



**CONSTRUCTING AN ENGLISH TEST MODEL OF
READING, VOCABULARY, AND STRUCTURE
FOR MID TERM TEST FOR SIXTH GRADE STUDENTS OF
*BBC ENGLISH TRAINING SPECIALIST***

A final project submitted in a partial
fulfillment of the requirements for
the degree of *Sarjana Pendidikan* in English Language Education

by
LULUT HARCITO

2201404573

**ENGLISH DEPARTMENT
LANGUAGES AND ARTS FACULTY
SEMARANG STATE UNIVERSITY**

2011

APPROVAL

This final project was approved by the Board of the Examination of the English Department of the Languages and Arts Faculty of the Semarang State University on 18th August 2011

Board of Examination

1. **Chair Person**

Prof. Dr. Agus Nuryatin, M.Hum
NIP. 196008031989011001

2. **Secretary**

Dra. Rahayu Puji Haryanti, M.Hum
NIP. 196610201997022001

3. **First Examiner**

Drs. Amir Sisbiyanto, M.Hum
NIP. 195407281983031002

4. **Second Advisor as Second Examiner**

Drs. Ahmad Sofwan, Ph.D
NIP. 196204271989011001

5. **First Advisor as Third Examiner**

Dr. Dwi Anggani Linggar Bharati, M.Pd
NIP. 195901141989012001

Approved by
Dean of Languages and Arts Faculty

Prof. Dr. Agus Nuryatin, M.Hum
NIP. 196008031989011001

PERNYATAAN

Dengan ini saya :

Nama : Lulut Harcito
NIM : 2201404573
Prodi : Pendidikan Bahasa Inggris S1
Jurusan : Bahasa dan Sastra Inggris
Fakultas : Bahasa dan Seni

Menyatakan dengan sebenarnya bahwa skripsi/ tugas akhir/ final project yang berjudul :

CONSTRUCTING AN ENGLISH TEST MODEL OF READING,
VOCABULARY, AND STRUCTURE FOR MID TERM TEST FOR SIXTH
GRADE STUDENTS OF *BBC ENGLISH TRAINING SPECIALIST*

yang saya tulis dalam rangka memenuhi salah satu syarat untuk memperoleh gelar sarjana pendidikan ini benar-benar karya saya sendiri, yang saya hasilkan setelah melalui penelitian, bimbingan, diskusi, dan pemaparan/ujian. Semua kutipan baik langsung maupun tidak langsung, baik yang diperoleh dari sumber kepustakaan, wahana elektronik, wawancara langsung, maupun sumber langsung lainnya, telah disertai keterangan mengenai identitas sumbernya dengan cara sebagaimana yang lazim dalam penulisan karya tulis ilmiah. Dengan demikian walaupun tim penguji dan pembimbing penulisan skripsi ini / tugas akhir/ final project ini membubuhkan tanda tangan keabsahannya, seluruh karya ilmiah ini tetap menjadi tanggung jawab saya sendiri. Jika kemudian ditemukan pelanggaran terhadap konvensi tata tulis ilmiah, saya bersedia menerima akibatnya.

Demikian, harap pernyataan ini dapat digunakan seperlunya.

Semarang, 28 Juli 2011
Yang membuat pernyataan

Lulut Harcito
NIM. 2201404573

live ur life lively....



Dedicated to:

- My beloved parents
- My dearest brothers
- bung
- Big family of LSI
- All my brothers in GARIS
- My best friends in PKM FBS

ACKNOWLEDGMENT

First of all, I would like to thank to Allah SWT for blessing, health and inspiration which is given to me during the writing the final project. I would also express my special thank to my parents and family who always support during my study at UNNES.

I would also like to thank to Dr. Dwi Anggani Linggar Bharati, M. Pd. as the first advisor for guiding me in the process of writing final project from beginning until its completion with great patience. My appreciation also goes to Drs. Ahmad Sofwan, Ph. D as second advisor for his kindness in giving useful suggestions for its finalization.

I would also thank to the Chief and all the teachers of *BBC English Training Specialist* who helped me to do observation, and all the sixth grade students for the cooperation and helps. In addition, my gratitude goes to all my lecturers of the English Department of UNNES for guidance during the study.

Finally, I also express my special thanks to my beloved bung for her invulnerable support, my brothers and sisters in GARIS, LSI, and big family of PKM FBS (luar biasa!!!). Thank you for the friendship and thanks for all people who will accompany me now and ever after.

Semarang,

Lulut Harcito

ABSTRACT

Harcito, Lulut. 2011. *Constructing an English Test Model of Reading, Vocabulary, and Structure for Mid Term Test for Sixth Grade Students of BBC English Training Specialist*. Final Project. English Education. English Departement Languages and Arts Faculty, Semarang State University (First advisor : Dr. Dwi Anggani Linggar Bharati, M. Pd., Second Advisor : Drs. Ahmad Sofwan, Ph. D).

Key words: Item Analysis, Validity, Reliability, Difficulty Level, Discrimination Power.

It is important to analyze the competencies to be achieved in the curriculum of mid term, before the teachers administered the model test as the mid test and find out whether the test is good or not, based on the terms of validity, reliability, discrimination power, and difficulty level of English questions items. Using the result of the analysis it is assumed that the teachers could build a good test for each grader.

The general objectives of the study are to obtain an objective description of the English test model administered to the sixth grade students of *BBC English Training Specialist*. The method used in analyzing the data is a quantitative approach. The population was the sixth grade students of *BBC English Training Specialist*. There were three classes, and each class consists of 20 students. Thus, the total number of the students is sixty students. The writer took twenty four students as the sample by using random sampling technique.

In writing this final project, the writer will conduct two activities. The first is library activity, in this activity; the writer will select some books for reference, which give information and supporting his writing. Then, the field activity is used to collect the data.

From the result of the analysis, this test is not valid and still needs some revisions; it is not reliable, poor in terms of discrimination power, and easy in terms of difficulty level. Based on the result, the writer would like to offer some suggestions: first, the test constructors should know about the characteristics of good language test, especially procedure of determining difficulty levels and discrimination power. Second, items that still can be used should be revised and save. Finally, the writer suggests this test should not be used in the English final test. It can be used if it has been revised and the writer hopes that the result of this item analysis could be used as an example in analyzing other test items and encourage others teachers to do research on the same subject.

TABLE OF CONTENTS

ACKNOWLEDGEMENT.....	iv
ABSTRACT.....	v
TABLE OF CONTENTS.....	vi

CHAPTER

I. INTRODUCTION

1. 1	Background of the Study.....	1
1. 2	Reason for Choosing the Topic.....	3
1. 3	Statement of the Problem.....	3
1. 4	Objective of the Study.....	4
1. 5	Significance of the study.....	4
1. 6	Limitation of the study.....	5
1. 7	Outline of the study.....	5

II. REVIEW OF RELATED LITERATURE

2. 1	Review of Preview Studies.....	7
2. 2	Review of Theoretical Background.....	9
2.2.1	Curriculum.....	9
2.2.2	Definition of Assesment.....	11
2.2.3	Types of the Test.....	12
2.2.3.1	Essay.....	13
2.2.3.2	Multiple Choice.....	13
2.2.4	Characteristic of a Good Test.....	14
2.2.4.1	Validity.....	14
2.2.4.2	Reliability.....	15
2.2.5	Summative and Formative Assesment.....	17
2.2.5.1	Summative Assesment.....	18

2.2.5.2 Formative Assesment.....	18
2.2.6 Constructing Test.....	19
2.3 Framework of Analysis.....	24
III. METHOD OF INVESTIGATION	
3.1 Research design.....	27
3.2 The Main Sources Data.....	30
3.3 Population and Sample.....	30
3.4 Sampling Technique.....	31
3.5 Identification of the Problems.....	31
3.6 Technique of Data Collection.....	32
3.7 Techniques of Data Analysis.....	32
3.7.1 Difficulty Level Analysis.....	33
3.7.2 Discrimination Power Analysis.....	34
3.7.3 Analysis of Reliability.....	35
3.7.4 Analysis of Validity.....	36
IV. ANALYSIS AND DISCUSSION	
4.1. Analysis.....	38
4.1.1 Analysis of item Difficulty Level.....	39
4.1.2 Analysis of item Discriminative Power.....	40
4.1.3 Analysis of item Validity.....	41
4.1.4 Analysis of item Reliability.....	42
4.2. Discussion.....	42
V. CONCLUSION AND SUGGESTION	
5.1 Conclusion.....	72
5.2 Suggestion.....	73
BIBLIOGRAPHY.....	75

CHAPTER I

INTRODUCTION

In the first chapter, the writer would like to discuss background of the study, reasons for choosing the topic, statement of the problems, objective of the study, significance of the study, and the outline of the study.

1.1 Background of the Study

Language is a means of communication which is used among human beings in order to establish social relation. By using language, people are able to convey their message, to express their idea in oral, written or sign language so that they can cooperate and get along with each other. Generally, people in community understand speak their language well enough to carry out every activity (Hasanah, 2008: 1). This is line with the statement in dictionary that “language is system of communication in speech and writing used by people of a particular country”. (Hornby, 1995: 51)

English is spread all over the world that makes it become an international language. As international language, many countries use English in many fields of life such as communication, commerce, economy, politics, education, science, technology, and etc.

Knowing that English has significant position, our government has included English into Indonesian educational system in its curriculum as a compulsory

subject in elementary grade and high school grade. Our government also states English as one of the subject that be tested in national examination. These are to stimulate students' skill to improve their English to face the challenge in international world. As an addition, students require guidance from other source instead of governmental school, for example learning in nonformal institution.

During teaching learning process, it is expected that teachers will teach the subjects according to its objectives and the learners will get significance results. To assess learners' achievement Airasian and Russel (2008:9) give the following definition of assessment, test, measurement, and evaluation.

Assessment is a process of collecting, synthesizing and interpreting information in order to make decision. A test is a formal, systematic procedure for gathering information. Measurement is the process of quantifying or assigning a number to a performance or trait. Evaluation is the process of judging the quality or value of a performance or a course of action.

Any test, which is tested to measure student's achievement, must have the characteristics of a good test. According to Brown (2004:19) there are five cardinal criteria for "testing a test" practicality, reliability, validity, authenticity, and washback.

It is important to analyze the competencies to be achieved in the program outline of mid term, before the teacher administered the model test as the mid term test and find out whether the test is good or not, based on the terms of validity, reliability, discrimination power, and difficulty level of English questions

items. Using the result of the analysis it is assumed that the teachers could build a good test for other grade too.

1.2 Reasons for Choosing the Topic

In this study I would like to focus the research on the English Test Model of Mid Term Test, which is administrated to sixth grade students of *BBC English Training Specialist*. The Test model will give additional knowledge for teacher to construct a good test for the same situation and condition.

