AN ANALYSIS OF ESSAY TEST ON ENGLISH FINAL TEST FOR GRADE ELEVEN STUDENTS OF SMAN I, LAMONGAN

THESIS

"Presented as Partial Fulfillment of the requirement for Attainment the Degree of English Education"



By: ITA FARADILLAH D05208116

ENGLISH EDUCATION DEPARTMENT
FACULTY OF TARBIYAH
STATE INSTITUTE OF ISLAMIC STUDIES
SUNAN AMPEL
SURABAYA
2012

DECLARATION

Yang bertanda tangan di bawah ini:

Nama : Ita Faradillah

NIM : D05208116

Jurusan : Pendidikan Bahasa Inggris

Fakultas : Tarbiyah

Judul Skripsi : AN ANALYSIS OF ESSAY TEST ON ENGLISH FINAL TEST FOR

GRADE ELEVEN STUDENTS OF SMAN I, LAMONGAN.

Menyatakan bahwa karya ilmiah ini adalah hasil karya penulis sendiri dan sepanjang

pengetahuan penulis tidak berisi materi yang ditulis orang lain sebagai persyaratan

penyelesaian studi di perguruan tinggi ini maupun di perguruan tinggi lain, kecuali bagian

tertentu yang penulis gunakan sebagai acuan dengan mengikuti tata cara dan etika penulisan

karya ilmiah yang lazim.

Demikian pernyataan ini penulis dibuat dengan sebenar-benarnya. Apabila ternyata terbukti

bahwa pernyataan penulis ini tidak benar, maka sepenuhnya menjadi tanggung jawab penulis.

Surabaya, September 3th, 2012

Penulis,

Ita Faradillah

APPROVAL SHEET

This thesis by ITA FARADILLAH entitled "AN ANALYSIS OF ESSAY TEST ON ENGLISH FINAL TEST FOR GRADE ELEVEN STUDENTS OF SMAN I, LAMONGAN" has been approved by the thesis advisors for further approved by the Board of Examiners.

Surabaya, September 14th 2012

Advisor I

Masdar Hilmy, MA. Ph.D NIP 197103021996031002

Advisor II

Fitriah, MANIP 197610042009122001

EXAMINERS APPROVAL SHEET

Thesis Entitled:

"AN ANALYSIS OF ESSAY TEST ON ENGLISH FINAL TEST FOR GRADE ELEVEN STUDENTS OF SMAN I, LAMONGAN" has been accepted and approved by the broad examiners of English Department of Tarbiyah Faculty State Institute for Islamic Studies Sunan Ampel Surabaya

Dr. H. Nur Hamim, M.Ag

The Board Examiners Advisor/Chair Person,

Masdar Hilmy, MA. Ph.D NIP. 197103021996031002

Secretary,

Sigit Pramono Vati, M.Pd NIP. 198605182011011005

Examiner I,

Mokhammad Syaifuddin, M.Ed

NIP. 197310131997031002

Exammer II

M! Hanaff, MA

ABSTRACT

An Analysis of Essay Test on English Final Test for Grade Eleven Students of SMAN I, Lamongan.

Name : Ita Faradillah NIM : D05208116

Advisor : Masdar Hilmy, MA.Ph.D

Key Words: Analysis of Essay Test, Essay Test, Content Validity, Item Analysis, Index Difficulty, and Index Discrimination.

Testing is a tool to evaluate students' achievement and to determine the progress of objective of a course. All teaching process should be followed by evaluation indeed. Without testing, the result of teaching learning process will be fail. That is why testing is so beneficial to be applied during teaching learning activities. This case shows that teaching and testing are closely integrated that is one of them is almost impossible to work its self in achieving course objectives without working together. As a constructor, teachers should conduct a good test, so the test will be valid and reliable. Moreover, the test should reflect the criteria of good test such as content validity, reliability, index difficulty and index discrimination.

This study aims to know what is Essay Test on the English Final Test for Grade Eleven of SMAN I, Lamongan like in the term of content validity, index difficulty, and index discrimination. The process of design this study is through documentation and it used descriptive research to describe the quality of English Final Test for Grade Eleven Students of SMAN I, Lamongan. It only focuses on the quality of Essay Test. this study is also uses numerical calculation to compute the data. While the samples are XI IPA-5 and XI IPA-6, they are taken by random sampling.

The result of this study describes that essay test used by Grade Eleven Students of SMAN I, Lamongan has good validity of content. While index of difficulty and index of discrimination describe different result for those two classes. Index of difficulty for those XI IPA-5 is acceptable, but for XI IPA-6 the test is recognized too difficult. Besides, index of discrimination for XI IPA-5 class is satisfactory since the result is distant from zero point. The malfunction result shows for XI IPA-6 class, it means that the test for this class should be revised to distinguish between the more able students to the less able.

Finally, the writer expects that this study is significant for the teacher, it is as guide to create good English essay test based on good characteristics of essay test. And for further researcher, this study as references for other researchers which have the same concern on analyzing essay test.

TABLE CONTENT

COVER	
APPROVAL SHEET	
EXAMINERS' APPROVAL SHEET	iii
DECLARATION	iv
MOTTO	v
DEDICATION	vi
ABSTRACT	vii
ACKNOWLEDGMENT	viii
	ix
CHAPTER I: INTRODUCTION	
A. Background of Study	1
	6
	 7
	<mark>7</mark>
	7
	8
CHAPTER II: REVIEW AND RELATED	LITERATURE
A. Teaching and Testing	11
	13
	14
• •	ıge15
	15
b. Achievement Test	15
	17
<u> </u>	18
	t Makers19
	19
b. Standardize Test	20
3. Types of Test Based on Form	22
a. Objective Test	22
<u> </u>	25
<u> </u>	27
-	29
	30
<u> </u>	31
	32
	32

a. Content Validity	33
b. Criterion-Related Validity	
1) Concurrent Validity	35
2) Predictive Validity	36
c. Construct Validity	37
d. Face Validity	37
2. Reliability	38
3. Practicality	39
G. Item Analysis	39
a. Index of Difficulty	
b. Index of Discrimination	42
H. Previous Study	43
CHAPTER III: RESEARCH METHODOLOGY	
A. Research Design	46
B. Setting of Study	
C. Data and Source of Data	
D. Data Collection Techniques	
E. Research Instrument	
F. Data Analysis Techniques	
1. Analyzing Content Validity	
2. Analyzing Index Difficulty	
3. Analyzing Index Discrimination	54
CHAPTER IV: DATA ANALYSIS	
A. Analyzing Content Validity	56
B. Analyzing Index of Difficulty	58
1. Analyzing Index of Difficulty on Essay Test for XI IPA-5 Class	
2. Analyzing Index of Difficulty on Essay Test for XI IPA-6 Class	
C. Analyzing Index of Discrimination	
1. Analyzing Index of Discrimination on Essay Test for XI IPA-5 Class	
2. Analyzing Index of Discrimination on Essay Test for XI IPA-6 Class	65
CHAPTED W CONCLUDION AND SUCCESSION	
CHAPTER V: CONCLUSSION AND SUGGESTION	
A. Conclusion	
B. Suggestion	/(

BIBLIOGRAPHY

APPENDICES

Appendix 1 (The Result of Analyzing Content Validity)

Appendix 2 (The Score of Upper and Lower Group in English Essay Test for XI IPA-5)

Appendix 3 (The Score of Students' English Final Test for XI IPA-5 Class)

Appendix 4 (The Score of Upper and Lower Group in English Essay Test for XI IPA-6)

Appendix 5 (The Score of Students' English Final Test for XI IPA-6 Class)

Appendix 6 (The Result of Analyzing Index Difficulty in XI IPA-5)

Appendix 7 (The Result of Analyzing Index Difficulty in XI IPA-6)

Appendix 8 (The Result of Analyzing Index Discrimination in XI IPA-5)

Appendix 9 (The Result of Analyzing Index Discrimination in XI IPA-6)

CHAPTER I

INTRODUCTION

This chapter presents the background of study, statements of the problem, purpose of study, significance of study, scope and limitation, and definition of key terms. Each section is presented as follows:

A. Background Of Study

Testing is a tool for evaluating students, curriculum, and teaching methods. According to Arikunto, test is a series of question measuring skill, knowledge, intelligences, and capacities of individual or group. 1 It measures how far the purpose of teaching has been achieved. Without testing, the result of teaching learning process will be fail.

Hughes said that all teaching process should be followed by evaluation. In this case, sometimes it is common that evaluation is considered has same meaning as testing, evaluation is the process of examining and rating the subject based on its important features. For educational purposes, it used to measure whether the students success or not during teaching learning activity for a particular period of time. While testing is only a component in the evaluation process. There are several types of evaluation, one of them is a test.

¹ Suharsimi Arikunto, *Dasar-Dasar EvaluasiPendidikan*, (Jakarta: Bumi Aksara, 1993), 29

As the candidates of English teacher, the term testing is not as strange as we have never known before. Testing in general is finding out how well something works and also tells what level of knowledge or skill has been acquired. It plays an important role in giving feedback for education in various aspects.

The existence of testing is important to be applied in teaching activities. Heaton stated that testing and teaching cannot be separated. He also added that test might be established as devices to reinforce learning and to motivate students as a method of assessing their performance in language.² In line of that, Hughes said that teaching and testing are proper relationship, they have an interrelationship in the backwash. Backwash is the effect of testing on teaching and learning. It can be harmful and beneficial.³ For example in testing writing skill, when the skill of writing is evaluated only by multiple choice, then students will be possible only focused on practicing such item rather than practicing the skill of writing itself.

Considering the statement above, testing is crucial in teaching English although sometime a test is suitable for one teacher have different result for another. Therefore, the teacher should consider some functions of testing. These are some functions of testing in language abilities:⁴

⁴ Ibid. 8

² J. B Heaton, Writing English Language Tests, (New York: Longman Group, 1988), 5

³ Athur Hughes, *Testing for Language Teachers*, (Cambridge: University Press, 2003),

- 1. To measure the language proficiency.
- To discover how successful students achieve the objectives of a course of study.
- 3. To diagnose students' strengths and weaknesses, to identify what they know and what they don't know.
- 4. To place students in certain level based on their ability.

According to the explanation above, one of the functions of testing is to measure language proficiency. Language proficiency is a measurement of how well an individual masters a language. There are many kinds of test in evaluating language proficiency. One of them is essay test. According to John M. Stalnaker as quoted on Christian M. reiner, an essay test is a test which requires response composed by the examinee, usually in the form of one or more sentences, the character is no single response or pattern of responses could be said as perfect response, the accuracy and the quality are judged subjectively.⁵

Christian also said that educators choose an essay as a form of assessment because essay test challenges students to create a response rather than to simply select a response.⁶ Furthermore, essay test is potentially to reveal students' abilities to give a reason, create ideas, analyze information and synthesize given discourses in the text. Other kinds of test such as multiple choices question,

⁶ Ibid, 5

⁵ Christian M. Reiner, *Preparing Effective Essay Question*, (New Forum Press, 2002), 6

matching exercise and true false items are examples all of selected response test items because they require students to choose an answer from a list of possibilities, whereas essay items require students to compose their own answer.

