An ANALYSIS ON THE ENGLISH final test items for
The second semester of twelfth grade students of
Sma negeri 5 surakarta in 2008/2009 academic year
(a descriptive study)

By:
Arie lestari septiannawati
K2205027

A THESIS
Submitted to the Teacher Training and Education Faculty of Sebelas Maret
University to Fulfill One of the Requirements for Getting the Undergraduate
Degree of Education in English

English department
TEACHER TRAINING AND EDUCATION FACULTY
SEBELAS MARET UNIVERSITY
2010
APPROVAL

This thesis has been approved by the consultants to be examined by the board of thesis examiners of the English Department of Teacher Training and Education Faculty, Sebelas Maret University Surakarta.

On: July 2010

By:

Consultant I

Consultant II

Drs. Suparno, M.Pd.
NIP. 19511127 198601 1 001

Drs. Martono, M.A.
NIP. 19600301 198803 1 004
This thesis has been examined by the Board of Thesis Examiners of Teacher Training and Education Faculty, Sebelas Maret University, and accepted as one of the requirements for getting an Undergraduate Degree of Education in English.

On : Surakarta
Date : July 2010

Board of Examiners:

1. Chairman
   Drs. Muhammad Asrori, M.Pd.
   NIP. 19601015 198702 1 001
   ........................................

2. Secretary
   Teguh Sarosa, S.S., M.Hum
   NIP. 19730205 200604 1 001
   ........................................

3. Examiner 1
   Drs. Suparno, M.Pd.
   NIP. 19511127 198601 1 001
   ........................................

4. Examiner 2
   Drs. Martono, M.A.
   NIP. 19600301 198803 1 004
   ........................................

Legalized by,
Teacher training and education faculty
Sebelas maret university
Dean

Prof. Dr. M. Furqon Hidayatullah, M.Pd.
NIP 19600727 198702 1 001
ABSTRACT


The objective of the study is to know whether or not English final test items for the twelfth grade students are considered appropriate. The motive of the study is the existing phenomenon in the teaching and learning process which emphasizes its measurement through tests. The concern of the study is the appropriateness of multiple choice and essay test items.

The study focuses on the description of the test items’ appropriateness based on the quantitative data. The subject of the study is the English final test items for the second semester of twelfth grade students of SMA Negeri 5 Surakarta in 2008/2009 academic year. The data were taken from 100 students in four classes. The appropriateness of the test items is analyzed using item analysis technique. The analysis comprises three aspects i.e. index of discriminating power, level of difficulty, and the effectiveness of distracters. The appropriateness of the three aspects must be fulfilled if the test item is multiple choice. Meanwhile, if the item is essay, the last aspect is not included.

The study results a description of each test item based on quantitative data proceeded in the item analysis. Global result shows that there are only 27.5% of the total test items in the type of multiple choice that fulfil criteria of a good test items analyzed from the three aspects. Meanwhile, the essay test items are satisfactory and able to fulfil the two criteria.

Suggestion for teacher, test publisher, and other researcher is a follow up for the result of the study. The test maker should consider students’ ability in the subject and construct the items with paying attention to the aspects required in good item criteria as well as good criteria of scoring the test.
MOTTO

So, verily, with every difficulty, there is relief

Verily, with every difficulty there is relief

(Al Insyirah verse: 5-6)
This thesis is dedicated to:

- My beloved family
- My beloved friends
- The readers
ACKNOWLEDGEMENT

Alhamdulillahirabbil’alamin. All praises and thanks be for Allah The Most Merciful and Gracious, The Almighty God that gives His help to the writer so that she can complete her thesis.

The writer would not be able to finish her study without others’ help. For that reason, she would like to express her gratitude to:

1. The Dean of Teacher Training and Education Faculty who has given approval and permission to the writer to write her thesis.
2. The Head of The Art and Language Department, who acts as the first consultant too, who has given her permission and helped her to write the thesis.
3. The Head of English Department of Teacher Training and Education Faculty, who is also the second consultant for the writer, who has helped her to write her thesis and has given her the required approval.
4. The English teacher of the twelfth grade students of SMA Negeri 5 Surakarta who has helped the writer collect the data during the study.
5. Her beloved parents and brother who have supported her in every way.
6. Sunarto, for all passionate challenge and support to accomplish her thesis.
7. Mbak Ninik who has given a surprising reinforcement.
8. Her special friends in ED’05: Dian, Andri, Dita, Iffan, Dyah, Agung, Anto and Ignas, for an everlasting friendship and support.

The writer will appreciatively accept any criticism, suggestion, or comments since nothing in the world is perfect including her writing. She hopes the thesis will be useful, helpful, and able to give contribution to the readers.

Surakarta, July 2010

Arie Lestari Septiannawati
TABLE OF CONTENT

PAGE OF TITLE ........................................................................... i
PAGE OF APPROVAL .................................................................... ii
SUBMISSION PAGE ..................................................................... iii
ABSTRACT ................................................................................ iv
MOTTO ...................................................................................... v
DEDICATION ........................................................................... vi
ACKNOWLEDGEMENT .............................................................. vii
TABLE OF CONTENT ................................................................ viii
LIST OF APPENDICES ......................................................... x
LIST OF TABLES .................................................................... xi
CHAPTER I. INTRODUCTION .................................................... 1
   A. Background of the Study ................................................... 1
   B. Problems Identification .................................................... 5
   C. Problems Limitation ....................................................... 5
   D. Problems Statement ........................................................ 6
   E. The Objective of the Study ................................................. 6
   F. Benefits of the Study ........................................................ 6
CHAPTER II. REVIEW OF RELATED LITERATURE .................. 8
   A. Review on Test ............................................................... 8
      1. The Definitions of Test .................................................. 8
2. Kinds of Test ................................................................. 11
3. Types of Test Item ......................................................... 15
4. Multiple-Choice Item ..................................................... 18
5. Essay Item ................................................................. 24

B. Review on Item Analysis ................................................. 27
   1. Basic Concepts of Item Analysis .................................. 27
   2. Formulations and Examples ....................................... 29

CHAPTER III. RESEARCH METHODOLOGY ......................... 40
   A. The Method of Research ........................................... 40
   B. Time and Place of The Study ..................................... 41
   C. Subject of the Study ................................................ 41
   D. Technique of Collecting the Data ............................... 43
   E. Technique of Analyzing the Data ............................... 43

CHAPTER IV. RESEARCH FINDINGS ................................. 47
   A. Analysis on Multiple-Choice Item ............................... 48
   B. Analysis on Essay Test Item ..................................... 85
   C. Research Findings .................................................. 90

CHAPTER V. CONCLUSION, IMPLICATION, AND SUGGESTION 94
   A. Conclusion ........................................................... 94
   B. Suggestion ............................................................ 95

BIBLIOGRAPHY .............................................................. 96

APPENDICES ................................................................. 98
LIST OF APPENDICES

Appendix 1. Worksheet ................................................................. 99
Appendix 2. Answer Key ............................................................... 106
Appendix 3. Students’ answer sheet ............................................. 107
Appendix 4. Letter of Permission .................................................. 160
LIST OF TABLES

Table 4.1. Result of Item Analysis on Multiple Choice Item ....................... 90

Table 4.2. Result of Item Analysis on Essay Test Item ............................ 92
CHAPTER I

INTRODUCTION

A. Background of The Study

Evaluation is one of key elements in the curriculum model beside initial planning procedures (consisting of data collection and learner grouping), content selection and gradation, methodology (including the selection of materials and learning activities), and ongoing monitoring, and assessment.

Shaw and Dowsett in Nunan (1997:7) say that evaluation is the final component in the curriculum model. Traditionally, evaluation occurs at the final stage in the curriculum development process. Evaluation is parallel with other curriculum activities and may occur various times during the planning and implementation phases, as well as during a specified evaluation phase.

Evaluation, including evaluation in the teaching of English as a second or foreign language, concerns primarily with decision making. Generally, the aim of second language evaluation is to improve the teaching and enhance the learning process. The decision made through the evaluation is based on informed judgment. According to Genesee and Upshur (1996:4) the three essential components of evaluation are information, interpretation, and decision making. Information about teaching and learning becomes meaningful when it is interpreted. Meaningful interpretations are necessary in order to decide the actions to take or the changes to make in the instruction. These will lead into different
kinds of decision as what Genesee and Upshur (1996:3) say in the following quotation:

...second language evaluation involves many different kinds of decisions: decisions about the placement of individual student in particular streams, levels, or courses of instruction; about ongoing instruction; about planning new units of instruction and revising units that have been used before; about textbooks or other materials; about student homework; about instructional objectives and plans; and about many other aspects of teaching and learning. There is more to evaluation than grading students and deciding whether they should pass or fail. In fact, decisions about students, although important, are few in number compared to the full range of decisions that are made daily in second language classrooms.

According to Ngalimun Purwanto (2006:3) an educational evaluation is the estimation of the development and progress of pupils toward objectives or values in the curriculum. The aim of evaluation is to gain data or information that shows the level of ability and the success of students in achieving the curricular objectives.

Different ways are available for collecting information. Tests are considered useful means for gathering information about student’s achievement but cannot be applied for gathering any of other types of information. Other techniques of collecting data are suitable for these. For example, classroom observation and student conferences can help teacher to collect information about the strategies used by the students to read or write in the second language; dialogue journals can shed light on students attitudes toward their learning experience in class. School records, curriculum documents, or other instructional materials reveal facts about the school system physical and personnel resources as
well as about the course of instruction itself. In short, there are two ways to collect information about student’s learning i.e., test and non-test evaluation.

Evaluation method that will be the focus in this study is evaluation with test, due to the fact that tests are considered the most familiar way used in evaluating student’s learning. Test as a widely-used technique in evaluating students learning is actually used not only in Indonesia, but also in many other countries such as in what Linn and Gronlund (2000:1) say as follows:

Accountability demands, state, national, and international assessment programs; national content and performance standards; and global competition— all contribute to increased demands for testing and assessment. These factors have both stimulated and reflected new trends in educational measurement. The increased reliance on testing and assessment as educational reform tools has also raised issues concerning the fairness of their uses and interpretation.

Test and assessment can be used to fulfil various aims such as monitoring students’ achievement, providing school or system accountability, reporting to parents, or making decisions about individual students e.g., grade-to-grade promotion or school graduation.

Teaching and testing go hand-in-hand. Teachers may often ask questions to make sure whether the given learning materials have been successfully understood by the students or not. In an equal manner, teachers sometimes ask questions to find out whether teaching a particular point is necessary or not. Later, teachers might know why they ask a question: whether it is to teach or to test something.

Test can be of great help in gathering information for teaching evaluation of English as a second or foreign language. Tests, however, are relatively limited
because they can only reveal certain aspects of student achievement; they cannot show us much about the other factors that often appear in the second or foreign language evaluation.

Testing has, traditionally, measured the results of student performance. Testing involves some steps: choosing the representative samples of language; measuring whether a student can use these samples; quantifying this by turning it into a mark or grade; keeping a record of these marks and use this to give an end assessment.

Various kinds of test items are applied in measuring students performance. Students seem to be familiar to objective test items e.g., true-false, completion, matching, short answer, and multiple-choice items. Another familiar type is essay test items which give them chance for answering questions in their own word or ideas.

Multiple-choice test item is the major test type used in Indonesia. There are pros and cons toward the use of this type. In fact, the test is still conducted, and the multiple-choice item, however, is still used. The opposite view toward the use of multiple-choice item can be neutralized by encouraging test publishers to supplement statistical item analysis with a more careful, logical analysis of test items. Sometimes, essay test item is also used in the test, but only in a few numbers (around five numbers). The quality of the test can be achieved when the items fulfil the criteria of a good test item.

The question is now: have the test items fulfilled the criteria? If not, how should the test items be constructed? The questions shown above motivate the
writer to carry out a study entitled “An Analysis On The English Final Test Items For The Second Semester Of Twelfth Grade Student Of SMA Negeri 5 Surakarta In 2008/2009 Academic Year (A Descriptive Study)”

B. Problems Identification

Based on the background of the study above, it would be better to identify the problems that will be investigated in this study. The writer identifies the problems as follows:

1. Have the test items used in final test fulfilled the criteria of a good and appropriate test item?
2. In what aspect does the test item fulfil the criteria?

C. Problems Limitation

The writer needs to determine the limitation of the study in order to conduct the research in a good focus.

1. The data taken and analyzed in this study are the test items and the test results.
2. The research subject of this study is the test items used as the second semester final test for twelfth grade students of SMA Negeri 5 Surakarta in 2008/2009 academic year.
3. The method used in this study is statistical descriptive method using item analysis technique.
D. Problems Statement

The problems in this study are formulated, as follows:

1. Have the English final test items for the second semester of twelfth grade student of SMA Negeri 5 Surakarta in 2008/2009 academic year fulfilled the criteria of a good test item?

2. In what aspect is the test item considered appropriate or inappropriate?

E. The Objective of the Study

The objective of this study is to describe the appropriateness of the English final test items for the second semester of twelfth grade student of SMA Negeri 5 Surakarta in 2008/2009 academic year. The appropriateness of the test items is focused on the standard determined in the item analysis including level of difficulty, discriminating power, responses and the effectiveness of an item.

