

**EVALUATION OF A READING TEST ADMINISTERED TO 1<sup>ST</sup>  
YEAR ENGLISH DEPARTMENT STUDENTS  
(A Case Study at Hasanuddin University)**



PERPUSTAKAAN BUKU UNP HASANUDDIN	
Tgl. Terima	2-6-09
Asal Buku	Sustan
Penyedia	lily
Merah	Wahid
No. Inventaris	05
	SKR-BO9
	POM
	2

*A Thesis*  
*Submitted in partial fulfillment of requirement*  
*for the Sarjana Sastra Degree at the Faculty Cultural of Sciences*  
*Hasanuddin University*

**DEMITRIANA PONTO LOLOK**

**F211 04 064**

**MAKASSAR**

**2009**

WHATEVER U DO, DO IT HEARTILY, AS TO THE LORD AND NOT TO MEN

—Colossians 3:23—



This thesis special dedicated to my beloved parents

# SKRIPSI

## EVALUATION OF A READING TEST ADMINISTERED TO THE 1<sup>ST</sup> YEAR ENGLISH DEPARTMENT STUDENTS (A CASE STUDY AT HASANUDDIN UNIVERSITY)

Disusun dan diajukan oleh:

**DEMITRIANA PONTO LOLOK**

No. Pokok: F211 04 064

Telah dipertahankan di depan panitia ujian skripsi  
pada tanggal 20 Februari 2009  
dan dinyatakan telah memenuhi syarat.

Menyetujui  
Komisi Pembimbing,

Konsultan I

  
Dr. Hj. Ety Bazergan, M.Ed

NIP. 131 287 221

Konsultan II

  
Sitti Sahraeny, S.S., M.A.

NIP. 132 205 465

Dekan Fakultas Ilmu Budaya  
Universitas Hasanauddin

  
Dr. H. Muhammad Darwis, M.S.

NIP. 131 411 591

Ketua Jurusan Sastra Inggris  
Fakultas Ilmu Budaya UNHAS

  
Drs. Husain Hasyim, M.Hum

NIP. 131 671 048

**UNIVERSITAS HASANUDDIN**  
**FAKULTAS ILMU BUDAYA**  
**JURUSAN SASTRA INGGRIS**

Sesuai dengan Surat Tugas Dekan Fakultas Sastra Universitas Hasanuddin Nomor: 2186/H4.11.1/ PP.27/ 2008, tanggal 11 Agustus 2008. Kami menyatakan menerima dan menyetujui skripsi ini untuk diteruskan ke Panitia Ujian Skripsi Jurusan Sastra Inggris Fakultas Ilmu Budaya Universitas Hasanuddin.

Makassar, 16 Februari 2009

Konsultan I,



Dr. Hj. Etty Bazergan, M.Ed

NIP. 131 287 221

Konsultan II,



Sitti Sahraeny, S.S., M.A.

NIP. 132 205 465

Menyetujui:

a.n Dekan

Ketua Jurusan Sastra Inggris,



Drs. Hysain Hasyim, M.Hum

NIP. 131 671 048

**UNIVERSITAS HASANUDDIN**  
**FAKULTAS ILMU BUDAYA**

Pada hari ini, 20 Februari 2009 , Panitia Ujian Skripsi menerima dengan baik skripsi yang berjudul :

**EVALUATION OF A READING TEST ADMINISTERED TO THE 1<sup>ST</sup> YEAR  
ENGLISH DEPARTMENT STUDENTS  
(A Case Study at Hasanuddin University)**

yang diajukan dalam rangka memenuhi salah satu syarat ujian akhir guna memperoleh gelar Sarjana pada Fakultas Ilmu Budaya, Jurusan Sastra Inggris Universitas Hasanuddin.

Makassar, 20 Februari 2009

Panitia Ujian Skripsi :

1. Dr.Hj.Etty Bazergan, M.Ed

Ketua

2. Sitti Sahraeny, S.S, M.A.

Sekretaris

3. Dra. Marleiny Radjuni, M.Ed

Penguji I

4. Drs. Simon Sitoto, M.A.

Penguji II

5. Dr.Hj.Etty Bazergan, M.Ed

Konsultan I

6. Sitti Sahraeny, S.S, M.A.

Konsultan II

The image shows handwritten signatures for each role listed on the left. The signatures are written in black ink on a white background. The roles and their corresponding signatures are: Ketua (Dr. Hj. Etty Bazergan), Sekretaris (Sitti Sahraeny), Penguji I (Dra. Marleiny Radjuni), Penguji II (Drs. Simon Sitoto), Konsultan I (Dr. Hj. Etty Bazergan), and Konsultan II (Sitti Sahraeny). Each signature is written over a dotted line that corresponds to the role name.

## ABSTRAK

**Demitriana Ponto Lolok.** 2009. *Evaluation of a Reading Test Administered to the 1<sup>st</sup> year English Department Students (A Case Study at Hasanuddin University)* (dibimbing oleh Dr. Hj. Etty Bazergan, M.Ed and Sitti Sahraeny, S.S, M.A)

Thesis ini mendeskripsikan tentang tes Reading yang diberikan kepada mahasiswa tahun pertama di Sastra Inggris, Universitas Hasanuddin, merupakan analisis dari isi tes Membaca apakah test tersebut cocok digunakan untuk mengukur kemampuan membaca mahasiswa, dan analisis dari hasil tes mahasiswa untuk mengukur pencapaian tujuan pembelajaran yang ada.

Penulis menggunakan metode deskriptif dalam menganalisa data melalui penelitian lapangan dan penelitian pustaka. Data yang diperoleh kemudian dianalisis dalam dua bagian yaitu analisis isi tes dan analisis nilai menggunakan teori-teori yang relevan.

Hasil analisis menunjukkan bahwa tes yang digunakan tidak sesuai untuk mengukur kemampuan membaca mahasiswa karena terdapat beberapa ketidakcocokan data. Materi Membaca yang digunakan hanya 6 topik dari 8 topik yang seharusnya dipelajari. Isi dari Tes Membaca kurang mewakili semua tujuan dari pembelajaran. Kriteria penilaian yang digunakan juga tidak sesuai dimana salah satu soal yang levelnya paling sulit hanya menempati level dengan tingkat kerumitan menengah.

## ABSTRACT

**Demitriana Ponto Lolok.** 2009. *Evaluation of a Reading Test Administered to the 1<sup>st</sup> year English Department Students (A Case Study at Hasanuddin University)* (supervised by Dr. Hj.Etty Bazergan, M.Ed and Sitti Sahraeny, S.S, M.A)

This thesis describes the Reading Test given to the 1st year English Department students of Hasanuddin University. The analysis of the Reading Test content will make us know the test appropriate or not in measuring the students' reading ability, while the analysis of students' score to examine how successful the students are in achieving the objectives of the Reading course.

In analyzing the data, the writer uses descriptive methodology through field research and library research. The data obtained are analyzed into two parts, test content analysis and score analysis by using relevant theories.

The result of analysis shows that the test used is not appropriate to measure students' Reading ability since there are some invalid data. The Reading material only covers 6 topics from 8 topics which should be learned. The Reading Test content also can not represent the objectives of Reading course. The scoring criteria used are also inappropriate where one of the most difficult questions only scored as in the middle level of difficulty.

## ACKNOWLEDGMENT

Sincere gratitude the writer offers to the Almighty Lord Jesus Christ for taking care of everything in this thesis process so it can be conducted well. This thesis is submitted in partial fulfillment of requirement for the Sarjana Sastra Degree at the Faculty Cultural of Sciences, Hasanuddin University. The writer faced many obstacles and frustrated things in her efforts finishing this thesis but thankfully it can be done through strenght and blessed from Him.

This thesis also goes to people who has interfered, helped, criticized, and made it into an acceptable one. They are:

- 1) Dr. Hj. Etty Bazergan, M.Ed and Sitti Sahraeny, S.S.,M.A as my first and second advisor who have already given many of their time dealing with this thesis;
- 2) Dr. H. Muhammad Darwis, M.S. as the Dean of Faculty of Cultural Sciences, Hasanuddin University; Drs. Husain Hasyim, M.Hum as the Head of English Department and Drs. Simon Sitoto, M.A. as the Secretary of English Department.
- 3) Lecturers in Faculty of Cultural Sciences especially in English Department who have taught her, share their knowledges, abilities in every subject she learnt;
- 4) Staffs and every employee in Faculty of Cultural Sciences, Language Centre and MKU of Hasanuddin University who have served perfectly;



- 5) The beloved parents, M.D. Ponto, S.Th and Heryana Ary, S.P., and family who not only give prayers, supports but also unconditionally love every single moment in writers' efforts achieving her ambition, especially for my beloved sista 'Atieq' (after what we've been through, we're able to show that friendship for last) ;
- 6) Friends, colleagues friends, CHAPTER '04, members of PMKOS-UH and best friends who could not be mentioned one by one (in case the writer will miss one of them could occur another problem) but an outstanding thanks the writer gives to K'echy as the leader of her bible study (you have no idea what deal have you given in my maturity process, sist!)

