ITEM ANALYSIS OF ACHIEVEMENT TEST IN FINAL TEST FOR THE SEVENTH GRADE STUDENTS OF SMP N I MOGA PEMALANG in the academic year of 2008/ 2009

A FINAL PROJECT

Submitted in Partial Fulfillment of the Requirements for the Degree of

***Sarjana Pendidikan* in English**

Ika Rini Nuryulia

2201404523

ENGLISH DEPARTMENT

FACULTY OF LANGUAGES AND ARTS

SEMARANG STATE UNIVERSITY

2009

PERPUSTAKAAN UNNES

PAGE OF APPROVAL

This final project was approved by the board of the examiners of the English Department of the languages and Arts Faculty of Semarang State University on December 2008

The board of Examiners:

1. Chair Person

   Dra. Malarsih, MSn
   NIP. 131764021 _____

2. Secretary

   Dr. Dwi Anggani LB, MPd
   NIP. 131813665 _____

3. First Examiner

   Drs. Suprapto M.Hum
   NIP. 131125925 _____

4. Second Examiner

   Sri Wahyuni, S.Pd., M.Pd
   NIP. 132320169 _____

5. Third Examiner

   Drs. Amir Sisbiyanto, M.Hum.
   NIP. 131281220 _____

Approved by

The Dean of FBS

Prof. Dr. Rustono
NIP. 131281222

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN UNNES

PERNYATAAN

Dengan ini, saya:

Nama : IKA RINI NURYULIA

NIM : 2201404523

Prodi/Jurusan : Pendidikan Bahasa Inggris S-1/Bahasa dan Sastra Inggris

Fakultas Bahasa dan Seni Universitas Negeri Semarang, menyatakan dengan sesungguhnya bahwa skripsi/ tugas akhir/ final project yang berjudul:

ITEM ANALYSIS OF ACHIEVEMENT TEST IN FINAL TEST FOR THE SEVENTH GRADE STUDENTS OF SMP N I MOGA PEMALANG in the academic year of 2008/ 2009

Yang saya tulis dalam rangka memenuhi salah satu syarat untuk memperoleh gelar sarjana ini benar-benar merupakan karya saya sendiri yang saya hasilkan setelah penelitian, pembimbingan, diskusi dan pemaparan atau ujian. Semua kutipan, baik langsung maupun tidak langsung, baik yang diperoleh melalui sumber kepustakaan, wahana elektronik, maupun sumber lainnya, telah disertai keterangan mengenai identitas sumbernya dengan cara sebagaimana yang lazim dilakukan dalam penulisan karya ilmiah.

Semarang, 16 Maret 2009

Yang membuat pernyataan

PERPUSTAKAAN
UNNES

IKA RINI NURYULIA

NIM.2201404523

What you're convinced of,

Definitely would happen in the future;

And the believe in what you're thinking about make it comes true.

(Frank lioyd wright)

THIS FINAL PROJECT IS DEDICATED TO:

- *Allah the Almighty for the greatest blessing*
- *My beloved Mom and Dad: Mr. Andi and Mrs. Mureni.*
- *My beloved brother: Ikmal*
- *My lovely boyfriend: Prasetyo Adi, A.pt*
- *My best Friends: Pranita, Ike, Iyum.*
- *My cheerful friends in the English Department 2004 Universitas Negeri Semarang*
- *My Friends at Ricky Boarding House K5-sada Banaran.*
- *All of my friends.*

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

ACKNOWLEDGEMENTS

First and foremost the writer would like to express her gratitude to Allah SWT for his blessing and inspiration give to her in completing her study.

The deepest gratitude and appreciation is extended to Drs. Amir Sisbiyanto, M.Hum, her first advisor, for his patience to guide, advice and encourage her since the beginning until the end of this final project writing has been completed. Furthermore, she would like to express her deepest gratitude to Sri Wahyuni S.Pd., M.Pd, her second advisor, for the advice, corrections and encouragement for her until this final project was completely done.

She is also very grateful to all of her lecturers of the English Department of UNNES for all guidance and knowledge during her study at UNNES. Her writer also expresses her special thanks to the three other members of the board of examiners. Dra. Malarsih, Msn, the first chairman of examination, Dr. Dwi Anggani LB, MPd, the secretary of the examination, and Drs. Suprapto M.Hum., the first examiner.

Her special gratitude is forwarded to her beloved Mom and Dad for their patience and love. She also thanks to all of her family, who gave her spirit in finishing this study. The last to people who helped me that cannot be mentioned one by one, I thank them very much. Hopefully, Allah gives his help and blessing to them.

PERPUSTAKAAN UNNES

Finally, the writer realizes that her study is still far form being perfect. The writer has a great expectation that her study will be beneficial and useful for anyone who is interested in reading this final project.

Semarang, 2009

Ika Rini Nuryulia

NIM: 2201404523

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

ABSTRACT

Nuryulia Ika Rini. 2009. *Item Analysis of Achievement Test in Final Test for The Seventh Grade Students of SMP N 1 Moga Pemalang In The Academic Year 2008/ 2009.* Final Project. English Education. English Department Language and Art Faculty, Semarang State University.

Key words: English, Achievement Test, validity, Reliability, Difficulty Level, Discriminating Power.

Entering global era, we need adequate knowledge in facing global world which is full of competition. In this case, the role of English is very important either in direct communication or interaction. As a means of communication, English has to be mastered actively in oral and also in written. Through education process, it is expected that there is a change of students. One way to know students' ability in using English we can use evaluation or test. In learning, test is a tool of evaluation which has important role to measure the teaching learning process in schools. While, the main purpose of this study is to analyze the English final test items which are administered to the seventh grade students of SMP N 1 Moga Pemalang. The problem of this study is ""How good is test items in final test prepared for the seventh grade students of SMP N I Moga Pemalang in the academic year of 2008/ 2009?"

Achievement test emphasizes past progress, whereas aptitude test primarily concerns with future potenticalities. Achievement test is used for assesing present knowledge and abilities. The primary goal of the achievement test is to measure past learning, that is, the accummulated knowledge and skills of an individual in a particular field or fields.

The data used in this study were taken from the test papers and students' answer sheets. The test papers consist of 40 items in the form of multiple choice. The students' answer sheets are needed for statistical analysis to find out the quality of the items based on item analysis, validity and reliability of the test.

From the result of the analysis, the mean of validity level is 0.3162. Then, the result was consulted to the value of r product moment formula at level of significance 0.05. Since the value of r calculation is more than of the table, it can be concluded that the test is valid. However, this test is reliable, with the coefficient of reliability of the whole test item is 0.751. The mean of the difficulty level is 0.60. So, the English summative test items are classified as Medium items in term of their difficulty level. Then the mean of the discrimination power is 0.48, meaning that the items are still able to discriminate the clever students and the poorer ones.

Based on the result, the writer suggests to the teachers as the test makers to prepare test items far in advance before they give it to the students. They should also pay attention to the writing of multiple choice items and the characteristics of a good language. Finally, the writer draws a conclusion that the items in the English final test for the seventh grade students of SMP N 1 Moga Pemalang could still be used as an instrument of evaluation with some revisions.

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS…… i

ABSTRACT…… iii

TABLE OF CONTENTS…… v

TABLE OF APPENDIX…… vii

CHAPTER

I. INTRODUCTION……..... 1

1.1 Background of the study……..... 1

1.2 Reason for choosing the topic……..... 3

1.3 Statement of the problem……..... 4

1.4 Purposes of the study……..... 5

1.5 Significance of the study……..... 5

1.6 Limitation of the study..... 6

1.7 Outline of the study…….. 6

II. REVIEW RELATED LITERATURE……..... 7

2.1 Testing, Evaluation, Measurement..... 7

2.2 Criteria of A Good Test 8

2.2.1 Validity 9

2.2.1.1 Types of validity ...……..... 9

2.2.1.2 Factors Influencing Validity….....…….. 10

2.2.2 Reliability .....……. 11

2.3 Types of Test .....……..... 14

2.4 Achievement Test .....……..... 15

2.4.1 Types of Achievement Test..... 15

2.4.2 Basic Principles of Achievement Test 16

2.5 Multiple-choice Test Item..... 17

2.5.1 The advantages of using the multiple-choice form..... 17

2.5.2 Suggestions for writing multiple-choice form..... 18

2.6 Item Analysis..... 19

2.6.1 Item discrimination power..... 21

2.6.2 Index difficulty..... 22

III. METHOD OF INVESTIGATION..... 24

3.1 Population and sample..... 24

3.2 Sampling technique..... 24

3.3 Identification of the Problems..... 25

3.4 Technique of collecting data..... 26

3.5 Technique of Data Analysis..... 26

IV. ANALYSIS AND DISCUSSION..... 31

4.1 Result of the Analysis 31

4.1.1 Analysis of Validity..... 31

4.1.2 Analysis of Reliability..... 32

4.1.3 Analysis of Difficulty Level..... 33

4.1.4 Analysis of Discrimination Power..... 34

4.2 Discussion..... 35

V. CONCLUSION AND SUGGESTION 76

5.1 CONCLUSION 76

5.2 SUGGESTION 77

APPENDICES..... 78

BIBLIOGRAPHY..... 100

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN UNNES

TABLE OF APPENDIX

Appendix: Page

1. The Computation of the Validity, Reliability, Difficulty level, ... 79
   and Discriminating Power of the Final Test
2. Analysis of Validity..... 84
3. Analysis of Reliability..... 87
4. Analysis of Discrimination Power..... 89
5. Analysis of Difficulty Level..... 91
6. The Recapitulation of Multiple-choice..... 93
7. The Test Item..... 95

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

CHAPTER I

INTRODUCTION

1.1 Background of the Study

Entering globalization era, which is more popular with free market, claims each individual to prepare reliable resource especially in IPTEK (Technology and Science). In mastering technology, we need adequate knowledge so that we can exploit it in facing global world which is full of competition. In this case, it means that the role of English is very important either in direct communication or interaction. As a means of communication, English has to be mastered actively in oral and also in written. The fast growth of technology claims us to be more proactive in answering global information as asset in fulfilling requirement of market. As an international language, English is not merely as academic requirement but also as global communication media.