According to Weir in Hasanah (2008:4), in evaluating students achievement, the test not only must fulfill some characteristics of a good test such as validity, reliability, difficulty level and discrimination power, but also the test should relate closely to both the content of what has been learned and also what things they have been taught.

By applying item analysis, we can indicate which items are reliable and valid for whatever reasons or check properly whether the test has good quality or not.

1.3 Statement of the Problems

The problem that the writer wants to find out the answer is the following question: "How appropriate is the test model conducted by the teacher in *BBC English Training Specialist*".

More specific, in constructing the test item, the writer limits his problems into the following questions:

- (1) how is structure test for mid term test constructed?

(2) how is vocabulary test for mid term test constructed?

(3) how is reading test for mid term test constructed?

1.4 Objective of the Study

The general objectives of the study are to obtain an objective description of the test model conduct for the sixth grade students of *BBC English Training Specialist*.

The objective above is then specified into the following goal:

(1) to describe constructing structure test for mid term test.

(2) to describe constructing vocabulary test for mid term test.

(3) to describe constructing reading test for mid term test.

1.5 Significance of the Study

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

(1) for the student: The result of the study can be used by student to make their study more effective with regard to the right materials

(2) for the teacher: The result of the study can be used as an additional knowledge for teacher when they want to construct test items.

(3) for the test maker: The test maker may use it as a supplement in constructing tests.

(4) for the writer: The writer himself especially can increase his skill in constructing test items.

1.6 Limitation of the Study

The teachers construct the English test model for the sixth grade students. This test is in the form of multiple-choice; because it is practical and the result is objective, in brief that:

- (1) multiple-choice items have only one correct answer each. Thus, the grader will grade the answer objectively.
- (2) by using the type of items analysis, the discrimination power and difficulty level of the test can be practically determined.

1.7 Outline of the Study

Chapter I is the introduction of the whole thesis. It includes the General Background, Reason for Choosing the Topic, Statement of the Problem, Objective of the Study, Significant of the Study, and the Outline of the Study.

Chapter II presents the related literature used in this study. The writer is of the opinion that it is important to review literature related to language testing. This chapter discusses program outline of the study (curriculum), types of test (essay and multiple choices), characteristics of a good test (validity and reliability), summative and formative test, and constructing test (structure, vocabulary, and reading).

Chapter III deals with Method of the Study, which presents the main sources of data, population and sample, sampling technique, identification of the problem, technique of data collecting, and technique of data analysis.

Chapter IV deals with analysis and discussion of research judging.

Chapter V gives the conclusion of the research and some suggestions on the basis of the research finding.



CHAPTER II

REVIEW OF RELATED LITERATURE

In this chapter, I will review some related previous studies and theories found in the references to support the research. It also presents framework of analysis which summarizes two of previous studies.

2.1 Review of Previous Study

In doing this research, the writer reviewed previous studies. The writer reviewed a number of studies which almost similar to this research who worked about assessment and or the study about theory of validity and reliability.

First is Pauline Fitzgerald who wrote a journal entitled *Enhancing Employability in a Final-Year Undergraduate Module Using Groupwork and Peer Assessment* in ALT Journal Number 10: Winter 2010. The result of this study showed that the module assessment appears to have achieved its purpose in engaging students in developing and demonstrating employability skills, especially groupwork and presentation skills. This technique of assessment should be applied to students who are mature enough to think objectively of college goal. The more mature students, appeared to understand what the assessment was aiming to achieve. The less mature students, struggled with the concept from start to finish, and were uncomfortable with the task set. However, this kind of assessment should be combined to persuade those students who were not happy

with this type of module to be a valid and useful activity, possibly by introducing them to some of the literature about employability.

Second is Wahid Kurniadi Setiawan (2009) Semarang State University, whose study entitled "*Constructing English Test Model of Listening and Reading Mid First Semester for Eighth Grade Students of SMP Negeri 2 Demak in the Academic Year of 2009/2010*". The research was done to obtain an objective description of the English test model administered to the eighth grade students of *SMP Negeri 2 Demak* in the academic year of 2009/2010. The method used in analyzing the data was a quantitative approach. The result of the analysis stated that the test was not valid and still needed some revisions; it was not reliable, poor in terms of discrimination power, and easy in terms of difficulty level. Therefore, it can be concluded that: the knowledge about the characteristics of good language test is important and there were items which needed revision to be used in the English final test.

As mentioned above, giving assessment to student was needed to enhance students' ability. i.e. the giving of peer assessment in speaking skill, such as groupwork and presentation. However, there is an area of study which has not been studied, i.e. the use of assessment in reading, especially in vocabulary and structure. Moreover, to reach the goal of giving assessment, constructing the good assessment is very important base to realize it. So, in this study the writer focused his research on the constructing a test model of reading assessment to improve students' skill in vocabulary and structure. This research introduced the use of reading assessment to English teachers, particularly the English teacher of sixth

graders of BBC English Training Specialist, in order to make the good learning process.

2.2 Review of the Theoretical Background

In this section I present three areas of studies, i.e. (1) Curriculum, (2) Assesment, and (3) Constructing Test.

2.2.1 Curriculum

The Oxford Advanced Learner's Dictionary of Current English, defines curriculum as the subject included in a course of study or taught at a particular school, college etc. (Hornby; 1995:287)

From the explanation above, it means that teaching and learning process in school, college, or any other formal and nonformal institutions should relevant to the curriculum.

According to Yunestia (2008 : 7) curriculum is the basic of doing teaching learning activities. A teacher should know the curriculum before he or she teaches his or her students. A curriculum has been made on the basis of students' need. In Indonesia, government is trying to improve the quality of human resources of its people.

Another definition of the term curriculum is written in "*Kamus Besar Bahasa Indonesia*" published by the Center of Language Development and Establishment which states that curriculum is: (1) A series of lessons taught in the institution of education and (2) A series of lectures about certain skills of study. (Depdikbud, 1995:546)

In BBC, curriculum is defined as program outline. The writer focuses on certain language competence in this section as follows:

In this section, the writer focuses on certain language competence.

Reading and Structure

Understanding meaning in written functional text and short simple essay which is formed in description and recount related with closer environment.

Language Competence :

- a. Responding the use of good structure in simple short written functional text accuracy, fluently, and acceptable related with the topics in the program outline of *BBC English Training Specialist*.
- b. Responding the proper vocabulary in simple short written text and picture related with the topic in the program outline of *BBC English Training Specialist*.

Responding meaning and rhetoric step in simple short essay accuracy, fluently, acceptable related with the topics in the program outline of *BBC English Training Specialist* in descriptive and recount.

2.2.2 Definition of Assessment

In current educational practice, the words test and assessment might be thought as synonymous term. However, test just measures a given domain or particular domain. It is given to the student in a certain time of the curriculum whose skills are being measured and evaluated. Brown (2004:4) explains that assessment is an ongoing process encompasses a much wider domain than that the intention of testing. Whenever student responds to a question, offers a comment, or tries out a new word or structure, the teacher subconsciously makes an assessment of the student's performance.

Similarly, Richard (2002: 17) states that assessment is emphasized on ongoing assessment of students' performance as a course progresses or what has commonly been called formative evaluation. There are some techniques of the formative evaluation such as performance-based assessment, portfolio development, oral production inventories, cooperative student-student techniques, and other authentic assessments.

Distinguishing it from evaluation, Chapelle and Brindley (2002: 267) define assessment as "the act of collecting information and making judgments on a language learner's knowledge of a language and ability to use it". Assessment is thus concerned with individual student learning. The term evaluation, on the other hand, is often used in the broader sense of program evaluation, and refers to the process of collecting information and making judgments on the quality of the total language program. This involves a consideration of a range of elements in

addition to student learning, such as teacher and students attitudes, teaching methods and materials, as well as administrative systems and resources.

Parallel with Chapelle and Brindley, Coombee and Hubley (2003:5) state that assessment is an ongoing, cyclic process in which the techniques and instruments used are constantly monitored and reevaluated before being used again. Just as an individual students' progress in learning English is tracked and documented at many different points, so are the means of assessing students and programs.

From all the definitions of assessment above, it can be concluded that assessment is an ongoing process which does not only focus on the result of the test, but also emphasizes on the process of measuring the students' skill.

2.2.3 Types of the Test

According to Brown (2004:5) assessment can be divided into two parts. They are informal assessment and formal assessment. Informal assessment can take a number of forms, starting with incidental, unplanned comments and responses, along with coaching and other impromptu feedback to the student. Teacher can give compliment to the students who have done the work, for example *Good job!* *Nice try!* or something else. While formal assessments means exercises or procedures specifically designed to tap into a storehouse of skills and knowledge. They are systematic planned sampling techniques constructed to give teacher and student an appraisal of student's achievement. Formal assessments are the conventional method of testing that we are all very familiar with our school days.

Formal assessments are further broken down into separate groups; norm referenced tests, and criterion referenced tests.

In Indonesia, many schools and other institutions assess their student achievement using essay and multiple choice items.

2.2.3.1 *Essay*

An essay is usually a short piece of [writing](#). It is often written from an author's personal [point of view](#). Essays can be [literary criticism](#), political [manifestos](#), learned [arguments](#), observations of daily life, recollections, and reflections of the author.

(<http://en.wikipedia.org/>)

An essay has advantages and disadvantages:

(1) **Advantages**

- a. No guessing
- b. Allows divergent thinkers to demonstrate originality, creativity
- c. Reduced lead time required to produce
- d. Less work to administer for smaller number of students
- e. Can be rich in diagnostic information

(2) **Disadvantages**

- a. Impossible to mark objectively
- b. even different times of day make a difference

2.2.3.2 *Multiple Choice*

Multiple choice items consist of a stem and a set of options. The *stem* is the beginning part of the item that presents the item as a problem to be solved, a

question asked of the respondent, or an incomplete statement to be completed, as well as any other relevant information. The options are the possible answers that the examinee can choose from, with the correct answer called the *key* and the incorrect answers called *distracters*. (<http://en.wikipedia.org/>)

2.2.4 Characteristic of a Good Test

Harris (1969:13) points out “three characteristics of a good test called : reliability, validity, and practicality. In this section the focus is on validity and reliability.

2.2.4.1 Validity

According to Airasian and Russel (2008 : 16), the single most important characteristic of good assessment is its ability to help the teacher make appropriate decisions. This characteristic is called validity. Validity is concerned with whether the information being gathered is relevant to the decision that needs to be made.

Airasian and Russel give the key aspects of assessment validity:

- (1) Validity is concerned with this general question: To what extent is this decision based on appropriate assessment information?
- (2) Validity refers to the decisions that are made from assessment information, not the assessment approach itself. It is not appropriate to say the assessment information is valid unless the decisions, purpose, and groups for which it is valid are identified. Assessment information that is valid for

one decision or group of students is not necessarily valid for other decisions or groups.

