although most teachers also wish to evaluate individual performance, the aim of the classroom test is different from that of the external examination. While the latter is generally concerned with evaluation for the purpose of enabling teachers to increase their own effectiveness by making adjustments in their teaching to enable certain groups of students or individuals in the class to benefit more. Too many teachers gear their teaching toward an ill-defined average group without taking into account the abilities of those students in the class who are at either end of the scale.

A good classroom test will also help to locate their precise areas of difficulty encountered by the class or by the individual student.

The test should also enable the teacher to ascertain which parts of the language programme have been found difficult by the class. In this way, the teacher

¹⁰ Ibid. p. 5

can evaluate the effectiveness of the syllabus as well as the methods and the materials he or she is using. The test result may indicate, for example, certain areas of the language syllabus which have not taken sufficient account of foreign learner difficulties or which, for some reason, have been glossed over. In such case the teacher will be concerned with those problem areas encountered by groups of students rather than by the individual student. ¹¹

A test which sets out to measure students' performances as fairly as possible without in any way setting traps for them can be effectively used to motivate them. A well-constructed classroom test will provide the student with an opportunity that details of their performance are given as soon as possible after the test, the student should be able to learn from their weaknesses. In this way a good test can be used as a valuable teaching device.

2.2 The Kinds of Test and Testing based on the purposes

a. Proficiency tests

Arthur Hughes said in his book that "Proficiency tests are designed to measure people's ability in a language, regardless of any training they may have had in that language". The content of a proficiency test, therefore, is not based on the content or objectives of language courses that people taking the test may have followed. Rather, it is based on a specification of what candidates have to

¹¹ Ibid, p. 7

be able to do in the language in order to be considered proficient. This raises the question of what we mean by the word 'proficient'. 12

b. Achievement tests

Most teachers are unlikely to be responsible for proficiency tests. It is much more probable that they will be involved in the preparation and use of achievement tests. In contrast to proficiency tests, achievement tests are directly related to language courses, their purpose being to establish how successful individual students, groups of students, or the courses themselves have been in achieving objectives. They are of two kinds: final achievement tests and progress achievement tests. 13

Final achievement tests are those administered at the end of a course of study. They may be written and administered by ministries of education, official examining boards, or by members of teaching institutions. Clearly the content of these tests must be related to the courses with which they are concerned, but the nature of this relationship is a matter of disagreement amongst language testers. 14

In the view of some testers, the content of a final achievement test should be based directly on a detailed course syllabus or on the books and other materials used. This has been referred to as the syllabus-content approach. It has

Arthur Hughes, *Op Cit*, P. 11
 Ibid, p. 12

¹⁴ *Ibid*, p. 13

an obvious appeal, since the test only contains what it is thought that the students have actually encountered, and thus can be considered, in this respect at least, a fair test. The disadvantage is that if the syllabus is badly designed, or the books and other materials are badly chosen, the results of a test can be very misleading. Successful performance on the test may not truly indicate successful achievement of course objectives.

The alternative approach is to base the test content directly on the objectives of the course. This has a number of advantages. First, it compels course designers to be explicit about objectives. Secondly, it makes it possible for performance on the test to show just how far students have achieved those objectives.

Progress achievement tests, as their name suggests, are intended to measure the progress that students are making. They contribute to formative assessment. Since 'progress' is towards the achievement of course objectives, these tests, too, should relate to objectives. But how? One way of measuring progress would be repeatedly to administer final achievement tests, the (hopefully) increasing scores indicating the progress made. This is not really feasible, particularly in the early stages of a course. The low scores obtained would be discouraging to students and quite possibly to their teachers. The alternative is to establish a series of well-defined short-term objectives. These should make a clear progression towards the final achievement test based on

course objectives. Then if the syllabus and teaching are appropriate to these objectives, progress tests based on short-term objectives will fit well with what has been taught. If not, there will be pressure to create a better fit. If it is the syllabus that is at fault, it is the tester's responsibility to make clear that it is there that change is needed, not in the tests. 15

In addition to more formal achievement tests that require careful preparation, teachers should feel free to set their own 'pop quizzes'. These serve both to make a rough check on students' progress and to keep students on their toes. Since such tests will not form part of formal assessment procedures, their construction and scoring need not be too rigorous. Nevertheless, they should be seen as measuring progress towards the intermediate objectives on which the more formal progress achievement tests are based. They can, however, reflect the particular 'route' that an individual teacher is taking towards the achievement of objectives. 16

c. Diagnostic tests

Diagnostic tests are used to identify learners' strengths and weaknesses. They are intended primarily to ascertain what learning still needs to take place. At the level of broad language skills this is reasonably straightforward. We can be fairly confident of our ability to create tests that will tell us that someone is particularly weak in, say, speaking as opposed to reading

¹⁵ *Ibid*, p. 14 ¹⁶ *Ibid*, p. 15

in a language. Indeed existing proficiency tests may often prove adequate for this purpose. 17

The lack of good diagnostic tests is unfortunate. They could be extremely useful for individualized instruction or self-instruction. Learners would be shown where gaps exist in their command of the language, and could be directed to sources of information, exemplification and practice. Happily, the ready availability of relatively inexpensive computers with very large memories should change the situation.¹⁸

d. Placement tests

Placement tests, as their name suggests, are intended to provide information that will help to place students at the stage (or in the part) of the teaching programme most appropriate to their abilities. Typically they are used to assign students to classes at different levels. Placement tests can be bought, but this is to be recommended only when the institution concerned is sure that the test being considered suits its particular teaching programme. No one placement test will work for every institution and the initial assumption about any test that is commercially available must be that it will not work well. One possible exception is placement tests designed for use by language schools,

¹⁷ Ibid, p. 16

¹⁸ *Ibid*, p. 16

where the similarity of popular text books used in them means that the schools' teaching programmes also tend to resemble each other. 19

The placement tests that are most successful are those constructed for particular situations. They depend on the identification of the key features at different levels of teaching in the institution. They are tailor-made rather than bought off the peg. This usually means that they have been produced 'in house'. The work that goes into their construction is rewarded by the saving in time and effort through accurate placement.²⁰

e. Direct versus indirect testing

Testing is said to be *direct* when it requires the candidate to perform precisely the skill that we wish to measure. If we want to know how well candidates can write compositions, we get them to write compositions. If we want to know how well they pronounce a language, we get them to speak. The tasks, and the texts that are used, should be as authentic as possible. The fact that candidates are aware that they are in a test situation means that the tasks cannot be really authentic. Nevertheless every effort is made to make them as realistic as possible.²¹

Direct testing is easier to carry out when it is intended to measure the productive skills of speaking and writing. The very acts of speaking and writing

¹⁹ *Ibid*, p. 16 ²⁰ *Ibid*, p. 17 ²¹ *Ibid*, p. 17

provide us with information about the candidate's ability. With listening and reading, however, it is necessary to get candidates not only to listen or read but also to demonstrate that they have done this successfully. Direct testing has a number of attractions. First, provided that we are clear about just what abilities we want to assess, it is relatively straightforward to create the conditions which will elicit the behaviour on which to base our judgements. Secondly, at least in the case of the productive skills, the assessment and interpretation of students' performance is also quite straightforward. Thirdly, since practice for the test involves practice of the skills that we wish to foster, there is likely to be a helpful backwash effect.²²

Indirect testing attempts to measure the abilities that underlie the skills in which we are interested. Perhaps the main appeal of indirect testing is that it seems to offer the possibility of testing a representative sample of a finite number of abilities which underlie a potentially indefinite large number of manifestations of them.²³

The main problem with indirect tests is that the relationship between performance on them and performance of the skills in which we are usually more interested tends to be rather weak in strength and uncertain in nature. We do not yet know enough about the component parts of, say, composition writing to predict accurately composition writing ability from scores on tests that measure

²² *Ibid*, p. 18 ²³ *Ibid*, p. 18

the abilities that we believe underlie it. We may construct tests of grammar, vocabulary, discourse markers, handwriting, punctuation, and what we will. But we will still not be able to predict accurately scores on compositions (even if we make sure of the validity of the composition scores by having people write many compositions and by scoring these in a valid and highly reliable way).²⁴

f. Objective testing versus subjective testing

The distinction here is between methods of scoring, and nothing else. If, no judgement is required on the part of the scorer, then the scoring is objective. A multiple choice test, with the correct responses unambiguously identified, would be a case in point. If judgement is called for, the scoring is said to be subjective. There are different degrees of subjectivity in testing. The impressionistic scoring of a composition may be considered more subjective than the scoring of short answers in response to questions on a reading passage.²⁵

