AN ITEM ANALYSIS OF ENGLISH SUMMATIVE TEST FOR THE FIRST YEAR OF JUNIOR HIGH SCHOOL

(An Observation at the Seventh Grade of SMP YMJ Ciputat)
A’Skrupsi’
Presented to the Faculty of Tarbiya and Teachers’ Training in a partial Fulfillment of the Requirements for the Degree of S.Pd. (Bachelor of Art) in English Language Education

By
MUHAMAD TOHA
203014001576

ENGLISH EDUCATION DEPARTMENT
THE FACULTY OF TARBIYAH AND TEACHERS’ TRAINING
SYARIF HIDAYATULLAH STATE ISLAMIC UNIVERSITY
JAKARTA
2010
ENDORSEMENT BY THE EXAMINATION COMMITTEE

A ‘Skripsi’ titled: “AN ITEM ANALYSIS OF ENGLISH SUMMATIVE TEST FOR THE FIRST YEAR OF JUNIOR HIGH SCHOOL” was examined at examination session of the State Islamic University (UIN) Syarif Hidayatullah Jakarta on March 18th, 2010. This ‘Skripsi’ has fulfilled the requirements for the Degree of Strata-I (S-1) at English Education Department.

Examination Committee

The Head of Committee,
Drs. Syauki, M.Pd.
NIP. 196412121991031002

The Secretary of Committee,
Neneng Sunengsih, S.Pd.
NIP. 150 293 236

The Examiner I,
Drs. Nasrun Mahmud, M.Pd
NIP. 150041070

The Examiner II,
Drs. Zaenal Arifin Toy, M.sc
NIP. 150031215

Acknowledged by
Dean of Tarbiya and Teachers’ Training Faculty

Prof. Dr. H. Dede Rosyada, M.A.
NIP. 150 231 356
ACKNOWLEDGEMENT

All praises be to Allah, the Beneficent and the Merciful by thyne blessing the writer could complete this skripsi. May sovereignty and blessing be upon our Prophet Muhammad S.A.W., his families and his followers to the end of time.

Upon the completion of this paper, the writer would like to express his great honor and gratitude to Mr. M. Farkhan as the lecturer and advisor who has accompanied and guided him in completing this “skripsi” so the skripsi entitled “An Item Analysis of English Summative Test for the First Year of Junior High School” (An Observation at the Seventh Grade of SMP YMJ Ciputat) has finally completed and acknowledged to meet the requirement for Strata-1 (S1) degree.

The writer also gave his greatest gratitude to:

1. All lecturers in the English Teaching Department, the staff and officers of UIN library and central administration office, who have given permission to use their books as source of information.

2. Drs. Syauki, M. Pd., the Head of English Department.

3. Prof. DR. Dede Rosyada, the Dean of Faculty of Tarbiyah and Teachers’ Training.

4. His parents Mr. Yahya and Mrs. Hasnah who always gave him support and moral enhancement in compiling this skripsi.

5. All his family and relatives who never stops in giving moral and material supports to the writer.
6. Special thanks to Mrs. Hidayati, M.Pd who guide him in his every steps no matter how difficult the challenge is.

7. Mrs. Neneng Sunengsih, Ms. Aida, and all staffs in English Education Department administration office.


*May Allah the Almighty give thyne blessing upon them all, Amin.*

Ciputat, March 02\(^{nd}\), 2010

(The Writer)
# TABLE OF CONTENTS

**ACKNOWLEDGMENT**

**TABLE OF CONTENTS**

**CHAPTER I INTRODUCTION**

| A. Background of Study                          | 1 |
| B. Limitation and Formulation of Problem       | 3 |
| C. The Objective of Study                      | 4 |
| D. The Use of Study                            | 4 |
| E. Method of Study                             | 4 |
| F. Organization of Writing                     | 4 |

**CHAPTER II THEORITICAL FRAMEWORK**

| A. The definition of test                      | 5 |
| B. Types of test                               | 6 |
| C. Types of test item                          | 8 |
| D. The definition of item analysis             | 13 |
| E. Points of item analysis                     | 14 |

**CHAPTER III THE EVALUATION**

| A. Evaluation Definition                       | 18 |
| B. Types of Evaluation                         | 20 |

**CHAPTER IV RESEARCH METODOLOGY AND FINDING**

| A. Research Methodology
| 1. Place and time of study                     | 22 |
| 2. Technique of sample taking                 | 22 |
| 3. Technique of data collecting               | 22 |
| 4. Technique of data analysis                 | 23 |
| B. Research Finding
| 1. Description of data                        | 23 |
| 2. Data analysis                              | 24 |
| 3. Data interpretation                        | 29 |
| 4. Revision                                   | 29 |
CHAPTER V CONCLUSION AND SUGGESTION

A. Conclusion........................................................................................................33
B. Suggestion.........................................................................................................33

APPENDIXES
CHAPTER I
INTRODUCTION

A. The Background of Study

When people think of evaluation of teaching, they probably think of end-of-term or year-end questionnaires, which distributed to all the students in the class. The main purpose of such a summmative rating of instruction is to provide information on a persons’ performance as a teacher. This type of evaluation is useful for both teacher and the supervisor to examine. It gives an overview of the students' impression of the entire course, their learning, and the teaching feedback for the teachers. The feedback the teacher gets from these questionnaires may help him or her prepare for the next class he or she teaches.

Evaluation is something that has an important role in some of people activities, it is systematic determination of merit, worth, and significance of something or someone using criteria against a set of standards, It emphasizes on acquiring and assessing that involves collecting and sifting through data, making judgments about the validity of the information and inferences that derive from it.