Hopefully all the helps the writer gained will be returned in times by God. Last but not least the writer hopes this thesis will give a valuable contribution on the teaching learning processes and the testing of Reading skills in English Department, Hasanuddin University, Makassar particularly to motivate and help the lecturers and the students to design an appropriate Reading achievement test.

The writer realizes that this thesis far from perfect, mistakes occurred from the writer's limited knowledge and experience. Therefore, the writer accepts any comment, criticize which can be used to improve this thesis' accurateness. It will give a help in knowledge development particularly in the evaluation area where the writer concern on.

Makassar, April 2009

The writer

## CONTENTS

	Pages
HALAMAN JUDUL.....	i
HALAMAN PERSEMBAHAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PENGESAHAN PEMBIMBING.....	iv
HALAMAN PENGESAHAN TIM PENGUJI.....	v
ABSTRAK.....	vi
ABSTRACT.....	vii
ACKNOWLEDGEMENT .....	viii
CONTENTS.....	x
TABLE LIST.....	xiii
FIGURE LIST.....	xiv

### CHAPTER I INTRODUCTION

1.1	Background.....	1
1.2	Identification of Problem.....	2
1.3	Scope of Problem.....	3
1.4	Statement of Problem.....	3
1.5	Objective of Writing.....	3
1.6	Significance of Study.....	4
1.7	Methodology.....	4

1.7.1 Method of Collecting Theories and Data.....	4
1.7.1.1 Library Research.....	4
1.7.1.2 Field Research.....	5
1.7.2 Population and Sample.....	5
1.7.2.1 Population.....	5
1.7.2.2. Sample.....	5
1.7.3 Method of Analyzing Data.....	6
<b>CHAPTER II LITERARY REVIEW.....</b>	<b>8</b>
2.1 Previous Study.....	8
2.2 Evaluation.....	9
2.2.1 Definition of evaluation.....	9
2.2.2 Language Test.....	9
2.2.3 Kinds of Test.....	10
2.2.4 Characteristic of Test.....	12
2.2.4.1 Test Validity.....	13
2.2.4.2 Validation Procedures.....	16
2.2.4.3 Test Specification.....	19
<b>CHAPTER III PRESENTATION AND ANALYSIS.....</b>	<b>22</b>
3.1 Presentation and Analysis of Data.....	22
3.1.1 Reading Test Presentation.....	28
3.1.2 Analysis .....	35

A. The content.....	35
B. The score.....	43
2.1 Difficulty Level.....	49
2.1.1 The difficult item.....	50
2.1.2 The middle level of difficulty.....	51
2.1.3 The low level of difficulty.....	52
2.2 Students' score.....	53
2.2.1 Range.....	54
2.2.2 Distribution.....	54
2.2.3 Median.....	55
2.2.4 Mean.....	55
2.2.5 Standard Deviation and Normal Curve..	55
<b>CHAPTER IV CONCLUSION AND SUGGESTION.....</b>	<b>57</b>
4.1 Conclusion.....	57
4.2 Suggestion.....	58
<b>BIBLIOGRAPHY.....</b>	<b>59</b>

## Table List

Table	Page
1. The Reading Course syllabus.....	23
2. The Possible Learning hours .....	26
3. The Content of Reading Final Achievement Test.....	37
4. Recommend Test specification.....	42
5. Table score	
5.1. Part A.....	44
5.2. Part B.....	45
5.3. Part C.....	46
5.4. Part D.....	47
5.5 Total Score from each part.....	48

## Figure List

Figure	Page
1. Distribution List.....	54
2. Normal Curve.....	55

## CHAPTER I

### INTRODUCTION

#### 1.1 Background

In language testing area, many kinds of tests are designed to measure certain purposes i.e.: to measure ones English proficiency, to examine how successful the lecturers and the students in achieving the objective of the course, to assign in what level the students should go to and many more. Of course this is an important thing to make sure that the test made by the teacher or lecturer is suitable with its objectives. Therefore, test evaluation is needed to evaluate the test and to keep it focused.

There are several criteria for a good test. They are successful in measuring what they intended to measure, i.e. valid, and could measure consistently, i.e. reliable.

In fact, evaluating the validity and reliability of a test is to find out how test administrator or test maker design the test. It is because among experts there is no agreement or certain standardization through test that must be fulfilled.

As we know that reading is one of the skills that should be mastered by students. This is because most of the books they should read are written in English. Many reliable tests such as the TOEFL and the IELTS used to examine one's proficiency are used as a requirement by many scholarship foundations, universities outside Indonesia or even multi-national companies. It would be misleading if the test that should be used to help them in their effort to master reading skills in the course can not measure what it is intended to measure.

The researcher also conducted a small research on English language Testing course in the seventh semester, 2007, she has identified the lack of vocabulary of English Department students particularly the 2<sup>nd</sup> year students when they have taken the Reading Test writer's made. The result of the test shown that the part B, vocabulary part, was the most difficult for them to answer. It could be seen from the percentage of students who are able to answer correctly just 23,6%, the lowest percentage among 3 types of question. It raises many further questions such as, how have they been taught and how valid and reliable the test are in measuring their Reading skills?

Therefore, I am interested in carrying out a research on language testing to evaluate a Reading Test administered in the 1<sup>st</sup> year English Department students of Hasanuddin University.

## **1.2. Identification of Problem**

To evaluate a particular language test, it is necessary to know the purpose of the test. First, is the test testing the ability or the knowledge of the students? Second, is the test designed to define the course objectives? Or is it used to stimulate student progress? Based on those purposes the writer finds some problems related to the topic as follows:

1. Whether the test is designed to test the reading ability.
2. Whether the test is designed to analyze several reading tasks.
3. Whether the test is in accordance with the teaching and learning material.



### **1.3 Scope of Problem**

Based on the identification of problem, I will focus on the content of the test (texts and tasks) to evaluate a Reading Test administered to the 1<sup>st</sup> year English Department students, Hasanuddin University. The result of my research will determine the appropriateness of the Reading test.

### **1.4 Statement of Problem**

Following what was mentioned above, the writer has several questions to ask in this reasearch. The questions are:

1. What Reading Test is given to the 1<sup>st</sup> year of English Department students?
2. How appropriate the test to measure the 1<sup>st</sup> year English Department students' reading ability?
3. How successful are the students in achieving the objectives of the Reading course?

### **1.5 Objective of Writing**

The objectives of this writing are:

1. to describe the Reading Test given to the 1<sup>st</sup> year English Department students in Hasanuddin University.
2. to analyze the content of the Reading Test whether or not it is an appropriate measure of the students' reading ability.

3. to examine how successful the students are in achieving the objectives of the Reading course.

## **1.6 Significance of Study**

It is expected that the significance of the study is to give useful information on the teaching learning processes and the testing of Reading skills in English Department, Hasanuddin University, Makassar. The writer hopes that the result of this research will motivate and help the lecturers and the students to design an appropriate Reading achievement test.

## **1.7 Methodology**

This chapter will describe the method of collecting data including population and sample, and the method of analyzing the data as can be seen in the following sub-method,

### **1.7.1. Method of Collecting Theories and Data**

#### **1.7.1.1 Library Research**

In library research, the relevant theories are collected by finding out any reading materials connected to the topic discussed. The reading materials are in the form of textbook, and thesis. The purpose of this method is to find out more information, in the form of theoretical issues that may support the writing. Besides, information on the topic is also discovered through browsing at the internet center.

University Library, Faculty of Cultural Sciences Library and Language Center Library are the place where the writer takes the data accordingly.

#### **1.7.1.2 Field Research**

In carrying out the field research, the writer collected data using the following techniques:

- **Note taking**, the writer collected the syllabus of the Reading Course, the materials used in the course, and the final achievement test.
- **Interview**, the writer interviewed the lecturer about the teaching material and the test material.

#### **1.7.2 Population and Sample**

##### **1.7.2.1 Population**

The population of this research is taken from the test material of a Reading Test administered to the first year English Department students, Faculty of Cultural Sciences, Hasanuddin University.

##### **1.7.2.2. Sample**

The Reading Tests administered to the six Reading classes of first year English Department students, Faculty of Cultural Sciences,

Hasanuddin University are different test. Therefore, the writer only takes one test to represent the population.

### 1.7.3 Method of Analyzing Data

In analyzing the data, the writer uses descriptive method. The data that have been collected and presented are analyzed descriptively. In this method, the writer uses some steps namely:

1. Classifying data into table form. The writer takes the outline of the course based on the syllabus and handbook of the course as an adequate description of the whole material.
2. Matching the outline of the course with the Reading material and the Reading test.
3. Analyzing the content validity of the Reading Test through comparing the outline of the course with the test material.