Objectivity in scoring is sought after by many testers, not for itself, but for the greater reliability it brings. In general, the less subjective the scoring, the greater agreement there will be between two different scorers (and between the scores of one person scoring the same test paper on different occasions). However, there are ways of obtaining reliable subjective scoring, even of compositions.²⁶

²⁴ *Ibid*, p. 18 ²⁵ *Ibid*, p. 22

²⁶ *Ibid*, p. 22

2.3 Evaluation

Evaluation is the process of determining the value or worth of a program, course, or other initiative, toward the ultimate goal of making decisions about adopting, rejecting, or revising the innovation. It should not be confused with assessment, which encompasses methods for measuring or testing performance on a set of competencies. Evaluation is the more inclusive term, often making use of assessment data in addition to many other data sources.²⁷

Pauline Rea-Dickins and Kevin Germaine said that evaluation is a natural activity, something that is very much part of our daily existence. It is something that can be very formal or informal. It is also something that may not always be made explicit but may actually be undertaken unconsciously. When you listen to a lecture, a radio interview, or a political speech, you are making judgments on many different levels about the speakers. These range from the content of what is being said to thoughts of whether the speaker is kind and generous, honest and reliable, or harsh and uncaring, dishonest and unreliable. Further, as Rowntree points out speakers are constantly responding to what they take to be the emerging attitudes and understandings of other people, and deciding what to say in consequence. In other words, we are constantly making evaluations of kind or another and modifying our behaviour accordingly.

²⁷ http://www.edtech.vt.edu/edtech/id/eval/eval.html

Making evaluative judgments is therefore a feature of social life, but evaluation is not always something that we do in a *principled* and *systematic* way.

The criteria we use in making judgements may sometimes be vague and ill-defined.²⁸

2.4 Evaluation in Education

When we evaluate different aspects of the teaching and learning process, it becomes important to make explicit the criteria used in our judgments, and to be principled in our evaluations. Ill-prepared and ad-hoc evaluations are likely to be unreliable, unfair, and uninformative. They are not a suitable source on which to base educational decisions. Making decisions about a teacher at work in the classroom will involve a different set of evaluation criteria from those needed to evaluate a set of learning tasks. An evaluation of teacher performance will also be structured differently from an analysis of learner performance. In many way this is nt so different from the judgements we make other matters, which also involve different criteria at different times for different purposes.²⁹

2.5 Evaluation in Teaching English Process

It is important to be sure when we mentioned the need to evaluate our language teaching methods, our materials, our effectiveness as teachers and so on, that we actually know what it is we are evaluating. How materials are presented to

²⁸ Pauline Rea-Dickins & Kevin Germaine, Evaluation, (Oxford University Press: New York, 2011)

p.4 ²⁹*Ibid*, P. 4

learners, the types of learning tasks used, and the way that we design or courses, all from different aspects of our work as teachers. They are full part of the curriculum, of the full range of activities which take place both prior to and during the implementation of a learning programme. And they must be evaluated.

Evaluation has a different overall focus and several different purposes from student assessments. While evaluation may be seen as a 'means' analysis, student assessment has a much more limited perspective with a focus on the 'ends' f learning it terms of what the learner has achieved at particular points. ³⁰

2.6 Evaluation in Teaching English Process in Secondary school

An Adult Education Program in a large US city offers a series of conversations courses for immigrants wanting to improve their ability to participate in small talk. An exit test is needed to measure students' ability to converse with native speakers about topics included in small talk. The Adult Education Program will use this test to assign grades, as well as provide a certificate indicating the students' profile of language ability. Test takers might use the certificate to help obtain employment, for admission to another program, or simply for personal satisfaction.³¹

-

³⁰ Ibid, P. 5

³¹ Lyle F. Bachman and Adrian S.Palmer, Language Testing In Practice, (Oxford University Press: New York, 2009), P. 298

2.7 Definition of MCQ test

Multiple choice is a form of assessment in which respondents are asked to select the best possible answer (or answers) out of the choices from a list. The multiple choice format is most frequently used in educational testing. The items of a multiple choice test are often colloquially referred to as "questions," but this is a misnomer because many items are not phrased as a question. For example, they can be presented as incomplete statements or mathematical equations. Thus, the more general term "item," is the most appropriate label. Items are stored in an item bank.

Multiple choice testing has proliferated as a result of attempts to use multiple choice items to perform tasks for which they were never intended. Moreover, since the multiple choice items is one of the most difficult and time consuming types of items to construct, numerous poor multiple choice tests now abound. Indeed, the length of time required to construct good multiple choice items could often have been better spent by teachers on other more useful tasks connected with teaching or testing.³²

From reading J.B Heaton's book, the writer got some information about Multiple Choice Question (MCQ) test, such as:

a. MCQ test good for:

1. particularly at the lower levels,

³² J.B. Heaton, *Op Cit*, p. 27

2. test knowledge of grammar, vocabulary, rather than the ability,

3. measuring students' ability to recognize connects grammatical form, to make

important discriminations in the target language, and

4. multiple choice items offer a useful introduction to the construction of

objective test.³³

b. Multiple Choice Items.

Multiple choice items take many forms, but their basic structure is as follows.

1. there is a stem:

The primary purpose of the stem is to present the problem clearly and

concisely. The testee should be able to obtain from the stem a very general idea of

the problem and the answer required. At the same time, the stem should not

contain extraneous information or irrelevant clues, thereby confusing the problem

being tested.34

2. option:

For normal purpose of testing, this should be clearly the correct or best

option: thus, it is most important that each item should be checked by another

person.

³³ *Ibid*, p. 27 ³⁴ Ibid, p.30

3. distractor:

Each distractor or incorrect option should be reasonably attractive and plausible. It should appear right to any testee who is unsure of the correct option.

It is the candidate's task to identify the correct or most appropriate option. Perhaps the most obvious advantages of multiple choices, referred to earlier in the book is that scoring can be perfectly reliable. Scoring should also be rapid and economical. A further considerable advantage is that, since in order to respond the candidates has only to make a mark on the paper, it is possible to include more items than would otherwise be possible in a given period of time.³⁵

a. The Characteristics Of Good MCO test are:

- 1. There are should be one answer
- 2. Only one feature at a time should be tested
- 3. Each option should be grammatical correct when placed in stems
- 4. Be efficient about using word or phrase and sentence
- Chronological (in many tests, items are arranged in rough order of increasing difficulty)
- 6. It should not be dependent on other question
- 7. The stems should not give clues or answer to other question
- 8. Be careful on capital letters

³⁵ Arthur Hughes, Op Cit, p. 76

9. All multiple choice items should be at a level appropriate to the proficiency level of the testees.

2.8 Validity

A test is said to be valid if it measures accurately what it is intended to measure. We create language tests in order to measure such essentially theoretical constructs as 'reading ability', 'fluency in speaking', 'control of grammar', and so on.³⁶ J.B Heaton also said in his book that the validity of a test is the extent to which it measures what it is supposed to measure and nothing else. Every test, whether it be a short, informal classroom test or public examination, should be as valid as constructor can make it. The test must aim to provide a true measure of the particular skill which is intended to measure, to the extent that it measures external knowledge and other skills at the same time, it will not be a valid test.³⁷

Face Validity

According J.B Heaton, if attest items looks right to other testers, teachers, moderators, and testess, it can be described as having at least face validity. As constructors of the test, we can become so involved in the test that we sometimes fail to stand back and look at the individual test items objectively. Only it the test is

³⁶ *Ibid*, p 26 ³⁷ J.B Heaton, *Op Cit*, p 159

examined by other people can some of the absurdities and ambiguities then be discovered.38

The concept of face validity is far from new in language testing but the emphasis now placed on it is relatively new. In the past, face validity was regarded by many test writers simply as a public relations exercise. Today, however, most designers of communicative tests regard face validity as the most look valid even as far as the reproduction of the material itself is concerned.³⁹

b. Content Validity

The first form of evidence relates to the content of the test. A test is said to have content validity if its content constitutes a representative sample of the language skill, structure etc. with which it is meant to be concerned. It is obvious that grammar test, for instance, must be made up items relating to the knowledge or control of grammar, but this in itself does not ensure content validity. The test would have content validity only if it included a proper sample of the relevant structures.⁴⁰