According to Lyle F. Bahman, evaluation defined as the systematic gathering of information in the purpose of making decisions.\(^1\) In terms of education, evaluation gives information about how successful an effort has

been. In English Language Teaching for example, an evaluation can tell the teacher the effectiveness of instruction, the student’s progress, strengths, and weakness in teaching and learning process.

Many ways can be employed to evaluate the teaching learning of English language, everything is possible as long as in line with principles of a good language testing. According to Anas Sudjiono, one of the evaluation ways is a test\(^2\). A test is a set of question, each of which has a correct answer, which examinees usually answer orally or in writing\(^3\). Tests are administered for different reasons but the main one is to inform decisions. Tests compilers should therefore always strive to compile tests that are relevant to their objectives but also based on the principles of good tests, in this case is a good language tests.

The principles of good language testing mainly contain three major aspects, validity, reliability, and practicality. Validity in language testing means a test has to be able to give dependable score of the test taker. While, practicality concerned with a wide angle of factors of economy, convenience, interpretability\(^4\).

The other characteristic that is more specific and important is the quality of the test items. Teachers need to do an item analysis. The item analysis that is made by the teachers is intended to give benefits to the teachers and the students. For the teachers it can give essential information about the quality of the test items they made. Furthermore, the good quality test will measure the students’ mastery in lesson properly.

The analysis of student responses to objective test items is a powerful tool for the test improvement. Reexamining each test item to discover its strength and flaws is known as item analysis.\(^5\) It usually concentrates three vital features; they are level of difficulty, discriminating power and the

---

\(^3\) Wilmar Tinambunan, *Evaluation of Students Achievement*, (Jakarta: Depdikbud, 1988), p.3
effectiveness of each distracter. Item analysis provides data whether an item is easy, moderate, or difficult, whether the test can discriminate the students’ ability or not and whether all the alternatives of items function well or not.

Summative test as one of the achievement tests are commonly used in classroom by the teacher. It is held at the end of the term of semester and it can measure the students’ achievement as a whole. In other word, it can cover all the subjects unit learned that thought in a semester, while formative test only covers a part of subject unit in syllabus or program.

Based on the statement above, the writer is interested to analyze the English Summative Test for the First Year of Junior High School of SMP YMJ Ciputat.

B. Limitation and Formulation of the Problem

As point of discussion, the writer intends to see the quality of test items only by doing item analysis that focuses on the discriminating power of test item index that shows the students’ ability to differentiate between pupils who have achieved well (the upper group) and those who have achieved poorly (the lower group). If the test items are given to the students who have studied well, the score will be high and if they are given to those who have not, the score will be low. On the contrary, if the test items yield the same score when they are given to the two groups, or even to upper group yield the low score and to the lower group yield the high score, so they are not good test items.

The test item will be analyzed is objective test only of English summative test used at the first year students of SMP YMJ Ciputat. The writer formulates the problems as follows: “Are the English summative test items tested at the first year students of SMP YMJ Ciputat qualified as a good test item or not?”
C. The Objective of Study

The objective of this study is to find out the discriminating power of English summative test items tested at the first year of SMP YMJ Ciputat.

D. The Use of Study

With this study, the writer hoped that the teacher can get clear description about the quality of English summative test items at the first year of SMP YMJ Ciputat so that it can be known which one of the test items should be discarded and revised so they can be used for the next evaluation.

E. Method of Study

To support the discussion, the writer takes descriptive research by analyzing English summative test paper and the students’ answer sheet of the test items seen from the discriminating power of point of view.

F. Organization of Writing

This paper consists of four chapters. The first chapter is introduction which describes the background of study, the limitation and formulation of problem, Objective of the study, the use of study, the method of study and the. Organization of Writing

The second chapter is theoretical framework, which explains several matters. First, it explains some theories of test, from general view of test, types of test, to the definition of item analysis. In addition, it explains about kinds of item analysis.

The third chapter is research methodology and findings that consists of research methodology, place and time of study, population and sample, instrument of research, technique of data taking, and consists of research finding, description of data, analysis of data, interpretation of data and discussion of finding. The writer tries to give conclusion, discussion and suggestion in the fourth chapter.
CHAPTER II
THEORETICAL FRAMEWORK

A. The Definition of test

One of the evaluation methods that are commonly used in education field is a test. In other words, evaluation is a more inclusive term than test and test is only part of evaluation. Test is one of a good instrument for collecting data, which is latterly used for making various decisions about the student’s behavior or achievement in the teaching-learning process.

According to the Longman dictionary test is “any procedure for measuring ability, knowledge, or performance”\(^1\), Meanwhile, according to David G. Armstrong and Tom V. Savage, “Measurement refers to the information gathering. It is nonjudgmental. When the term applied to instruction, information that is collected generally relates to students performance. Evaluation refers to the process of drawing conclusion from a study of data gathered as a result of measurement.”\(^2\)

While the difference between measurement and evaluation in term of education can be, find in the practice of assigning final marks to students at

---


the end of unit of work. The teacher may have made measurements of student’s prior achievement before the unit began and the teacher may add his reaction to this students’ attitude, motivation etc. The result is generally a single letter or numerical mark that indicates the relative success of the students’ as they experienced the unit of work. When the teacher has done this, he has made evaluation.

According to Anthony J. Nitko, “Test is a systematic procedure for observing and describing one or more characteristic of a person with aid of either a numerical scale or category system.”

Based on the opinion above, we can conclude that a test is a systematic procedure for providing information about the student’s achievement that is relevant in teaching-learning process.

B. Types of test

There are many types of test that can be used by the teacher to measure students’ achievement, Jack C. Richards and Richard Schmidt in their book say that the overall description of a test in terms of the purpose it serves and its objectivity or subjectivity.