Many teachers consider essay test as an ideal form of testing since essay requires more effort from the students than other types of test. Students neither can answer an essay test correctly by simply recognizing the correct answer, nor can they study from the essay exam by memorizing factual memorial. Regarding the crucial of essay test Stalnaker defined some criteria for effective essay test:⁷

- 1. Essay test requires examinee to compose rather than select their response.
- 2. Essay test elicits students' responses that must consist of more than one sentence.
- 3. Essay test allows different responses from students.
- 4. Essay test requires subjective judgment by the teacher to judge the accuracy and quality of responses.

There are two major purposes for using essay test. *Firstly* essay assesses students understanding of and ability to think the subject matter content. *Secondly* essay assesses students writing abilities.⁸ Essay test is just another name of subjective test, it requires students to produce accurate information and may be as simple as listing facts about a particular topic. While the objective test as J.

⁷ Ibid, 6.

⁸ Ibid, 4.

B.Heaton said it can be made as easy or as difficult as the test constructor wishes, it is frequently criticized simpler to answer than subjective test.

However, there are a lot of things need to be considered in creating essay test such as having a good validity of content, analyzing index difficulty and index discrimination. In creating a test, it is also crucial when it tests English language, test writers do not only write a test just it is but also try to conduct an applicable test.

There are several efforts of testing techniques that need to be defined before writing an acceptable test. A test can be said a good test and satisfactory if it contains the criteria of validity, reliability and the items of test are not too difficult and not too easy. As Hughes said that language ability is not easy to measure. There are some ideas to create good essay, a teacher should recognize some characteristics of good test, but it is not enough to measure whether an essay items are good or not, or when the test should be constructed. Another term that should be considered is the way of judging whether it is appropriate or not, it will be analyzed through index difficulty and index discrimination.

The writer chooses senior high school students since they are used to express their critical thinking or ideas in responding on essay test. Moreover the students are at grade eleven. The writer also thinks that SMAN I, Lamongan has international standard curriculum (RSBI), so the test of this school is made by the

teacher itself and another thing to support this research shows from the school, SMAN I Lamongan becomes one of favorite schools in that town because several elements such as the teacher, students and also good facilities.

Most of the teachers in this school are undergraduate degree, even some of them have post graduate degree which their background of study are suitable with the lesson they teach. Another thing to support the element is academic achievement in English such as the students of SMAN I, Lamongan often win English competition and achievement competition.

This study tries to analyze an essay test which focuses on the content validity, index of difficulty and index of discrimination. The writer will analyze the English Final Test for Grade Eleven of SMAN I, Lamongan.

B. Research Question

- What is Content Validity of English essay tests used by Grade Eleven Students in SMAN I, Lamongan?
- 2. What is Index Difficulty of English essay test used by Grade Eleven Students in SMAN I, Lamongan?
- 3. What is Index Discrimination of English essay test used by Grade Eleven Students in SMAN I, Lamongan?

C. Objective of Study

This study aims to analyze a form of English final test which focuses on essay test, the study tries to comply the following objectives:

- Explain the appropriateness of content validity with English essay test used by Grade Eleven Students of SMAN I, Lamongan.
- Explain the Index Difficulty of English essay test used by Grade Eleven Students of SMAN I, Lamongan.
- 3. Explain the Index Discrimination of English essay test used by Grade Eleven Students of SMAN I, Lamongan.

D. Significance Of The Study

The writer expects this study gives significant to:

- 1. The teacher: it is used as guide for teacher to create good English essay test based on good characteristics of essay test.
- 2. The further researcher: it is used as references for other researchers which have the same concern on analyzing essay test.

E. Scope and Limits Of Study

The scope and limits of this study are:

This study is implemented to the grade eleven students of SMAN I,
 Lamongan. Furthermore, it focuses on essay test.

 This study focuses on the quality of essay test for grade eleven students of SMAN I, Lamongan based on the standard of test include standard content, content validity, index difficulty and index discrimination.

F. Definition Of Key Term

There are some terms used to give clear explanation related to this study.

Those terms are defined to give readers better understanding and avoid misunderstanding. They are:

1. Essay Test

Essay test is a test requires students to create a critical thinking by their own words. It may be in the form of one or more sentences, no single response or pattern of responses can be listed as correct answer. In this study, the essay test is a kind of test used for Grade Eleven of SMAN I, Lamongan. It is for English final test on the year 2011-2012.

2. An Analysis of Essay Test

An analysis of essay test means an analysis that will be done by the researcher to know the quality of English essay test which has been constructed by English teacher for grade eleven students in SMAN I, Lamongan based on the content validity, characteristic of good test and item analysis (index difficulty and index discrimination).

⁹ Christian M. Reiner, Writing Effective Essay Question, (New Forum Press, 2002), 6

3. Validity

Validity of a test is the extent to which it measures what is supposed to measure. If the test is supposed to be a test for grade eleven students, the items of the test should be appropriate with their level.¹⁰

4. Content Validity

It refers to the test that represents all materials in the course that are being tested.¹¹ In this study, the writer focuses on the content validity of essay test and it will be analyzed whether it fulfills with the *standard content* of the 2006 English curriculum for grade eleven of senior high school or not.

5. Reliability

Reliability is the consistency of scoring by two or more scorers. It means that if the test scoring is done by two or more raters, the reliability of their evaluations can be easily checked by comparing the scores for the same student responses on different occasions and then it shows same result. Then the test is reliable.¹²

6. Item Analysis

Item analysis indicates whether the test item is acceptable, less acceptable or poor. According to Arikunto that its purpose to get information about the weakness of a test and become references to revise the other next

¹⁰ J. B Heaton, Writing English Language Tests, (New York; Longman Group, 1988), 172

¹¹ Ibid, 160

¹² H. Douglas Brown, Teaching by Principles: An Interactive Approach To Language Pedagogy, (San Francisco State University: Longman, Inc, 2001), 387

tests.¹³ In addition, for essay test this term involves a variety methods summarizing and analyzing the responses of students to test items.

7. Index of Difficulty

Heaton stated that index difficulty shows how easy or difficult the particular item proved in the test. ¹⁴ In this study, index of difficulty of a test defined as how hard the item of test is. It also distinguishes from the students' score. For example, if a maximum score of each question on essay test is nine point and all students obtain this score, then it could say that the test item is easy one. If all students obtain zero score, then the test item is difficult.

8. Index of Discrimination

Heaton stated that item discrimination indicates the extent to which the item discriminates between the testees, separating the levels of them.¹⁵ Discrimination of essay can be said as the degree of students' level in achieving differently on an item or question of the test.

¹³ SuharsimiArikunto, *Dasar-Dasar EvaluasiPendidikan*, (Jakarta: Bumi Aksara, 1993), 211

¹⁵ Ibid. 179

¹⁴ J. B Heaton, Writing English Language Tests, (New York; Longman Group, 1988), 178

CHAPTER II

REVIEW RELATED LITERATURE

In this chapter, the writer will review several theories of literature related to discussion of this study. It presents some important terms to comprehend detail information concerned to this discussion. This subtitle will define the discussion of teaching and testing, definition of test, types of test, definition of essay test, types of essay test, characteristic of good test, item analysis, and previous study.

A. Teaching And Testing

Testing is an important part of every teaching and learning experience. Both testing and teaching are closely integrated that one of them is almost impossible to work its self in achieving course objective without working together.¹

In teaching process, testing is required to know how far some purposes of teaching have been stated. As it has been explained in the previous chapter, there may be situation where teaching process is potentially good but the testing does not have same result. Therefore, as the measurement device, a test must be good designed to comply with the criteria of good test. This situation led Davies in 1986 to confine testing as the role of servant to the teaching.

¹ J. B Heaton, Writing English Language Tests (New York; Longman Group, 1988), 5.

To recognize the importance of testing for students in learning English, there are two major purposes.² *Firstly* test can help to create positive attitudes toward the class. It means that teachers aim to provide positive classroom experience by giving students a sense of accomplishment and a feeling that teachers' evaluation matches to what they have been taught.

Secondly test can benefit students to help them master the language. When several tests are given, learning can also be enhanced by students' awareness of course objectives. That is why, good English test also can help students to learn the language by requiring them to study hard, emphasizing course objectives and showing them which parts they need to be improved.

Language tests have many uses in educational programs, and quite often the same test will be used for two or more related purposes. In accordance with purposes of test in the paragraph above, Arikunto said that testing has several purposes in measuring student ability such as: a) testing is able to select good student. b) testing is able to diagnostic the weakness and strength of students. c) testing is able to take a place as appropriate level of students. d) testing is able to measure the effectiveness of achievement of the course.³

The test should enable the teacher to ascertain which parts of the language programme have been found difficult by the class. In this way, the teacher can evaluate the effectiveness of the syllabus as well as the methods and the materials

² Harold S. Madsen, *Techniques in Testing* (New York: Oxford University Press, 1983), 3.

³ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 1993), 9.

he or she uses. The test result may indicate, for example certain areas of the language syllabus which does not have sufficient account of foreign learner difficulties, in this case the teacher will be concerned with that problem areas encountered by group of students rather than by an individual student.

Meanwhile, Hughes said that one who has an authority to make good test is a teacher, teaching profession can make three contributions to the improvement of testing: teachers can write better tests themselves, teachers can enlighten other people who are involved in testing process, and teachers can put pressure on professional testers and examining boards to improve their tests.⁴

B. Definition of Test

Generally, test is an evaluation to measure person's ability in a certain purpose. It is a set of techniques, procedures, and items that constitute an instrument of some sort that requires performance or activity on the part of the test taker. Furthermore, test is as motivator for students that stimulates them to study and provide them with feedback about their achievement.

Another definition of test, according to Kubizyn and Borich stated in their book, test is just a tool that can contribute importantly to the process of measuring students, the curriculum, and the teaching method.⁵ While as Quoted on

⁴ Athur Hughes, *Testing for Language Teachers* (Cambridge: University Press, 2003), 5.

⁵ Tom Kubiszyn and Gary Borich, *Educational Testing and Measurement* (Singapore, John Wiley & Sons, INC, 2003), 1.

Encyclopedia of Educational Evaluation, a test is comprehensive assessment of an individual or to an entire program evaluation effort.