F. Benefits of the Study

The research result is expected to be able to give some benefits, such as the following:

1. For the teacher, it is hoped that this study will provide information for the improvement of the quality of evaluation process especially through testing using multiple-choice and essay types.

2. For the test-publisher, it is hoped that this study will be an input on how a good test item should be created and is considered as a way of reflection before creating test items for the next evaluation.
3. For the researcher, this research of course will improve the writer’s writing and sharpen the intuition and ability to analyze a test.

CHAPTER II
REVIEW OF RELATED LITERATURE

A. Review on Test

1. The Definitions of Test

Before coming to the specific term of test, it will be better to think of a case dealing with the test; sometimes the term ‘assessment’ and ‘test’ are confusing since those possibly are included in a single process. According to Linn and Gronlund (2000:31) assessment is a broad term which covers the full range of procedures applied to obtain information about students’ learning (observation, ratings of performances or projects, paper-and-pencil tests) and the value judgments formation focusing the progress of learning. A test is a specific type of assessment which typically consists of a set of questions conducted during a fixed period of time under fairly comparable conditions for all students. People sometimes put the two terms as a similar definition although test is a specific type of assessment.

In Indonesian education system, a test seems to be the basis of the decision making of student graduation such as UAS (Ujian Akhir Semester) or UN (Ujian
Unfortunately, protests against the use of tests as the criterion of graduation aroused when some known smart students failed. They protested that the test is an unfair way to make an important decision such as the student graduation.

On the other hand, some students gave no complaint toward the use of the test. They still considered the test as an objective way and there was nothing wrong about the test.

The phenomenon above is in line with the pros and cons toward the test. The ‘cons’ side consider tests representative enough in measuring human potential. The opposite view on testing can be figured through the article from ETS Developments in Linn and Gronlund (2000:18) as follows:

Ability and academic achievement occupy an Olympian perch on the prestige ladder. Yet it is widely agreed that motivation, creativity, personal honesty, intuition, even the degree of social consciousness play significant roles in the struggle for the most cherished of American ideals—“success in life”. Admission tests thus measure a relatively narrow segment of the human potential. (ETS Development, 1979).

While the ‘pro’ side of testing gives an argument toward the opposite view above by stating it in this writing:

If the use of educational tests were abandoned, the encouragement and reward of individualized efforts to learn would be more difficult. Excellence in programs of education would become less tangible as a goal and less demonstrable as an attainment. Educational opportunities would be extended less on the basis of aptitude and merit and more on the basis of ancestry and influence, social class barriers would become less permeable. Decisions on important issues of curriculum and method would be made less on the basis of solid evidence and more on the basis of prejudice or caprice. These, it seems to us, are likely to be the more harmful consequences, by fair. Let us not forgo the wise use of good test. (Ebel, 1980:34-35)
Although pros and cons exist in the use of test, the government still considers test as an effective way to measure students’ learning. Based on KTSP curriculum the evaluation on the students’ learning is conducted by the school by means of tests, but the schools or academic institutions are allowed to make their own test item. The test items should be in line with the curriculum, and the content should not deviate from the syllabus.

The term of test has often been used in the discussion above. Generally, people suppose test as a way of evaluation. Some might say that test has various types based on the purpose, the time, or any other reasons.

Baxter (13:1997) states that the relationship between testing and evaluation is the same as the relationship between the curriculum and its syllabus as described in the figure below:

![Diagram showing the relationship between curriculum, syllabus, testing, and evaluation]

Baxter considers evaluation as wider than testing. Test is considered a useful tool of evaluation.

Genesee and Upshur (1996:154) define a language test as a set of tasks requiring observable responses to language or in language that can be scored and interpreted with reference to norms, domains or instructional objectives. Therefore, the selection of testing method is paramount in reflecting the language
skills intended to test and trying not to contaminate test performance through avoiding irrelevant skills or knowledge.

Further, Genesee and Upshur (1996:141) describe particular essential aspects to consider when devising and using the test due to the difficulty in distinguishing tests from other method of collecting information. A test is, first of all, about something i.e., intelligence, or European history, or second language proficiency. In educational terms, tests possess subject matter or content. Second, a test is a task or set of tasks that elicits observable behaviour from the test taker. Different test tasks have different methods of eliciting performance. Third, tests yield scores that show attributes or characteristics of individual. To interpret the test scores a frame of reference is necessary, and the process is called measurement. Thus, the tests are a form of measurement. In conclusion, tests have three aspects: content, method, and measurement.

Cronbach in Suke Silverius (5:1991) define test as a systematic procedure for observing and describing one or more characteristics of a person with the aid of either a numerical scale or a category system.

Based on the definitions above, test is a specific type of assessment which typically consists of a set of questions conducted during a fixed period of time under fairly comparable conditions for all students.

2. Kinds of Test

There are some kinds of test in measuring the students learning. Different experts give different classification of tests as will be shown below.
Baxter (1997:8) says that there are many ways of assessing students learning and the most common method is a test with the following classification:

a. Proficiency tests: tests which examine a general standard in ability, regardless of the teaching program.

b. Achievement tests: tests which examine whether students can do what they have been taught, either by testing specific syllabus items or general objectives.

c. Placement tests: tests which are the mixture of proficiency and achievement tests, depending on what criteria we use to place the student.

d. Diagnostic tests: tests which use proficiency or achievement tests to analyze strengths and weaknesses in the student or teaching program itself.

There are some other classifications of tests given by Genesee and Upshur (1996:152). Although they do not propose the classification of tests, they summarize the second language tests as the division below to alert the reader to the alternative terms for talking about tests so that we will recognize them in the independent reading and understanding.

The second language tests have been classified according to whether they focus on:

a. Underlying linguistic competence

b. Specific aspects or sub skills of language

c. A specific testing method
Based on the underlying linguistic competence, there are many tests that can be held to measure the competence that second language theoreticians and testers have been interested in. This may include grammatical competence, pragmatic competence, sociolinguistic competence, and communicative competence.

Linguistic sub skills are often described in terms of the structural or grammatical features of language – spelling, vocabulary, grammar (or syntax), pronunciation, etc. These kinds of tests refer to the test content and consist of discrete-point tests and integrative tests. Discrete-point tests focus on one specific aspect of language. Integrative tests call on a number of sub skills operating in concert.

Dealing with the method of testing, tests can be classified into open-ended, closed-ended, and limited-response formats.

Based on types of information provided by the tests, there are four types of tests: achievement tests, proficiency tests, performance tests, and diagnostic tests. Achievement tests give information on students’ accomplishment dealing with a certain instructional objectives. Proficiency tests provide information about student’s ability to use language in a particular way. Performance tests obtain information about students’ skill in applying the language to perform authentic task. Diagnostic tests yield information about students’ relative strengths and weaknesses.
Another category of test types describes the kinds of decisions that can be made using the results such as placement tests, final examinations, summative tests, and formative tests.

A simpler and broader classification of tests is given by Thorndike and Hagen (1977:5). They classify tests based on the aspects of the individual relevant to decisions. The type of ability test that describes what a person has learned to do is called an achievement test. Tests concerned with mastery of defined skills are often called content-referenced or criterion-referenced tests, because the focus is solely upon reaching a standard of performance on the specific skill called for by the test exercises. A test that calls for information about what a person can learn to do and is used as a predictor of some future performance is spoken of as an aptitude test. A test which point of reference is the performance in a group is called norm-referenced test because the norm of acceptable performance is set by group comparison.

Test can also be classified by its use as stated by Orlich et al. (1998: 366) as follows:

a. Placement tests (pre-tests), to determine if student has prerequisite skills to begin instruction.

b. Diagnostic tests, to determine causes (physical, intellectual) of persistent learning problems.

c. Formative tests, to monitor learning progress, provide feedback to reinforce learning, and correct learning errors.
d. Summative tests, to determine final achievement for assigning grades or certifying mastery.

There are test types used in evaluating the students with various categorization i.e., based on its use, its focus, and its aspect of the individual relevant to the decision. Dealing with this study, test which becomes the concern is an achievement test which is conducted at the end of course (summative test at once) using norm-referenced criteria and involving some linguistic competence.

3. Types of Test Item

Students find various types of test in the class. In the weekly assessment, for instance, the students may find a test which instructions are like the following: fill in the brackets, answer briefly, match the words, or choose the best answer. Students also possibly meet such instructions say the words in the correct pronunciation, discuss the problem with your friend, present your idea, or practice the dialogue in your own idea. In short, there are various types of test item. These common types of test are actually the application of concept given by the experts in education measurement.

According to Linn and Gronlund (2000:167) the major categories of objective tests or performance assessments may be subdivided further into some basic types of test items and assessment tasks as follows:

**Objective Test**

a. Supply type
   1. Short answer
   2. Completion
b. Selection type
   1. True-false or alternative response
2. Matching
3. Multiple choice

**Performance Assessment**

a. Extended response
b. Restricted response

Objective tests provide students a highly structured task which limits their response to filling a word, brief phrase, number, or symbol to choose the answer from among a given number of alternatives.

The ‘supply’ type requires an exactly specific answer and is usually in the form of short answer. It can be filling in a missing element, completing a sentence or answering a question with a short answer or a brief answer which the students are supposed to find by themselves.

The ‘selection’ type provides the students with the intended answer. The students are asked to select the correct or the best answer. The selection type consists of true-false, matching, and multiple choice. True-false item gives the students a task to choose whether a statement is true or not. In the matching type students are asked to match a statement with a certain word, phrase, picture or even a statement which have the intended characteristic required by the instruction. The multiple-choice item provides several options to answer the question. It can be 3, 4 or 5 options provided in the item.

Performance assessments allow students to answer by selecting, organizing, and presenting or performing ideas in a way they thought to be appropriate.

Almost similar to the previous explanation, Thorndike and Hagen (1977:210) put the different types of items into two major categories: (1) the type
in which the student produces his own answer and (2) the type in which the student selects his answer from answer choices provided by the test maker.

The first type includes the essay question requiring an extended answer from the student; the short-answer question requiring no more than one or two sentences for an answer; and the completion item requiring only a word or phrase for an answer. The second type consists of alternative response item such as the true-false statement; the multiple choice item; and the matching item.

Meanwhile, Genesee and Upshur (1996:172) classify the types of test item based on the response given to the tasks into the following categories:

Closed-ended
- Cloze
- Elicitation
- Judgment (e.g., true-or false)
- Multiple-choice

Limited-response
- Dictation
- Imitation
- Transformation
- Rearrangement
- Matching

Open-ended
- Interviews
- Essays
- Information gap
- Oral or written reports

On the other hand, Orlich et al. (1998:368) categorize types of test item into the following classification:

a. Short-answer, matching, and true-false items for measuring knowledge-level outcomes.

b. Multiple-choice items for measuring both knowledge-level and more complex learning outcomes.
c. Interpretive items for assessing complex, higher-level objectives
d. Essay items for assessing higher-level outcomes.
e. Performance items for measuring complex achievement.

The briefer classification of test item is given by Henson and Eller (1999:470) into two general types: objective and essay items. Some of the more familiar objective tests include multiple-choice, matching, true-false, and filling in the blank.

It can be summarized that test item types consist of objective tests and performance assessment. Objective tests include short answer, completion, true false or alternative response, matching, and multiple-choice. While performance assessment consists of essay with restricted and extended response. The test items used in this study are multiple-choice and essay item with restricted response.

4. Multiple-Choice Item

Multiple choice can be one of the various types of test item. This type of test item is considered the most useful and familiar type and widely used of objective test item. It also can measure both knowledge and level of understanding and is free of many limitations of other forms of objective item. Related to the types of test item, Hoffman in Linn and Gronlund (2000:18) contended that multiple-choice items penalized more intelligent, original thinkers. While Frederiksen, also still in Linn and Gronlund (2000:18) has argued that multiple-choice items place too much emphasis on “well-structured problems” when problems of greatest interest both in and out of school are often “ill-structured”
and when skills such as problem identification and hypothesis generation are often as important as problem solution.

Linn and Gronlund (2000:193) explain that a multiple-choice item consists of a problem and a list of suggested solutions. The problem can be in the form of a direct question or an incomplete statement. This is called stem of the item. The list of suggested solutions may consist of words, numbers, symbols, or phrase and refer to alternatives (also called choices or options). The student is usually asked to read the stem and the list of alternatives and to choose the one correct, or best, alternative. The correct alternative is called the answer, and the other alternatives are called distracters (also called decoys or foils). The term distracters came from their intended function – to distract those (students) who are unsure about the correct answer.

Almost similar to the previous opinion, Ahmann and Glock (1971:95) give their statement about multiple choice test items as follow:

A multiple-choice test item is one in which a direct question or incomplete statement is presented and a number of possible responses or options are given. The pupil chooses the response that is the correct (best) answer to the question or that is the correct (or best) expression for completing the statement. The question or incomplete statement introducing the test item is known as the stem. Any undesired answer is called a distracter or foil. Generally four or five responses are listed and all but one is a distracter.