In Indonesia, English is included into curriculum and considered as a first foreign language to be taught in Junior high schools, Senior high schools, and universities. Now, English has been given since elementary school, moreover English has been introduced at kindergarten and play group. In elementary school, English has been taught as local content.

Through education process, it is expected that there will be changes on the part of the learners. One way to know students' ability in using English we can use evaluation or test. Test is a procedure or an appliance used to know or measure

something, with determined procedures (Arikunto, 2005:53). In learning, test is a tool of evaluation which has an important role to measure the teaching learning process in schools. In this case, test has double functions, i.e. measuring the ability of student and measuring the efficiency of teaching learning process. Evaluation appliance which is used in school in the end of semester is final test. A test is an assessment, often administered on paper intended to measure the students.

Recognizing that evaluation is very important in school, teachers have to know the quality of a good test or criteria of a good test. There are some characteristics of a good test (Arikunto, 2005:53):

a. Tests have to have high validity. A validity represents all important condition in making of test. An evaluation technique is considered has high validity if the test measures what actually to be measured
b. Tests should be reliable or can be trusted. It gives consistent result, if it is tested several times. A test is reliable if the test shows constancy.
c. Tests must be objective. It means that, a test has objectivity if there is no subjective factor in doing the test especially in scoring system.
d. Tests must be practical and has clear instruction.

Curriculum, teaching learning process and evaluation are three of important parts in education. The three parts are closely related each other. Curriculum is a description of the goals of education as a base of teaching-learning process. Meanwhile, teaching-learning process is an effort to reach the goals in the curriculum that is done by the teachers.

In the year of 2006 / 2007, there is a new curriculum that is KTSP (Kurikulum Tingkat Satuan Pendidikan). KTSP is an operational curriculum, arranged by school which relates to content standard which is published by BSNP (Badan Standar Nasional Pendidikan) held by each education based on potenciality and requirement of education.

In this new curriculum, school has authority to make and develop the curriculum. In this case, the teacher must be creative and active in determining steps of making and developing curriculum. In KTSP, teacher can fully determine assessment to students' abilities.

Therefore, researcher tries to analyze the item test of evaluation in final test based on validity, reliability, discrimination power, and index difficulty because the researcher hopes that she can make a good test in each level.

1.2 Reasons for Choosing the Topic

In this study, the writer would like to focus her research on the test items analysis, which is administered to seventh grade students of SMP N I Moga Pemalang in the academic year of 2008/ 2009. The reasons of choosing the topic are as follows:

a. The evaluation of test items in final test in the academic year of 2008/ 2009 has just been administered to the student even though it has not been analyzed in terms of difficulty and discrimination power.

b. In teaching learning process, evaluation is carried out every term. If test constructors do not pay attention in selecting the items, the validity and the reliability of each test will be less guaranted. For this reason, every test constructor must be careful in constructing the test items, so that the result will meet the desired goal.

c. By applying item analysis, we can indicate which items may be reliable and valid. We can check properly whether the test has a good quality or not.

1.3 Statement of the Problem

The writer would like to find the answer of the following question: "How good is test items in final test prepared for the seventh grade students of SMP N I Moga Pemalang in the academic year of 2008/ 2009?"

Specifically, the writer will limit her problem into the following questions:

a. Do the test items in final test in SMP N I Moga Pemalang have validity ?

b. Do the test items in final test in SMP N I Moga Pemalang have reliability?

c. What is the difficulty level of the test items in final test in SMP N I Moga Pemalang?

d. Do the test items in final test in SMP N I Moga Pemalang have discriminating power?

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN UNNES

1.4 Purposes of the Study

The purposes of this study are to obtain an objective description of the achievement of test items in final test for the seventh grade students in the academic year of 2008/ 2009.

The purposes are then specified into the following:

a. To describe the validity of each test item in final test in SMP N I Moga Pemalang.
b. To describe the reliability of each test item in final test in SMP N I Moga Pemalang.
c. To describe the value of the difficulty level of each test item in final test in SMP N I Moga Pemalang.
d. To describe the value of the discriminating power of each test item in final test in SMP N I Moga Pemalang.

1.5 Significance of the Study

PERPUSTAKAAN UNNES

The advantages that can be required from this study are as follows:

a. For student: students can use the result of the evaluation to measure their ability in mastering the material.
b. For teacher: teacher can use the result to know the students' ability in mastering the material.
c. For test maker: the test maker may use it as a supplement in constructing test.
d. For writer: the writer herself can increase her skill in constructing test items.

1.6 Limitation of the Study

The writer argues to analyze the English test of the seventh grades of SMP students in the academic year of 2008/2009 in the form of multiple choices, in brief that:

a. Multiple- choice items have only one correct answer each. Thus, the grader will grade the answer objectively.
b. By using this type of item analysis the discrimination power and difficulty level of the test can be practically determined.

1.7 Outline of the Study

The study is outlined into five chapters:

Chapter I discusses the background of the study, reasons for choosing the topic, research questions, purposes of the study, and significance of the study.

Chapter II gives review literature, which consists of the definitions of testing, evaluation, measurement and criteria of a good test, types of test, achievement testing, multiple – choice test item, analysis item.

Chapter III deals with method of investigation which consists of population and sample, sampling technique, technique of data collection and technique of data analysis.

Chapter IV covers analysis and discussion of the research finding. Then, chapter V presents conclusion of the research and suggestions.

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

CHAPTER II

REVIEW RELATED LITERATURE

2.1 Testing, Evaluation, Measurement

Testing, Evaluation, and Measurement are three basic related concepts that we need to understand. The similarity among them is to assess the students' ability in mastering language. Test and measurement are parts of evaluation. The difference between test, evaluation and measurement can be found in the practice of assigning final marks to students at the end of a unit of work. Evaluation as one of the components of the curriculum is also one of the activities to stimulate the students so that they will learn and master the material that have been taught. In a test, the students will review the lesson, which their teachers have taught before. Measurement is a method to measure ability according to the rules. Here the definitions of Test, Evaluation and Measurement.

According to Heaton (1975:1), tests maybe constructed primarily as devices to reinforce learning and to motivate the student, or primarily as a means of assessing the student's performance as the language. Meanwhile, Valette (1977:3) argues that testing is a topic of concern to language teachers, both those in the classroom and those engaged in administration or research. As said by Oller (1979:1) a language test is a device that tries to assess how much has been learned in a foreign language course or some parts of a course. A test is a set of questions, each of which has a correct answer, that examines usually answer orally or in

writing (Tinambuan,1988:3).

Based on Gronlund (1976:6), evaluation may be defined as a systematic process of determining the extent to which instructional objectives are achieved by pupils. Measurement is the processes of assigning numbers to individuals or their characteristics according to specific rules (Ebel and Frisbie,1979:25).

From the statements above, the writer can conclude that test is a device to asses the student's ability in teaching learning process. Through a test, teachers can get information about students achievement. Evaluation is one of the activities to measure and assess the level of achievement of students. It is important to have a good evaluation. Measurement is a method to assess the student's based on the rules.

2.2 Criteria of a Good Test

A test can direct students attention either toward or away from the objectives of instruction. A test can encourage students to focus on limited aspect of the course content or direct their attention to all important areas. The result of a test can be functioned in educational decisions. Only a good test that can be used to sustain or enhance class morale and learning. Based on Brown (2004:19-30), there are five criteria for testing a test: Practicality, reliability, validity, authenticity and washback. Here, the focus is on validity and reliability because the validity and reliability level is very significant. With the reliability and validity of a test, it may be expected that the students' test result can be used for many kinds of necessities.

2.2.1 Validity

Validity is the most important variable of a measurement instrument. Brown (2004:22) states that Validity is the most complex criterion of an effective test and arguably the most important principle. A test's validity may focus on the sequences beyond measuring the criteria themselves of a test, or even the test-taker's perception of validity.

A. Types of validity

Basically, there are many types of validity according to some experts. According to Brown (2004:22-30), validity is divided into five types of evidence:

a. Content-Related Evidence

If a test actually samples the subject matter about which conclusions are to be drawn, and if it requires the test- taker to perform the behavior that is being measured, it can claim content-related validity, often popularly referred to as **content validity** (Mousavi:2002, Hughes:2003 quoted by Brown,2004:22).

b. Criterion-Related Evidence

A second form of evidence of the validity of a test may be found in what is called criterion-related evidence. Also reffered to as **criterion-related validity,** or the extent to which the "criterion" of the test has actually been reached.

c. Construct-Related Evidence

Construct-Related Evidence commonly referred to as **construct validity**. A construct is any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perceptions. Construct validity is a

major issue in validating large-scale standardized tests of proficiency.

d. Consequential validity

Consequential validity encompasses all the consequences of a test, including such consideration as its accuracy in measuring intended criteria, its impact on the preparation of test-takers, its effect on the learner, and the (intended and unintended) social consequences of a test's interpretation and use.

e. Face validity

Gronlund (1998:210) quoted by Brown (2004:26) says an important facet of consequential validity is the extent to which "students view the assessment as fair, relevant, and useful for improving learning." or what is popularly known as **face validity.** "Face validity refers to the degree to which a test *looks* rights, and *appears* to measure the knowledge or abilities it claims to measure, based on the subjective judgement of the examinees who take it, the administrative personnel who decide on its use, and other psychometrically unsophisticated observers" (Brown adapted from Mousavi, 2002:244).

B. Factors Influencing Validity

In the construction of the achievement test, the teachers should be aware of some factors that tend to influence the validity of the results (Tinambuan,1988:13-14).

a. ***Unclear Direction***
b. ***Word and sentence structure of the items should not be too difficult.***
c. ***Inappropriate level of difficulty of the test items.***
d. ***Avoid test items which are poorly constructed.***
e. ***Ambiguity***
f. ***Improper arrangement of the test items.***
g. ***Pattern of answers which is easily identified.***
h. ***Some factors in administering and scoring a test***.