There are several terms used to measure students' ability, those are totally common in teachers' mind; *Evaluation, Measurement, and Testing*. Evaluation is a process wherein the parts, processes, or outcomes of program are examined to see whether they are satisfactory, particularly with reference to the program's stated on the objectives of a course. While the assessment of program's outcomes or results are facilitated by measurement. Essentially, tests may be used constructively in the process of evaluation and as a useful device in a number of processes such as evaluation, diagnosis, or monitoring.⁶

Those are several definitions about test stated by some experts. Although those definitions are created in different sentences, but it shows the same meaning that test is a tool to evaluate students, curriculum, and the successful of teaching method.

C. Types of Test

Although the measurement of English language has several types but they have same purpose to measure person's competence either in performance or written form.

⁶ Bruce W. Tuckman, *Measuring Educational Outcomes, Fundamentals of Testing* (Harcourt Brace Jovanovich, Inc, 1975), 12.

1. Types of test based on the usage

Based on the usage test is divided into five types:

a. Proficiency Test

Proficiency test is examining student's language proficiency with reference to a particular task which will be required to perform. Regarding the term of proficiency, it means that having sufficient command of the language for a particular purpose. The content of a proficiency test is not based on the content or objectives of language courses that people taking but rather it is based on a specification of what candidates have to be able to do in the language in order to be considered proficient.

Besides that, proficiency test is concerned simply with measuring a student's control of the language in the light of what he or she will be expected to do with it in the future performance of a particular task. This test is commonly consisted of standardized multiple-choice items on grammar, vocabulary, reading comprehension and sometime a sample of writing. A typical example of proficiency test is Test of English as Foreign Language (TOEFL).

b. Achievement Test

Different with previous test, achievement test is directly related to language course which is intended to measure achievement on a large

⁷H. Douglas Brown, Teaching by Principles : An Interactive Approach To Language Pedagogy (San Francisco State University: Longman, Inc, 2001), 11.

⁸ J. B Heaton, Writing English Language Tests, (New York; Longman Group, 1988), 173.

scale. The purpose of this test is to establish the successful of individual or group of students in achieving the objective of learning in period of time.

In line of that, Heaton said that this test is designed to measure student's ability based on what they have leant. This kind of test is divided into two: final achievement test and progress achievement test.⁹

First, final achievement test as the name of this test, it is conducted on the end of course study. It may be written and administrated by ministries of education, official examining boards, or by members of teaching institution. Briefly, the content of test should be based on a detailed course syllabus or on the books and other materials used. In fact, since the test only contains what is taught that students have actually learned in the class then at least the test can be considered as fair test. Whereas, referring to that statement Hughes added that the disadvantage is that if the syllabus is badly designed or the books and materials are badly chosen, the result of test can be foolish.

Second, Progress achievement test is intended to measure the learners' language and skill progress in relation to the syllabus they have been following. It relates to the final achievement test because measuring 'progress' would be repeatedly to administer final achievement test. The aim of progress test is to stimulate learning and to reinforce what

⁹ Ibid. 171.

¹⁰ Athur Hughes, *Testing for Language Teachers* (Cambridge: University Press, 2003), 13.

¹¹ Ibid 13

have been taught. Therefore, a good achievement test should reflect the particular approach to learning and teaching that has previously been learnt.

c. Placement Test

Placement test is placing students into an appropriate level or section of language curriculum or school. 12 This test is administered to assign students to the classes at different levels. It typically includes a sampling of material to be covered in the curriculum and provides an indication thus the students will find a level or class to be neither too easy nor too difficult, but appropriately challenging.

Placement test is recommended only when the school or course is sure that the test being considered suits its particular teaching programme. This test aims to gather students which have at least similar level of ability in one classroom because if they are placed in the wrong class or do not suit their capability, it might be difficult to gain the effectiveness of teaching learning process.

d. Diagnostic Test

While placement test is designed to show how good a learners' English to the appropriate of their ability, diagnostic test is designed to assess learners' difficulties, gaps in their knowledge and skill deficiencies

¹² H. Douglas Brown, *Teaching by Principles: An Interactive Approach To Language Pedagogy* (San Francisco State University: Longman, Inc, 2001), 390.

during a course.¹³ Generally, it identifies learners' strengths and weaknesses. Hughes said that this test is primarily used to determine what learning still need to take place. Thus when the teachers know what the problems are, they can do something about them.¹⁴

The difficult part considered to be diagnosed such as test should give an appropriate remedial test which can be taken later. For instance, a teacher gives a test of writing skill to create a spoof based on students' experience. The result of their writing shows the part of students' difficulty on differentiating between regular and irregular verb to create past sentence. Because one of the characteristic of spoof is past tense so they must understand past sentence in deep. It thereby needs to have special focus.

e. Aptitude Test

Brown said this test can predict a person's success; it is designed to measure person's capacity or general ability to learn a foreign language and to be successful in that undertaking. ¹⁵ Language learning aptitude is complex matter, consisting of such factors as intelligence, age, motivation,

 13 Jeremy Harmer, *The Practice Of English Language Teaching* (New York: Longman, 2011), 321.

Athur Hughes, *Testing for Language Teachers* (Cambridge: University Press, 2003), 15.

¹⁵ H. Douglas Brown, *Teaching by Principles: An Interactive Approach To Language Pedagogy* (San Francisco State University: Longman, Inc, 2001), 389.

memory, phonological sensitivity and sensitivity to grammatical patterning.¹⁶

Generally, an aptitude test indicates individual's facility for acquiring specific skills and learning. This test seeks to predict the students' probable strengths and weaknesses in learning a foreign language by measuring performance in an artificial language. In other word, artificial language is the ability to learn new phonemic distinction and also to use language patterns in an unfamiliar but systematic manner.

2. Types of test based on the test makers

All types of test are same function, which are to measure the success of students' learning. The following explanation will discuss about the types of test based on the test maker. It is divided into two types:

a. Teacher-Made Test

As its name, teacher made test is a test made by the teacher of that classroom itself. The purpose of this test is to measure the degree of students' success in achieving instructional objectives of particular lesson or course are taught in the classroom.

Generally, the structure of test items are completely depends on the teacher who teaches the lesson itself because the teacher has an authority to formulate the indicators, materials, teaching learning process and

¹⁶ J. B Heaton, Writing English Language Tests (New York; Longman Group, 1988), 173.

assessing students' outcomes.¹⁷ In this case, the teacher knows what have been taught in his or her own classroom.

That statement above means that it is possible when a school has some teachers teaching the same subject. For instance, the teacher in class A is good at delivering the materials for students, but another teacher in different class is not. Thus, teacher made a test can only be applied in his or her own classroom and not for another class.

Nurgiyantoro stated that this test is only focuses on the score. The score of students can be comprised only with their classmates. He added that this test is primarily aims to know (i) the standard of learning objectives, (ii) the degree of students' comprehending to the materials, and (iii) to evaluate students as their report of learning during in the school.¹⁸

b. Standardize Test

In opposite to teacher made test, standardize test is a test that has been standardized. According to Nurgiyantoro, the term standard could be interpreted as the degree of particular achievement in which students must have in particular programs. The structure of standardize test is as same as the teacher made test which begun to create the materials and then arranging the test items. As stated on Nurgiyantoro's book, quoted by Tuckman,

¹⁸ Ibid. 72.

¹⁷ Burhan Nurgiyantoro, *Penilaian dalam Pengajaran Bahasa dan Sastra*, (Yogyakarta; BPFE-Yogyakarta, 2001), 71.

"The process of writing the standardize test is usually arranged by the team, selected materials, and based on the curriculum or national course books"¹⁹

Ebel said that this test refers to a test that has been expertly constructed, usually with try out, analysis, and revision; includes explicit instructions for uniform (standard) administration or scoring.²⁰ This statement means, this kind of test is unlike teacher made test because standardize test needs to be tried out before being tested. Its aim does not to measure the students' ability, but rather to know the appropriateness of the test items. The result of tried out will be analyzed through the item analyses to attain the coefficient of index difficulty and also index discrimination. The test item that is too difficult or too easy should be revised, whether it seems from the quality of index discrimination or index difficulty.

Furthermore, this test is called as standard because it is administrated and scored according to specific and uniform (standard) procedures. It includes an assistance of curriculum experts, teachers and school administrators.²¹ Thus it can be used in all different school and even it used many times. In other word, when standard test is

¹⁹ Ibid. 73.

²⁰ Robert L. Ebel and David A. Frisbie, Essential of Educational and Measurement (U.S.A, Prentice-Hall, INC, 1991), 288.

²¹ Tom Kubiszyn and Gary Borich, Educational Testing and Measurement (Singapore, John Wiley & Sons, INC, 2003), 343.

administrated and scored in Surabaya, it would exactly administrated and scored with the same manner in Lamongan, Malang, and etc.

3. Types of test based on the form

Considering that types of test are wide discussion, in this topic the writer defines some forms of test. There are two types of test based on the form: objective and subjective test. Those two types of test form can be used to measure almost any educational achievement. Heaton said those subjective and objective tests are terms used to refer to the scoring of tests. The following explanation will clarify about them:

a. Objective Test

An objective test can be said as short answer test. Derived from its name, it considers students to give the short answer, even though by selecting any particular code as alternative way which provided to make them easy to choose the answer.²² However items in objective test can be made either as easy or as difficult as the constructor wishes.

In fact that objective test looks easier because it does not need much time for students to think critically and encourage them for guessing. The result and answer of objective test are taken objectively; regarding that

²² Burhan Nurgiyantoro, *Penilaian dalam Pengajaran Bahasa dan Sastra*, (Yogyakarta; BPFE-Yogyakarta, 2001), 75

statement Arikunto added that its purpose to reduce some weaknesses as in essay test.

Objective test can be pre-tested before being administrated on a wider basis, this procedure enables the constructor to calculate the approximate degree of difficulty of the test. As a tool to measure the result of students' learning, an objective test has some strengths and weaknesses. The following explanations are about some strengths of objective test:

Strengths of Objective Test:

- a) It enables teacher to take the tested material better than an essay test.
- b) It enables to have one correct answer
- c) It can be easy in correcting because it only matches the answer with provided answer key.
- d) The result of objective test is believable because it is evaluated objectively in such item.²³

Furthermore, the explanation above could not be spared away from the existence of some disadvantage of objective test, the following explanation will clarify enough:

²³ Ibid, 76

Weaknesses of Objective Test:

- a) The arrangement of objective test takes longer time because it requires careful preparation, accuracy and ability from the teacher.
- b) There was a trend or willingness of teacher to emphasize students' attention for particular materials.
- c) It enables students do the test to guess.
- d) It sometimes might be long and costs much.²⁴

As indicated that objective test either tests the learner's ability to communicate the target language or evaluates an actual performance, whereby according to Ebel, there are four kinds of this test a) true-false items, b) multiple-choice items, c) short-answer items, and d) matching items.²⁵

In short, multiple-choice item is a test that most common used to test students. Multiple-choice as stated by Nurgiyanto is a test consists of incomplete sentence which required the testee to complete it with one of those provided answers. In case, to design multiple choice test, the test maker should know the primarily several terms used multiple choice. First is *a stem*, refers to an initial part of each multiple choice items. Second is *option/responses/alternatives* which refer to some choices from which the

²⁴ Ibid, 77

²⁵ Robert L. Ebel and David A. Frisbie, *Essential of Educational and Measurement* (U.S.A, Prentice-Hall, INC, 1991), 179.

students are able to select their answer. One option is the correct answer, while the other options are distractors. ²⁶

The illustration of the explanation above as follow:

Most cells in multi organism perform function.