According to Wilmar Tinambunan, there are two types of test used in determining a person’s abilities namely:

1) Aptitude test primarily designed to predict success in some future learning activities.

2) Achievement test is designed to indicate degree of success in some past learning activity that can be described broadly into four types achievement test which are very commonly used by teachers in the classroom, they are:

---

a. Placement test designed to determine the pupil performance at the beginning of instruction. It intended to know the pupil’s entry performance. That is whether or not the pupils has possessed the knowledge and skills needed to begin the planned instruction; to what extent has the pupil already mastered the objectives of the planned instruction.

b. Formative test intended to monitor learning progress during the instruction and to provide continuous feedback to both pupil and teacher concerning learning success and failure. It usually given in the end of unit of the course book or after a lesson designed to teach one particular point. The result of this test can be used for many purposes such as to provide information to the students about their progress in a term in particular material. It also will give immediate feedback to the students. In other words, feedback to pupils will reinforce successful learning and identify the learning errors that need correction. The last but not least, it can be directed toward improving learning instruction. That is to formulate a plan for remedial action, for example, for modifying instruction, prescribing group and individual work.

c. Diagnostic test is intended to diagnose learning difficulties during instruction. Thus, it is concerned with the persistent or recurring learning difficulties that are left unresolved by the standard corrective of formative evaluation.
Diagnostic evaluation is much more comprehensive and detailed because it searches for the underlying causes of those learning problems. The main aim of diagnostic test is to determine the cause of learning difficulties and then to formulate a plan for remedial action.

Summative Test is intended to show the standard which the students have now reached in relation to other students at the same stage.
C. Types of Test Item

Based on the manner of scoring, the type of test item is divided into two general types.

1. **Subjective Test**

Subjective test is a test which in its scoring requires judgment and evaluation of the scorer. Arthur Hughes stated, “If no judgment is required on the part of the scorer, then, the scoring is objective…if judgment is called for, the scoring is said to be objective”.6

In subjective test, the students are given a freedom to answer the question in their own words. The subjective tests that are commonly used in classroom are essay, short answer, and completion.

a. **Essay**

The essay item is the most complex of supply type item. It demands that the students compose a response, often extensive to a question for which no single response or pattern of response can be cited as correct to the exclusion of all the answer.7

The essay test usually consist of questions beginning with or including such directions as discuss, explain, outline, evaluate, define, compare, contrast, and describe.8

Example: Explain the definition of item analysis!

Thus, the distinctive feature of essay question is freedom of response it provides. In answering the question, the students are given freedom to select, relate, and present ideas in their own words. Because of the feature, the essay test usually scored differently by the same person on different occasion.

---

7 Wilmar Tinambunan, *Evaluation of Student...*, p. 86.
Meanwhile, there are some advantages of essay items, namely:

1) The essay question is useful for measuring the pupil ability to organize, integrate, and express their ideas.
2) Constructing essay items are relative easy.
3) The possibility of guessing requires less time than other measurement tool such as true false or on multiple choice items.

Besides the above advantages, the essay items have few disadvantages, namely:

1) Scoring essay items is difficult
2) Unreliability of scoring
3) Judgment of students; respond requires much time.
4) Limited sample of total instruction content.

b. Short Answer Question

The short answer item consists of a question, which can be answered with a word or short phrase. Generally, teachers prefer to use the short answer type of question probably because they think it has some advantages. It is relative easy to construct. It also gives the teacher some opportunity to see how well students can express their thought and it is also not difficult to score or mark than the essay questions. However, it is difficult to arrange the short answer question so that only one answer is correct. This type of question will be more useful only in testing knowledge of facts and quite specific information. Thus, when teachers are going to know the broader description about something, they are better to use the essay form.

c. Completion

The completion item is a written statement that requires the examinee to supply the correct word or short phrase in response to an incomplete
sentence, a question, or a word association. Completion test can be used effectively to measure the recall of terms, dates, and names.\textsuperscript{13} This type of test can be used at almost all levels. But it is extremely difficult to make the question so that only one answer is correct, and in making the questions, many clues should not be given to answer the questions. If too many clues are given, the items will be too easy, and if an insufficient number of clues are presented, the item will be ambiguous and may yield several possibility of correct answer.\textsuperscript{14}

2. **Objective Test**

An objective test item is any item that there is only a single predictable correct answer.\textsuperscript{15} The test is scored in such a manner that subjective judgment is eliminated when determining the correctness of pupil’s answer.\textsuperscript{16} Therefore, whether one teacher or another scores the item, today or last week, it will yield the same score.

The objective test items commonly used in classroom testing are true false, matching, and multiple choices.

a. **True False**

True false item is referred to alternative response item, the items asks the students to answer with “true” if it conform to the truth or “false” if it essentially incorrect.\textsuperscript{17} Thus, the item provides the students with a choice of two alternatives, so the students have a possibility to guess the answer and sometimes it will be the right answer. Because of the random guessing to produce the correct answer, the test becomes less reliable than the other types of test. However, these items are appropriate for occasional use, for example after the students choose the two alternatives between right

\textsuperscript{13} Wilmar Tinambunan, *Evaluation of Student…*, p. 61.
\textsuperscript{14} Wilmar Tinambunan, *Evaluation of Student…*, p. 62.
\textsuperscript{17} Wilmar Tinambunan, *Evaluation of Student…*, p. 70.
and wrong, correct or incorrect, etc, they are asked to explain by writing the sentences justifying their response.\textsuperscript{18}

Another advantage of constructing a true false item is that the students are able to respond to more true false items in a given time period than any other selection type items.\textsuperscript{19}

The most commons uses of true false items are:

1) To measure the ability to identify the correctness of statements of facts, definition of terms and statements of principles.