From the explanations above, the writer concludes that validity is one of the

most important criteria of a good test. There are types of validity. First,content-related evidence is concerned about what goes into the test.Second type is criterion-related validity, whenever test scores are used to predict future performance. Third type is construct related-evidence. If a test has construct validity, it is able to measure with specific criteria. Fourth type is consequential validity, one aspect of consequential validity has drawn special attention, test results may reflect conditions of the students being assessed. The last type is face validity, if a test item looks right to other, it can be described as face validity.

2.2.2 Reliability

According to Gronlund (1976:105) Reliability refers to the *consistency* of measurement. That is, how consistent test scores or other evaluation results are from one measurement to another. There are several ways of estimating the reliability of a test. The methods of estimating reliability are as follows:

a. Test-Retest Method.

To estimate reliability by means of the test-retest method the same test is administered twice to the same group of pupils with a given time interval between the two administrations of the test. The resulting test scores are correlated, and this correlation coefficient provides a measure of stability, it indicates how stable the test results are over the given period of time.

b. Equivalent-Forms Method.

Estimating reliability by means of the equivalent-forms method involves the use of two different but equivalent forms of the test (also called parallel or

alternate forms). The two forms of the test are administered to the same group of pupils in close succession and the resulting test scores are related. This correlation coefficient provides a measure of *equivalence*. The equivalent-method is sometimes used with a time interval between the administration of the two forms of the test.

c. Split – Half Method

The test is administered to a group of pupils in the usual manner and then is divided in half for scoring purposes. To split the test into halves which are most equivalent, the usual procedure is to score the even- numbered items and the odd- numbered items separately. This provides two scores two for each pupil which, when correlated, provides a measure of *internal consistency.* This coefficient indicates the degree to which the two halves of the test are equivalent.

d. Kuder- Richardson methods.

Another method of estimating the reliability of test scores from a single administration of a single form of a test is by means of formulas such as those developed by Kuder and Richardson. These formulas also provide a measure *internal consistency* but they do not require splitting the test in half for scoring purposes.

A less accurate but simpler formula to compute is the Kuder-Richardson formula 21. This formula can be applied to the result of any test which has been scored on the basis of the number of correct answer. A modified version of the formula is

$$\text{Reliability Estimate (KR21)} = \frac{K}{K-1}\left[1 - \frac{M(K-M)}{Ks^2}\right]$$

Where K = the number of items in the test

M = the mean (arithmetic average) of the test scores

S = the standard deviation of the test scores

In short, in order to be reliable, a test must be consistent in measurement. It means that in scoring the teacher has to be consistent. He/ she may not change the standard of scoring easily. The methods of estimating reliability are: first, test-retest method is a method to measure stability by giving the same test twice to the same group periodically. Second, equivalent-forms method is a method to measure of equivalence, it is held by giving two forms of the test to the same group in close succession. Third method is split-half method, it gives test once and score two equivalent halves of test. The last method is Kuder-Richardson method gives test once and score total test and apply Kuder-Richardson formula.

Considering the characteristics of a good test, there are five elements that have important roles. They are practicality, reliability, validity, authenticity and washback. In this section, the focus is on validity and reliability because both are very essential to effective testing and should be understood by anyone working with the test. The most important variable in judging the adequacy of measurement is validity. Every test should be as valid as the teacher can make it. Reliability is the extent to which the result can be considered consistent. A test should be reliable because unreliable test might produce different scores if it is taken again.

2.3 Types of Test

Tests motivate and direct student learning because tests guide student learning and help determine how students will prepare for a test. According to Vallete (1977:5-6) there are four types of test. They are:

a. The aptitude test

The aptitude test is conceived as a prognostic measure that indicates whether a student is likely to learn a second language readily.

b. The progress test

The progress test measures how much the student has learned in a specific course of instruction.

c. The achievement test

The achievement test is similar to the progress test in that it measures how much the student has learned in the course of second language instruction.

d. The proficiency

The proficiency test also measures what students have learned, but the aim of the proficiency test is to determine whether this language ability corresponds to specific language requirements.

According to the definitions, the writer concludes that types of test are aptitude test measures the student's performance in learning a foreign language, progress test measures the progress in mastering the material taught in the classroom. Another type is achievement test, it is to measure what has been taught and learnt. Proficiency is the last type, it measures the students' achievement to perform.

2.4 Achievement Test

Achievement test plays an important role in all types of instructional program. Achievement test emphasizes past progress, whereas aptitude test primarily concerns with future potentialities. The primary goal of the achievement test is to measure past learning, that is, the accummulated knowledge and skills of an individual in a particular field or fields. According to Gronlund (1982:1) an achievement test is a systematic procedure for determining the amount a student has learned. Brown (2004:47) states that an achievement test is related directly to classroom lessons, units, or even a total curriculum.

2.4.1 Types of Achievement Test

Tinambuan (1988:7-9) says there are four types of achievement test which are very commonly used by teachers in the classroom:

a. Placement test

A placement test is designed to determine the pupil performane at the beginning of instruction.

PERPUSTAKAAN
UNNES

b. Formative test

Formative test is intended to monitor learning progress during the instruction and to provide continous feedback to both pupil and teacher concerning learning successes and failures.

c. Diagnostic test

Diagnostic test is intended to diagnose learning difficulties during instruction.

d. Summative test

The summative test is intended to show the standard which the students have now reached in relation to other students at the same stage.

2.4.2 Basic Principles of Achievement Test

Gronlund (1982:8-13) says basic principles of achievement test are:

a. Achievement test should measure clearly defined learning outcomes that are in harmony with the instructional objectives.
b. Achievement tests should measure a representative sample of the learning tasks included in the instruction.
c. Achievement tests should include the types of test items that are most appropriate for measuring the desired learning outcomes.
d. Achievement tests should fit the particular uses that will be made of the results.
e. Achievement tests should be as reliable as possible and should then be interpreted with caution.
f. Achievement tests should improve student learning.

Based on the statements, the writer defines that achievement test is used for assesing present knowledge and abilities. There are some types of achievement test, they are: placement test is concerned to evaluate pupil performance in a sequence of instruction, formative test is used to monitoring pupil progress, diagnostic test is designed to evaluate the students' difficulties, and summative test is used to determine pupil achievement.

PERPUSTAKAAN UNNES

2.5 Multiple - choice Test Item

Heaton (1975:14) says that multiple-choice is now widely regarded as being one of the most useful of all objective item types. Although it is among the most

difficult of all objective item types to construct, it is simple to score and administer. Valette (1977:6) states that multiple choice test items are designed to elicit specific responses from the students.

Valette (1977:7) defines that, "in a multiple-choice item, the stem is the initial part: either a partial sentence to be completed, a question, or several statements leading to a question or incomplete phrase. The choices from which students must select their answers are known as *options, responses, or alternatives.* One response, the *key*, is distinctly correct or more suitable than the others. Incorrect responses are called *distractors* and should be so worded that they seem attractive to the uniiformed student. If a distractor is so obviously wrong that it is never selected, it plays no useful role in the item and should be eliminated or replaced by a new alternative."

2.5.1 The advantages of using the multiple- choice form

Tinambuan (1988:75-76) states that the advantages of using the multiple- choice form are:

a. The multiple-choice item is adaptable to subject matter content areas as well as different levels of behaviour.
b. The structure of premise with four or five alternatives provides less chance for guessing the correct response than the true-false item does.
c. Four or five options in the multiple-choice test provide more incorrect choices for selections of responses by the student who does not know the best or correct answer.
d. The difficulty of each multiple-choice item can be controlled by changing the alternatives.
e. Multiple-choice items are amenable to item analysis which enables the teacher to determine how well the items functioned with the students tested and how well each alternative functioned in discriminating between the higher achieving and lower achieving students.

2.5.2. Suggestions for writing multiple - choice item

Gronlund (1985) as stated by Tinambuan (1988:81-85) suggests how to construct multiple-choice items:

a. **The stem of the item should be meaningful by itself and should present a definite problem.**
b. **The stem should be free of irrelevant material.**
c. **Use a negatively stated item stem only when significat learning outcomes require it.**
d. **An item should contain only one correct or clearly best answer.**
e. **Items used to measure understanding should contain some novelty.**
f. **All distractors should be palausible.**
g. **Use sparingly alternatives such as 'all of the above' or 'none of the above'.**

Usually the achievement test that is held in many schools or other institutions are in the form of multiple-choice items because it can measure both knowledge outcomes and various types of intellectual skills. Since there are many advantages of using multiple-choice item, it has been the common way to measure students' achievement for years because the materials that are tested in the multiple choice items are relatively much and unlimited than other test forms. It can measure many levels of cognitive from memorizing until evaluation, it can be scored easily, faster and objective. The multiple-choice item consists of two parts. They are: stem or lead, which is either a direct question or incomplete statement. The students will have to answer or complete one alternative. Alternatives may consist of two or more choices or responses of which one is the answer and the others are distracters, that are, the incorrect responses. The function of distractors is to distract those students who are uncertain of the answer.

PERPUSTAKAAN UNNES

2.6 Item Analysis

The analysis of students response to objective test items is a powerful tool for test improvement. Ebel (1991:225) defines that Item analysis can indicate which items may be too easy or too difficult and which may fail, for whatever reasons, to

discriminate properly between high and low achievers. Item analysis begins after the test has been scored. According to Ebel (1991:225) a classroom teacher who chooses to complete the procedures by hand would follow these six steps:

a. Arrange the scored test papers or answer sheets in score order from highest to lowest.

b. Identify an upper group and a lower group separately. The upper group is the highest scoring 27 percent (one- fourth) of the group and the lower group is an equal number of the lowest scoring of the total group.

c. For each item, count the number of examinees in the upper group that choose each response alternative. Do a separate, similar tally for the lower group.

d. Record these counts on a copy of the test at the end of the corresponding response alternatives. (The use of colored pencils is recommended.)

e. Add the two counts for the keyed response and divide this sum by the total number of students in the upper *and* lower groups. Multiply this decimal value by 100 to form percentage. The result is an *estimate* of the index of item difficulty.

f. Substract the lower group count from the upper group count for the keyed response. Divide this difference by the number of examinees in one of the groups (either group since both are the same size). The result, expressed as a decimal, is the index of discrimination.