A good multiple-choice item is the most time consuming kind of objective test items to write.

b. Subjective Test

Since the objective tests have only one correct answer, it differs with subjective test. It enables students to show their ability in exploring knowledge through analyzing and evaluating a new given information. Since subjective test gives a freedom for students to arrange and deliver their answer, that is why this test is also conducted as an essay tests.

In this kind of tests, student must think what to say and then express their ideas as well as possible.²⁷ The answer of students shows their

²⁶ J. B Heaton, Writing English Language Tests (New York; Longman Group, 1988), 28.

quality of thought, the cognitive activity which does not only memorize and understand, but it's more about the way how they give reason and summary toward the test. In contrast with objective test, essay test are relatively easy to prepare, the difficult part of the job is usually grading students' answer. Furthermore, Sudijono mentions several characteristics of this test.

First, this test is as organized a question or imperative which demands students' responses descriptively. In other word, the response of this test is in long sentence.

Second, the imperative sentence in subjective test demands students to give explanation, comment, analysis, and distinction.

Third, the number of test items is generally definite which consists among five to ten items.

Finally, the items of this test are commonly begun with the words "explain...", "describe...", "how...", "why...". 28

As the name suggest, this test is commonly said essay test. the following subchapter is clear explanation about definition, strengths, and weaknesses of essay test.

²⁸ Anas Sudijono, *Penganter Evaluasi Pendidikan*, (Jakarta: PT. Raja Grafindo Persada, 1996), 100.

²⁷ Ibid. 25.

D. Definition of Essay Test

Essay test is another name of subjective test. It is also as a common methods for assessing students' learning, as written in the previous chapter that educators choose an essay question over other forms of assessment because it challenges students to create a response rather than to simply select a response.²⁹

Furthermore, this test provides students opportunity to structure and compose their own responses either in definite or extended limits. Essay test does not ask students to choose one of the responses as like objective test but essay test asks students to share their ideas by their own words. And the writer of essay test demonstrates that none of the wrong answer as a good as correct answer. As Norman stated essay test in his book,

"Some aspects of complex achievement are difficult to measure objectively Learning outcomes that indicate pupils to originate ideas . . . to organize and express ideas in a global attack on a problem which require the greater freedom of response provided by the essay test" 30

Based on the statement above, it explains that the focus of this test is to measure a variety of learning outcomes, from simple to complex and the interpretive exercise is especially useful for measuring complex achievement. The distinctive feature of essay is the freedom of response, it provides examinees the opportunity to structure and compose their own responses in the broad limits. Although this freedom enhances the value of essay test as a measurement of complex achievement, but it is difficult to score.

²⁹ Christian M. Reiner, Writing Effective Essay Question (New Forum Press, 2002), 4.

³⁰ Norman E Gronlund, *Measurement and Evaluation in Teaching* (New York: 1985), 213.

Essay test is essential part to measure students' cognitive skill because it conducts students' thinking and conscious mental process. Age is the influence for human cognition, it develops rapidly throughout the first sixteen years of life and less rapidly thereafter.³¹ This study concerns on grade eleven so it is also possible that average of students in this class are sixteen to seventeen years old.

The cognitive domain is the distinction between rote and meaningful learning. As cited by Ausubel on Brown's book that people all of ages have little need for rote that does not relate to existing knowledge and experience. Rather, most items in the essay are acquired by meaningful learning. Essay test may be well constructed or poorly constructed; the well constructed of essay test aims to test complex cognitive skill by requiring students to synthesize knowledge, solve problems and to be innovative in problem solving. While the poorly constructed essay test may require students to do no more than recall information as it was presented in the textbook. 33

In this discussion, the following explanation will define some strengths and weaknesses of essay test:³⁴

61.

³¹ H. Douglas Brown, *Principles of Language Learning and Teaching* (Longman, inc, 2000),

³² Ibid. 61

³³ Tom Kubiszyn and Gary Borich, *Educational Testing and Measurement* (Singapore, John Wiley & Sons, INC, 2003), 127.

³⁴ Burhan Nurgiyantoro, *Penilaian dalam Pengajaran Bahasa dan Sastra* (Yogyakarta; BPFE-Yogyakarta, 2001), 72.

Strengths of Essay Test

- a) It is appropriate to assess students' thinking process that involves good cognitive skill.
- b) It forces students to explore their responses by their own word.
- c) It forces them to use their thinking and does not give them a chance to guess.
- d) Easy to create which does not waste the time.

Weaknesses of Essay Test

- a) It has poor validity and reliability.
- b) Because of limited materials are being tested, sometimes the result seems unpredictable.
- c) It is difficult for teacher to assess the standard of students' responses.
- d) It needs longer time to examine students' work; it is not efficient when the number of students is many.

By the explanation above, Heaton added that a good classroom test will usually contain both subjective and objective test items.

E. Types of Essay Test

Essay test can be lengthy responses, either it is limited or restricted to one page or less. It will not change its function which measures general or specific

outcomes of instruction. This test is divided into two kinds: indefinite essay and definite essay.³⁵

1. Extended Essay

Extended essay, as the name suggests. The answers of this test are required indefinitely. It allows students to determine the length and complexity of response.³⁶ This type of essay is most useful to evaluate the levels of cognitive taxonomy. For a teacher it is a good option to determine whether the students can organize, express and evaluate information or ideas. In this case, the testee has a freedom and there is no limit to analyze and organize an essay response. For example:

1) Our parents have educated us since we were a child, we are as their children cannot calculate how much the cost that they gave either for our education or our daily needs. That is why, we should obey their command and always pray to God for giving them a healthy and blessed livelihood.

Explain what some evidences to obey our parents are?

To respond this question, students must be able to arrange relevant information, critically analyze it and apply it in essay. The complex item like this test will take time to develop and will be even more time consuming to score.

³⁶ Ibid. 130.

101a. 130

³⁵ Tom Kubiszyn and Gary Borich, *Educational Testing and Measurement*, (Singapore, John Wiley & Sons, INC, 2003), 129.

2. Restricted Essay

Restricted essay, it poses a specific problem for which the students must recall proper information, organize it in a suitable manner, derive a defensible conclusion and express it within the limit of the posed problem, the page or time limit.³⁷ In this case, the testee shows their ideas through essay depend on the limitation from the given instruction on the test. For example:

1) As we have learned grammar, it is common in our mind that sentence refers to a wishes or supposition called conditional sentence. This conditional sentence has three kinds.

Explain the second kind of conditional sentence and then give the example!

Restricting the form and scope of the answers to essay has both advantages and disadvantages. The advantages such as the question are prepared easily, related more directly to specific learning outcomes, and easy to give score. While the disadvantages are, it provides little opportunity for students to demonstrate their abilities to develop essentially new patterns of response.³⁸

In addition, an essay test provides a better indication of students' real achievements in learning, students are not given readymade answers but must have command of wide knowledge that enables them to relate facts and

³⁷ Ibid. 130.

³⁸ Norman E Gronlund, *Measurement and Evaluation in Teaching*, (New York: 1985), 133

principles, to organize into logical progression and then to do justice to these ideas in written expression.³⁹ Nevertheless, as Ebel stated that the length of students' answer may be closely related to the score it receives. It means that longer answers tend to receive higher score.

F. Characteristic of Good Test

All good tests definitely have good qualities, in this discussion there are three qualities to judge good test: validity, reliability and practicality. In other word we could say, any kinds of test must be appropriate with terms of four objectives (validity), dependable in the evidence it provides (reliability) and applicable in the particular situation (practicality). Without any one of them, a test would be poor investment in time and money.

Whether the teachers are constructing their own test or selecting a standard instrument to use both in their class and school, they should certainly understand what these concepts mean and how to apply.

1. Validity

As Hughes said that test is said to be valid if it measures accurately what is intended to measure. Different word in the same meaning, according

13

³⁹ Robert L. Ebel, *Essentials Of Educational Measurement*, (New Delhi; Prentice Hall, 1991).

¹⁸⁹David P Harris, *TestingEnglish as a second Language*, (New York: Mc Craw-ill, Inc, 1959).

to William, validity can be the best defined as the degree to which a test is capable of achieving certain aims.⁴¹

In language test, a validity is supported most convincingly by personal observation such teachers and peers. In other word, a test has validity evidence if we can demonstrate that it measures what it says it measures. For example, if the test is supposed to give a test for eleventh-grade students in English speaking ability, then it should measure eleventh-grade skill, not tenth-grade skill and not science skill.

Hence, there are several ways to decide whether the test has sufficient validity, they are: a) content validity, b) criterion-related validity, c) construct validity and d) face validity. This study is specialized in the content validity, the writer will conduct whether the content of test fulfils with the material taught or not.

a. Content Validity

In language testing, there are several ways to decide whether a test has sufficient validity. The simplest way is content validity evidence. A test is said to have content validity if it represents all materials taught in the class, it means the relationships between the test items and the course objectives are apparent.⁴³ For instance when the test is supposed to test some particular objectives of analytical exposition that are taught in the

⁴³ Ibid, 27.

⁴¹ William A. Mehrens, Standardized Test in Education, (New York: Winston, Inc, 1975), 62

⁴² Athur Hughes, *Testing for Language Teachers*, (Cambridge: University Press, 2003), 26

tenth-grade, but in fact it tests out of those objectives or whereas it tests other material of analytical exposition which has not been taught in the class, then the test is said to be poor content validity.