2) To measure the pupil’s ability to distinguish fact from opinion. To measure aspect of understanding, that is the ability to recognize cause and affect relationship. This type of item usually contains two true prepositions in one statement, and the pupil asked to judge whether the relationship between them it true or false.

3) To measure the simple aspect of logic as illustrated in the example.

b. Matching

The matching test item consists of two parallel columns with each word, number, or symbol in one column being matched to a word, sentence, or phrase on the other column.\textsuperscript{20} This type of item is employed widely in situation where relationship of more or less similar ideas, facts, and principles are to be examined or judged.\textsuperscript{21}

This kind of test is an effective way to test students’ recognition of the relationships between words, definitions, events, dates, categories, example, and so on.\textsuperscript{22}

Matching items are also useful in measuring students’ ability to make association, interpretations or measure knowledge of a series of fact. Besides that, the matching items can be used for a large quantity of


\textsuperscript{19} Wilmar Tinambunan, \textit{Evaluation of Student…}, p. 70.

\textsuperscript{20} Wilmar Tinambunan, \textit{Evaluation of Student…}, p. 64

\textsuperscript{21} Victor H. Noll, \textit{Introduction to Educational…}, p. 64.

\textsuperscript{22} Barbara Gross Davis , \textit{Tools for Teaching…}, p.243.
associated factual material to be measured in a small amount of space while the students’ time needed to respond is a relatively short.\textsuperscript{23}

c. Multiple Choice

A multiple-choice item consists of one or more introductory sentences followed by a list of two or more suggested responses from which the examinee chooses one as the correct answer.\textsuperscript{24}

The multiple choices item can measure a variety of learning outcomes from simple to complex, and it is adaptable to most types of subject matter content. The learning outcomes in the knowledge area that can be measured by the multiple choice items are:\textsuperscript{25}

1) Knowledge of terminology

For this purpose, pupils are requested to show their knowledge of a particular term by selecting a word that has the same meaning as the given term or by choosing a definition of the term. Special uses of a term can also be measured by having pupils identify the meaning of the terms when used in context.

2) Knowledge of principles

The multiple-choice items can also be constructed to measure knowledge of principles. In this case, the items appear a bit more difficult, this is because principles are more complex than isolated facts.

3) Knowledge of method and procedure

The multiple choice form can be used to measure the knowledge of method and procedure, such as knowledge of laboratory procedure, knowledge of methods used in problem solving, computational and performance skill.

Some advantages of using multiple choice items are: the multiple choice items are fast, easy and economical to score, they can be

\textsuperscript{23} Wilmar Tinambunan, \textit{Evaluation of Student...}, p. 65.
\textsuperscript{24} Anthony J. Nitko, \textit{Educational Test...}, p. 190.
\textsuperscript{25} Wilmar Tinambunan, \textit{Evaluation of Student...}, p. 76.
objectively so that they will be fairer and more reliable than subjectively scored tests.

Besides those advantages, the multiple choice items also have disadvantages such as: the technique of the test only recognition knowledge, so the students have no or little opportunity to express their own idea of a problem, pupils have much time to guess the answer and it may effect on their scores, it is difficult to write successful items, and cheating may be facilitated.  

D. The Definition of Item Analysis

An item analysis is a systematic procedure by which the teacher can get some information about the quality of the test item. The objective of the item analysis is to find which of the test item is good or need to be corrected. As stated by Suharsimi Arikunto, “Analisis soal antara lain bertujuan untuk membantu kita dalam mengidentifikasi butir-butir soal yang jelek, memperoleh informasi yang akan dapat digunakan untuk menyempurnakan soal-soal untuk kepentingan lebih lanjut, dan memperoleh gambaran secara selintas tentang keadaan yang kita susun”.  

There are several activities of item analysis. According to Anthony J Nitko, in his book, he states that: “Item analysis refers to the process of collecting, summarizing, and using information about individual test items especially information about pupils’ response to item.”

While according to J. Stanley Ahman and Marvin D. Glock, “Item analysis is reexamining each test item to discover its strength and flaws”. Item analysis is an important and necessary step in the preparation of good multiple choice test. Because of this fact; it is suggested that every classroom teacher who uses multiple choice test data should know something of item analysis. How it is and what it means.

---

28 Anthony J. Nitko, Educational Test..., p. 284
30 Jhon W. Other, Language Test at School, (London: Longman Group, 1979), p. 245
Based on explanation above, the writer concluded that the item analysis is a process of collecting information about pupil’s responses to the items to see the quality of test items. It is very useful for the teachers and the students. In analyzing the items, the writer analyzes the difficulty level of the items.

E. Points of Item Analysis

There are three points usually considered in the field of item analysis; they are level of difficulty, discriminating power and effectiveness of distracter. However, in this “skripsi” the writer limits the discussion on discriminating power.

1. The Difficulty Level of the Item

Level of difficulty means the percentage of pupils who answer correctly each test item. The percentage inversely related to the difficulty because the larger the percentage of correct answer, the easier the item and the more difficult the item is, the fewer will be the students who select the correct option.

A good test item should have a certain degree of difficulty. It may not too easy or too difficult because the tests that are too easy or too difficult will yield score distribution that make it hard to identify reliable differences in achievement between the pupil who have done well and those who have done poorly.

By analyzing the student’s response to the items, the level of difficulty of each item can be known and the information will be helpful for teacher in identifying concepts to re teach the study material and giving the students feedback about their learning.

According to Suharsimi Arikunto:

“Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa

31 J. Stanley Ahmann and Marvin D. Glock, Evaluating Pupil …., p. 245.

untuk mempertinggi usaha siswa untuk memecahkannya. Sebaliknya, soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya".  