Arthur Hughes said that the importance of content validity are, first, the great a test's content validity, the more likely it is to be an accurate measure of what it is supposed to measure, i.e. to be have construct validity. A test in which major areas identified in the specification are under-represented or not represented at all, is unlikely to be accurate. Second, such a test is likely to have a harmful backwash

³⁸ *Ibid*, p. 160 ³⁹ Ibid, p. 160

⁴⁰ Arthur Hughes, *Op Cit*, p. 26

effect. Areas that are not tested are likely to become areas ignored in teaching and learning. Too often the content of tests is determined by what is easy to test rather than what is important to test. The best safeguard against this is to write full test specifications and to ensure that the test content validity is a fair reflection of these. For this reason, content validation should be carried out while a test is being developed, it should not wait until the test is already being used.⁴¹

2.9 Reliability

Reliability is a necessary characteristic of any good test, for it to be valid at all, a test must first be reliable as a measuring instrument. If the test is administered to the same candidates on different occasions, then, to the extent that it produces differing result, it is not reliable. Reliability, measured in this way is commonly referred to as test/re-test reliability and the other kinds of reliability. This latter kind of reliability denotes the extent to which the same marks or grades are awarded if the same test papers are marked by: (i) two or more different examiners or (ii) the same examiner on different occasions. In short, in order to be reliable, a test must be consistent in its measurements.⁴²

Factors affecting the reliability of a test are:

a. The extent of the sample of material selected for testing, whereas validity is concerned chiefly with the concept of the sample, reliability is concerned

⁴¹ *Ibid*, p. 27 ⁴² J.B Heaton, *Op Cit*, p. 162

with the size. The larger the sample, the greater the probability that the test as a whole is reliable-hence the favouring of objective tests, which allow for a wide field to be covered.

b. The administration of the test, is same test administered to different groups under different conditions or at different times? Clearly, this is an production and listening comprehension.⁴³

One method of measuring the reliability of a test is to-administer the same test after a lapse of time. It is assumed that all candidates have been treated in the same way in the internal-that they have either all been taught or that none of them have. Provided that such assumptions can be made, comparison of the two results would then show how reliable the test has proved. Clearly, this method is often impracticable and, in any case, a frequent use of it is not to be recommended, since certain students will benefit more than others by a familiarity with the type and format of the test. Moreover, in addition to changes in performance resulting from the memory factor, personal factors such as motivation and differential maturation will also account for differences in the performances of certain students.⁴⁴

2.10 Item Analysis

The purpose of item analysis is to examine the contribution that each item is making to the test. Items that are identified as faulty or inefficient can be modified or

⁴³ *Ibid*, p.162 ⁴⁴ *Ibid*, p. 163

rejected. 45 For the result obtained from objective tests can be used to provide valuable

information concerning:

a) The performance of the students as a group, then informing the teacher about

the effectiveness of the teaching,

b) the performance of individual students, and

c) the performance of each of items comprising the test.⁴⁶

a. Item Difficulty

J.B. Heaton in his book said that the difficulty (or facility value) of an

item simply shows how easy or difficult the particular item proved in the test.

The index f difficulty (FV) is generally expressed as the fraction (or percentage)

of the students who answered-the item correctly. It is calculation by using the

formula:

$$FV = \frac{R}{N}$$

77'.1 D .

With: R =the number of correct answers

N = the number of students taking the test

A further argument or including items covering a range of difficulty

levels is that provided b motivation. While the inclusion of difficult items may be

necessary in order to motivate the good student, the inclusion o very easy items

will encourage and motivate the poor student. In any case, a few easy items can

⁴⁵ Arthur Hughes, *Op Cit*, p.225

provide a 'lead-in' for the student a device which may be necessary if the test it at all new or unfamiliar or if there are certain tensions surrounding the test situation.47

b. Item Discrimination

The discrimination index of an item indicates the extent to which the item discriminates between the testees, separating the more able testees from the less able. The index of discrimination (D) tells whether those students who performed well on the whole test tended to do well or badly on each item in the test. It is presupposed that the total score on the test is a valid measure of the student's ability (the good student tends to do well on the test as a whole and the poor student badly). On this basis, the score on the whole test is accepted as the criterion measure, and it thus becomes possible to separate the good students from the bad ones in performances on individual items. If the good students tend to do well on item and the poor students badly on the same item, then the item is a good one because it distinguishes the good from the bad in the same way as the total test score.48

4. Distractors Analysis

Each distractor or incorrect option should be reasonably attractive and plausible. It should appear right to any testee who is unsure of the correct option.

⁴⁷ *Ibid*, p. 179 ⁴⁸ *Ibid*, p.179-180

Items should be constructed in such us a way that students obtain the obviously incorrect options.

For most purposes, each distractor should be grammatically correct when it stands by itself: otherwise testess will be exposed to incorrect forms.

Unless a distractor is attractive to the student who is not sure of the correct answer, its inclusion in a test item is superfluous. Plausible distractors are best based on:

- a) mistakes in the students' own written work,
- b) their answer in previous test,
- c) these teacher's experience, and
- d) a contrastive analysis between the native and target languages.

Distractor should not be difficult nor demand a higher proficiency in the language than the correct option. If they are too difficult, they will succeed only in distracting the good student, who will be led into considering the correct option too easy (and trap)⁴⁹

2.11 Previous Studies

In this part, the writer present the result of the research concerning with the writer's research. There were five researchers who done this same research before, they were:

_

⁴⁹ Ibid, p. 32

- 1. Aan Kulistyono Ujiwandi, his research was about An Analysis Of The Teacher Made English Test Items In UAS Semester 2 2005 / 2006 for the first year students of SMAN 1 Puri Mojokerto. His research focused on the content validity, reliability, analysis of item difficulty and index of discrimination. The result of his research was SMA 1 Puri Mojokerto has good content validity, does not appropriate reliability, do not have acceptable item difficulty and a not adequate index of discriminaton. The differences of the writer study with this research are the writer focused just on MCQ test and the writer add face validity to be analyzed and measure the quality of the MCQ English test based on the criteria of constructing a good English MCQ test.
- 2. Sigit Eko Saputro, his research was about An Analysis of The Teacher Made Of The First Year Student Of SMA 2 Magetan. His research almost same with the first research, he focused on the content validity, reliability, index of difficulty, and index of discrimination. The result of his research was SMA 2 Magetan has high content validity, an adequate validity, not acceptable index of difficulty, not adequate index of discrimination. The differences of the writer study with this research are the writer more specific in MCQ test, and the writer try to measure the

⁵⁰ Aan Kulistyono Ujiwani, An Analysis Of The Teacher Made English Test Items In UAS Semester 2 2005 / 2006 for the first year students of SMAN 1 Puri Mojokerto, Unpublished S1 Thesis. (Surabaya: Universtas Negeri Surabaya, 2006)

- validity include face validity and content validity which did not analyze by the researcher before and the writer also add an analysis of distractor in her research.⁵¹
- 3. Millatul Islamiyah, her research was about Content Validity and Item Analysis Of Semester II English Final Test For Tenth Grade Students Of SMAN 3 Sidoarjo. Her research was focused on content validity and item analysis only. The result of her research reported that Semester II English Final Test for Tenth Grade of SMAN 3 Sidoarjo has good content validity. It also reported that the index of difficulty for X1 and X2 are acceptable, but they are recognized easy test for X3. Besides, the index of discrimination of this test is for X1, satisfactory for X2, add malfunction for X3. Moreover, the semester II English final Test for Tenth Grade Of SMAN 3 Sidoarjo has good distractor for X1, X2, and X3. The differences this research with the writer's research is lied on the writer focused on the English MCQ test of 7th grade SMP Buana Waru, so the result of this research can be used as an instrument to measure their quality in constructing English MCQ test, but the previous research did not mention who was the test maker. So her result did not give significant English system learning in this school especially for the teacher. 52
- 4. Ria Dewi Pratiktasari, her research with the title An Analysis of Semester II