According to Linn and Gronlund (2000:195) the multiple-choice is the most versatile type of test item available. It can measure a variety of learning outcomes from simple to complex, and is adaptable to most type of subject-matter content. The uses of multiple-choice items in measuring learning outcomes cover the following outcomes:
a. Measuring Knowledge Outcomes

1.) Knowledge of Terminology

2.) Knowledge of Specific Facts

3.) Knowledge of Principles

4.) Knowledge of Methods and Procedures

b. Measuring Outcomes at the Understanding and Application Levels

1.) Ability to Identify Application of Facts and Principles

2.) Ability to Interpret Cause-and Effect Relationships

3.) Ability to Justify Methods and Procedures

The following is a brief discussion dealing with the uses of multiple-choice item above. Some examples are also provided.

a. Measuring Knowledge Outcomes

1.) Knowledge of Terminology

To measure their knowledge, students are requested to show their knowledge of a specific term by choosing a word that has similar meaning to the given term or by deciding a definition of the term. Students can also be asked to identify the meaning of the term when used in context.

Example:
Which one of the following words has the same meaning as the word *egress*?

a. Depress
b. Enter

Which is the correct answer?

c. Exit ✓
d. Regress

(Linn and Gronlund. 2000:181)

2.) Knowledge of Specific Facts
Item which measures this knowledge outcome is important and it provides a necessary basis for developing understanding, thinking skills, and other complex learning outcomes. The most common questions are in the form of the *who, what, when, and where* variety.

Example:
When did a United States astronaut first orbit the earth in space?

- a. 1960
- b. 1961
- c. 1962 √
- d. 1963

(Linn and Gronlund. 2000:181)

3.) Knowledge of Principles

Multiple-choice items can be written to measure knowledge of principle as easily as those designed to measure facts.

Example:
The principle of capillary action helps explain how fluids
- a. enter solutions of lower concentration.
- b. escape through small openings.
- c. pass through semi permeable membranes.
- d. rise in fine tubes. √

(Linn and Gronlund. 2000:182)

4.) Knowledge of Methods and Procedures

This includes such diverse areas as knowledge of laboratory procedures; methods underlying communication, computational, and performance skills; methods used in problem solving; governmental procedures; and common social practices.

Example:
If you were making a scientific study of a problem, your first step should be to
- a. collect information about the problem. √
- b. develop hypotheses to be tested.
- c. design the experiment to be conducted.
- d. select scientific equipment.
b. Measuring Outcomes at the Understanding and Application Levels

1.) Ability to Identify Application of Facts and Principles

A common method to measure this ability is to ask students to identify a correct application in a situation that is new to the student. Application items measure understanding but they also include the ability to transfer learning to situations that have not been previously studied.

Example:
Directions: in each of the following sentences underline the word that makes the sentence correct.

1. This is the boy **that** asked the question.

2. This is the dog **that** he asked about.

2.) Ability to Interpret Cause-and Effect Relationships

Students’ understanding of cause-and-effect relationships can be measured by providing them with a particular cause-and-effect relationship and asking them to identify the reason that best accounts for it.

Example:
Bread will not become mouldy as rapidly if placed in a refrigerator because
a. cooling retards the growth of fungi. √
b. darkness retards the growth of mould.
c. cooling prevents the bread from drying out so rapidly.
d. mould requires both heat and light for best growth.

3.) Ability to Justify Methods and Procedures
A student might know the correct method or sequence of steps in carrying out a procedure, without being able to explain why it is the best method or sequence of steps. The concern of this case is the students’ ability to justify the use of a particular method or procedure. This can be measured with multiple-choice items that ask the students to choose the best of some possible explanations method or procedure.

Example:
Why do farmers rotate their crops?
- a. to conserve the soil. √
- b. to make marketing easier.
- c. to provide for strip cropping.
- d. it removes the brownish yellow colour.

(Linn and Gronlund. 2000:185)

The selection of form used in a stem, whether it is a direct question or incomplete statement, depends on several factors. The direct question form is easier to write, more natural for younger students, and more likely to present a clearly formulated problem. On the other hand, the incomplete statement is more concise, and if skilfully phrase, it also can present a well-defined problem.

Examples:

(Incomplete statement form)
1. The capital of California is in
   - a. Los Angeles
   - b. Sacramento
   - c. San Diego
   - d. San Fransisco

(Direct-question Form)
2. Which one of the following factors is given most consideration when selecting a city for a state capital?
   - a. Location
   - b. Climate
   - c. Highways
d. Population  

(Linn and Gronlund. 2000:179)

In the first example, there is only one correct answer. The capital of California is in Sacramento and nowhere else. All other options (alternatives) are wrong. This is known as the one-correct-answer type of multiple choice items.

When using questions containing who, what, when, and where variety, answers of varying degrees of acceptability are the rule. Questions with a number of possible reasons or procedures (as in example 2) are considered best-answer type of multiple choice items.

In short, multiple-choice item consists of stem and alternatives where the alternatives themselves have the answer and distracters. This type of test item may appear in two forms: incomplete statement and direct question. The answers can be the correct answer or the best answer.

5. Essay Item

Besides multiple-choice item, there is one familiar type of test item used by schools or academic institution i.e., essay test item. It usually consists of five numbers of questions. This type gives an opportunity for the students to answer the questions using their own words or based on their knowledge.

According to Linn and Gronlund (2000:235) open-ended questions may be the best way to measure some important learning outcomes. Essay questions give the freedom of response which is needed to sufficiently assess the ability of students in formulating problems; organizing, integrating, and evaluating ideas and information; and applying knowledge and skills.
Essay questions should be used primarily to measure learning outcomes that are not readily measured by objective test item. These include such outcomes as the ability to recall, organize, and integrate ideas; the ability to express oneself in writing; and ability to create rather than merely identify interpretations on data applications.

The response freedom provided by essay questions is not an all-or-nothing affair, but, rather, a matter of degree. At one side, the response is almost as restricted as that in the short-answer objective item, in which a sentence or two may be all that is required. At the other side, students are given almost complete freedom in constructing their responses.

Essay questions can be classified into two types: restricted-response questions and extended-response questions or assignments.

The following explanation dealing with those essay questions is given by Ahmann and Glock (1971:160).

a. Extended response

Extended response essay question can be very challenging for the students. In order to give a correct response, students must show their ability to organize, evaluate, write clearly, and be creative. Therefore, the responses of the test items display how well the students have achieved important educational goals. Free expression from the student can be analyzed to gain information about his motives, values, interests, modes of adjustment, and the like.

The following is an example of extended response test item: Write a short composition on the topic “What Conversation Means to Me”!
Students’ responses to such test item compared to the restricted one are more complex and varied. Presumably they provide helpful information about the student’s difficulties and patterns of thought that can be used as a partial basis for remedial instruction. In this type of test item, the scoring of student responses is difficult and the reliability is low. The practical value in this type of test item is limited by the inability of teacher to score them reliably.

b. Restricted response

This type of essay item differs from the first type i.e., extended response test item, in that the perimeter of pupil response is better defined. Further, Ahmann and Glock (1971:161) give more explanation about the features of this type in the following quotation:

… A specific problem is presented. It requires the pupil to recall the proper information, organize it in a suitable manner, arrive at a defensible conclusion, and express it in his own words. In several important respects, it requires the pupil to reveal abilities much like those required by a satisfactory answer to an extended response essay test item; however, this display must occur within well-defined restrictions. These restrictions simplify the scoring problem, thereby greatly improving the reliability of the scoring.

It can be concluded that essay item has two forms which require different way of answering the questions i.e., restricted response and extended response. Essay item involves student skills in formulating problems; organizing, integrating, and evaluating ideas and information; and applying knowledge and skills.
B. Review on Item Analysis

1. Basic Concepts of Item Analysis

An analysis toward student’s response to the test item determines the effectiveness of each test item. The analysis is familiarly known as item analysis. The quality of a test item could be analyzed through its power to discriminate the upper to lower students, it also has a certain level of difficulty. Besides the two requirements, an item should have alternatives or options that work. Some experts give concepts and methods in doing this analysis in various ways. They also propose the importance or benefit of doing the analysis.

A brief concept is given by Ahmann and Glock (1971:184) in the following quotation:

Reexamining each test item to discover its strengths and flaws is known as item analysis. Item analysis usually concentrates on two vital features: level of difficulty and discriminating power. The former means the percentage of pupils who answer correctly each test item; the latter the ability of the test item to differentiate between pupils who have done well and those who have done poorly.

Thorndike and Hagen (1977:251) state that an analysis of the test item serves two important purposes. First, it provides diagnostic information for studying the learning of the class and the failures to learn and for guiding further teaching and study. Second, it provides a basis for preparing better test another year.

Tabulation is needed in the basic analysis of each item on the test. We need to know how many students got each item correct, how many chose each of the possible incorrect answers, and how many omitted the item. In the tabulating
process, dividing the groups into three fractions i.e., upper, lower, and middle would be helpful.

Tabulation helps us to answer questions as the following for each item:

1. How hard is the item?
2. Does it distinguish between the better and poorer students?
3. Do all the options attract response, or are there some that are so unattractive that they might as well not be included?

Another discussion on item analysis is given by Linn and Gronlund (2000:362) who state that item analysis is usually associated with a norm-referenced perspective. The results of an item analysis help to select items of desired difficulty that best discriminate between high-and-low achieving students and to recognize faulty items. It also provides information about student misconceptions and topics that need additional works.

They also explain that item analysis is usually designed to answer such questions as the following:

a. Did the item function as intended?
b. Were the test items of appropriate difficulty?
c. Were the test items free of irrelevant clues and other defects?
d. Was each of the distracters effective (in multiple-choice item)?

There are a number of benefits of item analysis as follows:

a. Item-analysis data provide a basis for efficient class discussion of the test results.
b. Item-analysis data provide a basis for remedial work
c. Item-analysis data provide a basis for the general improvement of classroom instruction
d. Item-analysis procedures provide a basis for increased skill in test construction.

Based on the concepts above, item-analysis is a systematic procedure in analyzing the test item to find the strengths and weaknesses by computing the level of difficulty, discriminating power, and the function of distracters. This analysis provides two benefits i.e., to give diagnostic information and to give a basis for a better preparation in constructing the next test.

2. Formulations and Examples

The first steps of item analysis procedure are merely a convenient tabulation of student responses from which we can determine item difficulty, item discriminating power, and the effectiveness of each distracter. To obtain a precise estimate of item difficulty and discriminating power, relatively simple formulas to the item-analysis data are given. The following theories contain the formulas stated by Linn and Gronlund (2000:365).

**Computing Item Difficulty**

A test item difficulty that is scored right or wrong is indicated by the percentage of students who get the item right. Hence, we can compute item difficulty (P) by using the following formula, in which R equals the number of students who got the item right, and T equals the total number of students who tried the item.
Upper and lower group become the basis of calculation in computing the item difficulty from item analysis data.

**Computing Item Discriminating Power**

An item discriminates positively if more students in the upper group than the lower group get the item right.

The discriminating power of a test item refers to the degree to which it discriminates between high and low achievement students. Item discriminating power (D) can be obtained by subtracting the number of students in the lower group who get the item right (RL) from the number of students in the upper group who get the item right (RU) and dividing by one-half the total number of students included in the item analysis (.5T). The formula is summarized as follow:

\[
D = (RU - RL) / (0.5T)
\]

**Evaluating the Effectiveness of Distracters**

Generally, a good distracter attracts more students from the lower group than the upper group. Thus, it should discriminate between the upper and lower groups in a manner opposite to that of the correct alternative. The illustration is given in the data below:

(Alternative A is the correct answer)

<table>
<thead>
<tr>
<th>Alternatives</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Omits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper 10</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lower 10</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>
First, note that the item discriminates positively, because 5 in the upper group and 3 in the lower group got the item right. The index of discriminating power is fairly low, however (D=20), and this possibly partly caused by the ineffectiveness of some of the distracters. Alternative B is a poor distracter because it attracts more students from the upper group than from the lower one. This is most likely because of some ambiguity in the item statement. Alternative C is not a plausible distracter because it attracted no one. Alternative D is functioning as intended because it attracts a larger number of lower group students.

Application of Item-Analysis Principles With Performance-Based Assessments

Procedures of item-analysis have a quite restricted applicability with performance-based assessments, mainly because such assessments generally consist of a relatively small number of tasks. If the assessment has several tasks, however, the general principles can be readily adapted for use. One necessary modification results from the fact that scores on the performance-based tasks almost always involve more than a simple 0 or 1. For example, each task might have possible scores of 0, 1, 2, 3, or 4.

Still a comparison of the individual task scores for the 10 highest and 10 lowest scoring students can be useful.

Suppose, for example, that a performance assessment consisted of five separate tasks, each of which had possible scores ranging from a low 0 for no response or a response that was unrelated to the task to, a 4 for a complete and
well-elaborated response. Possible total scores for the set of five tasks would range from 0 to 20. The total scores then are ranked to identify the 10 lower and 10 higher. The average score is then calculated.