2. The Effectiveness of Distracter

One important aspect affecting the difficulty of multiple choice test items is the quality of distracters. Some distracters, in fact, might not be distracting at all, and therefore serve no purpose.  

A good distracter will attract more students who have not studied well (the lower group) than the upper group. On the contrary, a weak distracter: first, sometimes an item was drilled heavily in class, so almost everyone has mastered the item. So the answer is obvious. Second, sometimes a well recognized pair is used, such as this/these, is/are, etc. even though not everyone has controlled of these yet, students know that one of the two is the right answer, no other seems likely, and the third cause is the use of obviously impossible distracter.  

In a good test, the distracter must be functioned effectively; if the distracters are not functioned, they should be rewritten or discarded. And to know whether the distracters are functioned or not, distracter analysis is done that is by comparing the numbers of students in the upper group and the lower group who select each incorrect alternatives.  

3. The Discriminating Power

A good test item should have a discriminating power. The discriminating power of test item is an index that shows its ability to differentiate

33 Suharsimi Arikunto, Dasar-Dasar…, p. 207.
36 Wilmar Tinambunan, Evaluation of Student…, p. 141.
between pupils who have achieved well (the upper group) and those who have achieved poorly (the lower group). The discrimination index measures how well a test identifies differences in achievement levels of students. If the test items are given to the students who have studied well, the score will be high and if they are given to those who have not, the score will be low. On the contrary, if the test items yield the same score when they are given to the two groups, or even to upper group yield the low score and to the lower group yield the high score, so they are not good test items.

The tests that do not have discriminating power will not yield the proper description of the students’ ability, as stated by Nana Sudjana in his book: “... Tes yang tidak memiliki daya pembeda tidak akan menghasilkan gambaran yang sesuai dengan kemampuan siswa yang sebenarnya”. Therefore, it is very important to measure the discriminating power of test item to produce good test items.

The computation of the index of discrimination items as follows:

a. Collecting the students’ answer sheets, the test paper and the key answer of the test.

b. Checking the key answer of the test to see whether the key answer has been made correctly by the teacher. Then, the result of this checking becomes the reference to the marking or scoring.

c. Arranging the students’ answer sheets based on their rank of the test scores, from the highest score to the lower score.

d. Grouping the 27% of the papers with the highest score for the upper group and 27% of the papers with the lowest score for the lower group. In this case, the writer doesn’t use the rest of the students’ papers which are about 46%, because in doing the item analysis only the upper and lower group are needed.

---

38 Wilmar Tinambunan, Evaluation of Student..., p. 144.
e. Calculating and tabulating the students’ responses in the upper and lower group who responses each item correctly and put it in the format of tabulation of the item analysis.

f. Subtract the number in the lower group who answered the item correctly from the number in the upper group who answered the item correctly.

g. The result of the subtraction is divided by the number of total in each group.

h. The formula for computing item discriminating power is as follows:

\[ D = \frac{U - L}{N} \]

In which:
- \( D \) = the index of discriminating power
- \( U \) = the number of pupils in the upper group who answered the item correctly
- \( L \) = the number of pupils in the lower group who answered the item correctly
- \( N \) = number of pupils in each of the groups
CHAPTER III
EVALUATION

A. Evaluation Definition

Evaluation in general is systematic determination of merit, worth, and significance of something or someone using criteria against a set of standards. Evaluation is often used to characterize and apprise subjects of interest in a wide range of human enterprises, including the arts, criminal justice, foundations and non-profit organizations, government, health care, and other human services.

Educational evaluation is the process of documenting, usually in measurable terms, knowledge, skills, attitudes and beliefs. Evaluation can focus on the individual learner, the learning community (class, workshop, or other organized group of learners), the institution, or the educational system as a whole.

According to the Academic Exchange Quarterly: "Studies of a theoretical or empirical nature (including case studies, portfolio studies, exploratory, or experimental work) addressing the assessment of learner aptitude and preparation, motivation and learning styles, learning outcomes in achievement and satisfaction in different educational contexts are all welcome, as are studies addressing issues of measurable standards and benchmarks."

It is important to notice that the final purposes and evaluation practices in education depends on the theoretical framework of the practitioners and researchers, their assumptions and beliefs about the nature of human mind, the origin of knowledge and the process of learning.

A good evaluation has both validity and reliability, plus the other quality attributes noted above for a specific context and purpose. In practice, an evaluation is rarely totally valid or totally reliable. A ruler that is marked wrong will always give the same (wrong) measurements. It is very reliable, but not very valid. Asking random individuals to tell the time without looking at a clock or watch is sometimes used as an example of an evaluation which is valid, but not reliable. The answers will vary between individuals, but the average answer is probably close to the actual time. In many fields, such as medical research, educational testing, and psychology, there will often be a trade-off between reliability and validity.

A history test written for high validity will have many essay and fill-in-the-blank questions. It will be a good measure of mastery of the subject, but difficult to score completely accurately. A history test written for high reliability will be entirely multiple choices. It is not as good at measuring knowledge of history, but can easily be scored with great precision. We may generalize from this. The more reliable is our estimate of what we purport to measure, the less certain we are that we are actually measuring that aspect of attainment. It is also important to note that there are at least thirteen sources of invalidity, which can be estimated for individual students in test situations. Perhaps this is because their social purpose demands the absence of any error, and validity errors are usually so high that they would destabilize the whole evaluation industry.
B. Types of Evaluation

The term evaluation is generally used to refer to all activities teachers use to help students learn and to gauge student progress. Though the notion of evaluation is generally more complicated than the following categories suggest, evaluation is often divided for the sake of convenience using the following distinctions:

a. formative and summative

b. objective and subjective

a. Formative and Summative

Evaluation is often divided into formative and summative categories for the purpose of considering different objectives for evaluation practices.