In analyzing the data, the writer also makes the percentage of the data to give a clear description about the representative samples of the material presented in the test.

The formula is:

$$P = \frac{\text{Numbers of the question related to the topic}}{\text{Numbers of questions in the test}} \times 100\%$$

Since the writer objective is also to examine how successful are the students in achieving the Reading course objectives, she also analyzes their Reading Test scores into several steps followed:

1. Score their correct answer into table form
2. Describe the score using some statistics formula as follows,
  - a. Determine the range by subtract the highest with the lowest score
  - b. Make a list of the score distribution
  - c. Determine the median (the middle score)
  - d. Determine the mean score (average score)

$$(M) \text{ mean} = \frac{\text{sum of the score}}{\text{number of papers}} = \frac{\Sigma S}{n}$$

Valette (1977:51)

- e. Find the standard deviation

$$\text{SD (standard deviation)} = \frac{\text{sum of high sixth} - \text{sum of low sixth}}{\text{Half the number of students}}$$

Diederick in Valette (1977:52)

All the data gathered will determine whether the lecturer and the students are successful in achieving the objectives of the Reading course.

## CHAPTER II

### LITERARY REVIEW

#### 2.1 Previous Study

The writer has found out a research about the content validity. Haslinda Yusuf (2005), she did research about *Content Validity of an English Test Administered at SMP Negeri 12 Makassar (A Study Case)*. She found that the English test administered at SMP Negeri 12 Makassar is an achievement test. Based on the analysis she has conducted, the test can be said valid because it has representative samples of each material that are taught and suitable with the purpose of the test and the objectives of the course. Furthermore, she admitted that the test administered need to be improved in terms of its instructions and the test specification.

In accordance with this research, the writer uses the previous research as one of references where she takes some of the theories which can not be found in the library research. However, as a further study this research "Evaluation of a Reading Test Administered to the 1<sup>st</sup> year English Department Students (A Case Study at Hasanuddin University)" has a significant difference with the previous study above such as, the method in collecting and analyzing the data and certainly the content of the analysis.

## **2.2 Evaluation**

### **2.2.1 Definition of evaluation**

David Nunan (2004:138) said that evaluation is a broad, general set of procedures involving the collection and interpretation of information for curricular decision-making. This information will generally include data on what learners can and cannot do in the language. Procedures for collecting this learner data are referred to as 'assessment'. Assessment is thus a subset of evaluation. Testing is one form of assessment. It includes the more formal collection of data on learner performance.

He added, evaluation can take place at any time, and any aspect of the curriculum can be evaluated.

### **2.2.2 Language Test**

One factor that supports language learning program is language test. To know whether the process of learning languages have been achieved its objectives, a language test should be done. To test the students of the language they have learnt, many kinds of test can be done. It depends on the tests' purposes. Different purposes will usually require different kinds of test.

Teaching and testing can not be separated because they relate to each other. Heaton in Haslinda (2005:6) says that both teaching and testing are so closely interrelated that is virtually impossible to work in either field without being constantly concerned with the other. Therefore, when testing was devised which was directly meet the need, it can be called give beneficial backwash. Backwash is the effect of testing on teaching and learning Hughes (1989:1).

The use of test is mostly needed by all language teachers. As said in Valette (1977:4)

“Through frequent testing, the teacher can determine which aspects of the program are presenting difficulties for individual students and for the class as a whole. By analyzing the mistakes made on a given test, or more precisely on given items of that test, the teacher can determine where to concentrate extra class drill and how best to assist each student.”

Futhermore, Valette also said that testing also enables the teachers to discover whether the class objectives are being met or not, able to evaluate the effectiveness of a new teaching method, of a different approach to a difficult pattern or of new material.

### **2.2.3 Kinds of Test**

Test is administered based on its purposes. In accordance with this research the writer takes a final achievement test as her data because this kind of information used by lecturer to measure students reading ability and to establish whether or not they are successful in achieving the objectives of the course. According to Hughes (1989: 9-14), test consists of four types as follows:

#### **a. Proficiency Tests**

Proficiency tests are designed to measure people's ability in a language regardless of any training they may have had in that language. It is based on the specification of what candidates have to be able to do in the language in order to be considered proficient. Proficient means having sufficient command of the language for a



particular purpose. One of the examples of so-called proficiency test is TOEFL test.

b. Achievement tests

Achievement tests are directly related to language courses, their purpose being to establish how successful individual students, group of students, or the courses themselves have been in achieving the objectives. Achievement test consist of two kinds, final achievement test and progress achievement test. *Final achievement tests* are those administered at the end of the course of study. They may be written and administered by ministries of education, official examining boards or by members of teaching institutions. Whereas *progress achievement tests*, as their name suggests, are intended to measure the students' progress.

c. Diagnostic tests

Diagnostic tests are used to identify students' strength and weaknesses. They are focus on what will the further teaching be. We can be fairly confident of our ability to create tests that will tell us that the student is particularly weak in speaking than in reading a language.

d. Placement test

Placement test, as their name suggest, are intended to provide information which will help to place students at the appropriate stage

(or in the part) of the teaching programme based on their abilities.

Typically they are used to assign students to classes at different level.

For the term of how the tests are scored, Hughes (1989:19) points out that there are two kinds of test; objective testing and subjective testing. A test is classified into an objective testing if there is no judgement required in the method of scoring where as if the judgement is called for than it is called subjective test. An example for objective test is in multiple choice questions e.g. TOEFL® test where as an essay test is called subjective test e.g. in IELTS test.

#### **2.2.4 Characteristic of Test**

All tests should fulfill the characteristic of a good test. Since the test is a measurement device, it has to be qualified in order to achieve the aim of the test and the objectives of the course. The characteristic or qualities that must be fulfilled to be a good test are reliability, validity and practicality. Reliability concerns about the matter of how consistently a test does what it is supposed to do, while validity relates to how well a test measure what it is intended to measure and practicality of test that concern about the cost in terms of materials, time, and effort that it requires.

Based on the topic discussed in this writing that is content validity of a test, the writer describes about test validity and its kinds detailed as follows.

#### 2.2.4.1 Test Validity

Valette (1977:46) said that the objective tests are criticized for their failure to measure this or that aspect of subject. Thus, the teacher must assess the content validity of a test before administering it; they also have to evaluate the validity of the various items in relation to their course objectives.

Henning in Haslinda (2005:11) defines validity as follows:

“Validity in general refers to the appropriateness of a given test or any of its component parts as a measure of what it is supposed to measure. A test is said to be valid to the extent that it measures what it is supposed to measure. It follows that the term valid when used to describe a test should usually accompanied by the preposition for. Any test then may be valid for some purposes, but not for others”

Alderson, Wall and Clapham in Haslinda (2005:12), divide validity into:

1. Internal validity that relates to studies of perceived content of the test and its perceived effect, this consist of:

- a. Face Validity

A test is said to have face validity if it looks as if it measures what is supposed to measure (Hughes: 1989). If a test item looks right to other testers, teachers, moderators, and testees, it can be described as having at least face validity.

#### b. Content Validity

Content Validity, as stated by Kerlinger (cited in Alderson, Wall and Chapham (1995) ) is the representativeness or sampling adequacy of the content-the substance, the matter, the topic of a measuring instrument. It means that the sample of questions in a test should represent the important material content and skills of the course. The test should be so constructed as to contain a representative sample of the course. The relationship between the test item and the course objectives always being apparent

#### b. Responsive Validity

Responsive validity is intended to describe the extent to which examinees respond in the manner expected by the test developers which includes the way of the tester gives instructions and the test format. The instruction must be clear and the test format must be familiar to the students so their response of the test will reflect their true abilities.

2. External Validity relates to studies comparing students' test scores with measures of their ability viewed from outside the test, this consist of :

a. Concurrent Validity

Concurrent validity involves the comparison of the test scores with some other measure for the same candidates taken at roughly the same time as the test.

b. Predictive Validity

Predictive validity is most familiar with the proficiency test that is a test given to predict how well somebody will perform in the future

c. Construct Validity

Construct validity is the most difficult concept of validity. Hughes (1989) states that a test or a testing technique is said to have construct validity if it can be demonstrated that it measures just the ability which is suppose to measure. A test is said to have constructed validity if it is capable of measuring certain specific characteristic in accordance with the theory of language behavioral and learning.

From the explanation above, the writer's research concerning content-validity in a Reading test is a part of internal validity, clearly stated that the test content should has the representative of each objective of the Reading course.