The following is an example of the applied analysis above:

<table>
<thead>
<tr>
<th>Task 1</th>
<th>Score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper 10</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>3.4</td>
</tr>
<tr>
<td>Lower 10</td>
<td></td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task 2</th>
<th>Score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper 10</td>
<td></td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>Lower 10</td>
<td></td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1.5</td>
</tr>
</tbody>
</table>

The higher average score for the upper students than the 10 lower achievement students shows that the task discriminates between two groups. While the task 2 indicates that there is no discriminating power in the task.

Almost similar formulation is also given by Ahmann and Glock (1971:184). It is in line with their concept that item analysis concentrates on the level of difficulty and discriminating power. In fact, they also concern on the works of all options to the item.

**Item Difficulty**

It is easy to determine a test item’s level of difficulty. First, tabulation is made of the number of pupils who successfully answer the item. This figure is then divided by the total number of pupils attempting the item, and the quotient is multiplied by 100. These steps are summarized in the following pattern:
Where

\[
P = \frac{\text{NR}}{\text{NT}} \times 100
\]

\( P \): percentage of pupils who answer the best item correctly;
\( \text{NR} \): number of pupils who answer the test item correctly
\( \text{NT} \): total numbers of pupils who attempt to answer the test item.

Thorndike and Hagen in Anas Sudijono (2005:372) give an interpretation of difficulty index in three range i.e., less than 30% as very difficult, 30%-70% as satisfactory, and 70% as very easy.

**Item Discriminating Power**

The discriminating power of a test item refers to its ability to differentiate between pupils who have achieved well (the upper group) and those who have achieved poorly (the lower group). The aim is to specify the characteristics of the upper and lower group. We can use an independent criterion such as score from standardized test considered to measure the same achievement aspects or from a final mark in the similar achievement areas. On the other hand, an internal criterion may be used, such as the total scores from the classroom achievement test.

When the total test scores are used in this manner, a decision must be made as to which part of the distribution of scores is the upper group and which part is the lower group. Some choose the upper and lower halves, others the upper and lower thirds with the middle third discarded, and still others the upper and lower 27 percent with the middle 46 percent discarded.
The discriminating power of a test item is the difference between the number of correct and incorrect discriminations expressed as a percentage of the maximum possible correct discrimination. In simplified form, the formula to compute an index of discriminating power based on this idea is following:

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

Where

- \( D \): index of item discriminating power
- \( U \): number of pupils in upper group who answer the test item correctly
- \( L \): members of pupil in the lower group who answer the test item correctly
- \( N \): number of pupils in each of the two groups.

The maximum size of the index is +1.00 and the minimum size is -1.00. Any negative value means that the test item discriminates—to some degree—in the wrong direction and is not satisfactory. Positive values show that the test item discriminates in the desired direction, even though it may not be complete satisfactory. The larger the positive value the better. Any \( D \) values above +0.40 can be considered very good, any between +0.40 and +0.20 satisfactory, and any between +0.20 and zero poor.

**Examining all responses to an item**

The effectiveness of test items can be investigated by tabulating and comparing the entire responses of the pupils in the upper and lower groups.
multiple-choice test item, for instance, the responses to each of the distracters—as well as the correct responses—are counted.

To get a clearer understanding in this case, the following are examples of test item and the analysis of the responses of the item. This example is administered on English class concerning library skills.

If you wish to find quickly the page on which a particular topic or subject appears in a reference book, to which of the following would you refer?

(1) in the appendix  
(2) the subject index  
(3) the table of contents  
(4) the bibliography

The responses of the pupils in the upper and lower groups are as follows:

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

This test item is too easy; the estimate of P is 82 percent. The discriminating power is poor (D= +0.18).

Probably the weakness in test item is caused by the inclusion of the word “subject” in the second option. This may be a clue to the correct response. It is so obvious that students notice it easily and choose their answer accordingly. The uninformed student will profit more by the clue than the informed student who didn’t need one in the first place. If the word were deleted, it is conceivable that the differential attractiveness of the distracters would increase, thereby improving the difficulty level and discriminating power of the item.
Thorndike and Hagen (1977:253) also give some examples of a simple and brief analysis on test item given to 100 high school seniors who had had a course in current American problems. The top 25 papers represent the upper group and the last 25 papers represent the lower group. They give the examples as follows:

EXAMPLE
Item 1: “Everyone’s switching to Breath of Spring Cigarettes!” is an example of the propaganda technique called

<table>
<thead>
<tr>
<th></th>
<th>Upper</th>
<th>Lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. glittering generality</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>B. bandwagon</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>C. testimonial</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>D. plain folk</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

(Omit)

This is an easy item, since all 25 in the upper group and 20 in the lower group get it right. However, it does differentiate in the desired direction, since the wrong answers fall more in the lower group. The item is also good in that all of the wrong answer choices do work; that is, each wrong answer has been chosen by one or more persons in the lower group. Two or three easy items like this would be good ice-breakers with which to start a test.

EXAMPLE
Item 2: There were no federal income taxes before 1913 because prior to 1913

<table>
<thead>
<tr>
<th></th>
<th>Upper</th>
<th>Lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. the federal budget was balanced</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>B. regular property taxes provided enough revenue to run the government</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>C. a tax on income was unconstitutional</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>D. the income of the average worker in the U.S. was too low to be taxed</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

(Omit)
Item 2 is a difficult item but a very effective one. That it is difficult is shown by the fact that only 13 out of 50 got it right. The effectiveness of this item is shown by the fact that all 13 getting the item right were in the upper group. All of the wrong options attracted some choices in the lower group and all of the wrong options attracted more of the lower group than the higher group. Incidentally, an item such as this shows how faulty the idea of “blind guessing” often is when an item is effectively written. In this item the majority of the lower group concentrated upon one particular wrong option that was particularly plausible and appealing.

EXAMPLE

Item 3: Under the “corrupt practices act” the national committee of a political party would be permitted to accept a contribution of

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper</th>
<th>Lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. $10,000 from Mr. Jones</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>B. $1,000 from the ABC Hat Corporation</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>C. $5,000 from the National Association of Manufacturers</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>D. $500 from union funds of a local labor union</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>(Omit)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

This item turned out to be poor. First, the item is very difficult; only 8 out of the 50 or 16 percent of the students got it right. Second, the item is negatively discriminating, that is, correct answers were more frequent in the lower group than in the upper group. There are two possible explanations for the item analysis data: (1) the item was ambiguous especially for students who knew the most or (2) the students have not learned the provisions of the “corrupt practices act”. There are two things that point to the second as the more probable explanation; the concentration of responses of the upper group at option A and the
apparently random responses of the lower group that indicate random guessing. In order to arrive at the correct answer to the item the student would have to know (1) the limit placed on contributions to the national committee of a political party, (2) who is forbidden to make contributions, and (3) what kind of organization the National Association of Manufactures is. The teacher would have to discuss the item with the class to determine where the difficulty lies but one might guess that it is points 1 and 3 that are causing difficulty in the upper group.

EXAMPLE
Item 4: The term “easy money” as used in economics means

<table>
<thead>
<tr>
<th>Choice</th>
<th>Upper</th>
<th>Lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. the ability to borrow money at low interest rates</td>
<td>21</td>
<td>17</td>
</tr>
<tr>
<td>B. dividends that are paid on common stocks</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. money that is won in contests</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D. money paid for unemployment compensation</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>(Omit)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

This item shows some discrimination in the desired direction (21 versus 17), but the differentiation is not very sharp. The response pattern is one that is quite common. Only two of the four choices function well. Nobody selects either the B or C options. If we wished to use this item again, we might try substituting “wages paid for easy work” for option B and “money given to people on welfare” for option C. The repeat of the vowel “easy” in option B and the idea of getting money for not working in option C might take the item more difficult and more discriminating.
Item statistics such as these can be used not only for evaluating the items but to guide review and restudy of the material with a class. The item that proves difficult for the class as a whole provide leads for further exploration. Discussion of these items with the class should throw light on the nature of misunderstanding. The misunderstanding may in some cases be cleared up by a brief discussion, although in some cases a fuller review of the topic may be indicated. It is desirable, if local policies permit, to let pupils have their answer key available on them, so that they themselves can use the test as a guide o review and clarification of the points they missed. An examination should teach as well as test.
CHAPTER III
RESEARCH METHODOLOGY

A. The Method of Research

This study aims to describe the appropriateness of final test item used in the second semester of twelfth grade students in SMA Negeri 5 Surakarta. Based on the purpose of the study, the writer uses descriptive method with quantitative approach in conducting the research.

This decision is supported by the idea proposed by Winarno Surakhmad (1990:139):

Penyelidikan deskriptif tertuju pada pemecahan masalah yang ada pada masa sekarang. Karena banyak sekali ragam penyelidikan demikian, metode penyelidikan deskriptif lebih merupakan istilah umum yang mencakup berbagai teknik deskriptif. Di antaranya ialah penyelidikan yang menuturkan, menganalisa dan mengklasifikasi; penyelidikan dengan teknik survey, dengan teknik interviu, angket, observasi, atau dengan teknik test; studi kasus, studi komperatif, studi waktu dan gerak, analisa kuantitatif, studi kooperatif atau operasional. Tidak mengherankan apabila mudah timbul selang surup pengertian atau kekaburan batas istilah tersebut.

Fraenkel and Wallen (2000:13) give a short definition about descriptive study. They state that descriptive studies describe a given state of affairs as fully and carefully as possible.

The theories above are the reason with which the writer chooses the descriptive method to be applied in the study.
B. Time and Place of The Study

The research is conducted in 2008/2009 academic year, started in April and finished in September 2009. The study is conducted in SMA Negeri 5 Surakarta which is located in Jl. Letjen Sutoyo 18 Surakarta.

C. Subject of The Study

To simplify and reduce the complexity that might occur during the research, the writer needs to specify the population, sample, and the technique of sampling.

1. Population

Dalen (1979:128) defines population as a well-defined group or set of human beings or other entities. The size of populations may vary. After identifying the population clearly, a complete, accurate, and up-to-date list of the whole units in the population needs to be constructed.

According to Fraenkel and Wallen (2000:103) population is the group of interest to the researcher, the group to whom the researcher would like to generalize the result of the study.

In this study, the population is the test items and test results used in the second semester final test of twelfth grade students in SMA Negeri 5 Surakarta.

2. Sample

After the population is clearly determined, we need to take the sample in order to simplify the research procedures. Suharsimi Arikunto (2006:131) gives a
brief definition of sample as a part or representative of the investigated population.

Another concept of sample is also given by Fraenkel and Wallen (2000:104) saying that sample is a group in a research study on which information is obtained. The sample is taken in order to overcome the problem time, financial, and energy insufficiency of the researcher. It will be too complicated if all subject or whole population are used in this research. Although sample is just a part of population, it is expected to be representative.

The sample used in this research is the test items and test results taken from 100 twelfth grade students of SMA Negeri 5 Surakarta from class XII A 1, XII A 2, XII S 4 and XII S 5.

3. Sampling Technique

Based on Johnson and Christensen (2000:156) sampling is the drawing process of a sample from the population.

Dalen (1979:128) gives a further explanation that sampling does not consist in collecting data casually from any conveniently located units. He also mentions several steps involved in the process as follows:

a. define the population

b. procure an accurate and complete list of the units in the population

c. draw representative units from the list

d. obtain a sufficiently large sample to represent the characteristics of the population
This research applies stratified random sampling technique. The sample is taken from four classes i.e., XII A 1, XII A 2, XII S 4 and XII S 5.

D. Technique of Collecting The Data

The technique used in this study is documentary. This method is commonly defined as a way or technique used to collect data by taking the existing data or document. It has the same idea as what Suharsimi Arikunto (2006:231) states in the following quotation:

_Tidak kalah penting dari metode-metode lain, adalah metode dokumentasi, yaitu mencari data mengenai hal-hal atau variable yang berupa catatan, transkrip, buku, surat kabar, majalah, prasasti, notulen rapat, lengger, agenda, dan sebagainya. Dibandingkan dengan metode lain, maka metode ini agak tidak begitu sulit, dalam arti apabila ada kekeliruan sumber data masih tetap, belum berubah. Dengan metode dokumentasi yang diamati bukan benda hidup tetapi benda mati._

E. Technique of Analyzing The Data

In line with the purpose of the study, the writer chooses item analysis to be the technique of analyzing the data.

The following is the formula of item analysis suggested by Ahmann and Glock and will be used in the analysis:

**Computing Item Difficulty**

To find the difficulty level of the test item, the following formula is considered appropriate in the analysis.

\[
\text{P} = \frac{\text{NR}}{\text{NT}} \times 100
\]
Where

- **P**: percentage of pupils who answer the best item correctly;
- **NR**: number of pupils who answer the test item correctly
- **NT**: total numbers of pupils who attempt to answer the test item.