Summative evaluation is generally carried out at the end of a course or project. In an educational setting, summative evaluations are typically used to assign students a course grade. Summative evaluations are evaluative.

While formative evaluation is generally carried out throughout a course or project. Formative evaluation, also referred to as "educative evaluation," is used to aid learning. In an educational setting, formative evaluation might be a teacher (or peer) or the learner, providing feedback on a student's work, and would not necessarily be used for grading purposes. Formative evaluations are diagnostic.

Summative and formative evaluations are often referred to in a learning context as evaluation of learning and evaluation for learning respectively. Evaluation of learning is generally summative in nature and intended to measure learning outcomes and report those outcomes to students, parents, and administrators. Evaluation of learning generally

---

occurs at the conclusion of a class, course, semester, or academic year. Evaluation for learning is generally formative in nature and is used by teachers to consider approaches to teaching and next steps for individual learners and the class.

b. Objective and Subjective

Objective evaluation is a form of questioning which has a single correct answer. Subjective evaluation is a form of questioning which may have more than one correct answer (or more than one way of expressing the correct answer). There are various types of objective and subjective questions. Objective question types include true/false answers, multiple choice, multiple-response and matching questions. Subjective questions include extended-response questions and essays.

Some have argued that the distinction between objective and subjective evaluations is neither useful nor accurate because, in reality, there is no such thing as "objective" evaluation. In fact, all evaluations are created with inherent biases built into decisions about relevant subject matter and content, as well as cultural (class, ethnic, and gender) biases.
CHAPTER IV
RESEARCH METHODOLOGY AND FINDINGS

A. Research Methodology

1. Place and Time of Study

The research was held at SMP YMJ (Yayasan Miftahul Jannah) which was located on Jalan Limun Nomor 27 Ciputat.

The First English Summative test of academic year 2008/2009 was done on Tuesday, 02 June 2009. Then the writer took copies of test paper and the students’ answer sheets on 15 June 2009.

2. Technique of sample taking

The population of this research is the first grade students of the second semester of SMP YMJ (Yayasan Miftahul Jannah), Ciputat. There are 120 students that are divided into three classes. Because the population is homogenous, the sample is taken only in one class; consisting 33 students in 7.B class.

3. Technique of Data Collecting

To collect data based on the topic of discussion, the writer collected the students’ answer sheet and the test paper of the English Summative test at the second year students to get the data.
4. Technique of Data Analysis

The writer used quantitative method to analyze the discrimination power of English Summative test of academic year 2008/2009 by using statistic formula, namely:

\[
D = \frac{U - L}{N}
\]

In which:
- \(D\) = the index of discriminating power
- \(U\) = the number of pupils in the upper group who answered the item correctly
- \(L\) = the number of pupils in the lower group who answered the item correctly
- \(N\) = number of pupils in each of the groups

The classification of the index of discrimination power (DI) are:

\[
\begin{align*}
\text{DI} & \geq 0.40 \quad \text{= Very Good} \\
0.30 - 0.39 & = \text{Good} \\
0.20 - 0.29 & = \text{Revised} \\
\leq 0.19 & = \text{Discarded}
\end{align*}
\]

B. Research Findings

1. Description of Data

The data that the writer used in this research is the English Summative test that is Ulangan Umum Semester Genap (II) academic year 2008/2009 for the first year students of SMP YMJ Ciputat, Tangerang.

The total numbers of test items are 35 items which consists of 25 multiple choice items and 10 essay test items.
2. **Data Analysis**

Based on the procedures of the research, a number of the candidates who answered the items correctly found. The following tables show their scores and their groups.

**Table 1**
The Group Position of the First English Summative test Result for 33 Students of First Grade Students of SMP YMJ, Ciputat.

<table>
<thead>
<tr>
<th>No</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>87</td>
</tr>
<tr>
<td>5</td>
<td>85</td>
</tr>
<tr>
<td>6</td>
<td>85</td>
</tr>
<tr>
<td>7</td>
<td>82</td>
</tr>
<tr>
<td>8</td>
<td>82</td>
</tr>
<tr>
<td>9</td>
<td>82</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>11</td>
<td>80</td>
</tr>
<tr>
<td>12</td>
<td>80</td>
</tr>
<tr>
<td>13</td>
<td>80</td>
</tr>
<tr>
<td>14</td>
<td>75</td>
</tr>
<tr>
<td>15</td>
<td>75</td>
</tr>
<tr>
<td>16</td>
<td>75</td>
</tr>
<tr>
<td>17</td>
<td>72</td>
</tr>
<tr>
<td>18</td>
<td>72</td>
</tr>
<tr>
<td>19</td>
<td>70</td>
</tr>
<tr>
<td>20</td>
<td>70</td>
</tr>
<tr>
<td>21</td>
<td>70</td>
</tr>
<tr>
<td>22</td>
<td>70</td>
</tr>
<tr>
<td>23</td>
<td>70</td>
</tr>
<tr>
<td>24</td>
<td>70</td>
</tr>
<tr>
<td>25</td>
<td>65</td>
</tr>
<tr>
<td>26</td>
<td>65</td>
</tr>
<tr>
<td>27</td>
<td>65</td>
</tr>
<tr>
<td>28</td>
<td>60</td>
</tr>
<tr>
<td>29</td>
<td>60</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>31</td>
<td>55</td>
</tr>
<tr>
<td>32</td>
<td>55</td>
</tr>
<tr>
<td>33</td>
<td>45</td>
</tr>
</tbody>
</table>

27% **UPPER GROUP**

46% **MIDDLE GROUP**

27% **LOWER GROUP**
Table 1 shows that the students who took the test are classified into three groups: upper, middle and lower groups. For this analysis, the writer paid more attention on the upper group and the lower group. It is efficient to compare the students who performed well 27% (*the top*) with those who performed poorly 27% (*the bottom*). The next table (2) discriminates the students into upper and lower groups.