#### 2.2.4.2 Validation Procedures

There are some procedures to measure a test validity but there is no an agreement among the experts about this as it is stated by Carol & Hall (1895) in Haslinda (2005) that there is as yet a lack of agreement among expert about validation procedures. Carol & Hall (1985:129) describe about the validation procedures at it is shown below:

##### *Procedures of Looking for Test Validity*

Step	Procedures
Validation	<ol style="list-style-type: none"><li>1. Compare the test specification with the test content</li><li>2. Correlate test judgments with judgments of teachers + employers.</li><li>3. Intercorrelate items + subtest</li><li>4. Correlate test performance with performance on comparable test.</li><li>5. Relate features of test performance with characteristic of testees (aptitude).</li></ol>

The validation procedures above describe the process of validating a test generally. It does not explain clearly about the steps to evaluate each kind or validity, especially the content validity. But Alderson, Wall and Clapham in Haslinda (2005:15) describe the procedures for evaluating a test validity based on its kind as follow:

*Procedures for Evaluation*

Types of Validity

Procedures for Evaluation

*Internal Validity*

Face Validity

Questionnaires to, interviews with candidates, administrator, and other users.

Content Validity

1) Compare test content with specification or syllabus

2) Questionnaires to, interviews with 'expert' such as, teachers, subject specialist, applied linguists.

3) Expert judges rate test items and texts according to precise list of criteria.

Response Validity

Students introspect on their test-taking procedures, either concurrently or retrospectively.

*External Validity*

Concurrent Validity

1) Correlate students' test scores with their scores on other test

2) Correlate students' test scores with teachers' rankings.

3) Correlate students' test scores with other measures of abilities such as students' or teachers' ratings.

#### Predictive Validity

1) Correlate students' test scores with their scores on tests taken some time later

2) Correlate students' test scores with other measures of their ability taken some time later, such as subject teachers' assessments, language teachers' assessment.

3) Correlate students' test scores with success in final exam.

4) Correlate students' test scores with success of later placement.

#### Construct Validity

1) Correlate each subtest with other subtests.

2. Correlate each subtest with total tests.

3) Correlate each subtest with total minus self

4) Compare students' test scores with students' biodata and psychological characteristics.

5) Multitrait—multimethod studies.

6) Factor analysis.

In analyzing the data in this research, the writer uses the procedures of comparing the content of Reading test directly with the objectives of the course and



the material of the course as the replacement of the test specification as it is mentioned in the first step of Carol & Hall procedures and also mentioned in the Anderson, Wall and Clapham's procedures.

#### **2.2.4.3 Test Specification**

The specification of the test should be made first before developing the real test. A test specification is a detailed document about what the test maker should involve in the test, as stated by Alderson, Wall and Clapham in Haslinda (2005:18): "A test's specification provides the official statement about what the test tests and how it tests it. The specifications are the blueprint to be followed by the test and item writers, and they are also essential in the establishment of the test's construct validity"

A test specification is not only needed by the teacher or the test maker but also by those responsible for or interested in establishing the test's validity. The people here includes the outsiders or other independent individuals also the textbook writer. Alderson, Wall and Clapham in Haslinda (2005:15) states that for validation purposes, the test specification need to make the theoretical framework which underlies the test explicitly, and to spell out the relationship among its construction, as well as the relationship between the theory and the purpose for which the test is designed.

Test specification vary based on its necessary or its used. According to Hughes (1989) a test specification should provide information about content, format and timing, criteria levels of performance and scoring procedures. Alderson, Wall and Clapham in Haslinda (2005:19-20) give a list about what should a test specification provide, as follows,

1. The test's purposes
2. Description of the test taker
3. Test level
4. Construct (theoretical framework for test)
5. Description of suitable language course or textbook
6. Number of sections/paper
7. Time for each sections/paper
8. Weighting for each sections/paper
9. Target language situation
10. Text types
11. Text length
12. Language skills to be tested
13. Language elements to be tested
14. Test task
15. Test methods
16. Rubrics
17. Criteria for marking
18. Descriptions of typical performance at each level

19. Description of what candidates at each level can do in the real world
20. Sample papers
21. Sample of students' performance tasks.

All the points mentioned in the list above are things should be included in a test specification but since the test specification vary according to its necessary or its uses, then it does not have to cover all the points in the list.

## CHAPTER III

### PRESENTATION AND ANALYSIS

In this chapter the writer analyzes the content of a Reading Test administered at 1<sup>st</sup> year English Department students of Hasanuddin University. The Reading Test obtained is the representative sample of population, six Reading classes held in beginner semester 2008/2009. The test type is a final achievement test in order to examine the direct representative from each objective of the course.

#### 3.1 Presentation and Analysis of Data

After collecting all the data needed by using the method of collecting data in chapter 1, the writer first classified the data into table form which presented the reading material and the reading syllabus. The first data presented is the reading syllabus followed by its analysis. I have divide the analysis in two parts, first the analysis of the content and the second for the score analysis. The writer analyses the data by using some theories as has been described in chapter two. Unfortunately, when gathering all the data, the writer can not provide the test specification as the lecturer did not make it for limited timing excuses. Therefore, to provide more accurate information in analyzing the content of the test, the writer directly compare it with the objectives of the course. The writer also provide a recommend test specification as a tool to justify the content of the reading test whether or not it is valid to measure students' reading ability.

Table 1. The Reading Syllabus

No	Learning Goals	Topics/ Sub-topics of Discussion	Allocated time (minutes)	References (sub-part, page(s))
1.	Students are able to: a. complete the summary b. match 6-8 new vocabularies in every reading text	1. Instruction, How to do things a. Danger from fire b. Wear and care of contact c. How to shine at lenses a job interview d. American telephones	4 × 100	9b, page. 14-27
2.	Students are able to: a. summarize the main idea of a text b. define at least 5 new vocabularies in every text.	2. Description, what things are like a. A city is dying b. Loneliness c. Holiday in Scotland d. Zen and the art of motorcycle maintenance	4 × 100	9b, page.28-41
3.	Students are able to: a. summarize the text into a systematic draft b. explain the main preposition based on the draft they made	3. Process, How thing happen a. Drinking while driving b. Drink from fruit and grains c. The towns that kids built d. Inside Story	4 × 100	9b, page.42-55
4.	Students are able to: a. conclude a reading text orally b. write at least 4 sentences about the cases in every reading text	4. Narrative: what happened a. The Diamond b. Dear Editors c. Great Operatic Disaster d. The Poisoning of Michigan	4 × 100	9b, page.56-71

5.	Students are able to find at least 5 main idea from each reading text.	5. Persuasion: why you should do it a. The challenge b. Go a head—read this c. Save the children d. Two Letters	4 × 100	9b, page.72-81
6.	Students are able to find at least 4 differences of things in each reading text	6. Categories : how things are classified a. Elephants b. Airports c. The trials of an O in a world of Xs d. Wonder wander	4 × 100	9b, page.82-94
7.	Students are able to make a summary in an efective paragraph in each reading text.	7. Extracting main ideas a. Doodles b. Money is the only home	2 × 100	9a, page 6-13
8.	Students are able to rewrite the reading text correctly.	8. Understanding text organization a. Shielding Brook b. Down and Out in Paris and London	2 × 100	9a, hlm 14-25

From the table above, we can see that during the semester, students have to learn 8 main topics which are divided into several sub-topics. These topics are taken from 2 different books: *Authentic Reading* (Walter: 1984) and *Effective Reading* (Greenal: 1988). Six topics are from *Authentic Reading* and the two last topics are from *Effective Reading*. Each topic has its objective(s) and has equal time allocated, 100 minutes per sub-topics. There are 28 sub-topics so the total time in achieving all the

objectives are 28x100 minutes. The class itself has 32 meetings so they will meet twice a week in ±4 months. Two meetings will be used for progress achievement test and final achievement test. The first meeting is usually used for class-contract. It means they still have 29 meetings to do teaching and learning reading processes in the class. Particularly in the class I observed, I found that they only used one book, *Authentic Reading* as their material for the whole semester. The lecturer simply divided the class material into two parts, before and after the administered of progress achievement test. Three main topics, instruction, description, and process (Unit 1-12) were taught for the first part and after that three last subjects; narratives, persuasion and categories (Unit 13-24).

At this point the reading material used is closely related to the objectives of the reading syllabus for the sixth first topics in the syllabus are presented in the content of reading material. However, two points of objectives from the syllabus; point 7 and point 8 is clearly will not been achieved since the *Effective Reading* book is not used by the lecturer.

Therefore, I have made possible allocated time in learning each sub-topic. It is presented in Table 2 below,

Table 2.