Difficulty level of an item is judged through the following range:

<table>
<thead>
<tr>
<th>P</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 0.30</td>
<td>Very difficult</td>
</tr>
<tr>
<td>0.30 – 0.70</td>
<td>Satisfactory</td>
</tr>
<tr>
<td>More than 0.70</td>
<td>Very easy</td>
</tr>
</tbody>
</table>

**Computing Item Discriminating Power**

Before computing the discriminating power, we need to classify the students into three groups i.e., lower, middle, and upper. Upper and lower thirds are used in the computation, while the middle third discarded. The formula to compute an index of discriminating power is as follows:

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

Where

- **D**: index of item discriminating power
- **U**: number of pupils in upper group who answer the test item correctly
- **L**: members of pupil in the lower group who answer the test item correctly
- **N**: number of pupils in each of the two groups.

The maximum size of the index is +1.00 and the minimum size is -1.00. Any negative value means that the test item discriminates to some
degree—in the wrong direction and is not satisfactory. Positive values show that the test item discriminates in the desired direction, even though it may not be complete satisfactory. The larger the positive value the better. Any D values above +0.40 can be considered very good, any between +0.40 and +0.20 satisfactory, and any between +0.20 and zero poor.

**Examining responses and effectiveness of an item**

In multiple-choice item we also need to investigate the responses to an item and the effectiveness of the distracters. The item difficulty and discriminating power will be the underlying consideration in deciding the effectiveness of test item in multiple-choice type. The effectiveness of test items can be investigated by tabulating and comparing the entire responses of the pupils in the upper and lower groups. The responses to each of the distracters, as well as the correct responses, are counted.

**Application of Item-Analysis Principles**

**With Essay Test Item**

The scores of essay item range from 0, 0.5, 1, 1.5, and 2 for each item. Score 0 is given for no response, and score 2 is given for the exactly correct response. Each item consists of two tasks. To apply the item analysis in this kind of test item, the writer applies the procedure given by Linn and Gronlund.

A comparison of the individual task scores for the 25 highest and 25 lowest scoring students is the key in this analysis. First, the total score of each
item is calculated. The total scores then are ranked to identify the 25 lower and 25 higher. The average score is then calculated.

The following is an example of the applied analysis above:

Task 1

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper 25</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>13</td>
<td>1.72</td>
</tr>
<tr>
<td>Lower 25</td>
<td>2</td>
<td>8</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0.86</td>
</tr>
</tbody>
</table>

The higher average score for the 25 upper students than the 25 lower achievement students shows that the task discriminates between two groups.
CHAPTER IV
RESEARCH FINDINGS

In the first chapter, the writer proposes that problems of this study are as follows:

3. Have the test items used in final test fulfilled the criteria of good and appropriate test items?

4. In what aspects do the test items fulfill the criteria?

In the second chapter, the writer presents the related theory and the third chapter explains the research methodology used by the writer.

Dealing with the previous three chapters, this chapter deals with the writer’s effort to answer the questions proposed in the first chapter. This chapter presents the data analysis applying item analysis technique. The analysis results a justification whether an item is considered appropriate or not. To decide the appropriateness of an item, the writer focuses on three aspects i.e., level of difficulty, discriminating power, and the effectiveness of the distracter.

The first item type is of multiple-choice type. The criteria of good multiple-choice items consist of the three aspects. While the second type, essay item, does not include the last aspect i.e. the effectiveness of distracter. The item is analyzed based on the quantitative data from the test results.

This chapter presents the analysis in three sub-chapters. The first is item analysis on multiple-choice item. The next sub-chapter is analysis on the essay
items. The last one is the global finding of the analysis which is presented in research finding.

A. Analysis on Multiple-Choice Item

Text 1
This text for questions 1-3

*My Super Ex-Girlfriend* (*Comedy, 96 minutes*)

Superhero Jenny/G-Girl (*Uma Thurman*) uses her powers to make an ugly post-break-up world for her ex-boyfriend, Matt (*Luke Wilson*). The film exploits female irrationality and could well be called “Why a Woman Shouldn’t Be a Superhero?” and deserve a slap in the face for that.

*(Taken from: “The Jakarta Post”. 2006)*

1. My Super Ex-Girlfriend is a comedy film about …
   A. an ugly Post-break-up world
   B. a superhero played by Uma Thurman √
   C. a woman who is soft and humble
   D. a soft female who uses powers
   E. a G-Girl who is very ugly

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>B</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

This test item is considered satisfactory. The level of difficulty is 37.04% and shows that the item is neither too difficult nor too easy. It also discriminates upper and lower groups in positive value (+0.30). Some options do not work as intended although all distracting options were chosen by the students.
and predominantly distract the lower group. Option A is the best distracter since 19 students chose this option; 48.15% of the total lower group and 22% of the upper group chose it. Option C does not attract any students from upper group; it attracts only one student from the lower group. Option D is a poor distracter because it attracts more students in the upper group. Option E works quite well because more students in the lower group are attracted than those in the upper one. The responses to the distracting options represent the weakness of the distracter.

In short, this item has a good level of difficulty and discriminating power. This item actually has effective distracters except option D.

2. The writer thinks that My Super Ex-Girlfriend is …
   A. a very fantastic film
   B. a bad film √
   C. good enough to enjoy
   D. an unforgettable drama film
   E. super if we like an action film

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Omits</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The level of difficulty is 25.93% and shows that the item is too difficult since only a quarter of the total students got the answer right. Although it discriminates upper and lower groups in positive value, the index of discriminating power is poor (+0.07). Not all distracters work as intended. Option A and C attract more students than options D and E. Nevertheless, option A is the poor one because it attracts more students in the upper group than the lower one.
Option C works as intended, more students from lower group are attracted than the upper group. Option D and E work in the desired direction although only a small number of students chose the options. There is also 1 student who omitted the answer.

This item is poor because it does not have a good level of difficulty and discriminating power. Three distracters, except option A, are good.

3. Text 1 above belongs to a/ an …
   A. recount
   B. report
   C. review ✓
   D. explanation
   E. spoof

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

This item discriminates upper and lower group well with a very good index of discriminating power (D= 0.44). More students from upper group got the right answer than those from the lower group. Nevertheless, if it is examined based on the level of difficulty, the item is too easy, 92.6 % of total upper group students and 48% of total lower group students got it right. The level of difficulty is very easy one (P= 70.37%). The distracters function as intended because most of lower group chose them. The distribution is also evenly spread to the entire distracters.
In conclusion, this item has an appropriate discriminating power and effective distracters, but its level of difficulty does not fulfill the criteria of a good test item.

Text 2
This text for questions 4-6

Nuclear power is generated by using uranium, which is a metal mined in various parts of the world. The first large-scale nuclear power station was open at Calder Hall in Cumbria, England, in 1956.

Some military ships and submarines have a nuclear power plants for engines. Nuclear power produces around 11% of the world’s energy needed, and produces huge amounts of energy. It causes no pollution as you’d get when burning fossil fuels.

The advantages of nuclear are as follow:
1. It costs about the same as coal, so it’s not expensive to make.
2. It doesn’t produce smoke or carbon dioxide, so it doesn’t contribute to the greenhouse effect.
3. It produces huge amounts of energy from small amount of uranium.
4. It produces small amount of waste.
5. It is reliable.

On the other hand, nuclear power is very, very dangerous. It must be sealed up and buried for many years to allow the radioactivity to die away. Furthermore, although it is reliable, a lot money has to be spent on safety because if it does wrong, a nuclear accident can be a major disaster.

People are increasingly concerned about this. In the 1990’s nuclear power was the fastest growing source of power in many parts of the world. In 2005, it was the 2nd slowest growing.

4. The text discusses…
   A. nuclear in general
   B. nuclear biggest station
   C. the danger of radioactive
   D. the disadvantages of nuclear power
   E. the advantages and disadvantages of nuclear power √

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>
The item is too easy because 81.48% of total number of students got the answer right. Option A attracts the same number of students from the two groups. Option C attracts more students from lower group which means it works as intended. On the other way around, options B and D do not work as intended. They do not work at all since nobody chose them. The item discriminates poorly (D= +0.07). Possibly, due to the ineffectiveness of distracters, students are not attracted to the distracters.

It can be concluded that this item is not good because it does not fulfill the criteria of a good test item. The level of difficulty and the index of discriminating power are poor, and most of the distracters are not effective.

5. Which is NOT TRUE according to the text?
   A. The first large nuclear power station was lunched in 1956.
   B. Nuclear is needed to run the ships and submarine.
   C. It only needs a little amount of uranium to produce nuclear energy.
   D. To sweep away the radioactive waste is not a big problem √
   E. Most people think nuclear energy causes terrible catastrophe.

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>E</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The item is satisfactory in difficulty level. It is neither too difficult nor too easy (P= 68.52%). However, this item turns out to be poor in discriminating upper and lower groups. The index is only +0.04. The difference in the number of students from upper and lower groups who got the correct answer is only one. The discriminating power is clearly not sharp (19:18). The options do not seem to be
effective in distracting students’ answer. Option A does not work as intended because no students chose it. Although one student chose option B, this option functions quite well since the wrong answer comes from lower group. Although the same number of students from upper and lower groups got the wrong answer, option C and E also work quite well.

This item is good at the level of difficulty, but is less appropriate in the discriminating power and effectiveness of the distracters.

6. Why the nuclear power very dangerous?
Because …
A. its radioactivity lasts long √
B. uranium is renewable
C. it is reliable
D. it is cheap
E. it is safe

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The item discriminates students in positive and satisfactory value (D=+0.22). Options B and C function as intended. They attract more students in the lower group. Unfortunately, there are some options which are unable to attract any students i.e. options D an E. The two options do not work as intended because none of the students chose them. This test item is too easy, 48 of 54 students got the item right (P= 88.89%).
This test item cannot fulfill the required criteria of a good test item. It has good discriminating power, but it has poor level of difficulty, although 50% of the distracters are also ineffective.

**Text 3**
This text for questions 7-9

**What causes weather?**

Weather is the physical condition of the atmosphere at the particular time. It includes temperature, air pressure and water content.

Weather is produced when air moves place to place. This moving air is known as wind. Winds are formed when cooler air moves to replace the rising warm air. Warm air is usually less dense than cool air; therefore, it creates low air pressure. Cool air is more dense and creates high pressure.

Usually we have fine weather when the air pressure is high and there are clouds, rain and snow when the air pressure drops.

7. What does the text explain?
A. moving air
B. air pressure
C. low air pressure
D. types of weather
E. the formation of weather √

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>E</strong></td>
<td>23</td>
<td>20</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The item is too easy, 43 out of 54 students got the right answer and the index of difficulty level is 79.63%. The item also discriminates poorly (D=+0.11). Option A, B, and D work quite well, but option C does not work at all. Option C attracts no students. In general, the distracters do not work satisfactorily. This item does not represent appropriateness in the discriminating power and level of
difficulty. The distracters are less effective to be used as the half number of 
distracters do not function effectively.

8. “Warm air usually less dense than cool air …” (paragraph 1).  
The underlined word has the opposite meaning to … 
A. thin √ 
B. solid 
C. thick 
D. heavy 
E. compact

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The discriminating power of this test item is poor (D=+0.19). The level 
of difficulty is 27.78%, which means that this item is very difficult. Only 15 
students got the answer right, 10 comes from upper group and 5 from lower 
group. This results in a low or poor discriminating power. However, the other 
distracting options work as intended. All distracters function well because they 
attract students from the lower group more than those from the upper group.

However, this item is not a good one. It has good distracters, but it does 
not have appropriate level of difficulty and discriminating power.

9. The second paragraph mainly tells that … 
A. warm and cool air create air pressure 
B. winds are caused by warm air rising 
C. weather is produces by moving air √ 
D. the air moves from place to place 
E. the moving air is called wind
<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>24</td>
<td>19</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The item could not discriminate well. The index is only +0.19 and most of lower group students also got the answer right. The distracters do not work as intended. Option A attracts more students from the upper group while option B does not attract any students. Option D works well as most students from the lower group chose it. Option E works quite well, although it only attracts one student from lower group only. This item is also very easy that most students got the answer right (the percentage is 79.63%).

This item is not good at the level of difficulty and discriminating power. Only two out of 4 distracters are effective.

**Text 4**

**This text for questions 10-14**

Long, long ago Mosquitoes didn’t buzz, they talked and talked and talked and talked.

One day Mosquito was talking to Iguana, telling about his vacation, about every minutes of his vacation. Mosquito would not let Iguana say one word. Iguana was so annoyed that he walked away, leaving Mosquito still talking. Iguana grumbled and waved his tail.

He was still grumbling when he passed his friend snake and forgot all about saying hello. Snake was feeling hurt. He felt so sad that he slithered down rabbit’s hole.

“Help!” yelled Rabbit as she scurried out of the hole, terrified of Snake.
“What’s wrong?” cawed the Crow as he saw Rabbit racing. Danger must be near. “Run for your lives!” cawed Crow.

Monkey heard Crow’s warning and took of the treetops, leaping branch to branch. When Monkey landed on the Owl’s branch, high up in leafy tree, Owl’s
nest tipped of the branch and fell to the ground, breaking Owl’s eggs. The Owl was heartbroken so much that she didn’t hoot for the sun to come up.

The whole jungle was mad at Mosquito. Finally Owl hooted for the sun to come up and when it did, Mosquito lost his voice. All he could do was buzzing in everyone’s ears. “Zzzzzzz is everyone still mad as me.