<table>
<thead>
<tr>
<th>Number of Item</th>
<th>Number of correct answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
</tr>
<tr>
<td>22</td>
<td>4</td>
</tr>
<tr>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>24</td>
<td>9</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>26</td>
<td>9</td>
</tr>
<tr>
<td>27</td>
<td>9</td>
</tr>
<tr>
<td>28</td>
<td>9</td>
</tr>
<tr>
<td>29</td>
<td>9</td>
</tr>
<tr>
<td>30</td>
<td>9</td>
</tr>
<tr>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>32</td>
<td>8</td>
</tr>
<tr>
<td>33</td>
<td>8</td>
</tr>
<tr>
<td>34</td>
<td>5</td>
</tr>
<tr>
<td>35</td>
<td>5</td>
</tr>
</tbody>
</table>
Based on the table 2 on the previous page, the writer concluded the achievement of upper group students in their English test. The following description tells about the responses of each item.

1. There none of students who answered correctly for item number 1.
2. All students answered correctly for item number 5, 10, 11, 13, 14, 20, 24, 26, 27, 28, 29, 30 and 31.
3. There are 8 students who answered correctly for item number 2, 7, 8, 32 and 33.
4. There are 7 students who answered correctly for item number 3, 4, 16 and 18.
5. There are 6 students who answered correctly for item number 15 and 17.
6. There are 5 students who answered correctly for item number 9, 23, 25, 34 and 35.
7. There are 4 students who answered correctly for item number 12, 21 and 22.
8. There are 3 students who answered correctly for items number 19.
9. There is 1 student who answered correctly for item number 6.

The following table shows the tabulation of the test answer of lower group students.

### Table 3
The data of the Lower Group

<table>
<thead>
<tr>
<th>Number of Item</th>
<th>Number of correct answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 3 shows the achievement of lower group students in their English Test. The following description tells about the responses of each item.

1. There none of students who answered correctly for item number 1 and 27.
2. All students answered correctly for item number 7, 10, 11, 14 and 20.
3. There are 8 students who answered correctly for item number 8, 13, 28, 29, 30, 31, 32, and 33.
4. There are 7 students who answered correctly for item number 5 and 24.
5. There are 6 students who answered correctly for item number 3 and 18.
6. There are 5 students who answered correctly for item number 2, 15 and 34.
7. There are 4 students who answered correctly for item number 23 and 35.
8. There are 3 students who answered correctly for item number 4, 16, 21 and 26.
9. There are 2 students who answered correctly for item number 6, 9, 12, 22 and 25.
10. There is 1 student who answered correctly for item number 17 and 19.
After analyzing the data above, the writer describes the index of discriminating power of each item.

**Table. 4**

**The Tabulation of Discriminating Power Index**

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Total Correct Answer</th>
<th>U - L</th>
<th>DI = U - L</th>
<th>REMARK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Upper Group</td>
<td>Lower Group</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0,00</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>0,33</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>0,11</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>0,44</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>0,22</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>-1</td>
<td>-0,11</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td>-1</td>
<td>-0,11</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0,00</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0,33</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>0,00</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>0,00</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0,22</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>0,11</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>0,00</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>0,11</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>0,44</td>
</tr>
<tr>
<td>17</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>0,56</td>
</tr>
<tr>
<td>18</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>0,11</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0,22</td>
</tr>
<tr>
<td>20</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>0,00</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0,11</td>
</tr>
<tr>
<td>22</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0,22</td>
</tr>
<tr>
<td>23</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>0,11</td>
</tr>
<tr>
<td>24</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>0,33</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0,33</td>
</tr>
<tr>
<td>26</td>
<td>9</td>
<td>3</td>
<td>6</td>
<td>0,67</td>
</tr>
<tr>
<td>27</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>1,00</td>
</tr>
<tr>
<td>28</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>0,11</td>
</tr>
<tr>
<td>29</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>0,11</td>
</tr>
<tr>
<td>30</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>0,11</td>
</tr>
<tr>
<td>31</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>0,11</td>
</tr>
<tr>
<td>32</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0,00</td>
</tr>
<tr>
<td>33</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0,00</td>
</tr>
<tr>
<td>34</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0,00</td>
</tr>
<tr>
<td>35</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>0,11</td>
</tr>
</tbody>
</table>
3. Data interpretation

The table 4 above shows that there is more than 60% of item is discarded, it means that they are in less than 0.19 of discriminating power index.

While less than 7% is good and revised with discriminating power index between 0.30 – 0.39 that categorized as a good item and in the range between 0.20 up to 0.29 that categorized as a revised item, 7% left item is categorized as a very good item with discriminating power index equal or more than 0.40.

4. Revision

Items with Revised mark need to revise, There are four items with revised mark; no 5, 12, 19 and 22.

Copy item no 5

My Family

My name is Randi, I am an SMP student. I live on Jalan Suryakanta. My father’s name is Mr. Rahman. He works in a hospital. He is a doctor. My mother is a teacher.

My parents have three children. Ely, the eldest, works as a programmer in a private company. Wulan is my sister. We are students. She goes to SMA 3.