Content validity is dealing with the tests items tested whether those items reflect on the content of curriculum measured or not. A problem with content validity evidence is that it gives information about whether a test looks valid, but not whether the reading level of the test is too high or if the test items are poorly constructed. Therefore, a test sometime look valid but measures something entirely different than what is intended, such as guessing ability, reading skill that may have been acquired before instruction.⁴⁴

In summary, content validity evidence answers the question "does the test measures the instructional objectives?" In other word, a test with good content validity evidence matches or fits the instructional objectives.

b. Criterion-Related Validity

The second way to measure the validity evidence is criterion-related validity. The use of this test is said to be accepted when it has fulfilled as a predictor or not. In other word, it takes two forms: one relates to determine a present standing on criterion measure and the other relates to

⁴⁴ Tom Kubiszyn and Gary Borich, *Educational Testing and Measurement* (Singapore, John Wiley & Sons, INC, 2003), 300.

predict future performance on criterion measure.⁴⁵ There are two types of criterion-related validity evidence: concurrent and predictive.

i. Concurrent Validity

According to Hughes this kind of validity is established when the test and the criterion are administrated at the same time. It deals with the relationships between two sets of test scores. In other word, concurrent validity evidence as Ebel stated:

"Concurrent evidence would be useful to show that students appear in the same relative rank order on the two measures",46

For example, the teacher wants to measure the validity of test score for English formative test in one of senior high schools. The test is on 10 of may 2012 which consist of twenty students, then the teacher organizes the test once more on 20 of may 2012 with the same students. After doing those two tests, the teacher compares two scores students get. In fact both those two tests have positive correlation, then the test is called having high concurrent validity evidence because it shows correlation coefficient two different test measuring the same candidates in different time and has same relative score.

⁴⁶ Ibid. 106

⁴⁵ Robert L. Ebel and David A. Frisbie, *Essential of Educational and Measurement*, (U.S.A, Prentice-Hall, INC, 1991), 106.

ii. Predictive Validity

Predictive validity refers to indicate how well the test predicts some future behavior of the examinee. In addition it is particularly useful and important for aptitude test which attempt to predict how well the test taker will do in some future setting.⁴⁷ Furthermore, this validity indicates that there is a need to show positive relationship exists between scores on the test (the predictor) and scores on some acceptable measure of future performance (the criterion).

For instance: a given test for students before joining in the university is estimated test that able to predict the success of test' candidates to follow the teaching learning process later when they are accepted. The quality of some accepted candidates are hopefully will reflect their ability based on the result of test. It means that the upper candidates become success in their activities on the course.

On the contrary with statement above, the lower candidate is predicted not be able to follow the teaching learning process well. Finally, when in the reality the upper candidates who have high result during selection test, but in the progress test or final test they get bad result. Means, the selection test does not have predictive validity.

 $^{^{47}}$ Tom Kubiszyn and Gary Borich, $\it Educational\ Testing\ and\ Measurement\ (Singapore,\ John Wiley & Sons, INC, 2003), 301.$

c. Construct Validity

The term construct refers to a theoretical conceptualization about an aspect of human behavior that cannot be measured or observed directly. 48 Heaton added if a test has construct validity, it is capable to measure certain specific characteristics in accordance with a theory of language behavior and learning. This kind of validity intends as the validity which observes from the point of composition of the test.

d. Face Validity

Meanwhile, construct validity is appreciable from the specific characteristic in language leaning, but this validity is appreciable from its appearance. Hughes said that face validity is not a scientific notion and is not seen as providing evidence for construct validity. For example: a test which intended to measure the pronunciation ability but it did not require the testee to speak, then it might be said to lack face validity. ⁴⁹

Furthermore, face validity is very important to be considered of writing tests. A test which does not have face validity may not be accepted by the testee, teachers, education authorities or employers. The concept of face validity is closely related to content validity.

In short, this validity asks the question "does the test on the "face" of it, appears from the learner's perspective to evaluate what is designed to

⁴⁸ Robert L. Ebel and David A. Frisbie, *Essential of Educational and Measurement* (U.S.A, Prentice-Hall, INC, 1991), 108.

⁴⁹ Athur Hughes, *Testing for Language Teachers* (Cambridge: University Press, 2003), 33.

test?. Therefore, face validity is almost perceived in the terms of content. If the test samples cover what learners have achieved or expect to achieve, the face validity will be perceived.

2. Reliability

Another way to measure the quality of good test is reliability. Reliability is consistent and dependable.⁵⁰ In short, to be reliable, a test must be consistent in its measurement. In other word, it measures the test or re-test reliability to distinguish from mark or re-mark reliability. For example: the teacher who gives the same test to the same subject or matched subjects on different occasions, then the extent that it produces similar results, it is reliable test. But if it is different results, means not reliable. This case as Haris stated on his book,

"Test reliability is affected by a number of factors, chief among them being adequacy of the sampling of the task" ⁵¹

In line of that, Heaton said that it is primary importance in the use of both public achievement and proficiency test *and* classroom tests. For essay test, reliability is necessary to distinguish between three types of reliability: scorer, examinee, and examination. Each type of reliability emphasizes or accounts for certain types of errors.

14

⁵⁰ Douglas Brown, *Teaching by Principles : An Interactive Approach To Language Pedagogy*, (San Fransisco State University: Longman, Inc, 2001). 386

⁵¹ David P Harris, *TestingEnglish as a second Language*, (New York: Mc Craw-ill, Inc, 1959).

3. Practicality

A good test is practical. The term practical is dealing with financial limitations, time constraints, ease of administration, scoring and interpretation.⁵² Brown stated that a test is prohibitively expensive is impractical. A test of language proficiency that takes a student ten hours to complete is impractical. And also a test takes a few minutes for students to take and several hours for an examiner to evaluate is impractical for most classroom situation.

A test may be a highly reliable and valid instrument but still be beyond of facilities, thus as a teacher should keep in mind a number of practical consideration such as economy, ease of administration and scoring, and ease of interpretation. In addition, Arikunto determined some characteristics of practicality in test, they are: 1) Easy to do. 2) Easy to investigate. 3) Completed by some clear instructions.

G. Item Analysis

Evaluating the quality of essay test, those above terms are not adequate to analyze the test. It requires some procedures of item analysis; they are item discrimination and item difficulty. Item analysis is general term that encompasses

⁵²Douglas Brown, *Teaching by Principles : An Interactive Approach To Language Pedagogy*, (San Frnsisco State University: Longman, Inc, 2001). 386

a variety of methods for summarizing and analyzing the responses of students to test items.⁵³

In essay test, there are several certain patterns of responses can indicate desirable and undesirable features of the item or of the scoring procedure employed. Heaton also defined that item analysis can indicate which items may be too easy or difficult and which may be fail, for whatever reasons to discriminate properly between low and high achiever.

Those two terms of item analysis are general related to analyze an objective test item which its concepts are common in language testing. But the focus of this study is to analyze those two terms in essay test.

1. Index of Difficulty

This term can be interpreted as "how hard is this item?" To make a good test, it must be neither too easy nor too difficult. Furthermore, the easiest test does not stimulate students to enhance the effort in solving the test. Otherwise, the hardest test will cause students despondent and they do not have the spirit to try in finishing the test more and more because it is out of their reach.

Index of difficulty is used to show students' approach from the highest possible level of achievement on an item. The achievement level of students

⁵³ Preparing and Evaluating Essay Test Question, Unpublished paper (The University of Lowa), 13.

on question may be showed as the difference between their average item score and the lowest item score.

In essay test, this item may be illustrated as an example, if a maximum of eight points can be obtained on essay question and all examinees receives eight points, then we can say the item is an easy one. Just the opposite, if all examinees obtained a score of zero on the item, it will be a difficult one. There are two patterns to calculate an index difficulty of essay test:⁵⁴

$$X - X min$$

$$X - X min$$

$$X - X min$$

$$Where :$$

$$P = index of item difficulty$$

$$X = the average of item score for all students$$

$$X_{min} = the smallest item score$$

A formula **P** which gives identical values, but is simpler to use:

 X_{max}

= the highest item score

$$P = \underset{n (X \text{max} - X \text{min})}{\approx fX - nX \text{min}}$$

 $^{^{54}}$ Preparing and Evaluating Essay Test Question, Unpublished paper (The University of Lowa), $14\,$

Where:

 \approx fX = total number earned by all students on the question

n = the number of students

X_{min} = the smallest item score

 X_{max} = the highest item score

In addition for this difficulty, it is possible to estimate the index of difficulty for an item based on only a sample of the students.

2. Index of Discrimination

Discrimination may be thought as the capacity among the different candidates and reflect the differences in the performances of the individuals in the group. ⁵⁵ That is, the index of discrimination answers the question "do high achieving students usually produce better responses than lower achieving students?" If so, it could be judged the item functions properly in differentiating between high and low students.

Furthermore, in computing the index of discrimination, it is required the item scores of high and low students' score. For Positive discrimination for an item indicates when the high scoring student obtains a higher average score on the item than does the low scoring student. To separate the more able testee from the less able, here is the formula to calculate the index of discrimination in essay test:

⁵⁵ J. B Heaton, Writing English Language Tests, (New York; Longman Group, 1988). 165

⁵⁶ Preparing and Evaluating Essay Test Question, Unpublished paper (The University of Lowa). 15

$$\mathbf{D} = \mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{L}}$$

Where:

D = index of discrimination

 P_U = upper students

 P_L = lower students

A good method of grouping students for this computation is to take the highest and the lowest 27% of the examinee group. Nevertheless, this way is better to be applied for big class rather than small class. For classes with 20 to 40 students, it is practical to take the highest and the lowest 10 test scores of students. It aims to avoid distortion due to a single student's response.

H. Previous Study

Some theories mentioned in this study which analyze an essay test are based on some studies concerning on test. The writer also has read some previous thesis focused on the same case as writer does. One of the theories that the writer conducted was in the book of *Preparing and Evaluating Essay Test Question*, it explained some guidelines to create an essay tests. One of them is the question must represent the materials taught in the class. In analyzing this kind of test, it has similarity with an objective test which involves two items of analysis; they are index of difficulty and index of discrimination. But in this theory, to analyze the test does not involve the term of item distractors, they are usually used in analyzing an objective test item.

The second study which focuses on the same case was the thesis of English student Education Department of IAIN Surabaya. The first was done by Millatul Islamiyah (2010), she analyzed "Content Validity and Item Analyses of semester II English Final Test for tenth grade students of SMAN 3 Sidoarjo". The result reports that content validity of semester II English Final Test for tenth grade students of SMAN 3 Sidoarjo has good validity of content since the agreement of its content validity is 72%. It means that the item test covered almost all the material taught, but the index of discrimination of final test was poor since the test could not discriminate the upper level of students to the lower level.⁵⁷

The second thesis was done by Ema Rahmatun Nafsah (2011), the student of English Education Department of IAIN Surabaya. She analyzed "An Analysis of English Multiple Choice Questions Test for Seventh Grade of SMP Buana Waru Sidorjo". The result reports that the test is good test based on the characteristics of good test, good face validity and high content validity, high reliability, good index difficulty but poor index of discrimination. The test is also good since 84% items test represent all materials.⁵⁸

The third thesis was done by Ria Dewi Pratikasari (2006), the student of English Department and Art of Surabaya State University (UNESA). She

⁵⁷ Millatul Islamiyah, The content Validity and Item Analysis of semester II English Final Test for tenth grade students of SMAN 3 Sidoarjo, Unpublished S-1 thesis (Surabaya: IAIN Sunan Ampel-Tarbiyah 2010).