10. Why did iguana leave the mosquito talking alone? Because …
   A. long, long time ago mosquito did not buzz but talk.
   B. mosquito was talking about his annoying vacation.
   C. Iguana felt annoyed with what mosquito behavior. √
   D. Iguana liked grumbling and waving his tail.
   E. Iguana didn’t like talking about vacation.

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>22</td>
<td>18</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The table shows that most students are able to find the correct answer. 40 out of 54 students answered the question correctly. This item is too easy with level of difficulty in 74.07%. It seems that the distracter is not so attractive that only a few students got the wrong answer. Option A and B attract more students from lower group and both items work quite well. Option D seems less effective than option A and B as the same number of students chose it i.e. one student only. Option E is the worst of all distracters. It does not attract lower group students, it even attracts upper group students. This item also has a poor index of discriminating power ($D = +0.15$).

Viewed from three aspects, this item is not a good test item. The level of difficulty and discriminating power is poor. Only two distracters function effectively.
11. The snake felt uneasy? Because …
   A. his friend, Iguana passed him.
   B. Iguana didn’t greet him. √
   C. Iguana liked grumbling.
   D. Iguana must have forgot him.
   E. He slathered down a rabbit’s hole.

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>B</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

In this item, it seems that students were a bit confused to find the right answer. The distracters work well. All distracters attract students especially from the lower group. More students from lower group are attracted to choose the distracters to be their answer. Unfortunately, option D is quite disappointing, because more students from upper group chose it. Nevertheless, this item discriminates positively with a very good index of discriminating power (D=+0.52). The calculation also results in a good index of difficulty. The level of difficulty is satisfactory (P=40.74%).

In short, this item is good at discriminating power and level of difficulty, and there is one ineffective distracter only.

12. What made the rabbit rushed out quickly from his house?
   A. snake √
   B. mosquito
   C. iguana
   D. crow
   E. monkey
This item surprisingly seems very difficult for upper group students although in fact the level of difficulty has a satisfactory level (P= 31.48%). The discriminating power is poor. This item discriminates in negative value (D= - 0.48). More students from lower group i.e., 15 students, got the answer right, while on the contrary only two students from lower group got the right answer. Option B and C which function as distracters, are chosen mostly by upper group students. Option B ironically attracts more students from upper group (23 students). Option D and E work well due to their attractiveness to lower group students. It is clear that this item has a poor discriminating power and distracters.

13. The owl felt so sad because …
   A. The nest in which she out her eggs was on free top.
   B. The monkey leaped and landed on her nest.
   C. The monkey broke her eggs which fell into the ground.
   D. The nest fell down and her eggs were broken. √
   E. The branch where she put her nest were broken down.

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>22</td>
<td>8</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>
This item seems to be a good item. It discriminates upper and lower group students in positive value and has a very good index (D= +0.52). Most of upper group students and some lower group students look as if they find no complication in answering the question. The level of difficulty is satisfactory (P= 55.55%). Half of the number of distracter are not effective. Option C attracts more students from lower good and considered as a good distracter. Although only three students from lower group choose option B, this option functions as intended. Meanwhile, options A and E do not work at all.

14. “Mosquito was talking to iguana, telling about his vacation …”  
What is the synonym of the underlined word?
A. holiday √
B. spare time
C. opportunity
D. chance
E. occasion

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The table shows that most students do not get any confusion. Most of the students got the right answer. The level of difficulty is 85.19% and illustrates that the item is very easy. Although the distracters work in the desired direction, only a few students are attracted to choose them. The item also discriminates poorly although its discriminating power index is in positive value (D= +0.15). This item is not good at the level of difficulty and discriminating power although it has good distracters.
   Via : “I failed the interview.”
   Ana : “You should improve your English before you applied for the job.”

The underlined sentence is used to express …
A. warning
B. support
C. advice √
D. regret
E. ability

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>27</td>
<td>14</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The table shows that all of upper group and 14 of lower group students got the right answer. This results in a high percentage of the difficulty level (P=75.93%) and shows that this item is very easy. This item discriminates in positive value with a very good index (D= +0.48). All distracters function as intended, more students from lower group are attracted to the distracters, while none from upper group chose the distracters. This item has effective distracters and an appropriate discriminating power, but the discriminating power is poor.

16. Benny : “You know Luth, the violin concert last night was really wonderful.”
   Luth : “….., I was so tired last night that I couldn’t go.”
A. I think so.
B. You are right. √
C. It’s not good enough.
D. I am sorry to have missed it. √
E. It is not as good as I expected.
<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

This item is also one of some very easy items. Twenty six students of upper group chose the right answer while 24 lower group students got no difficulty in finding the right answer. The level of difficulty is 92.59%. Although option A and B do not work optimally and only a few students choose them, the two options are still able to attract lower group students. Option C does not function as intended since only one upper group student chose it while none of lower group student considered this option as the correct one. On the other hand, option E does not work at all, no students are attracted to this option. The unattractiveness of distracters results in a poor index of discriminating power (D= +0.07).

In conclusion, this item does not fulfill the criteria of a good test item because its level of difficulty and discriminating power is poor. The distracters are also less effective than the required standard.

17. Mother : “What’s wrong with you, Nit?”
   Nita : “I got a bad mark in my mathematic test this morning.”
   Mother : “Oh dear, don’t be so sad. ....., it will be better next.”
   What does Nita’s mother say to encourage her?
   A. Sorry to hear that.
   B. I can’t stand hearing that.
   C. How irritating.
   D. You always complain about your test.
   E. I have no doubt of your capability. √
This item is the good one. It can discriminate between upper and lower group students with a very good index of discriminating power (D=+0.48). The level of difficulty is satisfactory (P= 53.70%) since more upper group students have the right answer. The options also work as intended. The incorrect answers come from the other four options which distract the correct answer. The distracters are quite effective.

This item has a good standard in the three aspects i.e., level of difficulty, discriminating power, and effectiveness of distracters.

18. Mr. Hadi : “The Usman family is poor.”
   Mr. Kasim : “We all know it. . . . , they can send their children to universities.”
   A. despite the fact that they are poor √
   B. they are poor despite the fact
   C. however they are poor
   D. they are poor although
   E. because they are poor

Data in the table show that only a few upper group students find that the question is rather difficult to answer. Only 14 of 27 students from the lower group
find the right answer. The level of difficulty is 68.52% which means that the item difficulty is satisfactory. Option B, C, and D function as distracters and work well; more students from lower group chose them. Option E does not work as intended because no student chose it. The discriminating power is also satisfactory (D= +0.33).

This item is a satisfactory one. It has good level of difficulty and discriminating power. Most of the distracters also work effectively.

19. Jaya : “Why are you still here? Didn’t you tell me that you would go to Jakarta yesterday.”
Bowo : “I would have been in Jakarta if the bus had not an accident.”
It means . . .
A. Bowo is in Jakarta
B. The bus got an accident √
C. The bus was save
D. Jaya went to Jakarta
E. Bowo went to Jakarta

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>27</strong></td>
<td><strong>27</strong></td>
</tr>
</tbody>
</table>

It is clearly shown through the table that the item is not complicating for the students. The level of difficulty is 92.59%. Option A does not attract any students. Options C and E work in the desired path. While option D work in the opposite purpose, it attracts upper group student. The distracters are not so effective that only a few students are attracted to choose the wrong answer. The discriminating power is poor (D= +0.07).
This item is not a good one. The data represents the poorness of discriminating power and level of difficulty. The distracters are also less effective.

20. … He admitted breaking into the car, he denied stealing the tape.
   A. Although √
   B. In spite of
   C. Despite
   D. As
   E. Since

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>21</td>
<td>16</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The table shows that more students from upper group got the right answer than the lower group did. The item has level of difficulty in the index of 68.52%. Option C is a worse distracter since it attracts more students from the upper group. The other distracters work quite well. Option B and E work as intended, more students from the lower group are attracted to choose them. Option A attracts the same number of students from the two groups. The index of discriminating power is poor (D= +0.19).

This item has a good level of difficulty, but it does not have a good index of discriminating power although only one single distracter is considered ineffective.
Why The Sky Is High.

A long time ago, the sky was close to the earth. When you stood up and stretched your arms, you could touch it.

One day, a farmer brought home some coconuts for his wife. His wife wanted to drink the juice inside. She knocked one of the coconuts against a rock to break open the shell.

“That’s not the way to do it,” her husband said. “you should cut it open with a knife.”

He gave her a bolo or jungle knife. His wife raised the knife to cut the coconut shell. “Crack!” As she lifted the knife above her head, it spilt open the sky and there was heavy rain.

The woman told the sky, “Will you please go a little higher? I cant raise my knife high enough to cut this coconut shell.”

She raised the knife over her shoulder the second time. Once again, it hit the sky and there was thunder and lightening.

“You are still too low,” the woman said to the sky, “Please go higher”

She raised the bolo a third time. It still cracked the sky. There was a typhoon, and the woman’s house was almost blown away by the fierce wind.

“What are you doing?” cried her husband. “Can’t you even try to open coconut shell without causing a typhoon?”

“It’s not me,” his wife answered. “It’s the sky which is too close. I need more room to use the knife.”

Finally, the woman became very angry. She looked up and shouted at the sky, “You’d better do as you are told. Otherwise I shall cut you into pieces.”

The clouds rose and rose the sky went up higher and higher. That’s why it is now high.

21. Which statement is NOT TRUE according to the text?
A. Bolo is a jungle knife.
B. The woman cut her husband’s hand with the knife. √
C. The farmer’s wife wanted to drink some coconut water.
D. A very long time ago, the sky was very low.
E. The woman tore the sky for several times.

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>
From the data, it seems that upper group students do not find any difficulty. All of the upper group students got the right answer. Most of lower group students also got the right answer. This results in a high percentage of difficulty level (P= 88.89%) which means the item is very easy. The discriminating power is satisfactory (D= 0.22). Some distracters (Options A, C, and E) work as intended since all incorrect answers come from the lower group students. On the other hand, option D does not function as intended, no students were attracted to this option.

This item is appropriate in the level of difficulty and discriminating power. Most distracters are effective, except option D.

22. Why was the woman angry?
   A. Her husband forbade her to crack the coconut shell open.
   B. The sky went up to slowly.
   C. The sky gave her rain, thunder, lightening, and typhoon. √
   D. She regretted cutting the sky into pieces.
   E. Her husband took over her work.

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>C</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

This test item can be classified as a good test item. This item can discriminate upper and lower students well. It has a very good discriminating power index (D= +0.44). All distracters function as intended. All options were chosen by students and most of lower students were attracted to the distracters. 24
of 54 students got the answer right. Calculation of the level of difficulty is also satisfactory (P=44.44%). This index means the item is not too difficult nor too easy for students.

23. She knocked one of the coconuts against a rock to break open the shell.
   (paragraph 2)
   The word “knocked” can be substituted by …
   A. stroke √
   B. rang
   C. slid
   D. kicked
   E. threw

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>E</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

A satisfactory test item can be represented by this test item. The discriminating power is satisfactory (D= +0.41). Level of difficulty is also satisfactory (P= 38.89%). Among the distractors, only option B that does not work as intended. Option B does not attract any students at all. Option C, D, and E work as intended. Most of the wrong answers come from the lower group students.

Reading text 6
This text is for questions no 24 to 26.
Dear Sir or Madam,

We are interested in becoming your distributors for your software products in the republic of Indonesia. Would you please send us your latest catalogs, descriptive brochures, and terms?

We are a hardware company that would like to add software to our sales offerings. Our annual report is enclosed.

We look forward to hearing from you soon.

Sincerely yours,

Setiawan Djuharie
Director.

24. What does Lingua Comunica sell now?
   A. Printing services
   B. Financial reports
   C. Software ✓
   D. Computers √
   E. Wall and street tools

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>D</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

Similar to the previous test item (item no.23) this test item can be classified as satisfactory test item. The item is able to discriminate upper and lower students in positive value and in a good index of discriminating power (D=
It also seems that the item is neither too difficult nor too easy. The level of difficulty is satisfactory (P= 37.04%). Although the distribution of distracters selected by students is not spread evenly, they work effectively. They attract more students from lower group than from the upper one. Options A and B do work, although only a few students were attracted. Option C works well since most of the students chose it. Option C was chosen by most of the students of either upper or lower groups. Generally, the wrong answer comes from lower students. Option E does not work as intended because none of the students chose it.

25. Which items were not requested?
   A. Pricing information
   B. Catalogs
   C. Terms
   D. Samples √
   E. Brochures

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

At a glance, the item seems to be able to discriminate upper and lower group students (24 versus 12). It is clearly shown by the data calculation that the item has a very good discriminating power index (D= +0.48). Calculating the data, the percentage of difficulty level is 68.52%. The result of calculation illustrates that the level of difficulty is also satisfactory. But, the work of distracters is not optimal enough. The distribution of selection does not spread among all options. Option A was chosen by most students, especially lower
students. Option B works as intended. Option C does not work at all. Option E is the worst one, it does not work as intended. This option, in contrary to its function, attracts upper group students.