According to Gronlund (1981: 151-160) there are some criteria to determine which items that still can be used, revised or should be discarded.

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

a. An item is used if it has the following criteria.

1. Valid, reliable, good discrimination power and moderate difficulty level.
2. Valid, reliable, satisfactory discrimination power and moderate difficulty level.

b. An item is used with several revisions if it has the following criteria.

1. Valid, reliable, good discrimination power but the difficulty level is too easy.
2. Valid, reliable, satisfactory discrimination power but the difficulty level is too difficult and too easy.
3. Valid, reliable, poor discrimination power and moderate difficulty level.
4. Not valid, reliable, good discrimination power and moderate difficulty level.
5. Not valid, reliable, satisfactory discrimination power and moderate difficulty level.

c. An item should be discarded if it has the following criteria.

1. Valid, reliable, poor discrimination power and the difficulty level is too difficult or too easy.
2. Not valid, reliable, good discrimination power, but the difficulty level is too difficult or too easy.
3. Not valid, reliable, satisfactory discrimination power, but the difficulty level is too difficult or too easy.

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN UNNES

4. Not valid, reliable, poor discrimination power, and moderate difficulty level.
5. Not valid, reliable, poor discrimination power, and the difficulty level is too difficult or too easy.

2.6.1. Item Discrimination Power

The discriminating power of a test item is its ability to differentiate between pupils who have achieved well (the upper group) and those who have achieved poorly (the lower group). To estimate item discriminating power is by comparing the number of students in the upper and lower group who answered the item correctly.

According to Gronlund (1982:103) the computation of item discriminating index (D) for each item can be done by subtracting the number of students in the lower group who get the item right (L) from the number of students in the upper group who get the item right (U) and divided by one half on the total number of students included in the item analysis (1/2 T).

Tinambuan (1988:145) says that the discrimination index can take values from 0.00 +1.00. The higher the D value for an item, the better that item discriminated. Any item which has a D value of +0.40 or above is considered to be good in discriminating student differences. D values between +0.20 and +0.39 are usually considered to be satisfactory, but items with lower values in this range should be reviewed and revised to make them more effective discriminators.

2.6.2. Index Difficulty

The more difficult the item is, the fewer the students who select the correct option. The easier the item is, the more students select the correct one.

Nitko (1983:288) says that the index of item difficulty (p) can be computed by using two ways. First, dividing the number of students who answered an item correctly (R) with the total number of students tested (T), then multiply by one hundred. The second way is based on the upper group and the lower group only, by assuming the responses of students in the middle group would follow essentially the same pattern.

Heaton (1975: 172) states that the index of difficulty will run from 0.00 with 0.70 – 1.00 indicating the easiest possible item. The index of difficulty that runs from 0.30 – 0.70 is considered moderate item or acceptable item. The hardest will run from 0.00 – 0.30. In other word, the percentage answering the item correctly, the smaller the percentage figures the more difficult the item.

Item analysis covers the steps of an item analysis, item discrimination power and index difficulty. There are six steps of an item analysis and there are some criteria to determine the items. Item discrimination tells how well each item discriminates between better students from the the poorer students. The item of difficulty is the percentage of students who got each question right.

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

CHAPTER III

METHOD OF INVESTIGATION

In the third chapter, the writer would like to discuss population, samples, identification of the problems, techniques of data collection, and techniques of data analysis.

3.1 Population and Sample

Margono (2003:118) said that a population is defined as a complete set of individuals or subject having common observable characteristics. The population is the establishment of boundary condition that specify who shall be included in or excluded from the population (Tuckman, 1978:117).

The population of this study was the seventh year students of SMP N I Moga Pemalang in the academic year of 2008/2009. It has five classes of seventh year students. In order to study population more effectively, she had to select fifty students from the population as sample in this study.

3.2 Sampling Technique

In order to make this study effective, researcher had to select sample. Sample is part of population, which represent the population.

According to Brink (1974:33), random sampling refers to the process of drawing a random sample of individuals of some population. In this study, the

writer used random sampling technique to take samples. In the random sampling technique, each number has an equal chance of being selected for the sample. The writer took 10 students of each class to be taken as samples. So there were fifty students taken from five different classes.

In selecting 10 students of each class, the writer took the procedure called lottery method. This method is an objective selection. She did it by writing down the order number of the student's names list on a small piece of paper, and then the pieces of paper was rolled and let the ten rolls of paper drop out of the glass one after another. Although the steps in taking samples are very simple, many researchers have to adapt the random sampling techniques as one way to select the samples since it is not influenced by thoughts and feelings.

3.3 Identification of the Problems

The fact of the analysis result shows that the most of teacher SMP and SMU still do not know how to construct a good test. They do not know about the characteristics of the quality of a good test.

Based on that fact, there are four problems related to the English test items. The problems are:

a. The validity level
b. the reliability level
c. The difficulty level
d. The discrimination power

3.4 Technique of Collecting Data

The technique of data collection in this study involves several steps, those are:

a. Copying the answer sheet in final test of the subject.

b. Collecting the item test and analyzes the result of the item test to know the validity and reliability.

c. Finally, the writer computes the data.

3.5 Technique of Data Analysis

The data to be analyzed in this study were taken from the students' answer sheet in final test for the seventh grade of SMP students in the academic year 2008/2009. These students' answer sheets were used to analyze the quality of the test items. From this item analysis, we would know that the item is too easy or too hard. Here, the item analysis consists of the following:

a. Difficulty level analysis

A good test item is an item which is not too difficult or too easy. The difficulty of the test items is the percentage of students who get the right items. Here the index of item difficulty level (P) used Nitko formula to analyze. Then, the writer divided the level of items difficulty (P) into three categories. The criteria of item difficulty level could be seen in the table below:

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN UNNES

No	Index Difficulty level (P)	The categories
1.	0.00- 0.30	Difficult
2.	0.30- 0.70	Moderate
3	0.70- 1.00	Easy

(Heaton,1975:172)

The formula is:

$$P = \frac{R}{T}$$

Where: P = difficulty level or index of difficulty.

R = the number of students who respond correctly to an item.

T = the total number of students who respond to the item.

(Nitko, 1983:288)

b. Discrimination power analysis

The discrimination power of the test items tells how well the item performs in separating the upper group and the lower group. The formula to compute item discrimination power is as follows

$$D = \frac{RU - RL}{1/2T}$$

Where: D = the index of discrimination power

RU = the number of students in the upper group who answer the items correctly

RL = the number of students in the lower group who answer the items

correctly

½ T = one half of the total number of students included in the items analysis.

(Gronlund, 1982:103)

The criteria of item discrimination power could be seen in the table below which is proposed by Ebel and Frisbie (1991: 232) are:

Discrimination index	**Item evaluation**
0.70≤ DP≤ 1.00	Excellent
0.40< DP≤ 0.70	Good
0.20< DP≤ 0.40	Satisfactory
0.00< DP≤ 0.20	Poor

c. Analysis of reliability

In this study, the writer uses the Kuder- Richardson formula in estimating the reliability of the test. Kuder Richardson formula is:

$$\text{Reliability Estimate (KR21)} = \frac{K}{K-1}\left[1 - \frac{M(K-M)}{Ks^2}\right]$$

Where K = the number of items in the test

M = the mean (arihmetic average) of the test scores

S = the standard deviation of the test scores

The formula of Standard deviation is:

$$S = \frac{\text{sum of high sixth} - \text{sum of low sixth}}{\text{half the number of students}}$$

d. Analysis of validity

Validity refers to whether or not a test measures what it is supposed to measure. In this study, the writer used criterion- related validity (validity coefficient to determine whether the test items are valid or not). To find the coefficient validity, she used the Pearson Product – Moment formula. The formula is as follow:

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma Y^2 - (\Sigma Y)^2\}}}$$

rxy : correlation index
x : the score of each item
y : the total score
n : the total number of the respondent
∑ : the sum of the score

There are two ways to determine the value of correlation coefficient. First, by interpreting the r value of calculation with the following criteria:

0.81≤ rxy≤ = the test has very high validity

0.610 ≤ rxy≤ = high validity

0.410 ≤ rxy ≤ = moderate validity

0.210 ≤ rxy ≤ = very low validity

Second, by consulting the r (product moment) value on the table. The table is listed on the appendix 6. According to Arikunto (2005:75), if the r value of calculation is lower than the critical value on the table, it means that the correlation is not significant or we can say that the test item is not valid and so vice versa.

CHAPTER IV

ANALYSIS AND DISCUSSION

4.1 Result of the Analysis

The goal of this study is to analyze items of the English final test for the seventh grade SMP students in the academic year of 2008/2009. The analysis consists of four aspects, namely the difficulty level, discrimination power, validity and reliability of the test.

Item analysis aims to identify good, moderate, and poor items. Through item analysis, we get information about the shortcomings of the items and how to revise them. From data analysis of the English final test for the seventh students of SMP N 1 Moga Pemalang in the academic year 2008/2009, the writer obtained the following results.

PERPUSTAKAAN UNNES

3.1.1. Analysis of Validity

The Pearson's product moment formula is used to calculate the validity level of the test items since the value of r calculation is more than the r table ($r_c > r_t$), the item is valid and vice versa. For N= 50 with the significance level 0.05, the value of r on the table is 0.279 (see appendix 2). From the validity calculation, the writer got the results as follows.