⁵⁸ Ema Rahmatun Hafsah, An Analysis of English Multiple Choice Questions Test for Seventh Grade of SMP Buana Waru Sidorjo, Unpublished S-1 thesis (Surabaya: IAIN Sunan Ampel-Tarbiyah 2011).

analyzed "Semester II English Summative Test for the Eight Year Students of SMPN I, Slahung Ponorogo". She found out that the test could not discriminate the upper level of student from the lower level. In the other word the index of discrimination of the test was poor.⁵⁹

This thesis and those three theses differ on implementation of the multiple choice questions that are in the English final test. Thus, some writers of those theses applied the term of distractors to analyze the option of answering question based on the general principles of multiple choice items. While, the first theory has the same concern on this study which deals with two terms of item analysis to analyze the quality of essay test, they are index difficulty and index discrimination.

⁵⁹ Ria Dewi Pratikasari, An Analysis of Semester II English Summative Test for the Eight Year Students of SMPN I, Slahung Ponorogo, Unpublished S-1 thesis (Surabaya: UNESA 2006).

CHAPTER III

RESEARCH METHODOLOGY

In this chapter the writer will define the research methodology. The writer will develop the research design used by the writer in the way of analyzing the study, they are: Research Design, Setting of the Study, Data and Source of Data, Research Instrument, Data Collection Techniques, and Data Analysis Techniques.

A. Research Design

Each research must be planned, it requires a research design. Research design is a plan concerning the way how to collect and analyze the data which are suitable with the purpose of research.¹

There are four types of research method that are often used, they are: Historical research, explorative research, descriptive research and explanatory research.² This study uses a qualitative research and the process deals with documentation, analysis the collected data and explanation of the result of content validity, index difficulty and index discrimination in the numerical calculation. In other word, to do this research, the researcher also conducts quantitative which is used numerical calculation to calculate the data. Besides documentation, the researcher also does an interview with the teacher to support the data.

¹ Prof. Dr. S. Nasution, *Metode Research*, (Jakarta: Bumi Aksara, 2006), 23.

² Drs. Mardalis, *Metode Penelitian*, (Jakarta: Bumi Aksara, 1995), 25.

The study of "an Analysis of Essay Test on the English Final test for Grade Eleven (XI) of SMAN I, Lamongan" uses descriptive research as the way or technique to do a research in case, it will describe the content validity and also describe index difficulty and index discrimination of essay test on English final test for grade eleven (XI) of SMAN I, Lamongan. Those discussions will be explained clearly in the next chapter.

B. Setting of The Study

The subject of this study is the test and students' score of English final test for Eleventh grade of SMAN I, Lamongan. The writer uses random sampling as the technique to determine the sample. I take two classes; they are XI IPA-5 and XI IPA-6. The writer will take some students with the same number from each class. The reason of the writer takes these classes to get valid result of doing this research related to testing. I think that the more I get the sample the better I get the result.

C. Data and Source of Data

Typically data is divided into two types, it is called quantitative if it is numerical form and qualitative if it involves words rather than uses numerical form.³ In this research the writer uses a qualitative since it uses more words or

.

³ William M. K. Trochim, *The Research Method Knowledge Base*, (Ithaca, N.Y.: Cornell Custom Publishing, 1999), 11.

text. This research also combines with quantitative approach since it uses numerical calculation to calculate the data.

The source of data is the subject of research where data is placed. In this way, the researcher can observe about related matter of this research.⁴ The source can be person, place and documentation. From the source of data, the data of this study are:

- Sheets of English Final Test for Grade Eleven Students in SMAN I, Lamongan.
- 2. Students' score of English Final Test Grade Eleven in SMAN I, Lamongan.
- 3. The Content Standard of 2006 English curriculum for Grade Eleven in SMAN I, Lamongan.

D. Research Instrument

The activity of this research uses an instrument as a tool of measurement. By having instrument tool, it helps to collect and measure the data. This study uses documentation to measure the validity of collected data and also interview with the teacher to support the data. The target of those activities is to obtain the result of this research and as learning evaluation being candidate of English teacher.

_

⁴ Suharsimi Arikunto, *Prosedur Penelitian*, (Jakarta: Rineka Cipta, 2006), 222.

E. Data Collection Techniques

There are some techniques used to collect the data, they are test, questionnaires, interview, observation and documentation.⁵ The data of this study is collected through documentation. Documentation is a method to get anything in the form of notes, books, magazine, and etc. ⁶ By documentation; the data will be collected through these following steps:

- 1. Finding the sheets of English Final Test of Grade Eleven for each major in SMAN I, Lamongan.
- 2. Finding the syllabus and content standard of the 2006 English curriculum for the Grade Eleven of senior high school.
- 3. Finding students' score of English Final Test of Eleventh Grade for each major in SMAN I, Lamongan.

F. Data Analysis Techniques

The data must be analyzed to get the empirical evidence of the research. The result of collected data will be analyzed using descriptive. In this case, the data will be described as the way it is. These are some steps that the writer uses to analyze the data:

Suharsimi Arikunto, *Prosedur Penelitian*, (Jakarta: Rineka Cipta, 2006), 222.
 Ibid, 158.

1. Analyzing the Content Validity

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

- Make a list of standard competencies, basic competencies, indicators and learning experience for eleventh grade students of senior high school and the indicators of basic competencies given by SMAN I, Lamongan for eleventh grade students.
- Placing each of the test items in appropriate place with the standard competencies and basic competencies to identify whether those terms covered by the final test.
- Counting the percentage of test items of every language aspects.
- Concluding the result of analysis.

In order to make these procedures clear, the writer presents the illustration of the procedures as follows:

Table I: The example of analyzing Content Validity⁷

COMPETENCE COMPETENCE	Experience $\left \begin{array}{c c} \text{TEM} \\ \text{TEST} \end{array} \right \sum \left \begin{array}{c} \% \\ \% \end{array} \right $
	L L
MENDENGARKAN Memahami makna dalam percakapan transaksional dan kontek kehidupan sehari-hari Merespon makna dalam percakapan transaksional (to get things done) dan interpersonal (bersosialisasi) resmi dan tak resmi secara akurat, lancer dan berterima yang menggunakan ragam bahasa lisan sederhana dalam berbagai kontek kehidupan sehari-hari dan melibatkan tindak tutur: berterima kasih, memuji, dan mengucapkan selamat Merespon makna dalam yang di dengar, makna kat hubungan antar pembicara • Mengiden kasi kata yang di dengar, makna kat hubungan antar pembicara • Mengiden kasi kata yang di dengar, makna kat hubungan antar pembicara • Mengiden kasi kata yang di dengar, makna kat hubungan antar pembicara • Mengiden kasi kata yang di dengar, makna kat hubungan entindak tut berterima kasih, memuji, du ucapan selamat	ifi Mendeng arkan 10, 7, berbagai tindak tutur yang didengar melalui tape atau teman 4, 6, sikan berbagai tindak tutur yang didengar melalui tape atau teman 30, ata didengar melalui tape atau teman 4, 35 melalui tape atau teman 7,

⁷ Millatul Islamiyah, The content Validity and Item Analysis of semester II English Final Test for tenth grade students of SMAN 3 Sidoarjo, Unpublished S-1 thesis (Surabaya: IAIN Sunan Ampel-Tarbiyah 2010). Page 45

2. Analyzing the Index of Difficulty

The index of difficulty in essay test is defined as how hard the question or an item is. It is the extent to which a group of examinees approaches the highest possible level of performance of an item.⁸ The following steps are the procedures of analyzing the index of difficulty on essay test:

- a. Arranging students' score from the highest score to the lowest one.
- b. Taking the high and low students' score, as the upper and the lower groups. Then, classifying the rank order of total score into two scoring groups of equal size, the top half of students as the high score and the bottom half of students as the low score.
- c. Counting the index of difficulty by using the formula of: 9

$$P = \frac{\approx fX - nX \min}{n (X \max - X \min)}$$

Where:

 \approx fX = total number earned by all students on the question

n = the number of students

 X_{min} = the smallest item score

Xmax = the highest item score

⁸ Preparing and Evaluating Essay Test Question, Unpublished paper (The University of Lowa), 13.

⁹ Ibid, 14.

To make those procedures are clear, the following is the example of calculation the index difficulty in essay test:

Table 2: The example of analyzing Index of Difficulty 10

Item Score	`	No. of Students Earning Each Score (f)	Total Points Earned by f Students (fX)	
3		3	9	
2		2	4	
1		2	2	
0	6	5	0	
		n = 12	$\approx fX = 15$	

The explanation of that illustration above shows that essay test is completed by 12 students and the possible scores are provided 0, 1, 2, or 3. That result shows the first column is filled by the item scores (X), the second column is filled by the number of students earning each possible score (f), and then the third column is filled by the total points of students in each score level. It means that the item score is multiplied by the number of students who earned that score. For this item, $X_{max} = 3$ and $X_{min} = 0$. So:

$$P = \underbrace{15 - (12)(0)}_{12(3-0)} = \underbrace{15}_{36} = 0,42$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

¹⁰ Ibid, 14.

The result is about 42%, it indicates that the item is moderate (P=50%). This case means, if the result shows such number, it can discriminate properly between the levels of achievement.

3. Analyzing the Index of Discrimination

The index of discrimination indicates the degree of which students with varying levels of achievement perform differently on a question of essay test. Index discrimination is the difference in item difficulties between groups of students with high and low tests scores. This statement means that positive discrimination for an item result when the high scoring group obtains higher score on the item than does the lowing score.

To analyze the index of discrimination, it has the same steps as in analyzing index of difficulty. Then, to compute the index of discrimination, the writer uses a formula bellow:

$$\mathbf{D} = \mathbf{P}_{\mathrm{U}} - \mathbf{P}_{\mathrm{L}}$$

Where:

D = index of discrimination

 P_U = upper students

 P_L = lower students

¹¹ Ibid. 15.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

The following is the example of analyzing index discrimination; it defines how to calculate index discrimination in essay test:

Table 3: The example of analyzing Index of Discrimination 12

Item Score (X)	iber of Students ing Upper Score (fu)	Total Points fuX	Number of Students Earning Lower Score (f _{L)}	Total Points f _L X
4	5	20	1	4
3	3	9	1	3
2	1	2	4	8
1	1	1	4	4
	nu= 10	\approx fuX = 32	$n_{L=}10$	$\approx f \mathbf{i} X = 119$

In this calculation, the writer takes 10 highest and 10 lowest of test score. As the same reference with index difficulty, for a big class which consists more than 30 students, it is better to take 10 students of highest and lowest one. For this test $X_{max} = 4$ and $X_{min} = 1$

Pu =
$$\frac{32 - (10)(1)}{10(4 - 1)}$$
 PL = $\frac{19 - (10)(1)}{10(4 - 1)}$
= $\frac{22}{30}$ = $\frac{9}{30}$
= 0.73 = $0.73 - 0.30 = 0.43$

¹² Ibid, 16.