No	Topics and Sub-topics	Allocated time (minutes)	References
1.	Class-contract 1. Instruction, How to do things a. Danger from fire b. Wear and care of contact c. How to shine at lenses a job interview d. American telephones	100  100 100 100 100	9b, page. 14-27
2.	2. Description, what things are like a. A city is dying b. Loneliness c. Holiday in Scotland d. Zen and the art of motorcycle maintenance	100 100 100 100	9b, page.28-41
3.	3. Process, How thing happen e. Drinking while driving f. Drink from fruit and grains g. The towns that kids built h. Inside Story	100 100 100 100	9b, page.42-55
	Progress achievement test	100	Unit 1- Unit 12
4.	4. Narrative: what happened e. The Diamond f. Dear Editors g. Great Operatic Disaster h. The Poisoning of Michigan	100 100 100 100	9b, page.56-71
5.	5. Persuasion: why you should do it a. The challenge b. Go a head—read this c. Save the children d. Two Letters	100 100 100 100	9b, page.72-81



6.	6. Categories : how things are classified e. Elephants f. Airports g. The trials of an O in a world of Xs h. Wonder wander	100 100 100 100	9b, page.82-94
	Review Final achievement test	100 100	All Units All Units
Total learning hours		28 × 100 minutes	

From the table 2 we can see that after dividing each class activity with its time there are still 3 meetings (from 32 meetings-28 meetings) which can be used by the lecturer to give additional explanations or quizzes. One meeting is also added for review before final achievement test. The purpose of the progress achievement test is to find out their Reading progress and also about the problems found during the first part (Unit 1-12). Certainly it will help the lecturers have a control of achievement in their class and prepare for the solutions. Unfortunately, when I confirmed to the lecturer the progress achievement test were administered just  $\pm 4$  weeks before the final achievement test. Then I assume whether they have finished in covering all the Units left (Unit 13-24) or not. The purpose of the review before final achievement test in one side helps the students have a brief look of overall material before doing the final test. It can be done by administering the pilot test. For the test makers it will help them remove the weak items before making a judgment over students' reading ability. But in contrary, there is no pilot test administered before the real test.

In developing a test, the test writers should make a test specification based on the syllabus as their guideline before administered the real test. It will help the lecturer to co-ordinate the test content with its objectives and remove the weak items before make a judgment over the students' answers. However, when I confirm the Reading lecturer which her class I observed, the lecturer did not make the test specification. The lecturer excuses for the limited time to develop it. In order to make a fair judgment over the reading test, the writer then compare the content of the Reading test directly with its objectives in the syllabus.

The Reading Test material will be presented and analysed as follows:

#### **A. Reading Test Presentation**

FINAL TEST

Name :

MEMBACA BAHASA INGGRIS

Reg. No. :

Part A      **Dear editors**

Dear Editors,

Your readers may be interested in the following account of the behaviour modification of a small girl.

Kathy started at my nursery school at the age of two years nine months. She was small for her age but confident, competent and determined. She settled into the group easily, would be first on the slide and highest up the climbing frame. She could put on her coat without help and not only fasten her own buttons but fasten other children's too.

She was a lovely child but unfortunately a scratcher. If anyone upset her or stood in her way her right hand would flash out faster than Mohammed Ali's and score down the face of her playmates from forehead to chin. Children twice her age would fly in terror from her and cower screaming in corners.

This must have been very rewarding for Kathy but obviously it had to be stopped—and stopped quickly. All the usual ways failed and then I remembered an account by G.C.E. Atkinson of Highfield School, Haledown, Liverpool of how bullying in the playground had been stopped.

No punishment had been given, but the bullies had been ignored and the victims rewarded. So I decided that in future Kathy would be ignored and her victim given a sweet.

With a pocketful of Smarties I followed Kathy around. She was so quick that it was impossible to prevent her scratching but I was determined to stay within arms length all afternoon.

All was peaceful but then I saw Kathy's hand descend and heard the scream. Quickly and gently I gathered up the little hurt one and cradling her in my arms said 'Nice, nice sweetie' and I popped it into her mouth. Kathy open her mouth expectantly and then when she got nothing looked puzzled.

Minutes later another scream, this time from John and while cuddling him I said 'Look Kathy, a nice Smartie for John' and put the sweet into John's open mouth.

A smile of understanding flashed across Kathy's face and holding a finger tenderly came to me and wailed 'Hurt my finger'

'Never mind' said I coldly 'It will soon be better'. She stamped her small foot and shouted 'Give me a Smartie, I have hurt my finger!'

'No' I replied, 'You'll get a sweet if someone hurts you'

Deliberately she turned and scratched a child, waited quietly while I mothered and rewarded him, then walked away. She has never sratched a child since.

Parents who find older children bullying younger brothers and sisters might do well to replace shouting and punishment by rewarding and giving more attention to the injured one. It's certainly less exhausting.

Margaret Seekree,  
Belcombe Croft  
Nursery School,

Bradford-on-Avon,  
Wiltshire

1. Summarize the letter in one sentence.
  
2. Guessing unknown words
  - a. Kathy was a scratcher. So in the phrase 'score down the face her playmates,' (lines 25-26), score probably means.....
  - b. (line 70) Kathy was expecting a sweet and did not get one. A person who is *puzzled* is probably a person who does not.....
  - c. (line 82) Kathy was pretending her finger was hurt. So wailed probably means 'said in a .....way'.
  - d. Forehead (line 27) : the forehead is at the .....of a face.
  - e. Rewarding (line 32) : seeing the result of her scratching gave Kathy.....
  - f. Smarties (line 51, 75, 87) : This is a kind of .....
  - g. Cuddling (line 73) : You cuddle someone with your ..... . Some of the people you might cuddle are .....and .....
  - h. Exhausting (line 106) : Is it pleasant for something to be exhausting? ..... You cannot guess exactly from the text, but exhausted might mean .....

#### Part D

There was a man called Ephraim who lived in Johannesburg. His father was to do with diamonds, as had been his father. The family were immigrants. This is still true of all people from Johannesburg, a city a century old. Ephraim was a middle son, not brilliant or stupid, not good or bad. He was nothing in particular. His brother became diamond merchants, but Ephraim was not cut out for anything immediately obvious, and so at last he was apprenticed to an uncle to learn the trade of diamond-cutting.

To cut a diamond perfectly is an act like a samurai's sword-thrust. Or a master archer's centred arrow. When an important diamond is shaped a man spend a week, or even weeks, studying it, accumulating powers of attention, memory, intuition, till he has reached that moment when he finally knows that a tap, no more, at just that point of tension in the stone will split it exactly so.

While Ephraim learned to do this, he lived at home in a Johannesburg suburb; and his brothers and sisters married and had families. He was the son who took his

time about getting married, and about whom the family first joked, saying that he was choosy; and then they remained silent when others talked of him with that edge on their voices, irritated, a little malicious, even frightened, which is caused by those men and women who refuse to fulfill the ordinary purposes of nature. The kind ones said he was a good son, working nicely under his uncle Ben, and living respectably at home, and on Sunday night playing poker with bachelor friends. He was twenty-five, then thirty, thirty-five, forty. His parents became old and died, and he lived alone in the family house. People stopped noticing him. Noticing was expected of him.

Then a senior person became ill, and Ephraim was asked to fly in his stead to Alexandria for a special job. A certain rich merchant of Alexandria had purchased an uncut diamond as a present for his daughter, who was to be married shortly. He wished only the best for the diamond. Ephraim, revealed by this happening as one of the world's master diamond cutters, flew to Egypt, spent some days in communion with the stone in a quiet room in the merchant's house, and then caused it to fall apart into three lovely pieces. There were for a ring and earrings.

Now he should have flown home again; but the merchant asked him to dinner. An odd chance that—unusual. Not many....

1. Match each phrase in column A with the best description in column B

Column A

1. Ephraim as a child
2. A professional diamond cutter
3. How Ephraim's not getting married made some feel
4. Ephraim private live in Johannesburg
5. Ephraim's ability to cut diamonds
6. Ephraim's approach to cutting the rich merchant's diamond
7. The rich merchant, observing Ephraim's way of working

Column B

- a) difficult
- b) impressed
- c) uninteresting
- d) unexceptional
- e) mystical
- f) highly skilled
- g) exceptional
- h) bored



- i) violent
  - j) uncomfortable
2. Make connections between words and their meanings. Give the meaning of each word in italics. (The meaning can be in the same sentence or elsewhere in the text.)
1. His father was to do with diamonds, as had been *his father*. (line 2)
  2. While Ephraim learned to do *this*, he lived at home in a Johannesburg suburb;.....(line 17)
  3. The kind *ones* said he was a good son,.....(line 24)
  4. Then a senior person became ill, and Ephraim was asked to fly in *his* stead to Alexandria for a special job. (line 32)
  5. Ephraim, revealed by this happening as one of the world's master diamond-cutters, flew to Egypt, spent some days in communion with the stone in a quite room in the merchant's house, and then caused *it* to fall apart into three lovely pieces. (line 38)
3. Does the text say :
1. How many brothers Ephraim had?
  2. What Ephraim's brothers did?
  3. Who Ephraim worked for?
  4. How long Ephraim was an apprentice before cutting his first diamond?
  5. Whether Ephraim had any nephews or nieces?
  6. Where Ephraim was living when he was sent to Egypt?
  7. Whether Ephraim's friends thought he would become famous?
  8. How many children the rich merchant had?
  9. Where Ephraim was when he cut the diamond?
  10. Whether Ephraim met the girl before cutting the diamond?