(3) Validity is a matter of degree; it does not exist on an all-or-nothing basis.

Think of assessment validity in terms of categories: highly valid, moderately valid, and invalid.

(4) Validity is always determined by a judgment made by the test user.

Validity must be assured in terms of the following:

- (1) The content of the assessment
- (2) The way in which the assessment is constructed
- (3) A test's linkage with the way the items have been taught

2.2.4.2 Reliability

A reliable test is consistent and dependable. If you give the same test to the same students or matched students on two different occasions, the test should yield similar result (Brown, 2004 :20).

According to Airasian and Russel (2008 : 18), a second important characteristic of appropriate decisions is that they are based on assessment information that has consistency, or reliability. Reliability refers to the stability or consistency of assessment information, i. e., whether it is typical of a student's behavior.

Airasian and Russel give the key aspects of assessment reliability:

- (1) Reliability refers to the stability or consistency of assessment information and is concerned with this question: "How consistent or typical of the students' behavior is the assessment information I have gathered?"
- (2) Reliability is not concerned with the appropriateness of the assessment information collected, only with its consistency, stability, or typicality. Appropriateness of assessment information is a validity concern.
- (3) Reliability does not exist on an all-or-nothing basis, but in degrees: high, moderate, or low. Some types of assessment information are more reliable than others.
- (4) Reliability is a necessary but insufficient condition for validity. An assessment that provides inconsistent, a typical results cannot be relied upon to provide information useful for decision making.

Reliability must be assured in terms of the following:

- (1) Consistency across tasks
- (2) Consistency in scoring/grading

<http://eppi.ioe.ac.uk/>

The reliability of a test in this term is also one important factor of instrument to get valid data. A reliable is a consistent and dependable. If a test is reliable, then a students' score on it compared to the scores of his classmates should be similar this relative score another test measuring the same information.

The set of a whole test are checked in terms of their reliability. Reliability refers to what extent the composition test measures what is intended to be measured.

2.2.5 Summative and Formative Assesment

Summative Assessment is the formal testing of what has been learned in order to produce marks or grades which may be used for reports of various types. This is different from **Formative Assessment**, in which the emphasis is on on-going assessments of different types used to judge how best to help pupils learn further.

<http://www.highlandschools-virtualib.org.uk/>

Test model of mid term test consist of these two kinds of the test that are summative and formative assesment. The summative assesment is gained from the result of the study of a first mid term which this is a kind of achievement that shows on how much the student has learned in te course of second language instruction. Meanwhile the formative assesment is gained from the progress of the study in the mid of first term or we can say it as a progress test which measure how much the student has learned in specific course of instruction.

Both summative and formative assesment as general testing terminology are measured from the result of study in the mid first term in specific instructional purpose. The process and achievement both presented in this kind of test without any boundary.

2.2.5.1 Summative Assesment

Summative Assessments are given periodically to determine at a particular point in time what students know and do not know. Many associate summative assessments only with standardized tests such as state assessments, but they are also used at and are an important part of district and classroom programs.

<http://www.nmsa.org/>

Summative assessment at the district/classroom level is an accountability measure that is generally used as part of the grading process. It is usually given at the end of a marking period, in this case in the end of a mid first term of study and measured the total of the material covered. In this kind of test, students usually ranked and graded.

2.2.5.2 Formative Assesment

Formative Assessment is part of the instructional process. When incorporated into classroom practice, it provides the information needed to adjust teaching and learning while they are happening.

<http://www.nmsa.org/>

In this sense, formative assessment informs both teachers and students about student understanding at a point when timely adjustments can be made.

These adjustments help to ensure students achieve targeted standards-based learning goals within a set time frame. This test is given during the course instruction occasionally by the teacher, it's purpose is to show which aspects of the chapter the student has mastered and where the remedial work is necessary. The formative test is usually graded.

2.2.6 Constructing Test

Generally language testing can be divided into two types; language skills and language components. Language skills include the test of listening, speaking, reading, and writing. The last one is testing of language components, phonology/orthography, structure, vocabulary, rate and general fluency. In this section the writer focuses on reading tests.

There is more than one type of learning. A committee of colleges, led by Benjamin Bloom (1956), identified three domains of educational activities:

- **Cognitive:** mental skills (*Knowledge*)
- **Affective:** growth in feelings or emotional areas (*Attitude*)
- **Psychomotor:** manual or physical skills (*Skills*)

The cognitive domain (Bloom, 1956) involves knowledge and the development of intellectual skills. This includes the recall or recognition of specific facts, procedural patterns, and concepts that serve in the development of intellectual abilities and skills. There are six major categories, which are listed in order below, starting from the simplest behavior to the most complex. The categories can be thought of as degrees of difficulties. That is, the first one must be mastered before the next one can take place. According to Anderson & Krathwohl (2001:67-68), the new terms are defined as:

- (1) **Remembering:** Retrieving, recognizing, and recalling relevant knowledge from long-term memory.

- (2) **Understanding:** Constructing meaning from oral, written, and graphic messages through interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining.
- (3) **Applying:** Carrying out or using a procedure through executing, or implementing.
- (4) **Analyzing:** Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing.
- (5) **Evaluating:** Making judgments based on criteria and standards through checking and critiquing.
- (6) **Creating:** Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure through generating, planning, or producing.

Considering to the three domain activities that are cognitive, affective, and psychomotor, the writer conclude that an education institution must fulfill the three unequally. The cognitive have to be covered in a whole and the rest just need to be given less proportional than the first one, it is because Indonesian educational pointing the cognitive one as a target or we can call it as an objective of a teaching and learning process.

The teacher in this case need to give the students their dedication toward students' skills and ability progress according to the those six categories. When the teacher asses the student using that method, hopefully the student development will be easily measured and graded.

2.2.6.1 Reading Test

Reading is one of major skill which has distinctive characteristic from the others. According to Kitao (1997 :2) process of test construction are :

2.2.6.1.1 Choosing a Text

The text that is used to test reading can have an impact on the results. In this section, we will discuss some of the factors that need to be considered.

(1) Variety

It is useful to use a variety of texts when testing reading. It is traditional to use prose to test reading skills, but there are many other types of texts that can be used for testing. Some of these include schedules, newspaper articles, advertisements, and instructions.

(2) Type of Text

The type of text chosen should reflect the goals of the language teaching situation. If students are studying academic English, the texts should come from such sources. If they are studying English to use it in their daily lives, such texts as menus, leaflets, or letters might be useful.

(3) Reading and Background Knowledge

The text should not be about a subject that the testees are either very familiar with or very unfamiliar with. It must be considered that for a text for which they have considerable background knowledge, they will probably be able to answer the questions without even understanding the test. If, in contrast, it is about a subject for which testees have almost no background knowledge, they are likely to have a great deal of difficulty understanding, and particularly making inferences, even if

they are fairly good readers.

(4) Difficulty

In choosing a text, the test maker needs to take into account its difficulty. If the test is too difficult, few of the testees will do well; if it is too easy, too many of the testees will do well. However, the issue of difficulty is not a simple one. For subject areas for which they have good background knowledge, testees may be able to handle a more difficult text than for subject areas for which they have little background knowledge.

The consideration of choosing a simple short text based on the specific topic which was delivered to the students itself. Background of the topic and variety of the simple short text ensure the objective of the study. The type of the text was also based on the ability of an elementary students that would be automatically leading the test fit to them in terms of difficulty level.

2.2.6.1.2 Task to Test Reading

There are a wide variety of tasks that can be used to test reading ability. Since there are many more possible tasks than we can deal with in this article, we will present some of the more common and useful ones.

(1) Testing Low Level Skills

There are some tasks that can be used to test very low level skills such as word recognition, sentence recognition, and word and sentence comprehension. These types of tasks have a place when testing students who are just beginning to read English, but they should not be used to test students beyond that basic level. It is useful to have testees do these tasks under time pressure, since they are

intended to test skills that must be performed automatically in higher levels of reading.

Word recognition. Testees are presented with one word and a series of four or five words. They are instructed to mark the word that is the same as the first word. *Sentence recognition.* This is similar to the previous task. Testees are presented with sentences instead of words and indicate which sentences are the same. *Word and picture matching.* There are two variations on this task. The first is to give testees four similar pictures and one sentence. They are instructed to identify the picture that is described by the sentence. The other task is look at one picture with four similar sentences and decide which of the sentences accurately describes the picture. A variation on this is to give the testees, for example, ten sentences and have them choose five that accurately describe the picture.

(2) Tasks for Middle and Higher Level Students

True/False Questions. Probably the most common type of task is true/false questions. They are particularly useful for progress tests, since they can be constructed relatively quickly and easily, and they are also easy to mark. The drawback of true/false tests, though, are that testees have a fifty/fifty chance of getting the correct answer, meaning that the facility value is likely to be around 75 percent. This means that the test may not discriminate well between high and low level students, unless there are a lot of questions.

The table of six major categories:

Reading Multiple choice	C1	No. item	C2	No. item	C3	No. item	C4	No. item	C5	No. item	C6	No. item
LC1	1.1	13, 14, 16	2.1	2,11, 12,15 27,28	3. 1	3, 19, 21,26	4. 1	7, 20, 23,29	5. 1	8	6.1	-
LC2	1.2	9, 24	2.2	6,10, 18, 2	3. 2	25, 30	4. 2	1,5, 2	5. 2	4, 17	6.2	-
LC3	1.3	31, 34	2.3	32,35	3. 3	33	4. 3	-	5. 3	-	6.3	-
Total		7		12		7		6		3		

Notes:

C1 : Remembering

C6 : Creating

C2 : Understanding

LC 1 : Language competence 1

C3 : Applying

LC 2 : Language competence 2

C4 : Analyzing

LC 3 : Language competence 3

C5 : Evaluating

2.3 Framework of Analysis

The Construction of the test model in *BBC English Training Specialist* as I observed was based on the curriculum or outline which was conducted by this

institution. Many aspect was needed to make a good test before the teacher delivered the test. In conducting this research, I use some theoretical framework as the basic theories of my research as the following:

(1) Curriculum

According to Yunestia (2008 : 7) curriculum is the basic of doing teaching learning activities. A teacher should know the curriculum before he or she teaches his or her students. A curriculum has been made on the basis of students' need. In Indonesia, government is trying to improve the quality of human resources of its people.

(2) Reading

Hornby (1995:967) states that reading is the process of looking at and understanding the meaning of written or printed words or symbols. It means that when we do reading, we think about the thing that we read and understand the meaning of it.

Since reading is one of major skill which has distinctive characteristic from the others, students need to know reading a lot as their basic skill in learning English in wider concept.