No.	Criteria	Number	Percentage
1.	Valid Items	29	72.5%
2.	Invalid Items	11	27.5%

a. There are 29 test items, which fulfill the requirements of validity. They are items number 1,3,4,5,6,8,10,11,13,15,16,19,20,21,23,24,25,26,30,31,32,33, 34,35,36,37,38,39 and 40.

b. There are 11 test items, which do not fulfill the requirement of the validity. They are items number 2,7,9,12,14,17,18,22,27, 28, and 29.

From the table above, we found that the validity of the items is 0.3162. The example of the computation of item validity is listed in appendix 2.

3.1.2. Analysis of Reliability

As the writer has stated in the previous chapter, the coefficient of reliability of test items is found by applying the Kuder- Richardson 21 formula. From the computation, it is found that the coefficient of the test items is 0.751. The result is then consulted to the table of r product moment values at level of significance of 0.05. It is found that the value of r 0.279 for N 50. Since value of r calculation is more than that of table (r), it can be concluded that the test items used in English final test for seventh grade students in SMP N 1 Moga Pemalang in academic year 2008/2009 is reliable. The computation of the reliability coefficient is listed in apendix 3.

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN UNNES

3.1.3. Analysis of Difficulty Level

The difficulty of the test items is the percentage of students who got the right items. Here the index of item difficulty level (P) used Nitko formula to analyze. Then, the writer divided the level of items difficulty (P) into three categories. The criteria of item difficulty level could be seen in the table below:

No	**Index Difficulty level (P)**	**The categories**
1.	0.00- 0.30	Difficult
2.	0.31- 0.70	Moderate
3	0.71- 1.00	Easy

From the table in appendix 4, the result of the data analysis of the item Difficulty Level shows as follows.

No.	Criteria	Number	Percentage
1.	Difficult Items	4	10%
2.	Moderate Items	20	50%
3.	Easy Items	16	40%

In detail, the result of the data analysis is as follows.

a. Items that can be categorized as difficult are 8, 17, 23, and 40.

b. Items that belong to the moderate category are 1,2,4,5,10,11, 14,18,19,20,21, 22,28,33,34,35,36,37,38 and 39.

c. The following items are categorized as easy 3,6,7,9,12,13,15, 16, 24,25,26,27, 29, 30, 31, and 32.

From the test, the mean of their difficulty level is 0.60. So, the English final test items are classified as Medium items in terms of their difficulty level. Items, which are considered very easy, can still be used in a test to encourage and motivate the poor students. The example of computation of item difficulty level is listed in appendix 4.

3.1.4. Analysis of Discrimination Power

The discrimination power of the test items tells how well the item performs in separating the upper group and the lower group. The writer divided the discrimination power of the test items into four categories. The criteria are as follows:

Discrimination index	**Item evaluation**
0.70≤ DP≤ 1.00	Excellent
0.40< DP≤ 0.70	Good
0.20< DP≤ 0.40	Satisfactory
0.00< DP≤ 0.20	Poor

From the table in appendix 1, the result of data analysis can be seen on this table.

No.	Criteria	Number	Percentage
1.	Excellent Items	-	-
2.	Good Items	4	10%

3.	Satisfactory Items	24	60%
4.	Poor Items	12	30%

a. The items that belong to the good items are 1,5,11, and 33.

b. Items number2,6,8,13,14,15,18,19,20,21,23,24,25,26,28,30,31,32,34,35,36, 38,39,and 40 are categorized as satisfactory items.

c. Items number 3,4,7,9,10,12,16,17,22,27,29, and 37 are classified as poor items.

From the 40 test items, the mean of the item discrimination power value is 0.48. So as the whole the final test can be categorized as a good item, it means that this final test items can be used in final test. In this study, there is an item that has zero value of discrimination power. An item stated above is item number 17. This item is found with a discrimination power value 0.00. That item should be discarded.

4.2 Discussion

PERPUSTAKAAN UNNES

The goal of the writing of this fnal project is to identify the quality of each item, whether it can be classified as good, moderate, or poor item. It later can be determined which items can still be used, can be used with revision, or should be dropped. From the point of view of difficulty level, a good item is an item, which is not too easy or not too difficult. From the discrimination power point of view, a good item is an item that can be discriminate between the students from the upper group and the students from the lower group.

Based on the result of item analysis which includes the analysis of difficulty level, discrimination power, validity and reliability of the items (see appendix 1), this test items can be used in English final test with several revision. For more details will be explained as follows.

- Item number 1

Question	The text is a/ an . . . A. Announcement B. Greeting Card C. Invitation Card D. Short Message			
Result	**A**	**B**	**C**	**D***
Upper 27%	-	2	-	12
Middle 46%	3	8	-	12
Lower 27%	-	4	3	6
Total	3	14	3	30
Validity	Valid			
P value	0.60			
D value	0.48			

This item is moderate in terms of the difficulty level and it is good in discrimination power. This item is valid because instruction in the item is clear. If we look at the table above, we can see that option B is destructive, because it attracted more students than other option. In the writer's opinion this item can be used in final test.

- Item number 2

Question	What is the content of the text above? a. Berta celebrates her party

	b. Berta's birthday party c. Berta congratulates Sandra d. Berta invites Sandra			
Result	**A**	**B**	**C***	**D**
Upper 27%	-	-	12	2
Middle 46%	2	3	13	5
Lower 27%	1	1	9	2
Total	3	4	34	9
Validity	Invalid			
P value	0.68			
D value	0.24			

This item is moderate in terms of the difficulty level and satisfactory in terms of discrimination power. There were 12 students of the upper group who answered it correctly, and there were 9 students of the lower group who did the same. If we look at the table above, option D is said to be a good distractor, there were students choose this option than option A and B. This item has a good distribution in terms of students' response, but if we see from the validity, this item is invalid. This item needs many revisions, this option in this item should be answered easily so it can differentiate the upper group and the lower group.

- Item number 3

Question	Who wrote the SMS? a. Firda

	b. My mom c. Dinda d. Doctor			
Result	**A***	**B**	**C**	**D**
Upper 27%	13	-	1	-
Middle 46%	16	-	5	2
Lower 27%	10	-	3	-
Total	39	-	9	2
Validity	Valid			
P value	0.78			
D value	0.20			

This item is an easy one, because 39 students answered it correctly, while the difference of the upper group students and the lower ones who chose the correct answer is only 3. This item is inappropriate level of difficulty of the test item, this item is too easy. It causes this item has low discrimination power and poorly constructed, so the students can guess the answer easily. In writer's opinion this item should be discarded.

PERPUSTAKAAN UNNES

- Item number 4

Question	Who was sick? a. Dinda

	b. Dinda's mom c. Firda d. Firda's mom			
Result	**A**	**B**	**C**	**D***
Upper 27%	-	1	3	10
Middle 46%	2	2	14	5
Lower 27%	1	3	6	3
Total	3	6	23	18
Validity	Valid			
P value	0.36			
D value	0.16			

This item is moderate in terms of the difficulty level and poor in terms of discrimination power. From the table above we find out that 18 students chose the correct answer. Option C seems to be destructive, since more students choose this option than the right option. This item has low discrimination power. In the writer's opinion this item can be used in final test with several revisions, the text in this item should be understood by the students, so it is not too difficult to answer.

PERPUSTAKAAN UNNES

- Item number 5

Question	Firda sent the SMS because . . . a. She wanted to see a doctor

	b. She wanted accompany Dinda's mother c. She wanted to say sorry to Dinda d. She wanted to go to Dinda's house			
Result	**A**	**B**	**C***	**D**
Upper 27%	5	-	9	-
Middle 46%	12	3	7	1
Lower 27%	5	6	0	2
Total	22	9	16	3
Validity	Valid			
P value	0.32			
D value	0.48			

The P value of this item shows that it is moderate item and good in terms of discrimination power. If we see the option, option A is destructive, because the number of students who chose this option is larger than the correct answer. There were 9 students from upper group and there wasn't student from lower group. There is quite a large difference between these two groups, where the upper is larger than the lower one, which means it has good discrimination power. From the table above there is good distribution of students' answer, which shows that the item has good distractor. This item doesn't need any revision.

- Item number 6

Question	Where will the basket ball tournament be held? a. in field

	b. in the yard c. in the school d. in Kridanggo			
Result	**A**	**B**	**C**	**D***
Upper 27%	-	-	-	14
Middle 46%	-	-	1	22
Lower 27%	-	1	3	9
Total	-	1	4	45
Validity	Valid			
P value	0.90			
D value	0.28			

Item number 6 up to 9 is reading test; it is about reading an announcement. The vocabulary is familiar for students in the seventh grade students. Item number 6 is considered to be easy in terms of difficulty level and satisfactory in discrimination power, but it is valid in terms of the validity. It means that this item can be used in the final test with several revisions. In here, the text can be understood easily by the students. The writer suggests to give an announcement that is not too easy.

PERPUSTAKAAN UNNES

- Item number 7

Question	What time will the tournament start? a. At seven

	b. At eight c. At nine d. At ten			
Result	**A**	**B**	**C***	**D**
Upper 25%	-	-	14	-
Middle 46%	1		19	3
Lower 27%	-	-	12	1
Total	1	-	45	4
Validity	Invalid			
P value	0.90			
D value	0.20			

With a P value of 0.90 this item is classified as easy, because 45 of 50 students chose the correct answer. In terms of discrimination power, this item is considered poor, because the difference between the upper group and the lower group is only two. If we see from the validity, this item is invalid because of the answer can be identified. The students can guess the answer easily. This item should be discarded.

PERPUSTAKAAN UNNES

- Item number 8

Question	The announcement is for . . . a. All people

	b. girls c. boys d. students			
Result	**A***	**B**	**C**	**D**
Upper 27%	8	-	2	4
Middle 46%	5	3	4	11
Lower 27%	2	1	3	7
Total	15	4	9	22
Validity	Valid			
P value	0.30			
D value	0.28			

This item is said to be difficult because only 15 students chose the correct answer. The rest of the students chose the other option. If we look at the table above, we can see the distractors are good, especially option D can attract more students than other option. From the aspect of discrimination power, the item is considered as satisfactory item. From the P value is 0.30 which means that it is a difficult item. In the writer's opinion, this item can be used in final test.