#### Part B

Tosca: City Center, New York, 1960

This catastrophe is due, not to misunderstanding and incomplete, but entirely to ill-will between the stage staff and the soprano. With diabolical cunning they permitted her, after several stormy rehearsals, to complete her first performance without mishap until the very last moment, when Tosca throws herself off the battlements of the Castel Sant'Angelo. What normally happens is that on her cry 'Scarpia, davanti a Dio; she hurls herself off and lands on a mattress four feet below. This large young American singer landed not on a mattress, but—perish the thought—on a *trampoline*. It is said that she came up fifteen times before the curtain fell—sometimes upside down, then the right way up—now laughing in delirious glee, now screaming with rage..... Worse still, it seems that the unhappy lady was unable to reappear in New York because the Center's faithful audience, remembering the trampoline, would have burst into laughter. She had to remove herself to San Francisco, where of course no such grotesque incident could possibly occur...

1. Here is a summary of the first story. Fill the blanks with appropriate words.

In New York in 1960, the **a**.....in Tosca was not very friendly with the **b**..... workers. So instead of giving her a **c**..... to fall on when she **d**..... from the castle in the last act, they gave her a **e**..... Instead of disappearing she **f**..... up behind the battlements again and again. She had to **g**..... New York as a result of this incident.

2. Match each word in italics in column A with the meaning in column B that comes closest to it. Column B has some extra meanings

Column A

- a. This catastrophe is due.....entirely to ill-will between the stage staff and the soprano. (line 2)
- b. ....after several stormy rehearsals,..... (line 5)
- c. ....to complete her first performance without mishap.....(line 7)
- d. .... she hurls herself off and lands on a mattress.....(line 10)
- e. ....now laughing in delirious glee,.....(line 17)

Column B

- a) pretending to be shot
- b) joy
- c) in bad weather
- d) hand movements
- e) marches
- f) full of disagreement
- g) planned
- h) bad planning
- i) throws
- j) born
- k) suddenly and absolutely
- l) something going wrong
- m) bad feelings

Part C

These Cultural Sciences were received by a man who owed money from a doctor. The doctor had treated the man's wife.

G.B. BROWN M.D.  
4321-----, -----, Texas 77002  
November 5, 1981  
Mr. George -----  
3124 -----  
-----, Texas 77005  
Re: \$325.00  
Marion -----  
Account No. ----  
Dear Mr -----  
Two weeks ago we wrote to you asking you to contact us about the above account; as yet we have not heard from you.  
At the first time Mrs ----- needed medical attention she was treated immediately. We feel that you should at least extend the courtesy of replying to our several requests for payment by asking full or partial payments, beginning now. Even small payment will be accepted if they are regular ones.  
We expect to hear from you within the next two weeks, please.  
Very truly yours,  
*Helen Marques*  
Helen Marques  
Office Manager

G.B. BROWN M.D.  
4321-----, -----, Texas 77002  
November 17, 1981  
Mr. George -----  
3124 -----  
-----, Texas 77005  
Re: \$325.00  
Marion -----  
Account No. ----  
Dear Mr -----  
We cannot understand why you continue to ignore the requests for payment we have been mailing to you. We feel that you had ample time to send payment, call, or write regarding this account.  
If we have not heard from you within two weeks, we shall have no alternative than to turn this account over to a collection agency for payment. Surely you do not want to have this on your credit record.  
We expect prompt attention to this matter.  
Very truly yours,  
*Helen Marques*  
Helen Marques  
Office Manager

1. Mark T if the sentence is true according to the Cultural Sciences, and F if it is false according to the Cultural Sciences
  - a. The 5 November letter was the first one from Dr. Brown's office about the money the man owed.
  - b. the 5 November letter is more polite than the 17 November letter.
  - c. Helen Marques thinks that \$325.00 is small payment.
  - d. The 17 November letter contains a threat.
  - e. In two weeks Ms. Marques will write the same man another letter
2. Find words or phrases in two Cultural Sciences which have roughly the meanings given below.
  - a. Record of a patient's bills
  - b. Be polite, and answer
  - c. Pay no attention to.
  - d. Enough.
  - e. Other choice.
  - f. Company that specializes in getting money from people who owe it.



## **B. Analysis**

### **Part 1. The Content**

The Reading Test were administered to the 29 first year English Department students at the end of the course to measure their whole achievement of the Reading course therefore it is a final achievement test. It consists of 4 parts, each part is divided into several types of question. The total questions presented in each part are 54 questions. All the part of questions are taken directly from the Reading Material, *Authentic Reading*. Part A was taken from unit 14 "Dear Editors" page 60 consisted of 2 types of questions; make a summary from the passage and 8 guessing unknown words. Unit 15 "Great Operatic Disaster" page 68 consisted of 2 parts; 7 questions for fill the missing word and 5 questions for match the meaning in column A with column B are questioned in part B. "Two Cultural Sciences" from Unit 20 page 80 consisted of 2 parts; 5 questions for extracting the main idea and 6 guessing unknown words elicit in Part C. Lastly, Part D was consisted of 3 parts; 7 questions for match the phrase, 5 questions for match the meaning and 10 yes/no questions taken from Unit 13 "The Diamond" page 56.

It can be seen from the test that there is a wrong arrangement of the test which Part D directly followed questions in Part A but the lecturer already acknowledge the students about it when administered the test. However, I think it will be disadvantage for the students who come late and will spend most time for the lecturer has to tell them one by one. The unclear questions can be asked too to the lecturer so even it is not the matter which I concern on this research but it is hopefully not take for granted for it still influence the content validity.

Since this test only measures students achievement of the second part of objectives (Narrative, Persuasion and Categories), the test content will only compare with the objectives from the three topics. From the interviewed with the lecturer, she said that the three other topics' objectives (Instruction, Descriptions and Processes) already measured by progress achievement test therefore it should not be included in final achievement test. Futhermore, she admitted that the reading test she administered is not enough to measure the children's reading ability since it only takes 30% of her assessment. The assesment for her Reading class is devided into 4 parts of percentage; 20% for classroom assignment. Students' attendance are examined by counting how many students submit their homework assignments. 20% for homework assignment, 30% for progress achievement test and final achievement test which takes 30%. Therefore, the purpose of this final reading achievement test is only to fulfill the reading assessment.

In matching the test content with its objectives, the writer only compares between the content of Unit 13-Unit 24 as we can see from the following table 3 ,

Table 3

No	Topics, Sub-topics and type of questions	Number of questions	Percent -age	Reading Skills used
1.	Narrative: what happened Unit 13. The Diamond a. Match the meaning in two different columns b. Make connection between the words and their meanings c. Yes/No questions  Unit 14. Dear Editors a. summarize the letter in one sentence b. guessing unknown words  Unit 15. Great Operatic Disaster a. Fill the blanks with the missing word b. Match the meaning in two different column  Unit 16. The Poisoning of Michigan	7  5  10  1  8  7  5  -	40,7%          16,7%          22,2%          -          -          -          -          -          -          -	- guessing difficult words - guessing difficult words comprehension          - make a summarize - guessing difficult words          - completing the missing words - guessing difficult words          -          -          -          -          -          -          -          -          -
	TOTAL	54 questions	100%	5 kinds of reading skills

The test contents are taken directly from the handbook, in the one side it constitutes the representative sample for each sub-topic used, but in the other side students tend to use only their memory when learning the topics to answer the questions. Concerning with the equal time allocated for learning each sub-topics, the test should have equal number of percentage. However, the test content only covered 4 Units where mostly concern in 3 Units, sub-topics of narratives with 79% from . The second topic are presented by Units 20 "*Two letters*" with 20,4% . The last four Units has no representative sample in the test at all. But again the test writer said that there are some quizzes which can be taken into account as the alternative ways in measuring over the topic which has no representative in the test.

Based on the skills used, each part has different type of questions which acquired different kinds of Reading skills. As we can see from the table above, part A of the test content taken from Unit 14 has a number for making summary and 8 numbers for guessing difficult words. Part B which taken from Unit 15 acquired skills in completing the missing words and guessing difficult words whereas skills in extracting main idea and guessing unknown words are used in part C. For the Part D, the questions should be answered by using guessing difficult word and comprehension skills. In short, the skills used in doing the Reading test are making summary, guessing difficult words, completing the missing words and comprehension skills.

Comparing between the objectives of the course in Unit 13-24 about the skills which should be possessed by the students with the skills acquired in doing the reading test there are some compatibilities. It is only because the test are taken

directly from the reading material. However, when I compare with the reading skills covering in all Units, they are spread into some different units out of Unit 13-24. They are suitable with the objectives from *Instruction* and *Description* topic. Part A, question for making a summary and guessing unknown words suitable with the units' objectives in *description* where both of its objectives is to summarize the reading and to define at least 5 new vocabularies. Part B, to complete the missing words and to match the words in column A with the meaning in column B is suitable with in *Instruction* objectives. It can be said that the objectives of the test content nearly covered the objectives overall where in *Instruction* and *Description* out of *Narratives* and *Persuasion* but in the other point did not cover the objectives in two topics; *Processes* and *Categories*. Based on the those analysis, I think the test makers should put more attention to their test development so all the objectives can be covered in the test.