CHAPTER IV

RESEARCH FINDING AND DISCUSSION

This chapter discusses the research finding and discussion, the writer will present the process of analyzing data and also discuss the result of data analysis to obtain information about content validity, index of difficulty and index of discrimination of essay test on English Final Test for grade eleven of SMAN I, Lamongan.

All data that supports the writer are from item test, answer key, students' answer sheet, students' score and content standard of curriculum 2006.

A. Analyzing Content Validity

In analyzing content validity, it needs to explain the table of *content standard* that includes standard competencies and basic competencies of 2006 curriculum. To evaluate whether English essay test has content validity or not, the materials of test items should match with *content standard* of curriculum 2006. Furthermore, place each item of test number with appropriate component in the content standard table of 2006 curriculum. It uses to define the representative materials in English essay test.

Analysis table of this discussion is listed in the table 6 of appendix I, it makes the writer easy to analyze the content validity of essay test on English

Final Test for Grade Eleven Students of SMAN I, Lamongan. There are six columns in the table, first column consists of standard competencies, the second column consists of some basic competencies that represent standard competencies, the third column consists of several indicators that represent the basic competencies of English lesson, and the fourth column consists of learning experiences or the activities in the class during teaching learning process. While, the fourth, fifth and sixth column consist of the calculation of test items that appropriate and represent basic competencies. It also presents the percentage of total number of items related to basic competencies.

This analysis based on Hughes book, he said that the test has content validity if the test items represent the content of particular course of study. It means that the relationship between the test items and course objectives are apparent.¹ In this case, the writer conducts the result of analyzing content validity based on the content standard used by SMAN I, Lamongan for English lesson. The explanation of every feature in learning process will be explained as follow:

- There are 20% for speaking skill focuses on the discussion about the use of To Infinitive and Ving. This analysis is in number 34.
- 2. There are 20% for speaking skill focuses on discussion of how to express love, sadness and embarrassment. This analysis is in number 31.

¹ Athur Hughes, *Testing for Language Teachers* (Cambridge: University Press, 2003), 5.

- 3. There are 20% for reading skill focuses on the discussion of the generic structure of narrative text, hortatory exposition, and spoof. This analysis is in number 32.
- 4. There are 20% for writing skill focuses on the discussion of passive voice.

 This analysis is in number 33.

Meanwhile, English essay test used by Grade Eleven in SMAN I, Lamongan consists of five items. It is definitely that the calculation of percentage in each item eases the writer to do analysis. The writer concludes that English essay test is good since 80% items represent all materials. The test is said as high content validity if 50% of items or more represent all materials.

Based on the analysis above, there are four items representing the materials in the class. It means that an item does not cover the material, as stated in number 35. Although this item is tested but it does not match with indicator of standard and basic competencies.

B. Analyzing Index of Difficulty

To organize index difficulty value of English essay test for grade eleven students in SMAN I, Lamongan. The analysis is divided into two classifications; describing some scores for each number of test questions and computing the number of students got those scores.

In analyzing index difficulty, firstly the writer arranges the students' score from the highest score to the lowest one. Then, the writer classifies the number of students from that arrangement and divides it into two groups, upper group and lower group in equal size. It determines the students in the same amount because XI IPA-5 and XI IPA-6 class are equal which consist of odd numbers.

The writer uses Heaton's manner to classify the number of students in upper and lower group equally. Although those two classes consist of odd number, the writer can dispense a number of student's score, so the numbers of those sizes are same. As Heaton said in his book that if there is an odd number of a script, dispense with one script chosen in random.²

Having determined the upper and lower group of students, the writer computes index difficulty for each number of test questions by using the formula in journal of *Preparing and Evaluating Essay Test Question* as stated in chapter II and III. The writer uses a table to make the calculation efficient. The table is in appendix 6-7.

Analysis table of index difficulty consists of three columns, first column consists of some essay scores in each number, the second column consists of the total number of students who get those scores, and the last column consists of the multiplication result from essay scores and the total number of students who get those scores.

² J. B Heaton, Writing English Language Test, (New York: Longman Group, 1988), 178.

Here, the writer will present the result of computing index difficulty from two classes; XI IPA-5 and XI IPA-6 as the sample of her research.

1. The Result of Analyzing Index Difficulty on Essay Test for XI IPA-5 class

Having explained in previous sub chapter, XI IPA-5 class consists of an odd number, so it is obviously that the distribution of upper and lower group are not equal in size. The writer dispenses one score from upper and two scores from lower. So, the number becomes 14 students in upper and 14 students in lower. Based on the journal of *Preparing and Evaluating English Essay Question*, it is stated that if a maximum score of eight points could be earned on essay question and all examinee received eight points, it would say the item is easy one. If the value is about 40-50 %, it indicates that the item is moderate. This case means that the value shows such number can discriminate properly between the levels of achievement. ³

The result of analyzing index difficulty in this class is shown in table 11-15 appendix 6. There are 2 numbers of test questions are considered as too easy items since their values are 0,87 and 0,95. They are on number 31 and 32. Those two numbers are might be revised because they do not require students to organize their thought. Most of students might simply answer those numbers of question without spending to think hard. For number 34 and 35, they are moderate because their values are 0,42 and 0,51. It means that the

³ Unpublished paper, *Preparing and Evaluating Essay Test Question*, (The University of Lowa), 15.

test questions are work properly and available to apply for another test because they do not need to be revised. While, number 33 indicates as too difficult since its value is 0,28. This item of question is badly to rewritten because it cannot work as expected.

2. The Result of Analyzing Index Difficulty on Essay Test for XI IPA-6 class

Different with the previous class, XI IPA-6 class is unbalanced in acquiring the score among students. The students are mostly get low score, so it makes writer difficult to divide between the upper and lower group in equal size. The result of analyzing index difficulty in this class is shown in table 4 appendix 6.

In this class, the number of students in upper and lower group is 15 students. The writer only dispenses one script of student. The result of index difficulty for XI IPA-6 class indicates that there 2 test numbers are too easy, they are 31 and 32 which the results are 0,98 and 1. While, there are 3 test numbers shows as too difficult items which values are 0,1 up to 0,32. They are on number 33, 34, and 35. In other word, the test items for English essay test used by XI IPA-6 in SMAN I, Lamongan need to be revised to fulfill the standard of appropriateness for evaluating students.

Finally, after discussing the result of analyzing index difficulty for XI IPA-5 and XI IPA-6 class, generally this result can be concluded that most of items used for those classes are good since they have value around 0,31-0,87. However, difficult items test seem to make students spend more time because

they only focus to answer those items. At the end, from that their difficulty, students will probably utilize their time to cheat each other.

In addition, good test items may be used for the next test or they are able to be references for test maker to design a good test.

C. Analyzing Index of Discrimination

In analyzing index discrimination value, the writer uses similar reference for index difficulty. These two classes consist of less than 40 students. If a class with 20-40 students, it is better to take 10 students for each high and low score. Based on that statement, the writer finds difficulty on how to analyze index discrimination appropriately. This case is caused by some dispensed scores, so the writer decides to take the sample based on Heaton's calculation which is stated on previous discussion.

There are several steps to compute index discrimination. First, the writer classifies the amount of students for upper and lower group in equal size which based on English essay score, then the writer takes those calculations as the sample of analysis of index discrimination. Second, the writer makes a table to make clear explanation in analyzing index discrimination. It is shown in the table 21-30, appendix 8 & 9.

The analysis table of index discrimination consists of five columns, the first column consists of some scores of essay test in each number of test

⁴ Ibid. 15.

questions. The second and the fourth columns consist of the total number for upper and lower group obtained from each score of essay test. The third and the last columns consist of the multiplication result. The formula as stated in chapter II and III. Here, the writer also applies arikunto's classification to recognize index discrimination value for each test item:

D: $00 \text{ up to } 0.20 \longrightarrow \text{poor}$

D : 0.20 up to 0.40 \longrightarrow satisfactory

D : 0,40 up to 0,70 _____ good

D : 0,70 up to 1,00 _____ excellent

The next explanation will clarify the result analysis of index discrimination of XI IPA-5 and XI IPA-6 class.

1. The Result of Analyzing Index Discrimination on Essay Test for XI IPA-5 class

The calculation of index discrimination directly concerns on the level of students' achievement in essay test. The classification of those two groups becomes 14 upper and 14 lower students. The authentic score from upper is 16 students and 15 students from the lower group. Furthermore, the easiest way to obtain equal in size, the writer must dispense two scores from the upper group and one score from the lower group.

By understanding the classification, the table shows that there are 3 numbers of test question are categorized as satisfactory since their discrimination between 0,39-1 %. They are items number 33,34, and 35.

While 2 other numbers (31 and 32) are categorized as poor since their discrimination are 0,1-0,25.

The analysis of index discrimination defines that positive discrimination for an item is when the upper students obtain high score and the lower students obtain low score.⁵ Here, the students' scores are described as follow:

Table 4: Students' Scores to Compute the Result of Analyzing Index Discrimination in XI IPA-5 Class

	Upper Group			Lower Group	
No			No		
	Name	Score		Name	Score
1.	Khurunin Maulidia I.	25	15.	Dewi Nur Latifah	15
2.	Moh. Fatih Rohmatulloh	25	16.	D <mark>ha</mark> rma Favitri H.	13
3.	Muflichia Indri	25	17.	Q <mark>ud</mark> watun Niswah	13
	Khanifianah				
4.	Wahyu Apri Rhohati	25	18.	Aisyah	13
5.	Farih Jamil Rajabi	20	19.	Vawanda Dian Saputra	12
6.	Frans Kurniawan H.	20	20.	Fadilatul Lailah	10
7.	Heva Normalita P. M.	20	21.	Khoirul Huda	10
8.	Moh Adibhadiansyah	20	22.	Noto Mashari Joyo	10
9.	Nurhuda Hadi D.	20	23.	Sulthon Hans Prasetyo	10
10.	Sisca Juwita Kartika	20	24.	Try Indah Lestari	10
	Dewi				
11.	Winda Nurayu Octavia	20	25.	Vivi Viftiantika A.	10
12.	Desy Lukita Indriyani	18	26.	Ahmad Sulton Aria B.S	7
13.	Nur Arifah Nisa E. F.	18	27.	Ahmad Riyan Fahmi	5
14.	Rizki Vera Vernanda	18	28.	Muqorrobin Filosofi S.	5

⁵ Ibid, 15.