PERPUSTAKAAN
UNNES

- Item number 9

Question	The synonym of the word "tournament" is . . . a. lesson

	b. yard c. field d. competition			
Result	**A**	**B**	**C**	**D***
Upper 27%	1	-	-	13
Middle 46%	1	-	2	20
Lower 27%	-	-	1	12
Total	2	-	3	45
Validity	Invalid			
P value	0.90			
D value	0.12			

This item is an easy in terms of the difficulty level, and poor in terms of discrimination power. From the table above, we find out that 45 students chose the correct answer. Meanwhile, the difference between those who answered it correctly from the upper group and the lower group is only 1. This item should be deleted, since it has no discrimination power.

PERPUSTAKAAN UNNES

- Item number 10

Question	The notice suggest us not to . . . a. Make noise

	b. Park here c. Sit on the bench d. Walk on the grass			
Result	**A**	**B**	**C***	**D**
Upper 27%	1	4	8	1
Middle 46%	2	4	12	5
Lower 27%	2	5	2	4
Total	5	13	22	10
Validity	Invalid			
P value	0.46			
D value	0.20			

This item is not too easy and not too difficult. It can be classified as moderate in terms of the difficulty level, since there are 22 students chose the correct answer. From the table above we can see that option B and D are good distractors because they attract more students than option A. In terms of discrimination power, this is said to be poor. This item has a good distribution in terms of students' response, but if we see from the validity, this item is invalid. This item should be deleted.

PERPUSTAKAAN
UNNES

- Item number 11

Question	Mrs. Ketty wants to buy . . . a. Drink and fruit

	b. Food and drink c. Snack and vegetables d. Fruit and vegetables			
Result	**A**	**B**	**C**	**D***
Upper 27%	-	-	1	13
Middle 46%	3	2	4	14
Lower 27%	3	2	4	4
Total	6	4	9	31
Validity	Valid			
P value	0.62			
D value	0.68			

This item is considered to be moderate in terms of difficulty level, even though it is a good item in terms of discriminating power because 31 out of 50 students answered the item correctly. It means that the item is not too easy and too difficult because more than 50% of the students chose the correct answer. The option should be written in capitalize letter. From those 31 students, there were 13 students from the upper group and 4 students from the lower group. There is quite a large difference between those two groups, where the upper is larger than the lower group, which means it has good discrimination power. The item doesn't need to be revised and can be applied in the final test.

- Item number 12

Question	How much is 2kg s of bean? a. Ten thousand rupiahs

	b. Eleven thousand rupiahs c. Nine thousand rupiahs d. Five thousand rupiahs			
Result	**A***	**B**	**C**	**D**
Upper 27%	11	-	-	3
Middle 46%	17	-	1	5
Lower 27%	9	-	2	2
Total	37	-	3	10
Validity	Invalid			
P value	0.74			
D value	0.04			

This item belongs to an easy group in terms of the difficulty level because 37 out of 50 students answered it correctly while in terms of discrimination power, it is a poor item since the difference between those who answered it correctly from the upper group and the lower one is only 2 students. Option D is a good distractor because it attracts more students than option B and C. In the writer's opinion, this item should be rejected or discarded.

PERPUSTAKAAN
UNNES

- Item number 13

Question	How many kg s of bean will she buy? . . .kg s a. one

	b. two c. three d. four			
Result	**A***	**B**	**C**	**D**
Upper 27%	14	-	-	-
Middle 46%	22	1	-	-
Lower 27%	8	4	1	-
Total	44	5	1	-
Validity	Valid			
P value	0.88			
D value	0.32			

Item number 13 is considered to be easy in terms of difficulty level and satisfactory in discrimination power, but it is valid in terms of the validity. It means that this item can be used in the final test with several revisions, this item should be too difficult so the students do not answer easily. It can be changed with "How many g s of onions will she buy?. . kg s.

PERPUSTAKAAN UNNES

- Item number 14

Question	'Onion' means . . . a. Bawang merah

	b. Bawang putih c. slada d. sawi			
Result	**A***	**B**	**C**	**D**
Upper 27%	11	2	-	-
Middle 46%	15	9	-	-
Lower 27%	7	5	1	
Total	33	16	1	-
Validity	Invalid			
P value	0.66			
D value	0.28			

With reference to the P and D values, this item is said to be moderate and satisfactory. We can see that option B (incorrect) is distractor because it attracted more students than option C and D. In writer's opinion this item needs many revisions. In here, the students confuse to answer between option A and B. The writer suggests to change the options in this item.

PERPUSTAKAAN UNNES

- Item number 15

Question	What is the announcement about? a. To clean the class

	b. To join class meeting c. To vote the leader d. To join extracurricular			
Result	**A**	**B***	**C**	**D**
Upper 27%	-	14	-	-
Middle 46%	-	23	-	-
Lower 27%	1	9	-	3
Total	1	46	-	3
Validity	valid			
P value	0.92			
D value	0.24			

The number of students who answered this item correctly is 46, which consists of 14 students from upper group and 9 students from the lower group. We can conclude that the item is easy in terms of difficulty level and satisfactory in terms of discrimination power. Considering the criteria of difficulty level, discrimination power and the validity this item can be used in final test.

PERPUSTAKAAN UNNES

- Item number 16

Question	Where do you think you can hear the announcement?

	a. At school b. On television c. On the radio d. At home			
Result	**A***	**B**	**C**	**D**
Upper 27%	14	-	-	-
Middle 46%	23	-	-	-
Lower 27%	10	2	1	-
Total	47	2	1	-
Validity	Valid			
P value	0.94			
D value	0.20			

This item is easy in terms of the difficulty level because there 47 out of 50 students answered the question correctly, even though, in terms of discrimination power, this item is considered poor, because the difference between the upper group and the lower one is four. This item has low discrimination power. This item should be deleted, because it has low discrimination power.

PERPUSTAKAAN UNNES

- Item number 17

Question	The synonym of "join" is. . . a. together

	b. follow c. in group d. alone			
Result	**A**	**B***	**C**	**D**
Upper 27%	3	5	6	-
Middle 46%	5	5	12	1
Lower 27%	2	4	6	1
Total	10	14	24	2
Validity	Invalid			
P value	0.28			
D value	0.00			

The P value shows that the item number 17 is difficult since only 14 students answered it correctly, while more than 60% chose incorrect answer. Option A and C are good distractors than option D. The distribution is good in terms of students' response. Meanwhile, the difference between those who answered it correctly from the upper group and the lower group is only 1. And from discrimination power there is no discrimination between those two groups, because the D value is 0. This shows that this item is poor. In writer's opinion, this item should be discarded.

- Item number 18

Question	The underlined word refers to . . . a. teachers

	b. head masters c. students d. librarians			
Result	**A**	**B**	**C***	**D**
Upper 27%	-	-	12	2
Middle 46%	4	-	15	4
Lower 27%	5	-	5	3
Total	9	-	32	9
Validity	Invalid			
P value	0.24			
D value	0.64			

This item is moderate in terms of the difficulty level and satisfactory in terms of discrimination power. Option A and D are good distractors than option B. Considering the criteria on the difficulty level and the discrimination above, this item can be used in final test with several revisions. The writer suggest to revise the text in order to be valid.

PERPUSTAKAAN UNNES

- Item number 19

Question	We usually find this notice . . . a. At library

	b. At laboratory c. At teacher's room d. At school yard			
Result	**A***	**B**	**C**	**D**
Upper 27%	11	3	-	-
Middle 46%	3	13	7	-
Lower 27%	3	9	1	-
Total	17	25	8	-
Validity	Valid			
P value	0.34			
D value	0.28			

Up to this item, we can see from the table that this item is moderate in terms of the difficulty level and satisfactory in terms of discrimination power. Option B and C seem to be destructive, since these options attract more students than the correct option. From the criteria of the difficulty level, the writer suggests that this item could be applied in the test.

PERPUSTAKAAN UNNES

- Item number 20

Question	Today is 31st December and tomorrow will be the 1st January we usually say . . .

	a. Happy birthday b. Congratulation c. Merry Christmas d. Happy New Year			
Result	**A**	**B**	**C**	**D***
Upper 27%	2	-	2	10
Middle 46%	3	2	1	17
Lower 27%	3	2	3	5
Total	8	4	6	32
Validity	Valid			
P value	0.64			
D value	0.32			

From the table above, we can see those 32 students who chose option D consisted of 10 students from upper group and 5 students from the lower one. The D value shows us that, this item is satisfactory, since there is a difference in the number of students of the upper group and those of the lower group who answered it correctly. In the writer's opinion, this item could be applied in the test.

PERPUSTAKAAN
UNNES

- Item number 21

Question	Your friend borrows your English book. He/ she forgets to return it to you.

	What SMS you will to her/ him. a. Buy me an English book. OK? Thanks. b. I'll have a test. Bring my math's book. Thanks. c. Bring my English book tomorrow. I need it. Thanks. d. Come to my house. We have to talk. Thanks.			
Result	**A**	**B**	**C***	**D**
Upper 27%	1	-	12	1
Middle 46%	5	2	8	8
Lower 27%	1	2	3	7
Total	7	4	23	16
Validity	Valid			
P value	0.46			
D value	0.36			

With a P value of 0.46 this item is classified as moderate because 23 out of 50 students chose the correct answer. Meanwhile, from the D value the item is classified as satisfactory item, since 12 students from the upper group and 3 students from the lower group chose option C as the right option. The distribution is good in terms of students' response. If we look at the table, we can see that the distractors are good; especially option D can attract more students than other option. In writer's opinion this item doesn't need any revisions.