(3) Assessment

Brown (2004:4) explains that assessment is an ongoing process encompasses a much wider domain than that the intention of testing. Whenever student responds to a question, offers a comment, or tries out a new word or structure, the teacher subconsciously makes an assessment of the student's performance.

Meanwhile, Harris (1969:13) points out “three characteristics of a good test called : reliability, validity, and practicality.

The writer concluded that assessment is an ongoing process which does not only focus on the result of the test, but also emphasizes on the process of measuring the students’ skill and it needs some requirements to conduct a good test such as reliability and validity.

(4) *Research Design*

The data sources in a research are basically subjects from which a researcher gets data (Arikunto, 1986:102). The writer used both quantitative and qualitative method in doing the analysis research plus descriptive because he analyze the test and describe wether the test model is good or not to be delivered to the students.

The quantitative research will be measured by computing four problem related to the test model; reliability, validity, discrimination power and difficulty level. In other hand, the qualitative research is measured according to current status of phenomena. They are directed toward determining the nature situation of the study.

CHAPTER III

METHOD OF INVESTIGATION

In every research, there are data and method to study. The data can be taken from many sources, either people or things. The data sources in a research are basically subjects from which a researcher gets data (Arikunto, 1986:102). Related to this statement, the writer has to select the subject of the population of the study.

In this chapter the writer presents research design, the main sources of data, the subject that is the population and sample, and the method of collecting data which contain the instrument and the procedure of collecting data.

3.1 Research Design

Every research must have method of investigation to obtain a good result. This method is called research method. There were two methods of investigation; qualitative research and quantitative research. The research method that the writer used in this final project was descriptive quantitative qualitative research.

According to Brown (1988:82) quantitative is connected with amount of number of something rather than with how good it is. In this term quantitative data, refers to the use of statistical analysis to calculate the numerical data that were gathered and to analyze them by use of correlation analysis.

Referring to Christensen (2001:32) quantitative research is the research that collect some type of numerical data to answer a given research question. In quantitative research we classify features, count them, and even construct more complex statistical models in an attempt to explain what is observed.

Quantitative research which was used by the writer in analyzing the topic was based on the computation of the whole test including validity, reliability, discrimination power and difficulty level. From the result of the computation then the writer conclude wether was good or not.

Devy (1997:76) stated that qualitative research studies are designed information concerning the current status of phenomena. They are directed towards determining the nature of situation, as it exists at the time of the study. There is no administration or control of the treatment as it is found in experimental testing. The aim is to describe what exist with respect to variables or conditions or situations.

According to Nunan(1992:5) qualitative research advocates the use of qualitative methods, concern with the understanding of human behavior from the actor's own frame of reference explanatory, descriptive, and process oriented. It is also supported by Norman K Denzin(1994:1) that qualitative research is field of inquiry in its own right. He also give some descriptions for qualitative research. First, the world qualitative implies an emphasis on processes and meanings that are not rigorously examined or measured. It shows that in qualitative approach, the emphasis in on process of the research/study. It is not emphasis on examining

or measuring, like in quantitative approach. It is emphasis on describing or explaining. Second, qualitative research is multimethod in focus, involving an interpretive, naturalistic approach to its subject matter. Multimethod means that in qualitative approach use more than one method to do research. Third, qualitative researchers stress the socially constructed nature of reality. The intimate relationship between the researcher and what is studied and the situational constraint that shape inquiry. It means that in qualitative approach, the research is done naturally.

In short from the definitions above, it is clear that qualitative research is a kind of research with intent to describe and explain certain phenomena which occur within community. It means that qualitative approach implies and emphasize on processes that are not rigorously examined or measured. Qualitative research presented the data and research in the form of qualitative description. Analysis of this type was done with words to describe conclusions, so the qualitative study obtained the descriptive data either spoken or written.

In this case, the writer collected the data by doing research at the test model of mid term test for sixth grade students of *BBC English Training Specialist*. The writer analyzed the requirement of a good test which consist of validity, reliability, discrimination power, and difficulty level of the test. The analysis was about how the test met the requirement of a good test or not and whether the test could be used in the mid term test or should be discarded.

3.2 The Main Sources of Data

The writer needs the data in his efforts to find the solutions to the problems discussed in this thesis. The data and information required are obtained from two main sources.

3.2.1 Library Research

Library research refers to the activity of using library facilities. He tried to get the data, information and ideas by reading books on the related subject matters. The information and ideas from significant authors are the fundamental theories that support his ideas in finding the solutions to the problems in this final project.

3.2.2 Field Research

Field research refers to the activity of getting empirical data. He used a test which was administered to the sixth grades students of *BBC English Training Specialist*. The result of this case also shows the quality of the test, would be a source of the data.

3.3 Population and Sample

According to Margono (2003:118) population is defined a complete set of individuals or subjects which having common observable characteristic. The populations can vary widely depending on the research question and the purpose of the study. The population chosen by the writer in this research was sixth grade students of *BBC English Training Specialist*.

In order to study a problem more effectively, a researcher has to select a sample. Vockell (1983: 105) states that a sample has to be representative of the population about which we wish to generalize. Best points out that a sample is a small proportion of a population selected for observation and analysis. The process of taking the sample will be discussed in the following section.

3.4 Sampling Technique

In this case the writer selected the sample by using the lottery method. This method is an objective selection. He did it by writing down each of the students' numbers (from number 1 to number 24) on small pieces of paper. Then the pieces of paper were rolled and put into a drinking glass. After he mixed those rolls well, he let the required number of rolls of papers drop out of the glass one after another. The selected number, then, were matched with the students' names in the list. The answer sheets were taken from these students (in accordance with students' names in list). Although the steps in taking samples are very simple, many researchers have adapted the random sampling technique as one way to select the samples since it is not influenced by thoughts and feelings.

3.5 Identification of the Problems

There are four problems related to the teacher-made test items that will be analyzed in this study, they are:

- a) Difficulty level;
- b) Discrimination power;

- c) Reliability;
- d) Validity;

There are thirty five items are multiple-choice form that will be analyzed. The writer hopes that the teacher can use the result of this study or other test constructors in analyzing test items in the future.

3.6 Technique of Data Collection

In order to get the required data, the writer used the test technique for analyzing each test item. In this case, the intended test is the model of structure, vocabulary, and reading mid term for the sixth grade students. The data, here, are in the form of students' scores on their answer sheets.

Before the test was administered to the students, the writer had contacted the English teacher of the selected institution and doing research. Then he began to analyze them.

3.7 Technique of Data Analysis

The data to be analyzed in this study are taken from the students' answer sheets of the model of structure, vocabulary, and reading mid term test for the sixth grade students of *BBC English Training Specialist*. These students' answer sheets are used to analyze the quality of the test items.

The purpose of this item analysis is to identify the quality of each test item, whether they belong to good items, moderate items, or bad items. Through item

analysis, we can also find information about the weaknesses or the shortcomings of the items. Here, the item analysis consist of four things, they are:

3.7.1 Difficulty Level Analysis

A good test item is an item which is not too difficult and not to easy. While according to Arikunto (1995:210) a good test is an item which has index of difficulty between 0.30 and 0.70. The criteria of item difficulty level could be seen in the table below:

Index Difficulty Level	Categories
0.00 < P ≤ 0.30	Difficult
0.30 < P ≤ 0.70	Moderate
0.70 < P < 1.00	Easy

(Heaton, 1975: 172)

The formula is: $P = \frac{R}{T}$

Where P = difficulty level or index of difficulty

R = the number of students responding correctly to an item

T = the total member of students responding to the item

(Nitko, 1983 : 228)

To illustrate, if 40 students answered an item, and only 30 of them answered it correctly, then the difficulty level (P) of that item would be $30/40 = 0.75$. According to above criteria this item belongs to the easy item.

3.7.2 Discrimination Power Analysis

Item discrimination tells how well the item performs in separating the better students from the poorer ones. To get the index of item discrimination, the writer uses the following steps:

- 1) Arrange the papers in rank order of total score and divide them into three groups of equal size, approximately 25% of the papers in the top group and 25% in the lower group.
- 2) Count the number of those students in the top group answering the first item correctly and then count the number of the bottom group of the students answering the item correctly.
- 3) Subtract the number of correct answers in the bottom group from the number of correct answers in the top group; find the difference in the proportion passing in the top group and the proportion passing in the bottom group.
- 4) Divide the difference by the total number of candidates in one group.

The formula for computing item discrimination power is as follows:

$$D = \frac{U - L}{N}$$

In which D = the index of item discrimination power

U = the index of students in the upper group who answered the item correctly

L = the number of the students in the lower group who answered the item correctly

N = the number of students in each of the group

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

No	Discrimination index	Item evaluation
1	$0.70 < D \leq 1.00$	Excellent
2	$0.40 < D \leq 0.70$	Good
3	$0.20 < D \leq 0.40$	Satisfactory
4	$0.00 < D \leq 0.20$	Poor

3.7.3 Analysis of Reliability

Reliability refers to the consistency of test scores. That is, how consistent scores of the test or other evaluation results are from one measurement to another. A test is considered to have high reliability if the test can give a reliable result. If there is a change of the result, the change is worthless.

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

$$r = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum pq}{S^2} \right)$$

Where r = reliability coefficient of the test item

k = number of items in the test

p = the difficulty index

q = the proportion of the students who give the wrong answer ($q = 1-p$)

S^2 = the variance of the total test scores

(Phopam, 1981 :1430)

The formula to calculate variance is:

$$S^2 = \frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N}$$

The result of the reliability (r) calculation is consulted to the value of crucial (r) production product moment on the table. If the value of (r) calculation is lower than the value of (r) on the table, we can say that the item is not reliable. On the other hand, the item is reliable if the value of (r) calculation is more than the value of (r) on the table.