 English Summative Test for the Eighth Year Students of SMPN 1 Slahung

⁵¹ Sigit Eko Saputro, An Analysis Of The Teacher Made English Test Items In UAS Semester I 2005 / 2006 Of The First Year Students Of SMA 2 Magetan, Unpublished S1 Thesis. (Surabaya: Universtas Negeri Surabaya, 2006)

⁵² Millatul Islamiyah, Content Validity and Item Analysis Of Semester II English Final Test For Tenth Grade Students Of SMAN 3 Sidoarjo, Unpublished S1 Thesis, (Surabaya: Institut Agama Islam Negeri Sunan Ampel Surabaya, 2009)

Ponorogo. Her research focused on content validity and item analysis includes index of difficulty and index of discrimination. The result of her research was SMPN 1 Slahung Ponorogo has 75 % agreement with the curriculum, it means it has a good content validity but it as low reliability, acceptable index of difficulty and poor index of discrimination. The difference of this research with the writer's research is the writer more focuses on MCQ test. The writer analyzes about the validity includes face validity and content validity, reliability, item analysis includes index of difficulty and index of difficulty. The writer also measures the quality of English MCQ test based on the characteristic of a good MCQ test.⁵³

5. Erlik Widiyanistyati, her research almost same with the other researchers before. She analyzed about the content validity, reliability, index of difficulty and index of discrimination. Her result of her research An Analysis of the English Final Test for the First Semester of the Twelfth Year of SMAN 1 Surabaya was SMAN 1 Surabaya has low content validity, high reliability, appropriate index of difficulty, and poor index of discrimination. The differences of the writer study with this research are the writer more specific in MCQ test, and the writer measure the validity include face validity and content validity which did not analyze by the researcher before and the writer also add an analysis of distractor in her research. 54

53

⁵³ Ria Dewi Pratiktasari, An Analysis Of Semester II English Summmative Test For The Eighth Year Students Of SMPN 1 Slahung Ponorogo, Unpublih S1 Thesis, (Surabaya: Universitas Negeri Surabaya, 2006)

⁵⁴ Erlik Widiyanistyati, An Analysis of the English Final Test for the First Semester of the Twelfth Year of SMA NEGERI 13 SURABAYA, Unpublished S1 Thesis (Surabaya: Universitas Negeri Surabaya, 2008)

CHAPTER III

RESEARCH METHODOLOGY

3.1 Research Design

This study deals with An analysis of English Multiple Choice Questions (MCQ) Test of 7th grade at SMP BUANA Waru Sidoarjo. It tried to answer the question of how is the quality of English Multiple Choice Question (MCQ) test of 7th grade at SMP Buana Waru. In this research the writer used qualitative methods. This research it a descriptive qualitative research because this research gives us description about situation and condition of something. ⁵⁵ In SMP BUANA especially in 7th grade, the managing of the class is each class consists of 42-44 students with heterogenic ability. So, every class is a big class with the heterogenic students with different ability.

The qualitative methods is used because the writer want to know how is the quality of MCQ test which has been written by English teacher in 7th grade at SMP BUANA Waru Sidoarjo based on the charateristic of a good MCQ test, validity, reliability and items analysis. Qualitative method gives chance for writer to study about the isues, cases, and events that have been chosen before.⁵⁶

55 M. Nazir, Metode Penelitian, (Jakarta: Ghalia Indonesia, 2003), p. 55

⁵⁶ Michael Patton, Metode Evaluasi Kualitatif, (Semarang: Pustaka Pelajar, 2006), p. 5

3.2 Subject and Object of Research

The subjects of this study is English Multiple Choice Question (MCQ) test that has been constructed by an English teacher of 7th grade in SMP BUANA Waru Sidoarjo, especially in 7B, 7D and 7E, and the objects in this research are MCQ test paper and students' worksheet papers.

To choose the subjects of this research, the writer used to *purposive* sampling technique, it means choosing the subjects based on some characteristics that have close relationship with the previous population.⁵⁷ This technique also can be said as *judgemental sampling*, it means the choosing of sample based on the goal of research.

3.3 Research instruments

To collect the data that are needed for this study, the researcher would use two forms of is instruments. They were observation checklist, question test, and questionnaire. The detailed explanation of each instrument can be seen bellow:

a. Observation checklist

Observation is the primary technique which used by the writer to collect the data. There are two kinds of observation check list, those are participant and non participant. Participant observation means that the observer becomes a

⁵⁷ Cholid Narbuko dan Abu Achmadi, Metode Penelitian, (Jakarta: Bumi Aksara, 1997), p. 116

part of the situation being observed. On the other hand, non participant observation means that the observer is not directly involved in the situation being observed.

In this research included as non participant observation. When the researcher is being an observer, she would not take a part to teach student. Nevertheless when the researcher is being the teacher, she would not be an observer but the real teacher who will be a non participant observer. While participating as an observer, she would observe the application of the technique; the materials add the activities in the classroom. All of these points would be written down in the checklist.

Observation checklist used to identify some characteristics in MCQ test paper. Based on the explanation above, the researcher would get the qualitative data to answer the statements of problem in number 1 and 2.

b. Students' sheet paper

The data took from MCQ test paper and students' worksheet paper, and after that it would be analyze to measure the quality of MCQ test of 7th grade at SMP BUANA WARU SIDOARJO. After using this instrument the researcher would get statistic data and then the researcher would analyze it and get statistic data to answer the statements of problem in no 3.

3.4 Data and Source of Data

In this study the data was obtained from the items of UAS test, the key answer, the students' answer sheets and also the result of the students' score and the standard and basis competencies of 2010 English curriculum for Seventh grade of Junior High school in 2010-2011.

It is used to conduct data about the following points:

- 1. The test items and key answer of the students in Seventh grade of SMP BUANA Waru Sidoarjo.
- The students' answer sheets of the students in Seventh grade of SMP BUANA
 Waru Sidoarjo.
- The students' score of the Second semester in 2010- 2011, Final test of the students in Seventh grade of SMP BUANA Waru Sidoarjo.
- 4. The standard and basic competences of 2010 English curriculum for seventh grade of Junior High school.

3.5 Data Collection Techniques

The data in this study was collected by study documentation. Documentation is a method to get anything on the form of notes, transcripts, magazine, etc and the data by documentation would be collected through these following steps:

 The writer came to the school and met headmaster of the school to ask permission.

- The writer came to the English teacher to ask permission to analyze their MCQ test.
- The writer borrowed the MCQ paper test and the students' worksheet to be analyzed.
- 4. The writer analyzed the MCQ paper test based on the criteria of a good MCQ test, based on the validity, reliability and based on item analysis.
- 5. The writer analyzed the students' worksheet to measure the quality of the test based on statistic analysis.

3.6 Data Analysis

The result of collected data was analyzed by using descriptive, means that data would be described as the way it is.

1. Analyzing MCQ test based on the characteristict of a good MCO test.

In analyzing the characteristic of a good MCQ test, the writers used the observation check list, that contains of some characteristics of a good MCQ test, tas following:

Table I: table of analyzing the characteristicts of a good test

The characteristicts of a good MCQ test	Appropriate	Not appropriate
There should be one answer		

Only one feature at the time should be tested	
Each option should be grammatically correct when placed in stems	
Be efficience of using word, phrase, and sentences.	
Chronological	
It not should be dependent or other question	
The stems should not gives clues or question to other question	
Be carefull on capital letter	
All multiple choice items should be at a level appropriate to the proficiency level of the testess	

2. Analyzing the Validity of the test

A. Analyzing face validity

In analyzing the face validity, the writer just observed the performence of the test. Include the layout and the paper. After that the writer made a conclusion of these.

B. Analyzing the Content Validity

In analyzing the content validity, the writer collected it through the following steps:

- a. Making a list of the standard competencies, basic competencies, indicators and learning expirience for the seventh grade of SMP BUANA Waru.
- b. Placing each of the test items in the appropriate place with the standart competencies and basic competencies to identify whether or not the standart competencies and basic competencies covered by the MCQ test.
- c. Counting the persentages of the items test of every language aspects.
- d. Concluding the result of analysis.