Even though in analysing the content of the Reading test the writer did not use test specification as its measurement but the result can be trusted as for it is analysed directly based on the objectives of the course. But it will be much better if the test writer develops a test specification based on the Reading syllabus and Reading Material used before administering the test. Based on Alderson, Wall and Clapham in Haslinda (2005: 19-20) which give a list about what a test specification should provide, in my opinion the test specification should be arranged as follows:

### Test Specification

Institution	: English Department, Faculty of Cultural Sciences, Hasanuddin University
Subject	: Reading 1
Grade	: 1 <sup>st</sup> year English Department Students/ 1 <sup>st</sup> semester
Test Purpose	: To measure the students' achievement of the course during the semester
Time of administering	: Tuesday, 9th December 2008
Time Limit	: 90 minutes
Description of textbook	: Walter, Catherine. 1984. Authentic Reading. Cambridge: Cambridge University Press. Consist of six parts and 24 units

**Part 1. Instruction: How to do things, consist of 4 units;**

Unit 1. Danger from fire,

Unit 2. Wear and care of soft contact lenses,

Unit 3. How to shine at a job interview

Unit 4. American telephones

**Part 2. Descriptions: What things are like, consist of 4 units**

Unit 5. A city is dying

Unit 6. Loneliness

Unit 7. Holidays in Scotland

Unit 8. Zen and the art of motorcycle maintenance

**Part 3. Processes: How things happen, consist of 4 units**

Unit 9. Drinking while driving- How does the test work?

Unit 10. Drinks from fruits and grains

Unit 11. The town that kids built

Unit 12. Inside story

**Part 4. Narrative : What happened, consist of 4 units**

Unit 13. The diamond

Unit 14. Dear editors

Unit 15. Great operatic disaster

Unit 16. The poisoning of Michigan

**Part 5. Persuasion : Why you should do it, consist of 4 units**

Unit 17. The challenge

Unit 18. Go ahead- read this

Unit 19. Save the children

Unit 20. Two Cultural Sciences

**Part 6. Categories : How things are classified, consist of 4 units**

Unit 21. Elephants

Unit 22. Airports

Unit 23. The trials of an O in a world of Xs

Unit 24. Wonder wander

Type of the test : Essay test

Number of questions : 9 parts with 1-10 questions each part



Table 4

## Test Description

No	Activities	Topic Areas	Suitable test task/ Formats	Number of questions
1.	Match 6-8 new vocabulary from the passage	Instruction: How to do things	two columns A and B which should be matched	6-8
2.	Make a summary from the reading	Descriptions: What things are like?	Make a summary 2 or 3 sentences	1
3.	Guessing difficult vocabulary	Descriptions: What things are like?	the difficult words which are written in italic should be interpreted by their own words	5-7
4.	Make a scheme based on the reading completed with its explanation	Processes : How Things Happen	Draw or write a scheme from a passage followed by its explanation	1
5.	Write at least 4 cases each story	Narratives : What happened	2 short stories	8
6.	Write 5 main idea from each passage	Persuasion: Why you should do it?	5 short passages	5
7.	Find 6 differences from each passage	Categories: How things are classified	2 passages	6



The various number of questions are made based on the syllabus objectives and the Reading material.

## **Part 2. The score**

To provide more significant result of the final achievement test's content whether or not it has measure what it intends to measure, the writer has classify the students' results into table form. It will give us a brief look into each student's reading achievement. The table score presents as follows,

Table 5.1

NUMBER OF DATA	PART A									TOTAL
	SUMMARY	GUESSING UNKNOWN WORDS								
		a	b	c	d	e	f	g	h	
1	4	2	2	2	2	2	2	2	2	20
2	4	2	2	2	2	2	2	1	2	19
3	8	2	2	2	1	2	2	2	2	23
4	8	2	2	2	2	2	2	2	2	24
5	8	-	2	2	2	-	-	-	1	15
6	-	2	2	2	2	2	2	2	2	16
7	8	2	2	2	2	2	2	2	2	24
8	7	2	2	2	2	2	2	2	2	23
9	-	2	2	2	2	2	2	2	2	16
10	-	2	2	-	-	-	-	1	-	5
11	3	1	2	2	2	1	2	2	-	15
12	8	2	2	2	2	2	2	2	2	24
13	9	2	2	2	2	2	2	2	2	25
14	-	2	2	2	2	1	2	-	2	13
15	10	2	2	-	2	2	1	2	2	23
16	9	2	2	2	2	2	2	2	2	25
17	8	2	2	2	2	2	2	2	2	24
18	10	2	2	2	2	2	2	2	2	26
19	8	2	2	2	2	2	2	2	2	24
20	9	2	2	2	2	2	2	2	2	25
21	3	2	2	2	1	2	2	2	-	16
22	8	2	2	2	2	2	2	2	2	24
23	2	2	2	2	2	2	2	1	2	17
24	7	2	2	2	2	2	2	1	2	22
25	2	2	2	2	2	2	2	2	2	18
26	-	-	-	-	-	-	-	-	-	0
27	8	2	2	2	2	2	2	2	2	24
28	5	2		2	2	2	2	2	2	19
29	10	2	2	2	2	2	2	2	2	26
Total	166	53	54	52	52	50	51	48	49	

Table 5.2

NUMBER OF DATA	PART B												TOTAL
	FILL THE BLANKS							MATCH THE MEANING					
	a	b	c	d	e	f	g	a	b	c	d	e	
1	2	2	2	2	2	2	2	-	-	-	-	-	14
2	2	1	2	2	2	1	-	-	-	2	2	2	16
3	2	2	2	2	2	2	2	2	2	2	2	2	24
4	2	2	-	-	-	-	2	-	-	2	2	2	12
5	-	-	-	-	-	-	-	2	-	2	2	-	6
6	2	2	2	2	-	-	2	-	-	2	2	2	16
7	2	2	2	1	2	2	2	-	-	2	2	2	19
8	2	2	2	2	2	2	2	2	-	2	-	2	20
9	2	-	2	-	-	-	2	-	-	2	2	2	12
10	2	2	-	2	2	2	2	2	2	-	-	2	18
11	2	2	2	2	2	2		2	2	2	2	2	22
12	2	2	2	2	-	2	2	2	2	2	2	2	22
13	2	2	2	1	2	2	2	2	2	2	2	2	23
14	2	2	2	-	2	-	-	-	-	2	2	2	14
15	2	2	-	2	2	-	2	2	2	2	2	2	20
16	2	2	2	2	2	2	2	-	2	2	2	2	22
17	2	2	2	2	2	2	2	2	2	2	2	2	24
18	2	2	2	2	2	2	-	2	2	2	2	2	22
19	2	2	2	2	-	-	2	2	-	2	2	2	18
20	2	2	2	2	-	2	2	-	-	2	2	2	18
21	2	2	2	2	2	2	-	2	2	2	2	2	22
22	2	2	2	2	2	2	2	-	-	2	2	2	20
23	2	2	2	-	-	-	2	-	2	-	-	2	12
24	2	2	-	1	-	-	2	-	-	-	-	2	9
25	2	2	2	-	-	-	2	2	1	2	2	2	17
26	-	-	-	-	-	-	-	-	-	-	-	-	0
27	2	2	2	2	-	2	2	2	2	2	2	2	22
28	2	2	-	-	-	-	-	-	-	-	-	2	6
29	2	2	2	2	2	2	2	2	2	2	2	2	24
Total	54	51	44	39	32	33	42	30	27	46	44	52	

Table 5.3

NUMBER OF DATA	PART C											TOTAL
	T/F					FIND THE WORDS						
	a	b	c	d	e	a	b	c	d	e	f	
1	1	1	1	1	1	2	2	2	2	2	2	17
2	-	-	-	1	1	2	-	2	1	2	2	11
3	1	1	1	1	1	2	1	2	2	2	1	15
4	1	1	1	1	1	2	1	-	2	2	1	13
5	-	-	-	-	1	-	-	-	-	-	-	1
6	-	1	1	1	1	2	1	-	2	2	1	12
7	1	1	1	1	1	2	2	2	2	2	2	17
8	1	1	1	1	1	2	1	2	2	2	2	16
9	1	1	1	1	1	2	-	-	2	2	-	11
10	1	1	-	1	-	2	1	2	2	2	1	13
11	1	1	1	1	1	2	2	2	2	2	2	17
12	1	1	1	1	1	2	1	2	2	2	1	15
13	1	1	1	1	1	2	2	2	2	2	2	17
14	1	1	1	1	1	2	1	2	2	2	2	16
15	1	1	1	1	1	-	1	2	2	2	1	13
16	-	-	-	-	-	-	-	2	2	2	1	7
17	1	1	1	1	1	2	1	2	2	2	1	15
18	1	1	1	1	1	2	1	2	2	2	-	14
19	-	1	1	1	1	2	1	2	2	2	1	14
20	1	1	1	1	1	2	1	-	2	2	1	13
21	1	1	1	-	-	2	2	2	2	2	1	14
22	1	1	1	1	1	2	2	2	2	2	2	17
23	1	1	-	1	1	2	-	2	2	2	2	14
24	1	1	1	1	1	2	1	1	-	2	-	11
25		1	1	1	1	2	1	2	2	2	1	14
26	1	1	-	-	-	-	-	-	-	-	-	2
27	1	1	1	1	1	2	1	2	2	2	1	15
28	-	-	-	1	-	-	-	-	-	2	2	5
29	-	1	1	1	1	-	-	2	-	2	-	8
Total	21	25	22	25	24	46	27	43	47	54	33	