It is concluded that this item has a good level of difficulty and discriminating power, but not all distracters are effective.

26. Lingua Comunica wants to …
   A. manufacture computers
   B. design programs
   C. distribute software √
   D. receive an annual report
   E. purchase hardware

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The data shows that most of students, both from lower and upper group, find no difficulty in this item. The percentage of right answer is quite high. The item is very easy (P= 83.33%). The large number of students who got the right answer also influences the discriminating power. Both group easily got the right answer. The difference between students who got the right answer in the two groups is not significant (23:22). This result is a poor discriminating power (D= +0.04). The distracters are not able to attract students. Options A and D can work as intended, they attract students from lower group. Option B does not work at all. In addition, unfortunately, option E attracts students from upper group more than the lower group.
This item is not good at three aspects since its level of difficulty and discriminating power is poor, and the distracters are less effective.

Reading text 7
This text is for questions no 27 to 30

Should ads be banned from TV Programmers?

Dear editor,

We are writing to complain about ads on TV. There are so many ads, especially during our favorite programs. We think they should be stopped for a number of reasons.

First, ads are a nuisance. They go on for a long time and there are so many. Sometimes there seems to be too many ads in one program.

Second, ads are a bad influence on people. Then try to encourage people to buy unhealthy food like beer, soft drink, candy and chips. And they make people want things they do not really need and cannot afford. It can lead to consumerism.

Finally, the people who make ads have too much say in what programmer people watch. That because they want to put all their ads on popular programmer that a lot of people watch. Some programs that are not so popular get stopped because they do not attract enough ads, even though those programs may be someone’s favorite.

For those reason, we think your TV station should stop showing ads. They interrupt programs. They are a bad influence on people, and they are sometimes put a stop to people’s favorite shows. We are sick of ads, and now we mostly watch other channel.

27. What is the communicating purpose of the text?
   A. To persuade that something is the case
   B. To persuade that something should or should not the case √
   C. To inform about banning the advertisement
   D. To tell about pass event
   E. To explain about how to ban ads
The discriminating power is very good (D = +0.52). The data clearly shows that most right answers come to upper group students. Twenty out of 54 students got the right answer for this item. The level of difficulty is also satisfactory (P = 37.04%). The distracters work as intended but option A is an exception. Option A does not work as intended. It works in the opposite manner; it attracted more students from upper group than from the lower one. The other options work as intended; they attract lower group students more than the upper group students do.

This item has good aspects, except one distracter (option A) which is not an effective distracter.

28. Who is the letter addressed to?
   A. TV programs
   B. The director of TV
   C. The reader
   D. Masarani S. W
   E. Editor √

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>
The distribution of answers in this item does not spread among the entire options. This can be seen easily that most students find no difficulty in answering the question. The level of difficulty is 77.78%, meaning the item is very easy. The discriminating power is satisfactory (D = +0.22), but the distracters do not seem to work well. Although option A works as intended, it attracts only one student from lower group. Option B functions well as more students from lower group selected this option as their answer. Unfortunately, option C and D do not function at all.

This item is good at discriminating power only. The two other aspects (level of difficulty and effective distracters) are not fulfilled the required criteria.

29. Which is true according to the text?
   A. The writer doesn’t like ads on TV, especially during his favourite programs √
   B. Ads are needed by the editor
   C. The writer hates the editor
   D. Ads are interesting according to the writer
   E. The editor needed ads on TV

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The data clearly show that the item is not so difficult to be answered. The percentage of right answer is 85.19% which also figures the level of difficulty (P = 85.19%). Although the distribution of selected distracting option spreads evenly, it does not warrant that the discriminating power is good. Only a few students got wrong answers which also means that the distracters seemed not
attractive enough. The index of discriminating power is +0.15 and is categorized in poor.

Two aspects of this item are clearly unable to fulfill the required criteria of a good test item. The level of difficulty and discriminating power is poor although the distracters function in the desired direction.

30. How many reason did the writer mentioned to support his argument?
   A. 2
   B. 3√
   C. 4
   D. 5
   E. 6

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

Poorness in discriminating upper and lower students occurs in this item. The discriminating power is poor with low index (D= +0.11). From the data calculation, the level of difficulty is categorized as very easy (P= 72.22%). Options A, C, and D play their role as distracter quite well. More students from lower group chose them. Option E does not work well, it does not distract any students’ answer.

This item is also unable to fulfill all required criteria of a good test item. The level of difficulty and discriminating power does not accomplish the standard of a good item although the distracters are quite effective.
This text is for questions no 31 to 32

**What Young Published Women are Thinking, Saying, and Doing**
(By: Sherry’s Handle and Blue Jean Press)

It’s a wonderfully inspirational book on young women’s issues. The topics covered are so varied starting from how to boost self-confidence and self image, to struggle, and to get success. A collection of essays, article, and stories taken from a magazine by, from, and for a young women to be outspoken, independent and self sustainable. Considering that the pop culture has promoted unrealistic and artificial lifestyle which young women are often victims, this book is a worthwhile alternative.

31. What might the writer think about the book?
   A. is independent
   B. is too various in content
   C. will give a lot of information to young women √
   D. is the only alternative book
   E. will make young women become the victims

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>21</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

Based on the distribution or right answers, it seems that the item is rather difficult for lower group students. After calculating the data, the level of difficulty can be categorized satisfactory (P= 68.52%). Unfortunately, the discriminating power is not as satisfactory as the difficulty level. The item has a poor index of discriminating power (D= +0.19). Options A and B work in right way, they attract students from lower group. Option D seems not so optimal in distracting students
choice, it attracts the same number of upper and lower group students. Option E works as intended although it is only a single lower group student is attracted to this distracter.

This item has an appropriate level of difficulty and effectiveness of distracters. Unfortunately, the discriminating power is poor.

32. What the purpose of the text?
   A. To persuade the readers to buy the book
   B. To inform the readers about the book √
   C. To evaluate the book
   D. To retell the event in the book
   E. To explain the process involved in the making of the book

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>C</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

This item represents an extremely poor discriminating power. Most of the right answer comes from lower group students. The item discriminates upper and lower group students negatively (D= -0.33). On the other hand, the level of difficulty is satisfactory (P= 31.48%). Some distracters function well i.e., options A, D, and E. Option C does not function as intended because the wrong answer comes mostly from upper group students.

This item is poor at discriminating power although it has satisfactory level of difficulty. There is also one distracter which is extremely ineffective. This item cannot be considered good.
Reading text 9
This text is for questions no 33 to 37

It is hard to explain current interest today in the music to capture the past, but features like Naif, Junior, and Club’s 80s all search for a period long gone. One Bandung-based psychiatrist believes the new interest in the past is a retreat from reality and a depressed mood among Indonesian. The example of fuel price hike forces people to seek a time when life was more simple. The 1970s and 1980s represent an age when human relations seemed more real in spite of the hard times of the depression. So escapism is one reason for our interest in the old-fashioned days where we imagine—although this is not at all proven fact—life to be more true, more authentic.

33. What is the writer trying to explain?
   A. The current interest in the past
   B. Musical groups retreating the past √
   C. Moody Indonesian
   D. Untrue and fake reality
   E. Simpler escape of abnormal condition

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

This item also represents another extremely poor discriminating power. This item failed to discriminates upper and lower group students. The discriminating power has negative value (D= -0.11). Although the level of difficulty is satisfactory (P= 42.60%), the right answer comes from lower group students than from the upper group. The distracters do not function in a good way. Option A and C work in the opposite direction and deviate from their purpose;
distracting more lower group than the upper one. The fact is just the opposite; more students from upper group are attracted to choose the distracting option as their answer. Option D does not function at all since no student selects this option. Option E is the only distracter that works as intended; no upper group students choose it and one lower group student selected this option as the right answer.

This item is clearly classified as a poor item. The discriminating power is poor and the distracters are not effective although the level of difficulty is satisfactory.

34. How many reasons does the writer find to explain the popularity of music that capture the past?
   A. 2
   B. 3 √
   C. 4
   D. 5
   E. 6

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>19</td>
<td>9</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

After calculating the level of difficulty, the conclusion is that the difficulty level is satisfactory (P = 40.74%). On the other way around, the discriminating power is not as satisfying as the level of difficulty. The item can not discriminate upper and lower group students in the intended direction. The data clearly show that most right answers come from lower group students. The discriminating power has a negative value (D = -0.22). Most of the distracters do
not function well. Option A works in negative direction because more students from upper group choose it as their answer. Option C and E work as intended. Option D does not function at all. This item has the same problem to the previous item (item 33). It does not perform a good discriminating power and effective distracters.

35. Who is the source of reason which the writer explains?
   A. escapism
   B. Indonesian people
   C. depressed humans √
   D. groups of music
   E. a psychiatrist living in Bandung

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>E</td>
<td>22</td>
<td>8</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

The poor aspect of this item is represented by its poor discriminating power. Similar to the previous item, this item is also unable to discriminate upper and lower group students in the desired direction. It has low index and negative value of discriminating power (D= -0.11). The difficulty level is satisfactory (P=55.56%). Only a few students make correct answers and unfortunately, they come from lower group only. None of the upper group students gets the right answer. The distribution of selected option spreads evenly. Options A, B and D do their function as intended. More students from lower group were attracted to them. Option E does not work as intended. More students from upper group chose this option.
Two of the three criteria are not fulfilled. It is only the level of difficulty that has an appropriate value. On the other way around, two other aspects i.e., discriminating power and effectiveness of difficulty cannot be considered appropriate.

36. What is the cause the writer gives?
   A. old-fashioned days
   B. escapism √
   C. reality
   D. music features
   E. current interest

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

According to the distribution of students’ answers, all options seem to be attractive. Analyzing the data, the result shows that the item is very difficult (P=29.63%). Index of discriminating power is satisfactory and discriminates students in positive value (D=+0.30). The distracters function well. Options A, D, and E function as intended since the incorrect choices comes from lower group students. Option C does not work quite well because the same number of students from both groups chose it.

This item fulfilled some criteria of a good item comprising good discriminating power and effectiveness of distracters. On the other hand, this item does not present an appropriate level of difficulty.
37. Perdana: If the math teacher had given me ten more minutes, I would have been able to answer all the question items. Prabu: Well, just forget it. We just wait and see what scores we can achieve.

From the underlined expression we know that Perdana 

A. left few question items unanswered ✓
B. was able to solve the whole question items
C. could not answer all math question items
D. was granted the heist score
E. did not take the math test

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

Some satisfactory aspects possessed by this item comprise the satisfactory discriminating power and level of difficulty. The index of discriminating power is +0.22 while the level of difficulty is 40.74%. Most of the distracters are effective. Options B and E seem less attractive than option C although the three options work in desired direction i.e. distracting lower students’ answers. Option D seems not attractive at all since no students take it as the best or right answer.

38. My friend, Merry, … mother is a biologist will continue her study at UGM.
A. who
B. whom
C. whose ✓
D. of which
E. where
After observing the table, it seems the distribution of selected option in the lower group covers up all of the options. This test item has a good index of discriminating power (D= +0.22). It seems lower students find the item quite difficult to solve. The level of difficulty is satisfactory (P= 66.67%). The distracters do their job in the intended direction. Wrong answers mostly come from lower group students.

This item is good because it can fulfill the criteria of a good test item examined from the three aspects.

39. “Don’t forget to bring your guitar to my party”, asked Rossa.
   Rossa asked Robby … .
   A. don’t forget to bring your guitar to my party
   B. not to forget to bring his guitar to her party √
   C. not to forget to bring his guitar to my party
   D. not to forget to bring her guitar to my party
   E. don’t forget to bring his guitar to her party

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>26</td>
<td>10</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

Looking at a glance, the table shows that the item is not a trouble for most students. The item is very easy, 40 of 54 students got the right answer. This
results in a high percentage of difficulty level which reflects the ease of the item (P= 83.33%). Although some lower group students are able to find the correct answer, the discriminating power is satisfactory (D= +0.26). Options A, D and E work as intended, they attract more students from lower group than those from the upper one. Option C attracts the same number of student from both groups.

This item has a good discriminating power and effective distracters, but it does not have an appropriate level of difficulty.

40. I met a very pretty girl last night. My brother fall in love with her.
   The best combination of the sentences is … .
   A. I met a very pretty girl last night and my brother falls in love with her.
   B. I met a very pretty girl last night whom my brother falls in love with. √
   C. I who met a very pretty girl and my brother falls in love.
   D. My brother falls in love and I met a very pretty girl.
   E. I met a very pretty girl who my brother falls in love.

<table>
<thead>
<tr>
<th>Option</th>
<th>Upper third</th>
<th>Lower third</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>23</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Omits</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

A good test item can be labeled to this item. This item discriminates upper and lower group with a very good index (D= +0.70). The level of difficulty is also satisfactory (P= 55.56%). All distracters work as intended. Options A, C, and D work quite well, although only in a small number, they attract more students from lower group than students from the upper one. Option E is the best distracter as most choice comes to this option, especially from lower group students.
B. Analysis On Essay Test Item

41. - He is a successful businessman.
    - He always lacks money.