In order to make the steps clearly, the writer presents the illustration of the steps as following:

Table 2: Table for analysis the content validity

Standard of Competenci es	Basic of Competencies	Indicators	Learning Experience	Item Test	S	%
Mendengar kan 7. Memahami makna dalam percakapan transaksiona l dan interpersona l sangat sederhana	7.1. Merespon makna dalam percakapan transaksional (to get things done) dan interpersonal (bersosialisasi) sangat sederhana secara akurat, lancar & berterima untuk	Merespon ungkapan: • Meminta dan memberi jasa • Meminta dan memberi barang • Meminta dan memberi fakta	1. Tanya jawab tentang berbagai hal terkait materi 2. Membahas kosakata dan tata bahasa: noun phrase, adjective, adverb, verb yang terkait dengan meminta dan memberi			

untuk berinteraksi dengan llingkungan terdekat	berinteraksi dengan lingkungan terdekat yang melibatkan tindak tutur: meminta dan memberi jasa, meminta & memberi barang, serta meminta dan memberi fakta	jasa, meminta dan memberi barang, meminta dan memberi fakta 3. Mendengarkan percakapan sesuai dengan materi 4. Menjawab pertanyaan sesuai dengan isi dan struktur percakapan 5. Menjawab pertanyaan tentang		

3. Analyzing the index of difficulty

In analyzing index of difficulty, the writer do the following steps:

- a. finding the students' score.
- b. arranging te students' score from the upper to the lower.
- c. classifying the students into two groups, they are upper and lower group.
- d. counting the item difficulty by using the formula:

$$F.V = \frac{U+L}{\text{the number of student who taking test}} \quad \text{or} \quad F.V = R/N$$

Where:

F = index of difficulty

R = the number of students who answer correcly

N = the number of students who taking the test

4. Analyzing the index of discrimination

The index of discrimination indicates the extent to whichthe item discriminates between the testess, separating the more able students from the less able. To analyzing index of discrimination the writer do steps such as in index of difficulty. Then, to calculate the index of discrimination, the writer used the formula:

$$D = \frac{\textit{Correct U-Correct L}}{n}$$

Where:

D = the index of discrimination

Correct U = The number of students in upper group who answered the items correctly.

Correct L = the number of students in lower group who answered the items correctly.

n = the number of taking the test in one group (upper group/ lower group).

The index of discrimination value based on Arikunto:

D : 0.00 - 0.20 (poor)

D: 0.20 - 0.40 (satisfactory)

 $D : 0,40-0,70 \pmod{}$

 $D : 0.70 - 1.00 \quad \text{(excellence)}$

Table 3: table of analyzing index of difficulty and index of discrimination.

						Comment		
No Item	υ	L	U+L	U-L	F.V	ט	F.V	D
1	-							
2							- ""	
3								

5. Analyzing of effectiveness of distractors

Beside calculating index of difficulty and discrimination, it also important to analyze the items in very detail, moreover on those which cannot perform as expected. Analyzing the distractor aimed not only toknow which items that cannot work properly, but also to check why particular test taker failed to answer certain items correctly.

Distractors can function well if these are chosen by students from the lower level. According Arikunto, the distractor which is chosen at least by 5% students who taking the test is called dood test.

Table 4: the example of analyzing the effectiveness of distractors.

No Item	Option	Upper	Lower	Comment
1	A*			
	В			
	С			
	D			
	0			

CHAPTER IV

DATA ANALYSIS

In this chapter, the researcher explain the result of her research. The researcher do analysis about the quality of MCQ test that has been constructed by an English teacher of 7th grade in SMP BUANA Waru based on the characteristic of a god MCQ test, the validity include face validity and content validity, reliability, and item analysis include index of difficulty index of discrimination and analysis of distractors. The researcher took 3 classes of 7th grade in SMP BUANA Waru, they are 7B, 7D and 7E.

4.1 Analyzing of a Good MCQ Test

In analyzing of a Good Multiple Choice Question (MCQ) test, the researcher used the statement of J.B Heaton. He mentioned that there were 9 points of the characteristics of a good MCQ test, they were there are should be one answer, only one feature at a time should be tested, each option should be grammatical correct when placed in stems, be efficient about using word or phrase and sentence, chronological (in many tests, items are arranged in rough order of increasing difficulty), it should not be dependent on other question, the stems should not give clues or answer to other question, be careful on capital letters, and all multiple choice items should be at a level appropriate to the proficiency level of the testees.

Table analysis in the table 1 appendix I show if the MCQ test in SMP BUANA Waru can be categorized as a good MCQ test, because it has almost all the characteristics of a good test. First, in each item it has only one answer with clear question and clear instruction, so the students can understand the test easily. Second, this test only one feature at a time should be tested. Third, each option also has grammatical correct when placed in stem, so the stems not give clues or question to the question. Furthermore, it can be categorized that the distractor is useful because it seem correct. Fourth between a question with another questions be dependent and does not give clues to other question. And the last all multiple choice items should be at a level appropriate to the proficiency level of the testees.

4.2 Analyzing of Validity

In analyzing of validity, the researcher focused on face validity and content validity aspect.

a) Analyzing of Face Validity

Face validity here means the performance of the test, such as the paper and the layout. By looking the paper test, we can say that this test has a good layout with the good arranging. The teacher divided the paper test into 2 columns, it has some advantages. Firstly is more efficient because do not waste many space places. And the secondly, the students more easy to read

the text, it can help students to efficient the time. The teacher also consistent in using the capital letter, but there are some question which has not efficient stems, for example in question number 6, 12, 43, and 48. There are some repetition words and it can waste the students' time to do the test.

b) Analyzing of Content Validity

Table analysis in table 2 appendix II show us the syllabus of English subject in 7th grade SMP Buana Waru included standard of competencies and basic competencies 0f 2010 curriculum is used. Then, to categorize whether the English MCQ test of SMP Buana Waru fulfill the agreement of content validity or not, the *standar isi* table is matched with the materials in the test. Furthermore, place each item number in appropriate intersection of *standar isi* table of 2010 curriculum to identify the representative sample.

Table analysis in appendix 3 is used to ease the writer in analyzing content validity of English Multiple Choice (MCQ) test that have been constructed by English teacher in 7th grade in SMP BUANA Waru. There are six columns in that table. The first column contains standard competence, second column contains of basic competencies, the third column contains of indicators, the fourth column contains of learning experience, the fifth column contains of items test that appropriate with the basic competencies, the next column contains of the number of items test (S) and the last column contains

the percentage of total numbers of particular items represent the elated basic competence.

According to J. B Heaton, the test can be said has a good content validity if it covers all the contents as stated in the curriculum. Based on the result of analyzing content validity in appendix 3, the percentage of every aspect of the learning content is conclude as follows:

- 1. There are 8% or 4 items for speaking which focus on asking and giving services.
- 2. There are 4% or 2 items for speaking which focus on like and dislike.
- 3. There are 4% or 2 items for speaking which focus on giving response (sorry, thanks, please)
- 4. There are 6% or 3 items for speaking which focus on greetings and giving response.
- 5. There are 20% or 10 items for listening which focus on giving opinions and answer the question based on the text.
- 6. There are 12% or 6 items for reading which focus on announcement.
- 7. There are 30% or 15 items for reading which focus on descriptive and procedure text, with 16% or 8 items for procedure text and 14% or 7 items for descriptive text.
- 8. There are also 16% or 8 items test that did not cover the available materials.

Based on the result above, we can conclude English Multiple Choice Question (MCQ) test that have been constructed by English teacher in 7th grade in SMP BUANA Waru is good since 84% items test represent all materials. It is more than 50%, according to Bloom if the agreement of the test is 50% or more, it can be conclude that the test has high content validity.

Moreover, there are 8 items or 16% all of the test did not cover the materials, they are the items test number 18, 25, 34, 35, 36, 39, 46 and 47. Those items are not suitable with the indicators of standard and basic competencies and were not taught in this semester.

4.3 Analyzing of Reliability

In analysis the reliability, we can conclude from the students' score. Based on Arikunto in his book "Dasar-Dasar Evaluasi Pendidikan", an instrument to measure the reliability in the test can be seen by looking index of difficulty. And a test can be said has high reliability if the index of difficulty is high.⁵⁸

Thus, the result of index of difficulty of the MCQ test that have been constructed by an English teacher of 7th grade in SMP BUANA Waru can be said as appropriate index of difficulty, so we can conclude that this test also has high reliability. It can be seen from the students' score too. From the result of the students' score in 7B, 7D and 7E, the result is not so different, so this test can be categorized has high reliability.