3.7.4 Analysis of Validity

"Validity of a test is the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment"(Gronlund, 1998: 226)

In this study the writer uses content validity and criterion-related validity (validity coefficient) to determine whether the test items were valid or not. To find the coefficient validity, he used Pearson's Product Moment formula:

$$r_{pbis} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

Where: M_p = Mean of students score responding correctly to the item

M_t = Mean of total score

S_t = Standard deviation

p = the proportion of the students who give the correct answer

q = the proportion of the students who give the wrong answer

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

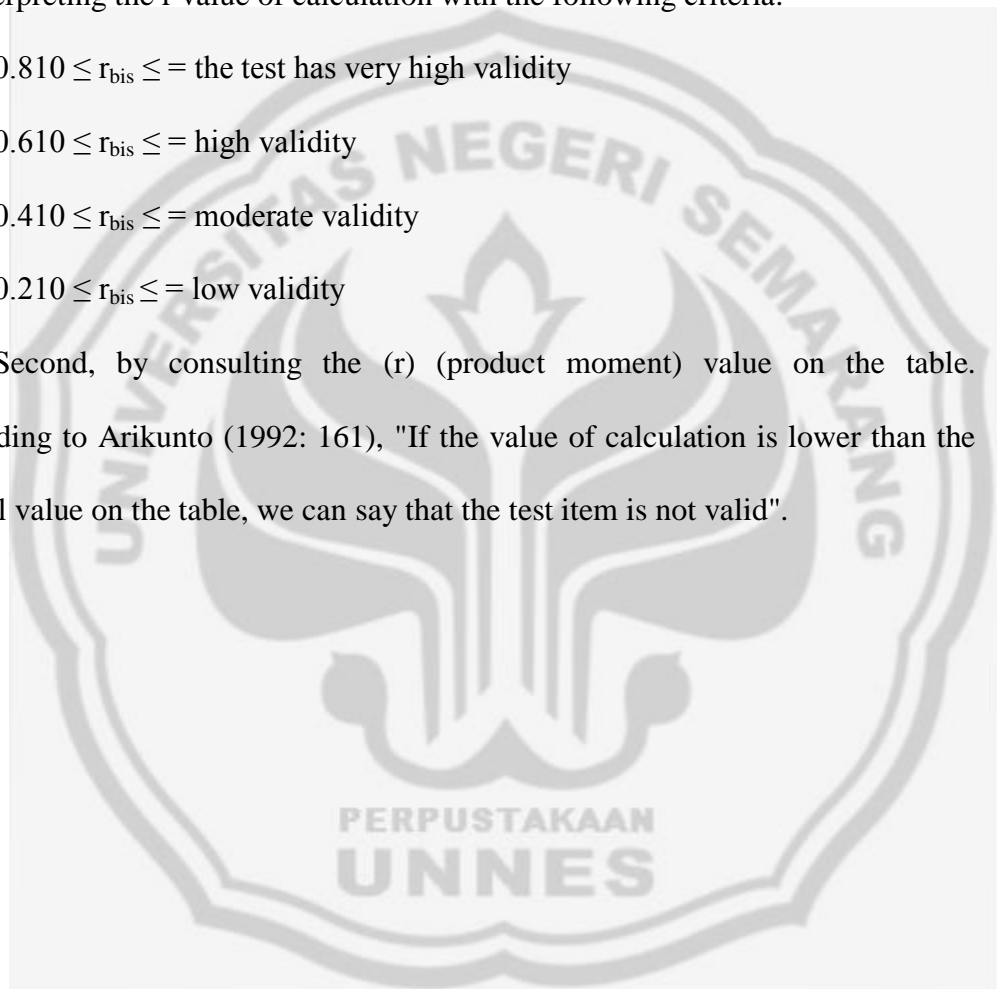
$0.810 \leq r_{bis} \leq 1$ = the test has very high validity

$0.610 \leq r_{bis} \leq 0.810$ = high validity

$0.410 \leq r_{bis} \leq 0.610$ = moderate validity

$0.210 \leq r_{bis} \leq 0.410$ = low validity

Second, by consulting the (r) (product moment) value on the table. According to Arikunto (1992: 161), "If the value of calculation is lower than the critical value on the table, we can say that the test item is not valid".



CHAPTER IV

ANALYSIS AND DISCUSSION

In the fourth chapter, the writer would like to discuss the result of analysis, analysis of difficulty level, analysis of discrimination power, analysis of validity, analysis of reliability, and discussion.

4.1 Analysis

The goal of this study is constructing English test model of structure, vocabulary and reading mid term for sixth grade students of *BBC English Training Specialist*. The analysis consists of four aspects, namely the difficulty level, discrimination power, validity, reliability of the test.

Item analysis aims to identify good, moderate, and poor items. From item analysis, we get information about the items are good or not and how to revise them. From data analysis of the English test model of structure, vocabulary and reading mid term for sixth grade students of *BBC English Training Specialist*, the writer obtained the following results.

4.1.1 Analysis of the Item Difficulty Level

From the table in the appendix 3 and computation in the appendix 6, the result of the data analysis of the Item Difficulty Level shows as follows.

No.	Criteria	Number	Percentage
1.	Difficult Items	0	0 %
2.	Moderate Items	8	23 %
3.	Easy Items	27	77 %

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

- (1) There are no item can be categorized as difficult.
- (2) Items that belong to the moderate category are 1, 8, 12, 16, 17, 22, 23, and 29.
- (3) The following items are categorized as easy; they are 2, 3, 4, 5, 6, 7, 9, 10, 11, 13, 14, 15, 18, 19, 20, 21, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34 and 35.

From those 35 test items, the mean of the difficulty level value is 0.80 taken from the subtraction of the number of difficulty level (P) by 35. So as a whole, the test items are easy in terms of their difficulty level. Items, which are considered very easy, can still be used in a test to encourage and motivate the poor students.

4.1.2 Analysis of Item Discriminative Power

From the table in appendix 3 and computation in appendix 7, the result of data analysis can be seen on this table.

No.	Criteria	Number	Percentage
1.	Excellent Items	2	6 %
2.	Good Items	6	17 %
3.	Satisfactory Items	5	15 %
4.	Poor Items	22	62 %

Discrimination index	Criteria
$0.70 \leq D \leq 1.00$	Excellent Items
$0.40 \leq D \leq 0.70$	Good Items
$0.20 \leq D \leq 0.40$	Satisfactory Items
$0.00 \leq D \leq 0.20$	Poor Items

- (1) Items that can be categorized as excellent items are 1 and 22.
- (2) Items number 6, 8, 15, 19, 23, 32 are categorized as good items.
- (3) Items number 3, 5, 16, 29, 34 are categorized as satisfactory items.
- (4) Items number 2, 4, 7, 9, 10, 11, 12, 13, 14, 17, 18, 20, 21, 24, 25, 26, 27, 28, 30, 31, 33, 35 are classified as poor items.

From the 35 test items, the mean of the item discrimination power value is 0,257. So as the whole the final test items are still need to be improved.

4.1.3 Analysis of Item Validity

$$r_{pbis} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

On a = 5% with N= 24 it is obtained = 0.404

In this study, the writer uses the product moment correlation formula in calculating the value of the item validity (r) of each item. The value of r calculation is then consulted to the table of r product moment values. If the value of (r) calculation is lower than the value of (r) on the table, the test item is invalid. For N 24 with significance level 0.05, the value of r on the table is 0. 404 (see appendix 2). The result of data analysis is shown on the following table.

No.	Criteria	Number	Percentage
1.	Valid Items	17	48.5 %
2.	Invalid Items	18	51.5 %

- (1) There are 17 test items, which fulfill the requirements of validity. They are items number 1, 3, 4, 5, 6, 8, 9, 11, 15, 19, 22, 23, 24, 25, 29, 32 and 33.
- (2) There are 18 items, which do not fulfill the requirement of the validity. They are items number 2, 7, 10, 12, 13, 14, 16, 17, 18, 20, 21, 26, 27, 28, 30, 31, 34 and 35.

From the table above, we found that only 48.5 % of items are valid. From the computation which is listed, the item which is found that the value of the validity is more than 0.404, it means that the item is valid and can be applied in the test. The example of the computation of item validity is listed in appendix 4.

4.1.4 Reliability

$$r_{11} = \left[\frac{35}{35 - 1} \right] \left[\frac{12.693 - 4.595}{12.693} \right]$$

$$r = 0.652$$

As the writer has stated in the previous chapter, the coefficient of reliability of test items is found by applying the Kuder-Richardson 11 formula. From the computation, it is found that the coefficient of the test is 0.657. The result is then consulted to the table of r product moment values at level of significance of 5%. It is found that the value of r 0.404 for N 24. Since value of r calculation is more than that of table (r), it can be conclude that the test items used in English test model of structure, vocabulary, and reading mid term test for sixth grade students *BBC English Training Specialist* is reliable. The computation of the reliability coefficient is listed in appendix 5.


4.2 Discussion

The goal of the writing of this final project is constructing English test model of structure, vocabulary, and reading mid term test for sixth grade students *BBC English Training Specialist* and to identify the quality of each item, whether

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

Based on the result of item analysis which includes the analysis of difficulty level, discrimination power, validity, and reliability of the items, this test items can be used in English mid first term with revisions. For more details will be explained as follows.

➤ Item number 1

Question	 <p>1. Bunga : Where is the security? Amel : He is . . . the bank A. in B. at C. behind D. in front of</p>			
Result	A	B	C	D*
Upper 25%	0	0	0	6
Middle 50%	3	0	0	9
Lower 25%	4	2	0	0
Total	7	2	0	15
Validity	Valid			

D value	1.00
P value	0.62

The item is difficult in terms of moderate level, because there are 15 out of 24 students answered the question correctly. This item has excellent discrimination power, proved with 1.00 values. This item is valid. This item does not need any revision.

➤ Item number 2

Question	2. Dana : Hello, Tika! How are you? Tika : Hello, Dana . . . , thank you. And how are you? A. we are fine B. you are fine C. I am fine D. She is fine			
Result	A	B	C*	D
Upper 25%	0	0	6	0
Middle 50%	0	0	12	0
Lower 25%	0	0	6	0
Total	0	0	24	0
Validity	Invalid			
D value	0.00			
P value	1.00			

This item is easy one, because all of 24 students answered the question correctly, while the difference of the upper group students and the lower ones who chose the correct answer is 0, which put this item in the group that should be discarded. Besides being too easy, the item has poor discrimination power and

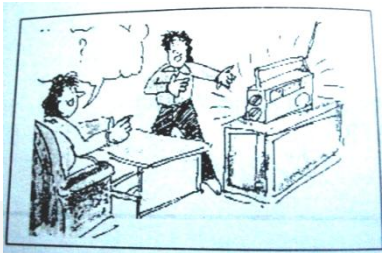
poorly constructed, so the students can guess the answer easily. In writer's opinion this item should be revised.

➤ Item number 3

Question	3. Lisa: What do you want to do, Adam? Adam: I want . . . a letter A. send B. to send C. letter D. sent			
Result	A	B*	C	D
Upper 25%	0	6	0	0
Middle 50%	4	8	0	0
Lower 25%	2	4	0	0
Total	6	18	0	0
Validity	Valid			
D value	0.33			
P value	0.75			

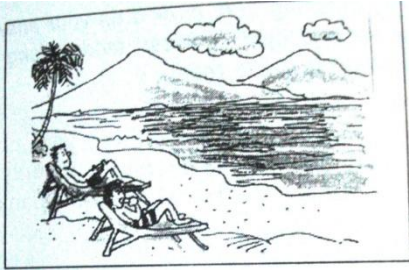
With a P value 0.75 this item is classified as easy, because 18 out of 24 students choosed the right answer. Meanwhile, from the D value the item is classified as satisfactory item. The distribution is good in terms of students' response. This item is valid. This item doesn't need any revision and can be applied in the mid term test.