- Item number 22

Question	Seven – I – to – go – usually – school – at 1 2 3 4 5 6 7

	The correct arrangement of the above words is. . . a. 2 5 4 3 6 7 1 b. 2 4 3 6 5 7 1 c. 7 1 4 2 3 6 5 d. 7 1 5 2 3 4 6

Result	**A***	**B**	**C**	**D**
Upper 27%	9	2	-	3
Middle 46%	8	8	4	3
Lower 27%	4	4	2	3
Total	21	14	6	9
Validity	Invalid			
P value	0.42			
D value	0.20			

The P value shows that this item is considered to be moderate, since 21 students chose it. But the D value shows it is a poor item, since there were 9 students of the upper group who answered it correctly, and there were only 4 students of the lower group who did the same. It means that the item has low discrimination power. If we see from the validity and the discrimination power above, this item should be discarded.

- Item number 23

Question	1. Bob : Great, thanks. And you? 2. Bob : he is getting well. 3. John: Hi, Bob. How is life? 4. John: Fine, too. By the way, how is your father? 5. John: Oh, thanks God. The correct arrangement is . . . a. 5 2 3 4 1 b. 3 1 4 2 5 c. 3 2 1 4 5 d. 5 1 4 2 3			
Result	**A**	**B***	**C**	**D**
Upper 27%	-	9	5	-
Middle 46%	2	3	14	4
Lower 27%	3	2	7	1
Total	5	14	26	5
Validity	Valid			
P value	0.28			
D value	0.24			

The P value shows that the item number 23 is too difficult since only 14 students answered it correctly, while more than 50% chose incorrect answer (option C). It means that option B is destructive. While the value D shows us that, the item is satisfactory, since there is a difference in the number of students of the upper group and those of the lower group who answered it correctly. In the writer's opinion, this item could be applied in the test.

- Item number 24

Question	What time should the students come? a. 12 o'clock b. 2 o'clock c. 20 o'clock d. 22 o'clock			
Result	**A**	**B***	**C**	**D**
Upper 27%	-	14	-	-
Middle 46%	1	22	-	-
Lower 27%	3	9	1	-
Total	4	45	1	-
Validity	Valid			
P value	0.90			
D value	0.28			

This item belongs to the easy group in terms of the difficulty level because 45 out of 50 students answered it correctly, while in terms of discrimination power, it is satisfactory item. The option should be written in capitalize letter. Based on the criteria above this item could be used in final test with several revisions, the students have understood about the content of this text, the text should be changed into rather difficult text.

- Item number 25

Question	Who should come to school? a. All students b. 9B students c. Girls d. Boys			
Result	**A**	**B***	**C**	**D**
Upper 27%	-	14	-	-
Middle 46%	-	23	-	-
Lower 27%	4	9	-	-
Total	4	46	-	-
Validity	Valid			
P value	0.92			
D value	0.24			

This item is considered easy in terms of the difficulty level, and satisfactory in terms of discrimination power. It seems that the option A is a good distractor, because it attracted students from the lower group to choose this option. This item could be used in the test with several revisions. This item is too easy to be answered, the option should be distract the students to answer.

PERPUSTAKAAN
UNNES

- Item number 26

Question	What subject does Mrs. Titin possibly teach? a. Mathematics b. Physics c. Geography d. Bahasa Indonesia			
Result	**A**	**B**	**C**	**D***
Upper 27%	1	1	-	12
Middle 46%	1	2	2	18
Lower 27%	5	2	-	6
Total	7	5	2	36
Validity	Valid			
P value	0.72			
D value	0.32			

This item is considered easy in terms of the difficulty level, and satisfactory in terms of discrimination power. From the table, we can see that 36 students answered it correctly. The distribution is good in terms of students' response. Meanwhile, the difference between the upper group and the lower group who answered it correctly is six. So the revision is needed. The item and the options can be understood, the dictionary will be brought in laguage subject.

- Item number 27

Question	Complete the dialogue with the suitable word or phrase. Dina : hello. I'm Dina Ari : 27) Dina : nice to meet you Ari Ari : nice to meet you too a. How do you do? b. Hello, Dina. I'm Ari c. How are you d. Good morning			
Result	**A**	**B***	**C**	**D**
Upper 27%	-	14	-	-
Middle 46%	-	22	1	-
Lower 27%	-	12	1	-
Total	-	48	2	-
Validity	Invalid			
P value	0.96			
D value	0.16			

The P value shows that the item number 27 is too easy since there are 48 out of 50 students answered it correctly, while more than 90% chose correct answer (option B). From the aspect of discrimination power, the item is considered as a poor item. The D value is only 0.16 which means that it is a poor item. This item is invalid; it is because of the pattern of the answers which is easily identified. This enables students to guess the answer more easily. In the writer's opinion, this item should be rejected or discarded.

- Item number 28

Question	Tika : Hi, Din . . .? Dina : I am very well, Tika. You look sad, why? Tika : I lost my book. a. How do you do b. Are you OK c. How are you d. What's your name			
Result	**A**	**B**	**C***	**D**
Upper 27%	3	1	10	-
Middle 46%	3	5	15	-
Lower 27%	4	1	7	1
Total	10	7	32	1
Validity	Invalid			
P value	0.64			
D value	0.32			

This item is said to be moderate because there are 32 students chose the correct answer. The rest of the students chose the other option. If we see the option A and B, they have the same meaning. It makes students confuse enough to think. Most of students chose option A and B; it means that option A and B are destructive. From the discrimination point of view, this item is satisfactory. In the writer's opinion, this item can be applied in the test with some revision. Option A,B, and C has same meaning, the option should has different meaning.

- Item number 29

Question	Father : Get me the coffee, please. Rani : 29) Father : Thank you a. I'm sorry. I can't. b. I'm sorry. I'm busy. c. What can I do for you? d. Okay, Dad.			
Result	**A**	**B**	**C**	**D***
Upper 27%	-	-	-	14
Middle 46%	-	-	1	22
Lower 27%	-	-	1	12
Total	-	-	2	48
Validity	Invalid			
P value	0.88			
D value	0.16			

The item is easy in terms of the difficulty level and poor in terms of discrimination power. Meanwhile, the difference between those who answered it correctly from the upper group and the lower group is only two. This item should be deleted, since it has low discrimination power.

- Item number 30

Question	Complete the dialogue with the suitable word or phrase. Teacher : please don't cheat! Students : 30)... a. Thank you b. You are welcome c. No way d. Yes, ma'am			
Result	**A**	**B**	**C**	**D***
Upper 27%	-	-	-	14
Middle 46%	1	-	2	20
Lower 27%	1	-	2	10
Total	2	-	4	44
Validity	Valid			
P value	0.88			
D value	0.32			

This item is easy, since 44 out of 50 students answered it correctly. And the difference between the upper group and the lower group indicates that it is a satisfactory one. Based on the criteria, this item can be applied in the test with revisions. In this item, the students can answer easily, because it is familiar for them.

- Item number 31

Question	Complete the dialogue with the suitable word or phrase. Ratih : Lis, meet Rico, my brother. Rico this is my friend Lisa. Rico : 31). . . Ratih : Hi Rico. Pleased to meet you. a. How are you, Lisa? b. Is your name Lisa? c. How are things, Lisa? d. Hi, Lisa. Pleased to meet you.			
Result	**A**	**B**	**C**	**D***
Upper 27%	-	-	-	14
Middle 46%	2	-	-	21
Lower 27%	3	1	-	9
Total	5	1	-	44
Validity	Valid			
P value	0.88			
D value	0.24			

The question is clear enough, 44 students out of 50 chose the correct answer. This item is easy, but the discrimination power is satisfactory because 14 students of the upper group and only 9 students of the lower group responded this answer. If we look at the table above, option A is said to be a good distractor, since there are students who choose this option than option B and C. This item needs revision. The dialogue is easy for the students, the writer suggests to rearrange it into a new one.

- Item number 32

Question	Complete the dialogue with the suitable word or phrase. Mia : Excuse me. Are you a new student? Rachel : 32). . . Mia : Hello, Rachel. I am Mia a. Please call me Rachel b. How are you? c. How do you do? d. Nice to meet you			
Result	**A***	**B**	**C**	**D**
Upper 27%	14	-	-	-
Middle 46%	18	3	-	2
Lower 27%	8	4	-	1
Total	40	7	-	3
Validity	Valid			
P value	0.80			
D value	0.32			

This item is considered easy in terms of the difficulty level and satisfactory in terms of discrimination power, since 14 students from upper group and 8 from the lower group chose option A. This item needs to be revised, it can be with change the option in order to distract the students. From the table above, option B is a distractor because it attracted more students than option C and D.

- Item number 33

Question	Complete the dialogue with the suitable word or phrase. Siska : . . . , Reni? Reni : I live in Pemalang. Siska : What's your address? Reni : At Pemuda strret No.2. a. Where are you from b. Where do you go c. Where do you come from d. Where do you live			
Result	**A**	**B**	**C**	**D***
Upper 27%	2	1	2	9
Middle 46%	7	2	5	9
Lower 27%	7	1	3	2
Total	16	4	10	20
Validity	Valid			
P value	0.40			
D value	0.48			

This item is moderate, neither too easy nor too difficult. There is quite large difference between the upper group and the lower group. The upper one is larger than the lower one. This means that it has good discriminate power. If we see option A and C, they have the same meaning. It makes students confuse enough to think. This item has a good distribution in terms of students' response. Option A and C are distractors because they attracted more students than option B. This item can be applied in the final test.

- Item number 34

Question	Dian : How do you go to school? Panji : . . . My house is not far from school. a. On foot b. By car c. By bus d. By truck			
Result	**A***	**B**	**C**	**D**
Upper 27%	11	-	3	-
Middle 46%	5	1	17	-
Lower 27%	1	5	7	-
Total	17	6	27	-
Validity	Valid			
P value	0.34			
D value	0.36			

From the P value, we can judge that this item is moderate, because 17 out of 50 students answered it correctly. It means that the item is neither too difficult nor too easy while the D value shows that the item number 34 is satisfactory item. 11 students from upper group answered it correctly, and a student from the lower group did the same. The difference between those two groups produced the D value 0.36. It means that this item discriminates the upper group students from the lower one quite well. We can say that in this item, the students from the upper group have shown their capability in answering the question. This item doesn't need any revisions and can be applied in the final test.