_

⁵⁸ Arikunto, Suharsimi. Op Cit. P. 206

4.4 Analyzing of Item Test

1. Analyzing Index of Difficulty

To get the data for analyze index of difficulty, the researcher divided each class into 2 groups. The first group was upper group, they were the number of students who got good score and the other group was lower group, they were the number of students who got the bad score. The researcher could get these groups after she arranged the students' score from the upper to the lower. After the researcher got the data, she done the analysis used the statistic method by using $F.V = \frac{U+L}{the\ number\ of\ student\ who\ taking\ test}$ formula to measure the index of difficulty of the test.

There are six columns in the table of analysis index of difficulty and index of discrimination. First columns contains the number of item. Second column contains the score of student in upper group who answer correctly of each item. Third column contains the score of lower group who answer correctly of each item. Fourth column contains the value of index of difficulty. The fifth column contains the value of index of discrimination and the last column contain of comment for index of difficulty and index of discrimination. (See table 3)

In addition, the researcher treated the same way in dividing class into two groups. In order to determine upper and lower level in each class. The researcher did this analysis of the MCQ test in 7th grade of SMP BUANA Waru, they are 7B, 7D, and 7E. Absolutely, there are 6 classes of 7th grade in SMP BUANA but the researcher just take half of all the class and this sample was taken randomly.

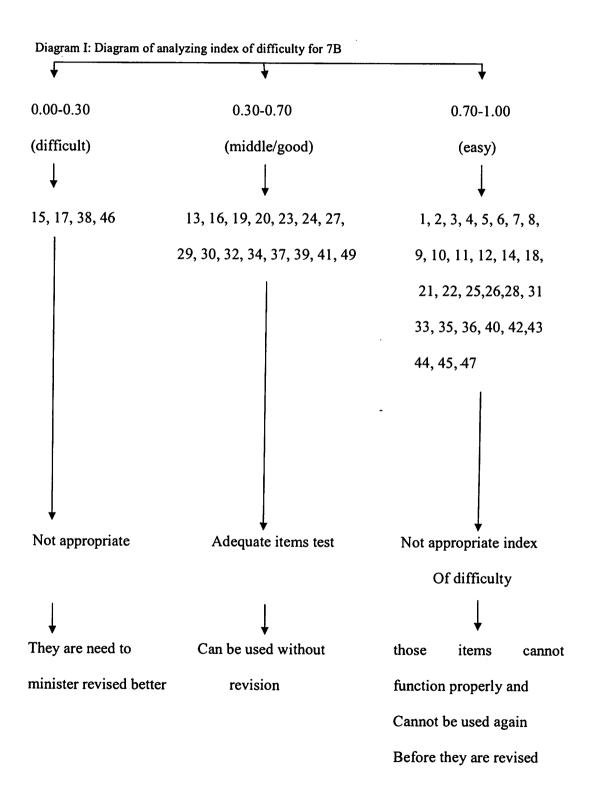
Thus, the number of students taking the test in 7B that used to measure index of difficulty and index of discrimination later are 44 students. So, the upper group consists of 22 students and lower groups consist of 22 students to. This condition is same with 7D class but different with 7E class. 7E class consists of 42 students who taking the test, so the upper group consists of 21 students and the lower group consist of 21 students to.

a) Analyzing of Index of Difficulty for 7B Class.

From the table of appendix 6, we can see the result of index of difficulty of 7B students in SMP BUANA Waru. From this table we can see that there are 62% of the items tests which can be answered by the students correctly, according Suharsimi Arikunto the item that shows value between 0.70 - 1,00 is called easy. We can say that this result is not appropriate index of difficulty and those easy items cannot function properly and cannot be used again before they are revised. Furthermore, the item numbers of the test which can be categorized as easy item test are the item test of numbers 1,

⁵⁹ Ibid, P.210

2, 3, 4, 5, 6, 7, 8, 9. 10, 11, 12, 14, 18, 21, 22, 25, 26, 28, 31, 33, 35, 36, 40, 42, 43, 44, 45 and 47, so the totals are 31 items. Moreover there are 8% of the tests that categorized as difficult item tests because just there are few students who can answer these items correctly with index of difficulty value 0,00 – 0.30. The item numbers of the test which can be categorized as difficult item test are the item test of numbers 15, 17, 38 and 46, so the totals are 4 items. Those items are needed to ministere revised better because they are not able to do well by the students. And the other 30% is categorized as middle or good item tests because has index of difficulty value between 0.30 – 0.70, they are the item tests of numbers 13, 16, 19, 20, 23, 24, 27, 29, 30, 32, 34, 37, 39, 41 and 49, so the totals are 15 items. Those items are concluded as adequate item test. it means those items can be fully utilized and become reference to design other items for other test. We can see the conclusion of index of difficulty in 7B by this following diagram.



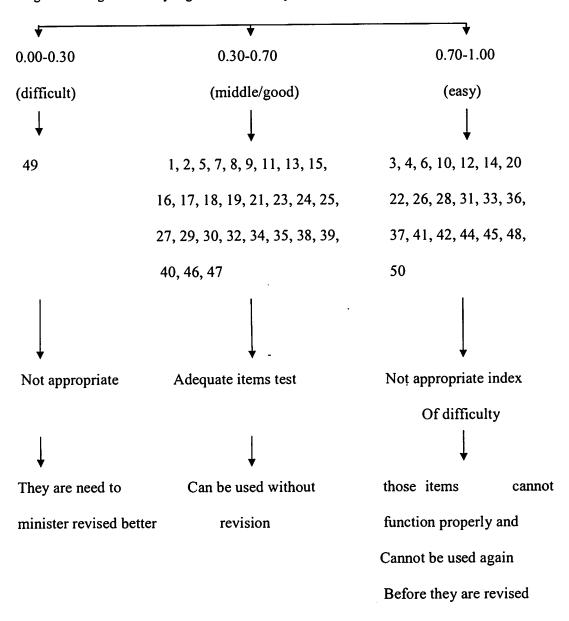
b) Analyzing of Index of Difficulty for 7D Class.

From the table of appendix 7, we can see the result of index of difficulty of 7D students in SMP BUANA Waru. From this table we can see that there are 42% of the items tests which can be answered by the students correctly, according Suharsimi Arikunto the item that shows value between 0.70 - 1,00 is called easy. 60 Furthermore, the item numbers of the test which can be categorized as easy item test are the item test of numbers 3, 4, 6, 10, 12, 14, 20, 22, 26, 28, 31, 33, 36, 37, 41, 42, 43, 44, 45, 48 and 50, so the totals are 21 items. We can say that this result is not appropriate index of difficulty and those easy items cannot function properly and cannot be used again before they are revised. This case also make the students do not have effort to study hard because they think the test is easy. Moreover there are 2% of the tests that categorized as difficult item tests because just there are few students who can answer these items correctly with index of difficulty value 0.00 - 0.30. The item numbers of the test which can be categorized as difficult item test are the item test of number 49, so the total is 1 item. Those items are needed to minister revised better because they are not able to do well by the students. And this case makes motivation of the student to study decrease because they think that although they have studied they still could to do the test well, and make them to be lazy. And the other 56% is categorized as

⁶⁰ Ibid, P.210

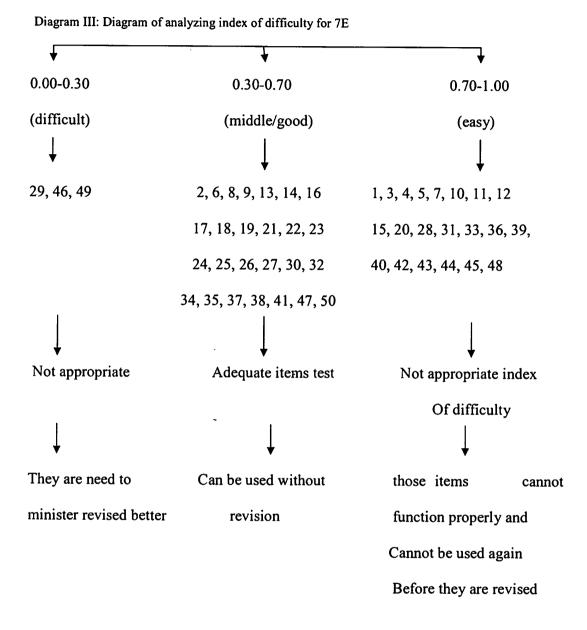
middle or good item tests because has index of difficulty value between 0.30 – 0.70, they are the item tests of numbers 1, 2, 5, 7, 8, 9, 11, 13, 15, 16, 17, 18, 19, 21, 23, 24, 25, 27, 29, 30, 32, 34, 35, 38, 39, 40, 46 and 47, so the totals are 28 items. Those items are concluded as adequate item test. It means those items can rewrite without revision. We can see the conclusion of index of difficulty in 7B by this following diagram.