➤ Item number 4

Question	<p>4. </p> <p>What is the woman saying?</p> <p>A. Can you help me to carry the radio?</p> <p>B. Can you take me to the hotel?</p> <p>C. Can you turn off the radio?</p> <p>D. Can you move the radio?</p>			
Result	A	B	C*	D
Upper 25%	0	0	6	0
Middle 50%	0	0	11	1
Lower 25%	0	0	5	1
Total	0	0	22	2
Validity	Valid			
D value	0.16			
P value	0.91			


This item is easy, which is showed from the number of difficulty index value. From the table above we find out that 22 students chose the correct answer. The item is poor item of discrimination power but valid in terms of validity. Based on the criteria this item can be used in the mid term test without any revision.

➤ Item number 5

Question				
	<p>5. B: What could you do there? A : I could , and I enjoyed it very much</p> <p>A. sun bath B. swim C. fish D. pay sand</p>			
Result	A*	B	C	D
Upper 25%	6	0	0	0
Middle 50%	11	1	0	0
Lower 25%	4	2	0	0
Total	21	3	0	0
Validity	valid			
D value	0.33			
P value	0.87			

The item is easy in terms of difficulty level, because there are 21 out of 24 students answered the question correctly. This item has satisfactory discrimination power. This item is valid, it is because the answer can be identified from the dialogue. In writer's opinion this item could be used in mid term test.

➤ Item number 6

Question	 <p data-bbox="635 651 1142 976">6. X: What are the tellers doing? Y: They are A. waiting for customer B. serving the customer C. standing in line D. talking on the phone</p>			
Result	A	B*	C	D
Upper 25%	0	6	0	0
Middle 50%	1	11	0	0
Lower 25%	4	2	0	0
Total	5	19	0	0
Validity	valid			
D value	0.66			
P value	0.79			

This item is an easy one, because 19 students answered the question correctly. Besides being easy, the item has good discrimination power and good constructed. In writer's opinion this item can be used in mid term test.

➤ Item number 7

Question	<p data-bbox="635 1753 1217 1962">7. We must our money in the bank. A. save B. pay C. put</p>
----------	---

D. take				
Result	A*	B	C	D
Upper 25%	6	0	0	0
Middle 50%	10	0	2	0
Lower 25%	5	0	1	0
Total	21	0	3	0
Validity	Invalid			
D value	0.16			
P value	0.87			

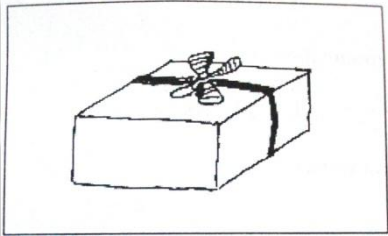
The item is easy in terms of difficulty level, because there are 21 out of 24 students answered the question correctly. This item has poor discrimination power. This item is invalid, it is because the answer can be identified easily from the dialogue. In writer's opinion this item should be revised.

➤ Item number 8

Question	8. The clerk: Can I help you? Indra : Which is the most appropriate respon? A. No, I need some stamps. B. Yes, I need some stamps. C. Of course, I want to buy some stamps. D. Because I want to buy some stamps.			
Result	A	B*	C	D
Upper 25%	0	6	0	0
Middle 50%	0	7	5	0
Lower 25%	0	3	3	0
Total	0	16	8	0
Validity	Valid			
D value	0.50			
P value	0.67			

This item is not easy or difficult in term of difficulty level, which is mean this item is moderate level. In term of discrimination power, this item is good. From the table above we find out that 16 students choose the correct answer. The distribution is good in terms of students' response. In writer's opinion this item can be used in mid term test without any revision.

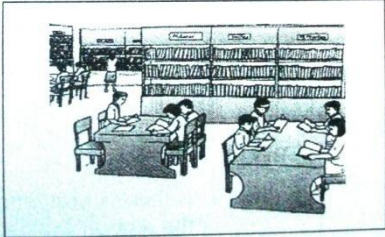
➤ Item number 9

Question				
9.	What do you call this? It is a A. aerogramme B. money order C. package D. postal order			
Result	A	B	C*	D
Upper 25%	0	0	5	1
Middle 50%	3	0	9	0
Lower 25%	2	0	4	0
Total	5	0	18	1
Validity	Valid			
D value	0.16			
P value	0.75			

This item is easy in terms of the difficulty level but it is poor in discrimination power, but it is valid in terms of the validity. From the table, we

can see that 18 students answered it correctly. It means that this item can be used in the mid term test.

➤ Item number 10

Question				
10. What is this? It is a				
A. beach				
B. library				
C. classroom				
D. shop				
Result	A	B*	C	D
Upper 25%	0	5	1	0
Middle 50%	0	12	0	0
Lower 25%	0	5	1	0
Total	0	22	2	0
Validity	Invalid			
D value	0.00			
P value	0.91			

With a P value 0.91 this item is classified as easy item, because 22 out of 24 students chose the right answers. Meanwhile, from the D value the item is classified as poor item. The distribution is just too good in terms of students' responses which make no difference in upper and lower who answer correctly. This item is invalid. This item needs to be revised.

➤ Item number 11

Question	11. You want someone to move the table. What do you say? A. Would you move the table, please? B. Should I move the table, please? C. Must you move the table, please? D. Do I have to move the table, please?			
Result	A*	B	C	D
Upper 25%	6	0	0	0
Middle 50%	11	0	1	0
Lower 25%	5	0	1	0
Total	22	0	2	0
Validity	Valid			
D value	0.16			
P value	0.91			

The item is easy in terms of difficulty level, because there are 22 out of 24 students answered the question correctly. This item has poor discrimination power. This item is invalid, it is because the answer can't be identified easily from the dialogue. In writer's opinion this item should be revised.

➤ Item number 12

Question	12. Nina : How was your weekend? Judi : It was . . . , we went to the mountain. A. beautiful B. interested C. great D. funny
-----------------	---

Result	A	B	C*	D
Upper 25%	1	2	3	0
Middle 50%	2	1	5	4
Lower 25%	3	0	2	1
Total	6	3	10	5
Validity	Invalid			
D value	0.16			
P value	0.41			

The item is moderate in terms of difficulty level, because there are only 10 out of 24 students answered the question correctly. This item has poor discrimination power. This item need to be revised before applied in the mid term test since this item is not valid because the validity index is negative (- 0.30)

➤ Item number 13

Question	13. What is the right word for it? T - T - S - I - R - O - U A. TSIROUT B. TROUSIT C. TOURIST D. TUROIST			
Result	A	B	C*	D
Upper 25%	0	0	6	0
Middle 50%	0	0	12	0
Lower 25%	0	0	6	0
Total	0	0	24	0
Validity	Invalid			
D value	0.00			
P value	1.00			

The item is moderate in terms of easy level, because all of the students answered the question correctly. Beside having poor discrimination power, this item is invalid, it is because the answer can be identified easily from the question. In writer's opinion this item should be revised.

➤ Item number 14

Question	14. People often go to the beach and mountain on A. holidays B. weekdays C. schooldays D. workdays			
Result	A*	B	C	D
Upper 25%	5	1	0	0
Middle 50%	9	3	0	0
Lower 25%	5	1	0	0
Total	19	5	0	0
Validity	Invalid			
D value	0.00			
P value	0.79			

This item is easy one. There upper one is larger than the lower one. From the table above we find out that 19 students choosed the correct answer. The item is poor item since there is no difference between upper and lower group. Based on the criteria and the distribution of students' response, this item should be revised.

➤ Item number 15

Question	15. A: Would you take me a glass of water?
-----------------	--

	I am ...			
	B: O.K.			
	A. hungry			
	B. angry			
	C. thirsty			
	D. dirty			
Result	A	B	C*	D
Upper 25%	0	0	6	0
Middle 50%	1	0	10	1
Lower 25%	3	0	2	1
Total	4	0	18	2
Validity	Valid			
D value	0.66			
P value	0.75			

This item is easy in terms of the difficulty level and good in discrimination power, and also it is valid in terms of the validity. From the table, we can see that 18 students answered it correctly. It means that this item can be used in the mid term test.


➤ Item number 16

Question	16. X: Where is Mrs. Anwar? Y: in the bank. A. he B. she C. him D. her			
Result	A	B*	C	D

Upper 25%	1	4	0	1
Middle 50%	1	10	0	1
Lower 25%	3	2	0	1
Total	5	16	0	3
Validity	Invalid			
D value	0.33			
P value	0.66			

From the table above we find out that 16 students choosed the correct answer. This item is moderate in term of difficulty level and satisfactory in term of discrimination power but it does not have test validity. In writer's opinion this item should be revised.


➤ Item number 17

Question	 <p>17. Andini : Is there a post office near the school? Soni : Yes, there is. It is . . . the police station.</p> <p>A. near B. beside C. next to D. across from</p>			
Result	A*	B	C	D
Upper 25%	1	1	4	0
Middle 50%	6	3	2	1
Lower 25%	1	3	2	0

Total	8	7	8	1
Validity	Invalid			
D value	0.00			
P value	0.33			

This item is moderate in term of difficulty level and poor in term of discrimination power. From the table above we find out that 8 students choosed the correct answer. The distribution isn't good in terms of students' response. Meanwhile, the difference between those who answered it correctly from the upper group and the lower group is 0. In writer's opinion this item should be revised.

➤ Item number 18

Question				
	<p>18. Will you? It is dirty.</p> <p>A. Wash the floor</p> <p>B. mop the floor</p> <p>C. sweep the floor</p> <p>D. brush the floor</p>			
Result	A	B	C*	D
Upper 25%	1	0	5	0
Middle 50%	0	0	12	0
Lower 25%	1	0	5	0
Total	2	0	22	0
Validity	Invalid			

D value	0.00
P value	0.91

This item is easy in terms of the difficulty level and poor in discrimination power and it is invalid in terms of the validity. From the table, we can see that 22 students answered it correctly. In the writer's opinion, this item should be revised.

➤ Item number 19

Question	19. You want someone to listen to you. What do you say? A. Do you listen to me? B. Do you hear me? C. Can I talk to you? D. Can you listen to me?			
Result	A	B	C	D*
Upper 25%	0	0	0	6
Middle 50%	3	0	0	9
Lower 25%	3	1	0	2
Total	6	1	0	17
Validity	Valid			
D value	0.66			
P value	0.70			

This item is moderate in term of difficulty level and good in term of discrimination power. From the table above we find out that 17 students choosed the correct answer. The distribution isn't good in terms of students' response. Meanwhile, the difference between those who answered it correctly from the

upper group and the lower group is good. In writer's opinion this item does not need any revision to be used in the mid term test.