- Item number 35

Question	Dialogue Mother : Winda, this is your pocket money. Winda : 35. . . , mom. Mother : 36. . . But remember. Don't buy candy. a. Thank you b. Yes, I do c. I am very well d. Let's go			
Result	**A***	**B**	**C**	**D**
Upper 27%	9	2	-	3
Middle 46%	7	4	2	10
Lower 27%	5	2	1	5
Total	21	8	3	18
Validity	Valid			
P value	0.42			
D value	0.28			

This item is considered moderate in terms of difficulty level and satisfactory in terms of discrimination power. It seems that option B and D are good distractors, because more students chose them than the other option. In the writer's opinion, this item could be applied in the test.

- Item number 36

Question	a. Go away b. Be careful c. You're welcome d. Back soon			
Result	**A**	**B**	**C***	**D**
Upper 27%	-	2	10	2
Middle 46%	-	7	12	4
Lower 27%	2	5	4	2
Total	2	14	26	8
Validity	Valid			
P value	0.52			
D value	0.24			

This item belongs to the moderate group in terms of the difficulty level because 26 out of 50 students answered it correctly, while in terms of discrimination power, it is a satisfactory item since the difference between those who answered it correctly from the upper group and the lower one is large, but this item has good distribution of students' response. The option should be written in capitalize letter. Option B and D are good distractors. Based on the criteria above, this item could be used in final test.

- Item number 37

Question	Bunga : Mother, . . . I broke the glass Mother : It's okay, dear. Next time will careful Bunga : I will, mum. I promise. a. I'm glad to see you. b. I'm happy to meet you. c. I'm very well thanks. d. I'm really sorry			
Result	**A**	**B**	**C**	**D***
Upper 27%	2	1	3	8
Middle 46%	4	3	9	7
Lower 27%	2	1	8	2
Total	8	5	20	17
Validity	Valid			
P value	0.34			
D value	0.20			

The P value shows that the item number 37 is moderate since 17 students answered it correctly, while more than 40% chose incorrect answer (option C). It means that option C is destructive. And from discrimination power there is no discrimination between those two groups, because the D value is 2. This shows that this item is poor. In the writer's opinion, this item could be applied with revision. The writer suggests to more explain about expression material.

- Item number 38

Question	In a meeting room. Betty : I feel cold. . . . , please? Father : O.K. a. Would you turn off the fan b. Would you turn on the fan c. Would you turn off the light d. Would you turn on the light			
Result	**A***	**B**	**C**	**D**
Upper 27%	11	2	1	-
Middle 46%	6	5	11	1
Lower 27%	3	1	7	2
Total	20	8	19	3
Validity	Valid			
P value	0.40			
D value	0.32			

The P value for this item is only 0.40. The P value shows that this item is moderate since there are 20 students who answered it correctly. The D value shows it is a satisfactory item. This item has good distribution in terms of students' response. Option C is distractor because it attracted more students than option B and D. This item can be applied in the final test.

- Item number 39

Question	Bella : Can I borrow your mobile phone, please? Sinta : 39) . . . I am using it Bella : No problem a. I am sorry b. Here it is c. Here you are d. Be careful			
Result	**A***	**B**	**C**	**D**
Upper 27%	10	3	1	-
Middle 46%	10	8	3	2
Lower 27%	5	4	3	1
Total	25	15	7	3
Validity	Valid			
P value	0.50			
D value	0.36			

This item is considered moderate in terms of the difficulty level, and satisfactory in terms of discrimination power. From the table above, we can see that 25 students answered it correctly. It means that the difference between the upper group and the lower one who answered it correctly is 5 students. If we see the option, option B and C seem to be good distractors, since more students choose this option than option D. In the writer's opinion this item doesn't need any revisions.

- Item number 40

Question	Bobi : How was the contest yesterdat? Tono : I won. I got a gold medal. Bobi : . . .! a. Oh, sorry b. Well done c. Oh, my God d. My condolence			
Result	A	B*	C	D
Upper 25%	2	9	3	-
Middle 50%	5	2	13	3
Lower 25%	2	3	7	1
Total	9	14	23	4
Validity	Valid			
P value	0.28			
D value	0.24			

This item can be classified as a difficult item, because there are only 14 students who answered correctly. More than 40% students chose the incorrect answer (C). It means that option C is destructive. From the discrimination point of view, this item is poor. In the writer's opinion, this item should be discarded and changed to easier one.

CHAPTER V

CONSLUSION AND SUGGESTION

5.1 Conclusions

Based on the result of the analysis of the 40 test items of SMP students in SMP N 1 Moga Pemalang in the academic year of 2008/2009, the following conclusions could be drawn.

a. In analysis of item validity, it was found that the value of validity of the whole test items is 0.3162. It means that the test is valid.

b. By applying KR – 21 formula, the writer found that the coefficient of reliability of the whole test is 0.751. It means that the test as a whole had reliability, and we can use the test items as the instrument of evaluation again if we want to.

c. The mean of the index difficulty for the sample examined was 0.60. It means that the P value is between 0.30 and 0.70, which put the test in the moderate position. As a result, on the whole the English final test items have met the requirements of a good test in terms of the difficulty level.

d. In analysis of the item discrimination power, it was found that the mean of the D value was 0.48, which put the English final test items in good categorized. The items of this test were able to discriminate the upper group (high achiever) from the lower one.

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

e. Finally, the writer draws a conclusion that the items in the English final test for the seventh grade students of SMP N 1 Moga Pemalang could still be used as an instrument of evaluation with some revisions.

5.2 Suggestions

Constructing good language test items (especially objective test items) is not an easy task. Based on the conclusions above, the writer would like to offer the following suggestions.

a. The test constructors should know about the characteristics of good language test, especially procedure of determining difficulty levels and discrimination power.
b. Items that can still be used should be revised and saved. The items that contain too many problems should be discarded.
c. The writer hopes the result on this item analysis could be used as an example in analyzing other test items, and encourages other researchers to do research on the same subject.

PERPUSTAKAAN UNNES

APPENDICES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

APPENDIX 1

The Computation of the Validity, Reliability, Difficulty level, and Discriminating Power of the Final Test

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

APPENDIX 2

The Computation of Validity Test

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

APPENDIX 3

The computation of the Reliability Test

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

APPENDIX 4

The computation of the Discriminating Power Test

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

APPENDIX 5

The computation of the Difficulty Level Test

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

APPENDIX 6

The Recapitulation of multiple-choice

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

Number of Item	A	B	C	D	∑
1	3	14	3	30	50
2	3	4	34	9	50
3	39	0	9	2	50
4	3	6	23	18	50
5	22	9	16	3	50
6	0	1	4	45	50
7	1	0	45	4	50
8	15	4	9	22	50
9	2	0	3	45	50
10	5	13	22	10	50
11	6	4	9	31	50
12	37	0	3	10	50
13	44	5	1	0	50
14	33	16	1	0	50
15	1	46	0	3	50
16	47	2	1	0	50
17	10	14	24	2	50
18	9	0	32	9	50
19	17	25	8	0	50
20	8	4	6	32	50
21	7	4	23	16	50
22	21	14	6	9	50
23	5	14	26	5	50
24	4	45	1	0	50
25	4	46	0	0	50
26	7	5	2	36	50
27	0	48	2	0	50
28	10	7	32	1	50
29	0	0	2	48	50
30	2	0	4	44	50
31	5	1	0	44	50
32	40	7	0	3	50
33	16	4	10	20	50
34	17	6	27	0	50
35	21	8	3	18	50
36	2	14	26	8	50
37	8	5	20	17	50
38	20	8	19	3	50
39	25	15	7	3	50
40	9	14	23	4	50

PERPUSTAKAAN UNNES

APPENDIX 7

The Test Items

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES

Bibliography

Arikunto, S. 2005. *Dasar- Dasar Evaluasi Pendidikan* (Revised Ed.). Jakarta: Bumi Aksara.

Brink, T.D.T. 1974. *Evaluation a practical guide for teacher.* University of Missouri: Mc Graw Hill.

Brown, H.D. 2004. *Language Assesement: Principles and Classroom Practices.* New York: Longman

Ebel, R.L and D.A Frisbie. 1979. *Essential of Educational Measurement.* New Jersey: Prentice- Hall: University of Lawn.

Ebel, R.L and D.A Frisbie. 1991. *Essential for Educational Evaluation.* Philipines. Addison- Wersley Publishing Company.

Gronlund, Norman E. 1976. *Measurement and Evaluation In Teaching.* New York: Macmillan Publishing Co., Inc, University of Illnois.

Gronlund, Norman E. 1982. *Constructing Achievement Test.* New Jersey: Prentice- Hall. Inc.University of Illnois.

Harris, D.P. 1969. *Testing as a Second Language.* New York: Mc Graw- Hill.

Heaton, J.B. 1975. *Writing English Language Tests.* England: Longman Group Limited.

Hinkle, D.E, Jurs, S.G, Wiersma, W.1979. *Applied Statistics for the Behavioral Sciences.* Hough company

Lado, Robert. 1961. *Language Testing.* New York: Oxford University Press.

Margono, Drs. 2003. *Metodologi Penelitian Pendidikan.* Jakarta: Rineka Cipta.

Nitko, A.J. 1983. *Educational Tests and Measurement an Introduction*. London: Harcourt Brace Jovanovich Inc.

Oller, Jr., W. 1979. *Language Test at School*. London: Longman Group Limited.

Tinambuan, Willmar. 1988. *Evaluation of Student Achievement*. Jakarta: Depdikbud.

Tuckman B.W. 1978. *Constructing Educational Research.* London: Harcout Brace Jacobovitz.

PERPUSTAKAAN UNNES

Vallete, R.M. 1977. *Modern Language Testing*. New York, Harcout Brace and Jovanovich. Inc.

UNIVERSITAS NEGERI SEMARANG
PERPUSTAKAAN
UNNES