*Join the two sentences above by using:*

a. Although  
b. Despite

Item 41.a

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>¼</th>
<th>½</th>
<th>¾</th>
<th>1</th>
<th>Omitted</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td>0.88</td>
</tr>
<tr>
<td>Lower</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>0.68</td>
</tr>
</tbody>
</table>

From the data above it seems that the sub item discriminates positively although it does not discriminate upper and lower group in a significant level. It also seems that the sub item is not too complicated for both groups. The level of difficulty is low. Most of the students got the maximum score (36 out of 54 students).

Item 41.b

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>¼</th>
<th>½</th>
<th>¾</th>
<th>1</th>
<th>Omitted</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0.60</td>
</tr>
<tr>
<td>Lower</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Item 41.b seems to be more difficult than the first sub item (41.a). The average score of this sub item is lower than the previous sub item. The difficulty level of this sub item is quite satisfactory. Upper group students get higher average score than the lower group. The task also discriminates well since the difference in average scores of both groups is quite sharp.
42. - Flight to Jakarta was cancelled yesterday.  
- The weather was bad.

*Join the two sentences above by using:*
  
a. Because  
b. Due to

Item 42.a

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>¼</th>
<th>½</th>
<th>¾</th>
<th>1</th>
<th>Omitted</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>0.96</td>
</tr>
<tr>
<td>Lower</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Concerning to the distribution of scores in this item, it can be concluded that the item is quite easy. The level of difficulty is low. The difference between upper and lower students’ average score shows that the item discriminates positively. The discriminating power is not too significant as the difference is quite small in score range. This item has low discriminating power as well as the level of difficulty.

Item 42.b

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>¼</th>
<th>½</th>
<th>¾</th>
<th>1</th>
<th>Omitted</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>0.89</td>
</tr>
<tr>
<td>Lower</td>
<td>8</td>
<td>3</td>
<td>10</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0.44</td>
</tr>
</tbody>
</table>

The second sub-item is better in discriminating upper and lower group. The comparison of average score is satisfactory. Upper group students get higher average score than the lower group students. This results in a good discriminating power. The level of difficulty is also satisfactory. The data shows that the item can be correctly answered by 27 out of 54.
43. Mr. Anton: “How can you determine students ability in writing?”
Miss Lina: “For start ask the students to write 150 words essay.”
Change the underlined sentence using causative have
A. Active
B. Passive

Item 43.a

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>¼</th>
<th>½</th>
<th>¾</th>
<th>1</th>
<th>Omitted</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>10</td>
<td>1</td>
<td>0.62</td>
</tr>
<tr>
<td>Lower</td>
<td>2</td>
<td>12</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.36</td>
</tr>
</tbody>
</table>

The average score is closer to the half score than to the full score of the sub item. It shows that the sub item is difficult for the students. Only 12 out of 54 students got the maximum score while on the other hand, three students did not give any answer. The task discriminates the two groups positively. The discriminating power is satisfactory.

Item 43.b

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>¼</th>
<th>½</th>
<th>¾</th>
<th>1</th>
<th>Omitted</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>1</td>
<td>0.58</td>
</tr>
<tr>
<td>Lower</td>
<td>3</td>
<td>12</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.34</td>
</tr>
</tbody>
</table>

The second sub item is also difficult for students. Less than half of the students gave the response correctly. Three students even omitted the task. They did not give any answer. This shows that the item is quite difficult. The task is also able to discriminate the groups. Only a few students from the lower group are able to solve the task.
44. a. The man were repairing the bridge when there was a big thunder.
   → Re-write in the passive voice
   b. The boy wanted to borrow my new novel. It was written by Marga T.
   → Combine those sentences by using “Relative Pronoun”

Item 44.a

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>¼</th>
<th>½</th>
<th>¾</th>
<th>1</th>
<th>Omitted</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>0</td>
<td>4</td>
<td>10</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>0.55</td>
</tr>
<tr>
<td>Lower</td>
<td>3</td>
<td>17</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.27</td>
</tr>
</tbody>
</table>

The task seems to be very difficult for the lower group. The lower group’s average score is the lowest compared to the other sub item. In fact, this item is very difficult for both groups of students. The average score is quite low. However, this task discriminates the two groups positively. Upper students’ average score is higher than the lower students’ score. This item works in the desired direction.

Item 44.b

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>¼</th>
<th>½</th>
<th>¾</th>
<th>1</th>
<th>Omitted</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>15</td>
<td>0</td>
<td>0.75</td>
</tr>
<tr>
<td>Lower</td>
<td>0</td>
<td>17</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>0.44</td>
</tr>
</tbody>
</table>

The second sub item discriminates the two groups well. The average score of upper group students is clearly higher than the lower group. Correct answer mostly comes from upper group students. Based on the difference of average scores between the two groups, the discriminating power is satisfactory. This sub item is neither too easy nor too difficult. The level of difficulty is satisfactory. The average score is higher than the previous sub item.
45. a. “Will you wait for me until I come to your house?”
Answer: Ima asked her boyfriend . . . 

b. I wish mother (not make) some “brownies” when my brother isn’t here
Correct the verb in bracket
Give the fact

Item 45.a

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>¼</th>
<th>½</th>
<th>¾</th>
<th>1</th>
<th>Omitted</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>2</td>
<td>1</td>
<td>15</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>0.56</td>
</tr>
<tr>
<td>Lower</td>
<td>4</td>
<td>15</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Based on the scores distribution, it seems that the first sub item is rather difficult for the students. Only five students from upper group answered the item correctly, none of the lower group students made the correct answer, and one student from upper group omitted the task. This task discriminates the groups well; it is shown by the distribution of maximum-score distribution and the differentiation of average score.

Item 45.b

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>¼</th>
<th>½</th>
<th>¾</th>
<th>1</th>
<th>Omitted</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>1</td>
<td>1</td>
<td>12</td>
<td>3</td>
<td>9</td>
<td>1</td>
<td>0.65</td>
</tr>
<tr>
<td>Lower</td>
<td>0</td>
<td>14</td>
<td>6</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>0.42</td>
</tr>
</tbody>
</table>

The data shows that the second sub item is a bit easier than the first sub item. But, in general, this item is not an easy item. This item is still classified as difficult. The task is also able to discriminate the two groups in the intended way. Mainly, the correct answer comes from the upper group.
C. Research Finding

This last sub-chapter presents the global finding of analysis on the multiple choice and essay items. The research finding is illustrated through the tables with brief discussion.

Based on the analysis on the multiple-choice items, the writer finds some items which are unable to fulfill the required aspects properly. The following table figures the data:

Table 4.1 Result of Item Analysis on Multiple Choice Items

<table>
<thead>
<tr>
<th>Item</th>
<th>D</th>
<th>P</th>
<th>ED</th>
<th>Item</th>
<th>D</th>
<th>P</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+0.30</td>
<td>37.04%</td>
<td>3</td>
<td>21</td>
<td>+0.22</td>
<td>88.89%</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>+0.07</td>
<td>25.93%</td>
<td>3</td>
<td>22</td>
<td>+0.44</td>
<td>44.44%</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>+0.44</td>
<td>70.37%</td>
<td>4</td>
<td>23</td>
<td>+0.41</td>
<td>38.89%</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>+0.07</td>
<td>81.48%</td>
<td>1</td>
<td>24</td>
<td>+0.30</td>
<td>37.04%</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>+0.04</td>
<td>68.52%</td>
<td>1</td>
<td>25</td>
<td>+0.48</td>
<td>68.52%</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>+0.22</td>
<td>88.89%</td>
<td>2</td>
<td>26</td>
<td>+0.04</td>
<td>83.33%</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>+0.11</td>
<td>79.63%</td>
<td>2</td>
<td>27</td>
<td>+0.52</td>
<td>37.04%</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>+0.19</td>
<td>27.78%</td>
<td>2</td>
<td>28</td>
<td>+0.22</td>
<td>77.78%</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>+0.19</td>
<td>79.63%</td>
<td>2</td>
<td>29</td>
<td>+0.15</td>
<td>85.19%</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>+0.15</td>
<td>74.07%</td>
<td>2</td>
<td>30</td>
<td>+0.11</td>
<td>72.22%</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>+0.52</td>
<td>40.74%</td>
<td>3</td>
<td>31</td>
<td>+0.19</td>
<td>68.52%</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>-0.48</td>
<td>31.48%</td>
<td>2</td>
<td>32</td>
<td>-0.33</td>
<td>31.48%</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>+0.52</td>
<td>55.55%</td>
<td>2</td>
<td>33</td>
<td>-0.11</td>
<td>42.60%</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>+0.15</td>
<td>85.19%</td>
<td>3</td>
<td>34</td>
<td>-0.22</td>
<td>40.74%</td>
<td>2</td>
</tr>
</tbody>
</table>
Based on the table, the writer concludes the results as follows:

1. Only 50% of the total items fulfilled criteria of a good discriminating power. From the discriminating power aspect, 25% of the test items have a very good index of discriminating power, 25% have a satisfactory index, and the rest are poor.

2. From the level of difficulty aspects, 21 items have satisfactory level, 3 items are very difficult, and the rest (16 items) are very easy.

3. From the effectiveness of distracter aspect, 22 items have effective distracters around 3-4 distracters function as intended, while the rest (18) have less effective distracters (around 1-2 distracters can work as intended).
While from the analysis on the essay test item, the result is presented by the following table:

**Table 4.2 Result of Item Analysis on Essay Test Items**

<table>
<thead>
<tr>
<th>Sub-item</th>
<th>Average score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Upper Group</td>
</tr>
<tr>
<td>41.a</td>
<td>0.88</td>
</tr>
<tr>
<td>41.b</td>
<td>0.60</td>
</tr>
<tr>
<td>42.a</td>
<td>0.96</td>
</tr>
<tr>
<td>42.b</td>
<td>0.89</td>
</tr>
<tr>
<td>43.a</td>
<td>0.62</td>
</tr>
<tr>
<td>43.b</td>
<td>0.58</td>
</tr>
<tr>
<td>44.a</td>
<td>0.55</td>
</tr>
<tr>
<td>44.b</td>
<td>0.75</td>
</tr>
<tr>
<td>45.a</td>
<td>0.56</td>
</tr>
<tr>
<td>45.b</td>
<td>0.65</td>
</tr>
</tbody>
</table>

In general, the essay test items are of satisfactory. The table shows that all items are able to discriminate positively. Some sub-items are very easy i.e., 41.a and 42.a. Sub-items 43.b, 44.a, and 45.b are very difficult, while the rest of the sub items have satisfactory level of difficulty.

Item analysis is helpful for the teacher especially for informal achievement tests or non-standardized tests which is constructed by the teacher. From the analysis, improvement of the items can be done since the analysis shows the weaknesses and strengths of the test items. Teacher can find students’ achievement, students’ difficulty in mastering a certain subject or topic. Teacher
can also consider which item is needed to be improved, discarded, or saved for the next tests.

Item analysis helps teacher to improve the multiple choice items since this type of test items are objective, easily quantified, and calculated using a certain formula. On the other hand, item analysis is quite difficult to be applied in the essay test items because the scoring is more subjective and it will affect the analysis.
CHAPTER V

CONCLUSION AND SUGGESTION

This chapter presents the conclusion and suggestion for the teacher, test publisher, and the other researchers.

A. CONCLUSION

The researcher makes some conclusion based on the data analysis in the previous chapter. The conclusions are as follows:

4. In general, not all English final test items fulfill the criteria of a good test item perfectly. Only 27.5% of the total items fulfill all criteria meaning only 11 items have a good index of discriminating power and level of difficulty with effective distracters. While the rest of the items i.e., 21 items (72.5%) still have some weaknesses, even nine of them do not fulfill any criteria of a good test item.

5. The essay items work in the intended direction. The items can discriminate positively although some of them are categorized as very difficult or very easy.

6. Item analysis works best if the item is multiple choices. If it is used to analyze essay types item then it is more difficult because the possibly occurred subjective scoring will take effect to its result.
B. SUGGESTION

The following are some suggestions which are hoped to be useful for the teacher, test publisher, and other researcher.

1. For the teacher,

   It is hoped that the teacher will consider student’s ability so that the teacher will be able to construct test items with a satisfactory level of difficulty. The items should contain what students have learned because, in this case, the test is a summative test that measures students’ achievement in the final semester.

   Teachers also should be careful to construct the test so that the test items will work in the desired direction, without negative result such as in the negative value in discriminating power.

   Teachers should make a clear scoring procedure or good criteria of scoring that will reduce subjectivity in giving score especially for essay items.

2. For the test publisher

   Test publishers should construct test items which are in line with what students have actually learned and is based on the syllabus. Test publishers should construct better test item especially if it is used as standardized achievement tests which is conducted in a large scale. Test publishers should have some research or conduct a try out before issuing the test items used for testing the students.

3. For other researcher

   Other researcher can use the result of this study as a reference or consideration before doing another research related to this study. It will be better if other researcher conduct a deeper research or analysis related to this study.
BIBLIOGRAPHY