➤ Item number 20

Question	20. Could you lend me your pen, please? I am going to . . . a letter. A. send B. receive C. write D. read			
Result	A	B	C*	D
Upper 25%	0	0	6	0
Middle 50%	1	0	11	0
Lower 25%	0	0	5	1
Total	1	0	22	1
Validity	Invalid			
D value	0.16			
P value	0.91			


The item is easy in terms of difficulty level, because there are 22 out of 24 students answered the question correctly. This item has poor discrimination power. This item is invalid, it is because the answer can be identified easily from the dialogue. In writer opinion this item should be revised.

➤ Item number 21

Question	21. A: . . . does the Andri family often go on holidays? B: The beaches or mountains. A. When B. Where C. What D. Which			
Result	A	B*	C	D
Upper 25%	0	4	2	0
Middle 50%	1	10	1	0
Lower 25%	1	3	2	0
Total	2	17	5	0
Validity	Invalid			
D value	0.16			
P value	0.70			

This item is easy in term of difficulty level and poor in term of discrimination power. From the table above we find out that 17 students choosed the correct answer. Since it is not valid, in writer's opinion this item should be revised before can be used in the mid term test.

➤ Item number 22

Question	
	<p>22. Teacher : What do you plan to do on weekend? Students: We plan to</p> <p>A. go camping B. go climbing C. go hiking</p>

	D. go hunting			
Result	A	B	C*	D
Upper 25%	0	0	6	0
Middle 50%	3	1	8	1
Lower 25%	3	2	1	0
Total	6	3	15	1
Validity	Valid			
D value	0.83			
P value	0.62			

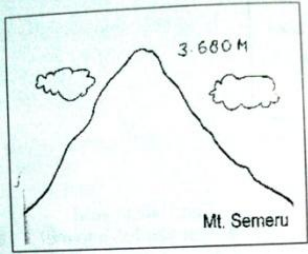
This item is moderate in term of difficulty level and excellent in term of discrimination power. From the table above we find out that there are 15 students choosed the correct answer. In writer's opinion does not need any revision.

➤ Item number 23

Question	23. X: What can you do in the river? Y: Many things, except A. swimming B. fishing C. go boating D. climbing			
Result	A	B	C	D*
Upper 25%	0	0	0	6
Middle 50%	2	3	1	6
Lower 25%	1	3	0	2
Total	3	6	1	14
Validity	valid			
D value	0.66			
P value	0.58			

This item is moderate in difficulty level. From the table above we find out that 14 students chose the correct answer. The item is good in terms of discrimination power and valid in terms of validity. Based on the criteria and the distribution of students' response, this item can be applied in the test.

➤ Item number 24

Question	<p>24. A: How is it? B: It's 3.680 m.</p> <p>A. tall B. high C. long D. deep</p> 			
Result	A	B*	C	D
Upper 25%	1	5	0	0
Middle 50%	1	11	0	0
Lower 25%	1	5	0	0
Total	3	21	0	0
Validity	valid			
D value	0.00			
P value	0.87			

This item is easy in terms of difficulty level from as we can see from the table above that 21 students choosed the correct answer. This item is poor in terms of discrimination power but it is valid, even if it just in low level of validity. In writer's opinion this item does not need any revision and can be used in the mid term test.

➤ Item number 25

Question	 <p>25. What are they doing?</p>
-----------------	--

	<p>A. They are playing</p> <p>B. They are studying</p> <p>C. They are having a picnic</p> <p>D. They are traveling</p>			
Result	A	B	C*	D
Upper 25%	0	0	6	0
Middle 50%	0	0	12	0
Lower 25%	1	0	5	0
Total	1	0	23	0
Validity	Valid			
D value	0.16			
P value	0.95			

From the table above, we find out that 21 students choosed the correct answer. This item is easy in terms of difficulty level and poor in term of discrimination power but it is valid because the $r_{bis}(0.414) > r_{tabel}(0.404)$. In the writer's opinion, it does not need any revision.

➤ Item number 26

Question	<p>26. There are many cars parking there because the bank has . . .</p> <p>A. a large parking lot</p> <p>B. a big yard</p> <p>C. no parking space</p> <p>D. no front yard</p>			
Result	A*	B	C	D
Upper 25%	6	0	0	0
Middle 50%	8	2	2	0
Lower 25%	6	0	0	0

Total	20	2	2	0
Validity	Invalid			
D value	0.00			
P value	0.83			

This item is difficult in term of difficulty level. From the table above we find out that 20 students choosed the correct answer. The difference of upper group and lower group is 0 which make this item poor in terms of discrimination power. In writer's opinion, the item should be revised.

➤ Item number 27

Question	27. It's hot in this room. Will you the window, please? A. shut B. close C. mend D. open			
Result	A	B	C	D*
Upper 25%	0	0	0	6
Middle 50%	0	0	0	12
Lower 25%	0	0	0	6
Total	0	0	0	24
Validity	Invalid			
D value	0.00			
P value	1.00			

The item is easy in terms of difficulty level, too easy probably, because all the students answered the question correctly. This item has poor discrimination power. This item is invalid, it is because the answer can be identified easily from the picture. In writer's opinion this item should be revised.

➤ Item number 28

Question	28. A: What are you doing, Dani? B: I am my homework. A. finishing B. translating C. explaining D. describing			
Result	A*	B	C	D
Upper 25%	6	0	0	0
Middle 50%	10	2	0	0
Lower 25%	5	0	1	0
Total	21	2	1	0
Validity	Invalid			
D value	0.16			
P value	0.87			

This item is easy in term of difficulty level and poor in term of discrimination power. From the table above we find out that 21 students choosed the correct answer. The distribution isn't good in terms of students' response. Meanwhile, the difference between those who answered it correctly from the upper group and the lower group is 1. In writer's opinion this item should be revised.

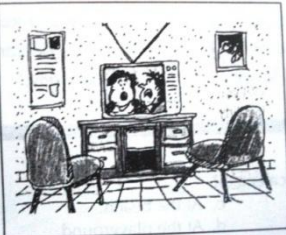
➤ Item number 29

Question	29. The two boys are spending the holiday. They are along the beach. A. cycle B. cycling
-----------------	---

	C. cycled D. to cycle			
Result	A	B*	C	D
Upper 25%	1	5	0	0
Middle 50%	4	7	1	0
Lower 25%	2	3	1	0
Total	7	15	2	0
Validity	Valid			
D value	0.33			
P value	0.41			

This item is moderate in term of difficulty level and satisfactory in term of discrimination power. From the table above we find out that 15 students choose the correct answer. The distribution is good in terms of students' response. In writer's opinion this item can be used in the mid term test.

Item number 30

Question				
30. Could you.....the television? No one is watching it.	<p>A. turn up</p> <p>B. turn down</p> <p>C. turn on</p> <p>D. turn off</p>			
Result	A	B	C	D*
Upper 25%	0	0	0	6
Middle 50%	0	0	0	12
Lower 25%	0	0	0	6
Total	0	0	0	24

Validity	Invalid
D value	0.00
P value	1.00

This item is easy in term of difficulty level and poor in term of discrimination power. From the table above we find out that all students choose the correct answer. The distribution isn't good in terms of students' response. In writer's opinion this item should be revised.

Questions for item number 31-35

The students of SD 8 Cicalengka are at Bandung Zoo. "Look, children," says Mr. Achmad. "There are lions. The lions are in a cage." "Look, are those monkeys?" asks Tono. "No, Tono" replies Mr. Achmad. "Those are not monkeys. They are orangutan."

"Look, children," says Mr. Achmad. "The zoo-keeper is coming to the cage. His name is Pak Mamat. He carries food to the animals. He works at the zoo every day."

➤ Item number 31

Question	31. Where are the students of SD 8 Cicalengka? A. At the conversation B. At the zoo C. At the beach D. At the playground			
Result	A	B*	C	D
Upper 25%	0	6	0	0
Middle 50%	0	12	0	0
Lower 25%	0	6	0	0
Total	0	24	0	0
Validity	Invalid			
D value	0.00			

P value	1.00
----------------	------

This item is easy in term of difficulty level and poor in term of discrimination power. From the table above we find out that all students choose the correct answer. The distribution isn't good in terms of students' response. In writer's opinion this item should be revised.

➤ Item number 32

Question	32. Where are the lions? A. inside the cage B. out of the cage C. next to the cage D. behind the cage			
Result	A*	B	C	D
Upper 25%	6	0	0	0
Middle 50%	11	1	0	0
Lower 25%	2	2	0	2
Total	19	3	0	2
Validity	Valid			
D value	0.66			
P value	0.79			

This item is easy in term of difficulty level but it is good in term of discrimination power. There are 19 students answered the question correctly. In writer's opinion this item does not need any revision and can be used in the mid term test.

➤ Item number 33

Question	33. Who is coming to the cage? A. The children B. The student C. Other visitor D. The zoo-keeper			
Result	A	B	C	D*
Upper 25%	0	0	0	6
Middle 50%	0	0	1	11
Lower 25%	0	1	0	5
Total	0	0	0	22
Validity	Valid			
D value	0.16			
P value	0.91			

This item can be classified as an easy item, because there are 22 students who answer it correctly. From the discrimination point of view, this item is poor but this item is valid in terms of validity. Because this item meets the requirement of validity, it does not need any revision to be used in the mid term test.

➤ Item number 34

Question	34. What is his name? A. Mr. Achmad B. Tono C. Mr. Mamat D. Cicalengka			
Result	A	B	C*	D

Upper 25%	0	0	6	0
Middle 50%	3	2	7	0
Lower 25%	2	0	4	0
Total	5	2	17	0
Validity	Invalid			
D value	0.33			
P value	0.70			

With a P value 0.70 this item is classified as moderate, because 17 out of 24 students chose the correct answer. Meanwhile, from the D value the item is classified as satisfactory item. But unfortunately the item is invalid because the $r_{bis} < 0.210$. This item needs to be revised.