Diagram II: Diagram of analyzing index of difficulty for 7D



c) Analyzing of Index of Difficulty for 7E Class.

From the table of appendix 8, we can see the result of index of difficulty of 7E students in SMP BUANA Waru. There are 42% of the items tests which have index of difficulty value between 0,70 - 1,00, so the item tests can be categorized as easy test. The item numbers of the test which can be categorized as easy item test are the item test of numbers 1, 3, 4, 5, 7, 10, 11, 12, 15, 20, 28, 31, 33, 36, 39, 40, 42, 43, 44, 45 and 48, so the totals are 21 items. Furthermore, there are 6% of the tests that categorized as difficult item test wit index of difficuly value 0, 00 - 0.30. The item numbers of the test which can be categorized as difficult item test are the item test number 29, 46, and 49, so the totals are 3 items. And the other 52% is categorized as middle or good item test and this test appropriate to use without revised, they are the item test of numbers 2, 6, 8, 9, 13, 14, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 30, 32, 34, 35, 37, 38, 41, 47 and 50, so the totals are 26 items. We can see the conclusion of index of difficulty in 7B by this following diagram.



Based on those result we can generally conclude that index of difficulty in SMP BUANA Waru are good with index of difficulty value 0.30-0.70. Although the result in 7B shown that most of items categorized as easy item test with 62%. Those

items may be used again for the next test without revised or able to references for the test maker to design other test.

However, there are so many numbers of items test that shown as easy test but there are some difficult items to. Both the difficult and the easy one must be revised because they cannot work as expected to measure students' progress and which students that able to catch the lesson that have been taught by their teacher well. But the difficult item tests can make the student confuse and spend their time only focus to answer those certain items and make them become lazy because they think they have studied hard but they still can do the test well, so their motivation to study decrease. In the other hand, the easy item test do not give chance to students to increase their thought to think hard, so they will to be lazy to study.

2. Analyzing Index of Discrimination

In analyzing the index of discrimination, the researcher used the same step like in analysis of index of difficulty. First, the researcher arranged the students' score from the high score to the low score. Second the researcher divided the students into 2 groups, they are upper group and lower group. Next the researcher counted the index of discrimination used D = $\frac{Correct U - Correct L}{n}$ formula. And the last step is concluded the result using the Arikunto's theory to interpret index of discrimination value,

D: 0,0 up to 0,20 means it is poor

D: 0,20 up to 0,40 means it is satisfactory

D: 0,40 up to 0,70 means it is good

D: 0,70 up to 1,00 means it is excellent

The researcher did this analysis of the MCQ test in 7th grade of SMP BUANA Waru, they are 7B, 7D, and 7E and started from 7B.

a) Analyzing of Index of Discrimination of 7B

Based on the calculating of index of difficulty in appendix 6, we can conclude that the big number of items tests in 7B that have discrimination value between 0.00 - 0.20 are 20 items or 40% of all items test. based on Arikunto in his book, the discrimination value between 0.00 - 0.20 is poor. So those items must be revised because it cannot separate the good students and the bad student. The number of the items test which have discrimination value between 0.00 - 0.20 are number 5, 6, 10, 13, 14, 15, 17, 19, 21, 22, 26, 29, 32, 33, 34, 40, 43, 44, 45 and 48.

The second numbers of items test are categorized as satisfactory with the discrimination value between 0.20 - 0.40 they are 10 items or around 20% all of the test. Although they are satisfactory, they still need to be redesigned because they are doubtful to use in the other test. The items test

which have discrimination value between 0,20 - 0,40 are number 1, 3, 9, 12, 18, 36, 39, 47, and 49.

From the appendix 6, we can also see that there are 10 items number or around 20% all of the test have the discrimination value between 0.40 - 0.70 and it is categorized as a good item test. It means these items are work properly to differentiate upper students from lower students and they could be utilized for the other test. the items number which have discrimination value between 0.40 - 0.70 are number 7.8, 11, 23, 24, 25, 27, 28, and 41.

In other hand, there are 4 items or around 8% which have negative discrimination value and it is categorized as wrong item test, they are number 4, 31, 35, 38, 42, and 50. Based on Arikunto, the items of the test which have negative discrimination value must be omitted. It means these items discriminate entirely in wrong way and need to be revised, it can make the purpose to discrimination test do not reach, it is to determine between upper students and lower students.

There are 6 items or 12% which have 0,00 discrimination value, it means these items did not work properly because they discriminate nothing. They are items number 4, 31, 35, 38, 42 and 50. To make it clearly, see this diagram following:

0 0.00-0.20 0.20-0.40 0.40-0.70 Negative 0.70-1.00 (satisfactory) (good) (poor) (execelent) (wrong) 4,31 5, 6, 10 1, 3, 9, 12, 7, 8, 11, 4, 31, 38 35, 38 13, 14, 15 18, 36, 39, 23, 24, 25 42, 50 41, 50 17, 19, 21 47, 49 27, 28, 41 22, 26, 29 32, 33, 34 40, 43, 44 45, 48 it work properly They must be need to must be Discrirevised redesign to differentiate upper omitted Minate and lower students nothing

Diagram 4: diagram of analyzing index of discrimination of 7B

b) Analyzing of Index of Discrimination of 7D

In appendix 7 we can see the result of calculating of index of discrimination of 7D. We can conclude that the big number of items tests in 7D are 24 items or 48% which have discrimination value between 0,20 – 0,40, it is categorized as satisfactory item test. It means they still need to be redesigned because they are doubtful to use in the other test. The items test which have discrimination value between 0,20 – 0,40 are number 1, 3, 5, 6,8, 10, 12, 13, 15, 16, 21, 22, 25, 28, 34, 36, 37, 42, 43, 44, 45, 47, 48 and 50.

The second, there are 18 items number or around 36% all of the test have the discrimination value between 0.40 - 0.70 and it is categorized as a good item test. It means these items are work properly to differentiate upper students from lower students and they could be utilized for the other test. The items number which have discrimination value between 0.40 - 0.70 are number 2.7, 9.11.17.18.19, 20, 23, 24, 27, 29, 30, 31, 32, 35, 41 and 46.

In other hand, there are 2 items or around 4% which have negative discrimination value and it is categorized as wrong item test, they are number 33 and 39. Based on Arikunto, the items of the test which have negative discrimination value must be omitted. It means these items discriminate entirely in wrong way and need to be revised, it can make the

purpose to discrimination test do not reach, it is to determine between upper students and lower students.

And there are 3 items or around 6% which have discrimination value between 0.00 - 0.20, it is categorized as poor item test, they are number 14, 26 and 38.

There are 3 items or 6% which have 0,00 discrimination value, they are number 4, 40 and 49. It means these items did not work properly because they discriminate nothing. To make it clearly, see this diagram following:

teacher should be aware to the other distractors, if they want to use again this test in the future

c) Analyzing the Effectiveness of Distractors in 7E

Like in 7B, the effectiveness of distractors of English Multiple Choice Question (MCQ) test that have been constructed by English teacher in 7th grade in SMP BUANA Waru has big number of not function distractors, there are 73 disrtactors. So almost half of the number of those distractors must be revised because no one or less of 5% from the number of students who taking the test choose them. Moreover, there are 8 adequate distractors because they have the same number amount of voters from upper students to lower students and 15 mal function distractors. In other hand there are 49 distractors are categorized as good distractors. It means these distractors work properly and more chosen by students in lower group that upper group.

As a result, from the explanation above we can conclude that the distractors of English Multiple Choice Question (MCQ) test that have been constructed by English teacher in 7th grade in SMP BUANA Waru used in 7E are not function, because almost half of the distractors are categorized as not function distractors and do not work properly because these are not attract the student both from upper group and lower group to chose them. So those distractors must be revised.