5. Who is Randi?
   a) He is an SMA student
   b) He is a teacher
   c) He is a doctor
   d) He is an SMP student (correct answer)

From the test item above, the writer assumed that the question is too simple and the answer purely copied the short-text. Briefly, it is too easy to answer it and it contained no distracter.

The first item is question number 5, where in this question there are 9 students from upper group who answer correctly, and there are 7 students from lower group who answers correctly, it means the question doesn’t have a good distracter because the two groups yield almost the same score and the question is too easy to be answered. And according to the formulas the discrimination index from that question is 0.22. It means it is classify as revised item

The test item would be better if it is revised into the example below.
My Family

My name is Randi, I am an SMP student. I live on Jalan Suryakanta. My father’s name is Mr. Rahman. He works in a hospital. He is a doctor. My mother is a teacher.

My parents have three children. Ely, the eldest, works as a programmer in a private company. Wulan is my sister. We are students. She goes to SMA 3.

5. Who is Randi?
   a) He is a Senior High School student
   b) He is going to be a teacher
   c) He is going to be a doctor
   d) He studied in Junior High School (correct answer)

Because the answer would not so easily guessed and the student would like to read the whole text to comprehend the question before answered it.

Copy item no 12

12) A mechanic works in a__________
   a) travel agency
   b) newspaper
   c) office
   d) workshop (correct answer)

The test item considered too difficult to understand, maybe because the lack of vocabulary or the student did not comprehend the question clearly. Maybe the teacher should add some clues about the vocabulary.

The second item is question number 12, where in this question there are only 4 students from upper group who answer correctly, and there are 2 students from lower group who answers correctly. And according to the formulas the discrimination index from that question is 0.22. It means it is also classify as revised item. It could be revised as follows.

Revised item no 12

12) A mechanic, or someone who repairs machine, works in a__________
   a) travel agency
   b) newspaper
   c) office
   d) workshop (correct answer)
Copy item no 19

19) A: Jane, this is Romeo. Romeo, this is Jane
    B: How do you do, Romeo?
    C: ________________
       a) How are you, Jane?
       b) How do you do, Jane? (correct answer)
       c) What do you do, Jane?
       d) I’m fine thanks.

The test item above need a revise because students may be confused about who is (C) speaker in the text. Maybe it would be better if the name of the speaker (Romeo, Jane …) replaced the initial name A), B), or C).

The third item is question number 19, where in this question there are only 3 students from upper group who answer correctly, and there is one student from lower group who answer correctly, it means the item is also too difficult for the two groups.. And according to the formulas the discrimination index from that question is 0.22. It means it is also classify as revised item. It would be better if it is stated as follows.

Revised item no 19

19) Marco : Jane, this is Romeo. Romeo, this is Jane
       Jane : How do you do, Romeo?
       Romeo : ________________
          a) How are you, Jane?
          b) How do you do, Jane? (correct answer)
          c) What do you do, Jane?
          d) I’m fine thanks.

In this revised items it is clear that the speaker C) is Romeo and he has to reply “How do you do, Jane?”
Copy item no 22

22) Susan has breakfast in_______
   a) The bathroom
   b) The dining room (correct answer)
   c) The bedroom
   d) The bathroom

This test item also revised because although it is clear that we had to have breakfast in the dining room or table. However, the answer maybe too difficult to understand because the lack of vocabulary or any other factors.

The fourth item is question number 22, where in this question there are only 4 students from upper group who answer correctly, and there are 2 students from lower group who answers correctly, it means the item is too difficult for the two groups. Just like the item number 12, probably because the students do not understand the vocabulary mentioned in the questions. And according to the formulas the discrimination index from that question is 0.22. It means it is also classify as revised item. Maybe it would be better if it is stated as follows.

Revised item no 22

22) Susan has dinner in_______
   a) The bathroom
   b) The dining room (correct answer)
   c) The bedroom
   d) The bathroom

From the description above, we can tell that those four questions are not good test item based on the discriminating power because some reasons. Of course the teacher need to explain, whether the question is too hard because those never been taught before or because something else.
CHAPTER V
CONCLUSION AND SUGGESTION

A. Conclusion

Based on the analysis the writer found that the English summative test that is administrated by SMP YMJ (Yayasan Miftahul Jannah), Ciputat can be categorized as discarded test items. It can be seen from the index shows by the result of analysis namely more than 60% of the test items are discarded test items. There is only 7% or 5 test items that can be categorized as a very good, good or revised.

The index found on the analysis shows that the test items can not differentiate between pupils who have achieved well (the upper group) and those who have achieved poorly (the lower group). The test items yield the same score when they are given to the two groups, or even to upper group yield the low score and to the lower group yield the high score, so they are not good test items.

B. Suggestion

After doing this research, there are some suggestions that can be given to the teachers in relation to the writer’s conclusion. They have to:

1. Give good techniques in answering the items. For instance, the teachers encourage the students to do the easier items and not to be stuck to the difficult items. The students should recognize this technique so that they will not waste their time.
2. Check the items many times, asked the other teachers to be corrector. Because for some cases we could not be an effective corrector to our own made test. Ask their opinion for the quality of the test.
BIBLIOGRAPHY


Davis, Barbara Gross, *Tools for Teaching* (San Francisco: Jersey Bass Publisher, 1993)


Other, John W., *Language Test at School*, (London: Longman Group, 1979)

Richards, Jack C. and Richard Schmidt, Longman Dictionary of Language Teaching And Applied Linguistics

Sudjana, Nana, *Penilaian Hasil Proses Belajar Mengajar*, (Bandung: Remaja Rosda Karya, 2001)

Tinambunan, Wilmar, *Evaluation of Students Achievement*, (Jakarta: Depdikbud, 1988)