To compute index discrimination, first the writer compute index difficulty from each group, and then from the calculation of those groups the writer uses the formula of index discrimination. The result of index discrimination directly relates to index difficulty, if the value of index difficulty is good means that the test items work properly to discriminate the level of students' achievement. In this case, English essay test which has been evaluated for XI IPA-5 has acceptable discrimination. It means that test items used by SMAN I, Lamongan for XI IPA-5 work properly to differentiate between upper students and lower students, and can be functioned for other tests.

2. The Result of Analyzing Index Discrimination on Essay Test for XI IPA-6 class

In accordance with previous class, the calculation of index discrimination is primarily compute index difficulty between highest and lowest score of students' achievement on essay test. Then the result of higher group is diminished by the result of lower group as shown in appendix 9.

The table 26-30 in appendix 9 reports that index discrimination of test items used to evaluate English essay test for XI IPA-6 class is poor. Almost the test numbers are categorized as too easy and too difficult. Their value are around 00-0,20. There are 4 test numbers are considered as poor discrimination and 1 number could be said as satisfactory. It is 0,39% in number 35. The result indicates that the essay test items are not accepted to

use in XI IPA-6. As Heaton stated when the test item is too difficult or too easy, it does not discriminate between the more able testee to the less able, and it can be said as wrong direction in evaluating candidates.⁶

When discrimination value approximates to zero, it indicates that test items do not relate to each other or they are too easy (or too difficult). In this case, it reports that test items used by XI IPA-6 of SMAN I, Lamongan does not discriminate between the more able students to the less able. The table bellow is the students' scores on essay test:

Table 5: Students' Scores to Compute the Result of Analyzing Index Discrimination in XI IPA-6 Class

No	Upper Group		No	Lower Group	
110	Name	Score	110	Name	Score
1.	Husnuz Zaka Ulin Nuha	25	16.	Ardhan Dwi Meitrika S.	15
2.	Luqman Aji Kusumo	23	17.	Arini Vidyastutik	15
3.	Ganang Dwi Kurniawan	20	18.	Dziky Arista Utomo	15
4.	Marda Inggar Setiawan	20	19.	Khoirus Viestaria	15
5.	A. Rizky Wahyu Sy.	18	20.	Sri Wahyuni Aries Shandy	15
6.	Ahmad Mahbub Junaidi	18	21.	Aisyah Cholifaur Rohmah	13
7.	Amirah Inas Sholihah	18	22.	Ella Putri Artifasari	13
8.	Eka Nurdiyana Rahmawati	18	23.	Fakhriatul Mila	13
9.	Fita Nur Rochma	18	24.	Fery Adhy Permana	13
10.	M. Fadli Ramadhan	18	25.	Isnaeni Agustin Zuhdy	13
11.	Ita Yuliani Endah Nur H.	16	26.	Mochammat Noor Effendi	13
12.	Anggie Putri Novita Sari	16	27.	Nasrul Ummah	13
13.	Qurrotu A'yuni Asfaara	16	28.	Nur Afifa Fatin	13
14.	Ummi Shobikhah	16	29.	Setiaji Utama	13
15.	Derry Bagus Akhmaddin	16	30.	Shofyan Ilhami	13

⁶ J. B Heaton, Writing English Language Test, (New York: Longman Group, 1988), 181.

.

Having discussed about index discrimination for students' of XI IPA-5 and XI IPA-6 in SMAN I, Lamongan, the writer concludes that test items made by the teacher in that school has acceptable discrimination for XI IPA-5 class because the test items are not too easy or too difficult. The result is distant from zero point thus it can discriminate between the levels of students in XI IPA-5. While for XI IPA-6, the result approximates to zero point. The test items might be too difficult and involves on default with the test making. In addition, for XI IPA-6 the test should be revised in appropriate way.

CHAPTER V

CONCLUSION AND SUGGESTION

This chapter will present the conclusion and suggestion about this research. In the conclusion, the writer will conclude the result organized by the previous chapter from research finding and discussion. This discussion also involves the result of analyzing content validity, index difficulty, and index discrimination of Essay Test in the English Final Test for Grade Eleven of SMAN I, Lamongan. While the suggestion presents how the writer suggests some matters needed to implement in creating good English essay test especially for the teacher and further researcher as stated in the first chapter.

A. CONCLUSION

After conducting the result of research finding and discussion in the previous chapter, in this discussion, the writer will summarize several matters concerning on the agreement of content validity, index difficulty, and index discrimination of Essay Test on the English Final Test for Grade Eleven of SMAN I, Lamongan. They are defined as follow:

From the research finding, it shows that essay test used by SMAN I,
 Lamongan has good content validity. It can be proved from the concurrence of its content validity which reaches 80%. It means that the test items almost

- represent all materials taught in the class. Even though there was an item of test does not match with the content standard of English lesson for grade eleven students, it could say that the content of test items deal with several materials taught.
- 2. The result of index difficulty of Essay Test on English Final Test for Grade Eleven of SMAN I, Lamongan used by XI IPA-5 is acceptable. Almost the result in each test items are around 0,28-0,80, it means that the test item is good to evaluate students' achievement. Thus, those test items are safe and proper to be used for future test without being rewritten. While, the result of index difficulty for XI IPA-6 in each test items are around 0,1-1,00 which indicates that the test items are both too difficult and too easy. The test items should be revised for XI IPA-6 to evaluate the levels student's achievement. Too easy and too difficult items do not discriminate students' achievement well because when too easy items occurred, most of students get correct answer. Meanwhile, when the items are too difficult, then a chance to get correct answer is only for upper students.
- 3. The result of index discrimination for XI IPA-5 is acceptable. Between the result of index difficulty the result of index discrimination shows a balance. Most of test items are categorized to comply the criteria of high discrimination. It could say as good discrimination since the result is distant from zero point and close to point 1. In fact, the discrimination value is around 0,39-1,00. The different result occurs for XI IPA-6, most of the test

items are considered as poor discrimination because the value is around 00-0,19. Those results mean that essay test used by Grade Eleven Students in SMAN I, Lamongan is only appropriate for XI IPA-5 class. While for XI IPA-6, the test should be revised to differentiate between the more able students to the less able.

B. SUGGESTION

The previous discussion explains how the writer concludes about her research. Here, the writer will share several matters to suggest for teacher and other researchers which have the same concerns of the study:

- 1. Although the result of analysis this test is good, but the teachers still should try to make pre-test before giving a real test, such as in the progress achievement test and final achievement test. It is useful to know whether the test is adequate to measure the progress of students during teaching learning activities.
- 2. The teachers should revise when they find that the test items are poor, too difficult or too easy. Thus, the test items could discriminate the students' level between the higher and the lower.
- 3. The teachers must understand how to give appropriate amount of items for test. It could say that the amount of test items is compatible with available time given to do the test. The amount of English Final Test for Grade Eleven Students in SMAN I, Lamongan is compatible with the time settled for doing

the test. It consists of 35 items and the time is 90 minutes. It helps students to review several difficult questions and they can check their answer carefully before the finished time.

4. For further researcher, there are several theories discussed the same case about testing. The researcher should recognize several tests that ignore the criteria of good test, such as validity, reliability, practicality, and item analysis. Thus, the researcher consider how to judge whether the test is good or not from analyzing several qualities of good test.

BIBLIOGRAPHY

- Arikunto, Suharsimi. Prosedur Penelitian. Jakarta: PT. Rineka Cipta. 1996.
- Arikunto, Suharsimi. Dasar Dasar Evaluasi Pendidikan. Jakarta: Bina Aksara. 1984.
- Bothell, Timothy W, Reiner, Christian M, Sudweeks, Richard R, and Wood, Bud. *Preparing Effective Essay Question*. New Forum Press. 2002.
- Brown, H. Douglas. *Teaching by Principles: An Interactive Approach to Language Pedagogy, Second Edition*. San Francisco State University: Longman Inc. 2001.
- ------ *Principles of Language Learning and teaching. Fourth Edition.* San Francisco State University: Longman Inc. 2000.
- Bulletin, Technical. *Preparing and Evaluating Essay Test Question*. Unpublished Paper. The University of Lowa.
- Ebel, Robert L. Essential of Educational Measurement. USA: Prentice-Hall Inc New Jersey. 1979.
- Gary Borich Tom Kubiszy. *Educational Testing and Measurement*. Singapore: John Wiley & Sons, INC. 2003.
- Gronlund, Norman E. *Measurement and Evaluation in Teaching*. New York: Macmillan, Inc. 1985.
- Gronlund, Norman E Waugh, C. Keith. *Assessment of Student Achievement. Ninth Edition*. USA: Pearson Education, Inc. 1968.
- Harmer, Jeremy. *The Practice of English Language Teaching*, New York: Mc Craw-ill, Inc. 2011.
- Harris, David P. *Testing English as Second Language*. New York: Mc Craw-ill, Inc. 1959.
- Hughes, Athur. Testing for Language Teacher. Cambridge: University Press. 2003.

- Heaton, J. B. Writing English Language Test. New York: Longman Group. 1988.
- Islamiyah, Millatul. Content Validity and Item Analysis of Smester II English Final Test for Tenth Grade Students of SMAN 3, Sidoarjo. Unpublished S-I Thesis. Surabaya: IAIN Sunan-Ampel (English Education Department). 2010.
- Madsen, Harold S. *Techniques in Testing*. New York: Oxford University Press. 1983.
- Mardalis, Drs. Metode Penelitian. Jakarta: Bumi Aksara. 1995.
- Mehrens, William A. Lehman, Irvin J. *Standardize Test in Education*. USA: Rinehart and Winston, Inc. 1969.
- Nasution, S. Metode Research. Jakarta: Bumi Aksara. 2006.
- Nurgiyantoro, Burhan. *Penilaian Dalam Bahasa dan Sastra*. Yogyakarta: BPFE Yogyakarta. 2001.
- Rahmatun, Ema Hafsah. An Analysis of English Multiple Choice Questions Test for Seventh Grade of SMP Buana Waru Sidoarjo. Unpublished S-I Thesis. Surabaya: IAIN Sunan-Ampel (English Education Department). 2011.
- Sudijono, Anas. *Pengantar Evaluasi Pendidikan*. Jakarta: PT. Raja Grafindo Persada. 1996.
- Thoha, Drs. M. Chabib. *Teknik Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada. 1991.
- Trochim, William, M.K. *The Research Method Knowledge Base*. Ithaca, N.Y: Cornell Custom Publishing. 1999.
- Tuckman, Bruce W. Measuring Educational Outcomes, Fundamental of Testing. USA: Harcourt Brace Jovanovich, Inc. 1975.