BAB V

CONLUSION AND SUGGESTION

In this chapter, the researcher present the conclusion and suggestion. In the conclusion, the researcher conclude the result of her research about "An analysis of English Multiple Choice Questions (MCQ) Test of 7th grade at SMP BUANA Waru Sidoarjo." and the researcher conclude the result from chapter IV, including analysis English MCQ test based on the characteristic of a good MCQ test, face validity, content validity, reliability, and item analysis include index of difficulty, index of discrimination and analysis of distractors. While in the suggestion, the researcher would like to recommend matters that are need to do and not dealing with design good test to the several peoples stated in the first chapter, including teacher, students and further researcher.

A. Conclusion

From data analysis in chapter IV, we can conclude that English Multiple Choice Questions (MCQ) test that has been constructed by English teacher in 7th grade in SMP BUANA Waru Sidoarjo, as follow:

Based on the characteristics of a good MCQ test by J. B Heaton, English Multiple Choice Questions (MCQ) test that has been constructed by an English teacher of 7th grade in SMP BUANA Waru Sidoarjo can be categorized as a good MCQ test, because it has almost all the characteristics of a good test. First, in each item it has only one answer with clear question and clear instruction, so the students

can understand the test easily. Second, this test only one feature at a time should be tested. Third, each option also has grammatical correct when placed in stem, so the stems not give clues or question to the question. Furthermore, it can be categorized that the distractors is useful because it seems correctly. Fourth between a question with another questions be dependent and does not give clues to other question. And the last all multiple choice items should be at a level appropriate to the proficiency level of the testees.

From the result of analysis the validity include face validity or the performance of the test and content validity, it is conducted that English Multiple Choice Questions (MCQ) test that has been constructed by an English teacher of 7th grade in SMP BUANA Waru Sidoarjo has good face validity. This test has a good layout with the good arranging. The teacher divided the paper test into 2 columns, it has some advantages. Firstly, it is no spent large place. And the secondly, the students more easy to read the text, it can help students to efficient the time. The teacher also consistent in using the capital letter, but there are some question which has not efficient stems, for example in question number 6, 12, 43, and 48. There are some repetition words and it can waste the students' time to do the test. Furthermore, from the result of analyzing content validity, we can conclude English Multiple Choice Question (MCQ) test that has been constructed by English teacher in 7th grade in SMP BUANA Waru is good since 84% items test represent all materials. It is more than 50%, according to Bloom if the agreement of the test is 50% or more, it can be

conclude that the test has high content validity. Moreover, there are 8 items or 16% all of the test did not cover the materials, they are the items test number 18, 25, 34, 35, 36, 39, 46 and 47. Those items are not suitable with the indicators of standard and basic competencies and were not taught in this semester.

From the result of analyzing of reliability, we can conclude that English Multiple Choice Questions (MCQ) test that has been constructed by an English teacher of 7th grade in SMP BUANA Waru Sidoarjo has high reliability, it is looked from the index of difficulty of the test.

Based on those results, we can generally conclude that index of difficulty in SMP BUANA Waru are good with index of difficulty value 0.30 – 0.70. Although the result in 7B shown that most of items are categorized as easy item test with 62%. Those items may be used again for the next test. However, there are so many numbers of items test that shown as easy test but there are some difficult items to. Both the difficulty and the easy one must be revised because they cannot work as expected to measure students' progress and which students that able to catch the material that has been taught by their teacher well. But the difficulty item test can make the student confuse and spend their time only focus to answer certain items and make them become lazy because they think that they have studied hard but they still cannot do the test well, so their motivation to study decrease. In the other hand, the easy item test do not give chance for students to increase their thought to think hard, so they will be lazy to study. Whereas, index of discrimination of English Multiple

Choice Questions (MCQ) test that has been constructed by an English teacher of 7th grade in SMP BUANA Waru Sidoarjo of 7B is poor with discrimination value between 0.00 – 0.20 and the percentage around 40%, so for 7B this items must be revised because it cannot separate the good students and the bad students. In the other hand, for 7D the big number is 48% is categorized as satisfactory item test with discrimination value 0.20 – 0.40. Although they are satisfactory items, they still need to be redesigned because they are doubtful to use in the other test. However, for 7E the big number is 48% is categorized as poor test, and those items test must be revised.

B. Suggestion

After recognizing the result of this study related to the characteristics of a good MCQ test, validity, reliability, and item analysis of English Multiple Choice Question (MCQ) test that has been constructed by an English teacher of 7th grade in SMP BUANA Waru, there are several matters that are seemed to be suggested.

As state in the first chapter, the suggestion due to the English teacher, the students and further researcher, as follows:

1. The teacher should try their own test out to know whether the tests are adequate before t is given to the students. So, the result of the test be used to measure the progress of students as the function of a good test.

- 2. Although the result of content validity of English Multiple Choice Question (MCQ) test that has been constructed by an English teacher of 7th grade in SMP BUANA Waru is good, but it is still need to be modified. So the items test can use in the future test.
- The teacher should revise the poor items test and difficult items test. So the teacher can measure their students' ability correctly.
- 4. The teacher should revise the items test that are poor, satisfactory, in order there will no items that discriminate in wrong way and could distinguish the better students from upper to lower students.
- For students, they are shall able to recognize which the test is good or not.So. They can do the test well and can measure their progress in the class.
- For further researcher, there are so many of tests that need to be analyzed and from doing analysis, you will know how to construct a good MCQ test.
- 7. By doing this research, the writer hopes there will be increasing from the quality of MCQ test aspect, so MCQ does not categorized as a weak instrument of test. For the teacher, the writer hopes they can improve MCQ test. For example, they can make an interesting MCQ test or change the model of MCQ test, so it is more variation and more effective. As a result, MCQ test can be an instrument to evaluation process and the result can be accountable.

BIBLIOGRAPHY

- Arikunto, Suharsimi. 2003, Dasar-Dasar Evaluasi Pendidikan, Jakarta: Bumi Aksara
- Bachman, Lyle F and Adrian S.Palmer. 2009, Language Testing In Practice, New York: Oxford University Press
- Heaton, J.B. 1989. Writing English Language Testing, United States Of America: Longman Inc
- http://dankfsugiana.wordpress.com/2008/07/08/populasi-dan-teknik-sampling/, accessed on 25 Desember 2009
- http://denijusman.files.wordpress.com/2008/10/lik-juni-2007.pdf accessed on 25 Desember 2009
- Hughes, Arthur. 2003. Testing For Language Teacher, Cambridge: Cambridge University Press
- Islamiyah, Millatul. 2009. Content Validity and Item Analysis Of Semester II English Final Test For Tenth Grade Students Of SMAN 3 Sidoarjo, Surabaya: Institut Agama Islam Negeri Surabaya Library. Unpublished S1 Thesis
- Narbuko, Cholid dan Abu Achmadi, 1997, Metode Penelitian, Jakarta: Bumi Aksara
- Nazir, M. 2003, Metode Penelitian, Jakarta: Ghalia Indonesia
- Patton, Michael. 2006, Metode Evaluasi Kualitatif, Semarang: Pustaka Pelajar

- Pratiktasari, Ria Dewi. 2006. An Analysis Of Semester II English Summmative Test For The Eighth Year Students Of SMPN 1 Slahung Ponorogo. Surabaya: Universitas Negeri Surabaya Library. Unpublished S1 Thesis
- Saputro, Sigit Eko. 2006. An Analysis Of The Teacher Made English Test Items In UAS Semester I 2005 / 2006 Of The First Year Students Of SMA 2 Magetan. Surabaya: Universitas Negeri Surabaya Library. Unpublished S1 Thesis
- Suryabrata, Sumadi. 1997, Pengembangan Tes Hasil Belajar, Jakarta: PT Raja Grasindo Persada
- Ujiwani, Aan Kulistyono. 2006. An Analysis Of The Teacher Made English Test Items In UAS Semester 2 2005 / 2006 for the first year students of SMAN 1 Puri Mojokerto. Surabaya: Universitas Negeri Surabaya Library. Unpublished S1 Thesis
- Widiyanistyati, Erlik. 2008. An Analysis of the English Final Test for the First Semester of the Twelfth Year of SMA NEGERI 13 SURABAYA. Surabaya: Universitas Negeri Surabaya Library. Unpublished S1 Thesis.
- Wijaya, Cece. 1994, Kemampuan Dasar Guru Dalam Proses Belajar Mengajar, Bandung: PT Remaja Pusdakarya