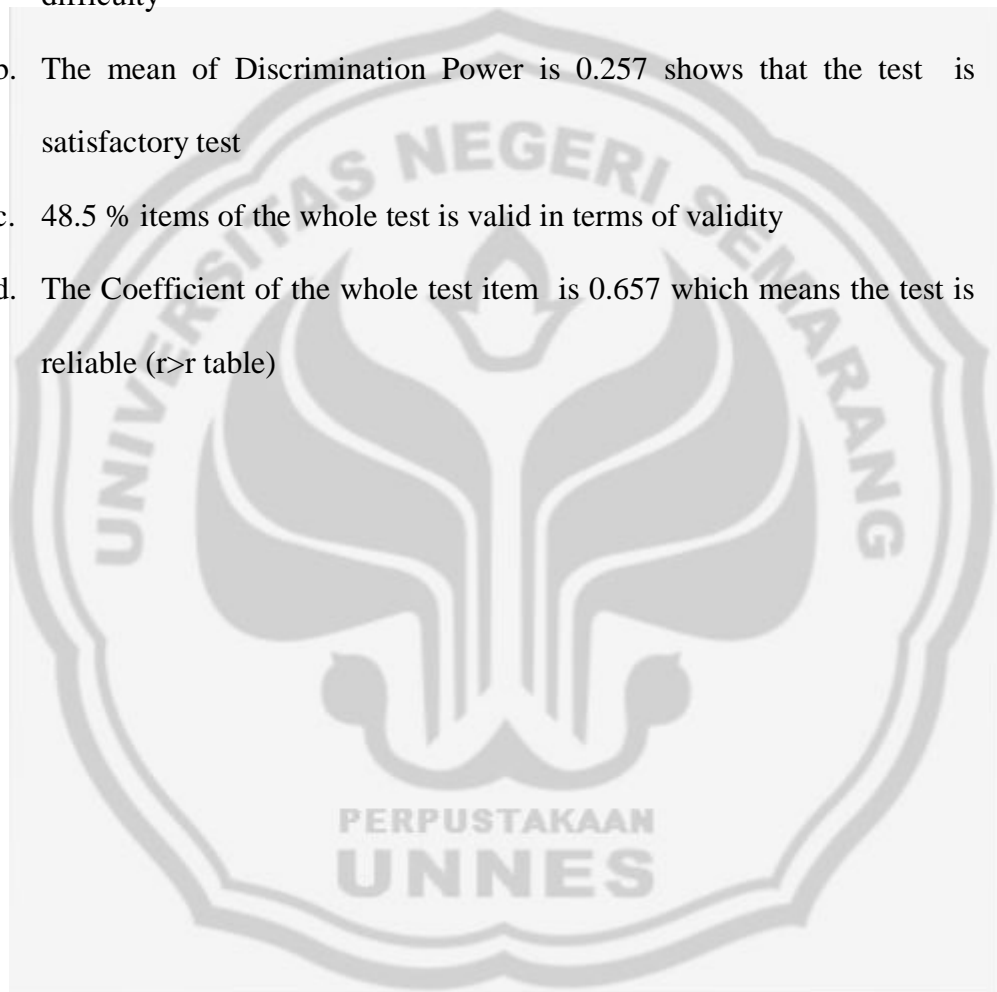
➤ Item number 35

Question	35. What does he bring for the animals? A. drink B. food C. vitamin D. medicines			
Result	A	B*	C	D
Upper 25%	0	6	0	0
Middle 50%	0	12	0	0
Lower 25%	0	6	0	0
Total	0	24	0	0
Validity	Invalid			
D value	0.00			
P value	1.00			

This item is easy in term of difficulty level and poor in term of discrimination power. From the table above we find out that all students choose the correct answer. The distribution isn't good in terms of students' response. In writer's opinion this item should be revised.

The computation result of the test model of midterm test for sixth grade students of *BBC English Training Specialist* in according to the formula of validity, reliability, discrimination power and difficulty level are:

- a. The mean of difficulty Level is 0.80 shows that the test has easy level of difficulty
- b. The mean of Discrimination Power is 0.257 shows that the test is satisfactory test
- c. 48.5 % items of the whole test is valid in terms of validity
- d. The Coefficient of the whole test item is 0.657 which means the test is reliable ($r > r_{table}$)



CHAPTER V

CONCLUSION AND SUGGESTION

The writer has already discussed the result of analysis in the fourth chapter. In the fifth chapter the writer would like to conclude the results and give suggestions.

5.1 Conclusion

According to the result of the analysis of the thirty five test items administrated to sixth grade students of *BBC English Training Specialist* had helped the writer to come the following conclusions.

The mean of the index difficulty of the test was 0.80 which means that the test has an easy level of difficulty. As a result, on the whole of English test model of structure, vocabulary, and reading mid term for sixth grade students of *BBC English Training Specialist* have met requirements of an easy test in terms of the difficulty level.

In analysis of the item discrimination power, it was found out that the mean of the D value was 0.257 which put the English midterm test items in the satisfactory test. In analysis of item validity, it was found that the value of validity of the whole test items is 0.256. It means that the test has low level validity.

By applying KR-11 formula, the writer found that the coefficient of reliability of the whole test item is 0.657. It means that the test as had low

reliability, and we can use the test items as the instrument of evaluation again if we want to with some revisions.

The construction of test model of structure, vocabulary, and reading is conducted by considering to the related problem, that is reliability, validity, discrimination power, and difficulty level. When it meets those requirement in a good criteria, then the test is considered as a good test.

The writer draws a conclusion that the items in the constructing English test model of structure, vocabulary, and reading for mid term for sixth grade students of *BBC English Training Specialist* has an easy level in terms of difficulty level, satisfactory in terms of discrimination power, and low level of validity. It could still be used as an instrument of evaluation with some revisions.

5.2. Suggestion

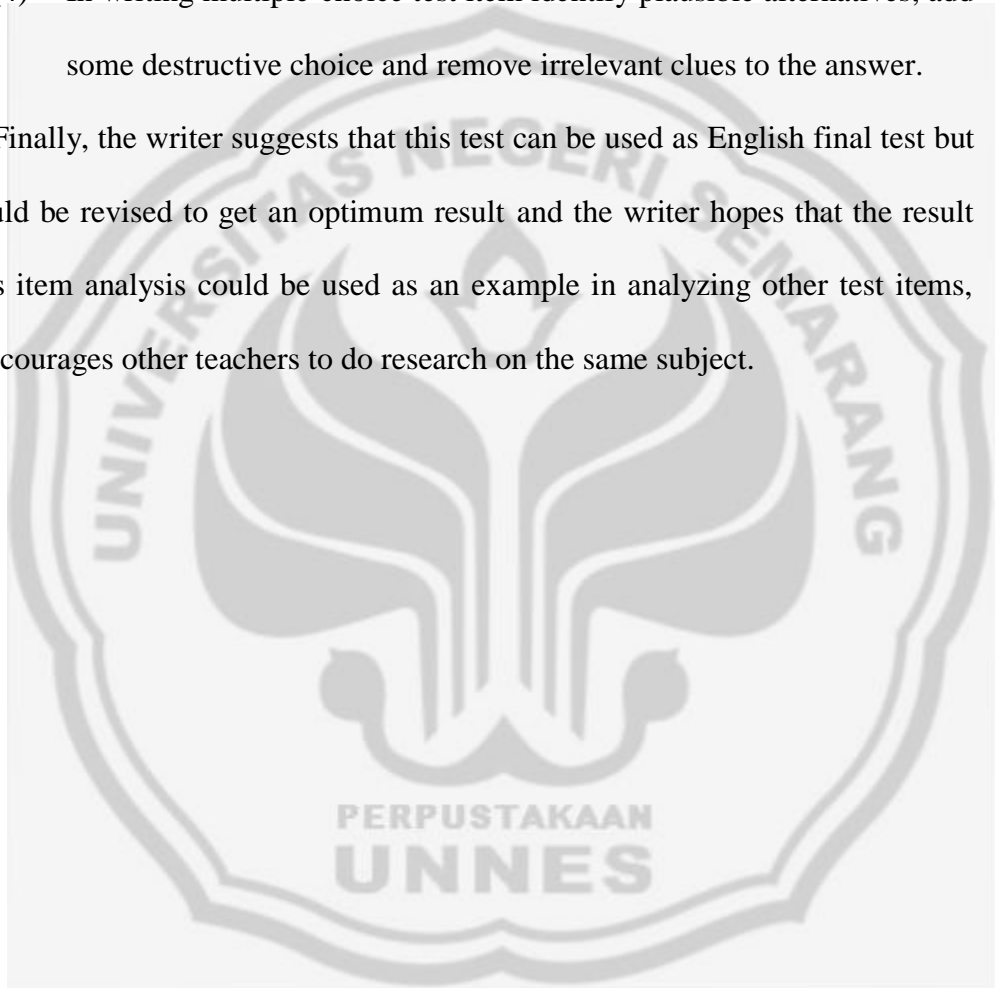
Constructing good language test items is not an easy task. Based on the conclusions above, the writer would like to offer the following suggestions: first, the test constructors should know about the characteristics of a good language test, especially procedure of determining difficulty levels and discrimination power. Second, items that still can be used should be revised and save.

There are some points to be considered in constructing test items.

- (1) Prepare the test items carefully, not in a rush, before they are administered to the students, it will helps the constructor to develop in making good test item.

- (2) Write each test item related to the intended learning outcomes to be measured.
- (3) In the construction of the test full care must be taken that involves formulating a clearly stated problem.
- (4) In writing multiple-choice test item identify plausible alternatives, add some destructive choice and remove irrelevant clues to the answer.

Finally, the writer suggests that this test can be used as English final test but it should be revised to get an optimum result and the writer hopes that the result on this item analysis could be used as an example in analyzing other test items, and encourages other teachers to do research on the same subject.



BIBLIOGRAPHY

- Airasian, Peter W. 2008. *Classroom Assessment : Concept and Applications*. New York : McGraw-Hill Companies.
- Anderson, L. W. and David R. Krathwohl (Eds.). 2001. *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives: Complete edition*. New York: Longman.
- Arikunto, Suharsimi. 1995. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta : Bumi Aksara.
- Brown, D. H. 2004. *Language Assessment: Principles and Classroom Practices*. San Fransisco: Longman.
- Departemen Pendidikan Nasional. 2005. *KBBI. Edisi Ketiga*. Jakarta : Balai Pustaka.
- Ebel, R.L and D.A Frisbie. 1991. *Essential for Educational Evaluation*. Manila: Addison-Wersley Publishing Company.
- Gronlund, N. E. 1981. *Measurement and Evaluation in Teaching: Forth Edition*. New York: Macmillan Publishing co., Inc
- Harris D. P. 1969. *Testing English as a Second Language*. New York: Mc Graw Hill.
- Heaton, J.B. 1975. *Writing English Language Test*. Manchester: Longman Group Limited.
- Hasanah, Millatin. 2008. *Item Analysis of A Teacher Made English Test for Seventh Grade Students of SMP N 02 Bandar in the Academic Year of 2007/2008*. Skripsi tidak diterbitkan. Semarang : Program Sarjana UNNES.
- Hornby. 1995. *Oxford Learning Dictionary*. New York : Oxford University Press.
- Margono, Drs. 2003. *Metodologi Penelitian Pendidikan*. Jakarta: Rineka Cipta.
- Nitko, A.J. 1983. *Educational test and measurement an Introduction*. London: Harcourt Brace Jovanovich Inc.

Phopam, James W. 1981. *Modern Educational Measurement*. London: Prentice Hall Inc., Englewood Cliffs

Vockell, Edward L. 1983. *Educational Research*. New York: Macmillan.