Table 5.4

D A T A N O	PART D																				T O T A L		
	MATCH PHRASE							MATCH THE MEANING					YES/NO QUESTIONS										
	1	2	3	4	5	6	7	1	2	3	4	5	1	2	3	4	5	6	7	8		9	10
1	-	2	2	-	-	-	-	2	2	2	2	2	1	1	1	1	1	1	-	-	1	1	22
2	2	2	-	-	-	2	-	-	1		1	2	-	1	-	1	-	1	-	1	1	1	16
3	-	2	2	-	-	-	-	2	1	2	2	2	1	1	1	1	-	1	1	1	1	1	22
4	-	2	2	-	-	-	-	2	1	2	2	2	1	1	1	1	1	1	1	1	1	1	23
5	-	2	2	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	14
6	-	2	2	-	-	-	2	2	2	2	2	2	1	1	1	1	-	1	1	1	1	1	25
7	-	2	2	-	-	-	-	2	2	2	2	-	1	1	1	1	1	1	1	1	1	1	22
8	-	2	-	-	-	-	-	2	1	2	2	2	1	1	1	1	1	1	-	1	1	-	19
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
10	-	2	2	-	-	-	-	2	1	2	1	2	-	-	1	1	-	1	-	1	1	1	18
11	-	2	2	-	-	-	-	1	1	-	2	2	1	1	1	1	1	1	1	1	-	1	19
12	-	2	2	-	-	-	-	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	24
13	-	2	2	-	-	-	-	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	24
14	-	2	2	-	-	-	-	2	2	-	2	2	1	1	1	1	1	1	-	-	1	1	20
15	-	2	2	-	-	-	-	1	2	1	1	2	-	1	-	1	-	1	-	1	1	1	23
16	2	2	-	-	-	-	2	2	2	2	1	2	-	1	1	1	1	1	1	1	1	1	24
17	-	2	2	-	-	-	-	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	24
18	-	2	2	-	-	-	-	2	1	2	2	2	1	1	1	1	1	1	1	1	1	1	23
19	-	2	2	-	-	-	2	2	2	-	2	2	1	1	1	1	-	1	1	1	1	1	23
20	-	2	2	-	-	-	-	2	1	2	2	2	1	1	1	1	1	1	1	1	1	1	23
21	-	2	2	-	-	-	-	2	1	2	2	-	1	1	1	1	1	1	1	1	1	1	21
22	-	2	2	-	-	-	-	2	2	2	2	-	1	1	1	1	1	1	1	1	1	1	22
23	-	2	2	-	-	-	-	1	2	2	2	2	1	1	1	1	1	1	1	1	1	1	23
24	-	2	2	-	-	-	2	2	-	-	2	-	1	1	1	1		1	1	1	1	1	19
25	-	2	2	-	-	-	2	2	1	-	2	2	1	1	1	1		1	1	1	1	1	22
26	-	2	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4
27	-	2	2	-	-	-	-	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	24
28	-	2	2	-	-	-	-	1	2	2	2	2	-	1	1	1	1	1	1	1	1	1	22
29	-	-	2	-	2	-	2	2	1	1	2	2	-	1	1	1	1	1	1	1	1	1	23
Total	4	5	4	0	2	2	1	4	3	3	4	4	1	7	5	8	9	8	1	6	7	7	

NUMBER OF DATA	TOTAL SCORE FROM EACH PART
1	73
2	62
3	84
4	72
5	36
6	69
7	82
8	78
9	39
10	54
11	73
12	85
13	89
14	63
15	79
16	78
17	87
18	85
19	79
20	79
21	73
22	83
23	66
24	61
25	71
26	6
27	85
28	52
29	81



## **Analysis**

Since this an essay test which its method in scoring the answer based on the test writer's judgment and there is no definite answer, therefore it is classified into subjective test. From the tables above we can see that the test score is vary from 1-10. The criteria test writer used in giving the score are based on the each items difficulty. The highest scale of point, 10, is given to part A, summary, which according to the test writer as the most difficult question to answer. But then we can see that almost everyone is able to provide the answer correctly except for 5 students. I assume they can not answer it correctly only because of the test presentation was not really good and almost can not be read (see appendix 2) or the students were coming late so they do not hear the additional explanation about the test instructions from the lecturer. For the middle level of difficulty, every question are related to vocabulary such as guess the meaning, fill the blank with appropriate words, etc. They were scoring between 2 or 1, it depends on the answer accuracy. For the easiest type of questions are scored only 1 as in part C, T/F questions and in Part D, yes/no question. Beside of the questions are easier, the test writer said that there is also a possibility for the students to tell each other about the answers.

### **2.1 Difficulty level**

Since the test maker scored each students' answers by determining its difficulty, thus the writer needs to examine whether the method of scoring is appropriate.

In analyzing the level of difficulty the writer only takes some parts or items as the representative sample from each level. Those items will determine whether the method used by the lecturer appropriate or not. To make a summary based on the reading in part A will represent the difficult level. For the middle level of difficulty the writer takes from part A (guessing unknown words), B (match the meaning) and C (find the words) which each part consist of 5 to 7 items. The low level is represented by items in part D (yes/no questions).

From 29 students, the writer divides their score into 3 parts; high, middle and low score. The high score is from 81 to 89, the middle score is from 69 to 79 and the low score from 6 to 66.

The writer analyzed the test items level of difficulty by using a formula taken from Madsen (1983:181) as follows:

$$\text{Difficulty level (DI)} = \frac{\text{High Correct (Hc)} + \text{Low Correct (Lc)}}{\text{Total Number in Sample (n)}}$$

It is considered too easy if > 90% answered the item correctly, too difficult if only < 35% and "about right" if the percentage fall between 35 and 90. The analysis can be seen as follows,

### 2.1.1 The difficult item

Summary (Part A, Table 5.1), the score given between 1-10 point so it is divided into correct and wrong answer;  $\geq 6$  point is correct and  $\leq 5$  point is wrong.

$$DI = \frac{Hc+Lc}{n} = \frac{9+2}{18} = 61\%$$



The result shows that this item fall between 35% and 90%, therefore, it is classify into "about right" item. It is clearly diferent with the lecturer's opinion where she divided it into the most difficult item.

### 2.1.2 The middle level of difficulty items

The score given by the lecturer between 1-2 so the writer also divide into correct and wrong answer; 2 is correct where as 0-1 is incorrect.

**Guessing unknown words (Part A, table 5.1):**

$$\text{item a : } DI = \frac{Hc+Lc}{n} = \frac{9+7}{18} = 89\%$$

$$\text{item b : } DI = \frac{Hc+Lc}{n} = \frac{9+7}{18} = 89\%$$

$$\text{item c : } DI = \frac{Hc+Lc}{n} = \frac{9+7}{18} = 89\%$$

$$\text{item d : } DI = \frac{Hc+Lc}{n} = \frac{9+7}{18} = 89\%$$

$$\text{item e : } DI = \frac{Hc+Lc}{n} = \frac{9+6}{18} = 83\%$$

$$\text{item f : } DI = \frac{Hc+Lc}{n} = \frac{9+6}{18} = 83\%$$

$$\text{item g : } DI = \frac{Hc+Lc}{n} = \frac{9+2}{18} = 61\%$$

$$\text{item h : } DI = \frac{Hc+Lc}{n} = \frac{9+6}{18} = 83\%$$

**Match the meaning (Part B, Table 5.2):**

$$\text{item a : } DI = \frac{Hc+Lc}{n} = \frac{7+2}{18} = 50\%$$

$$\text{item b : } DI = \frac{Hc+Lc}{n} = \frac{7+2}{18} = 50\%$$

$$\text{item c : } DI = \frac{H_c + L_c}{n} = \frac{9+4}{18} = 72\%$$

$$\text{item d : } DI = \frac{H_c + L_c}{n} = \frac{9+4}{18} = 72\%$$

$$\text{item e : } DI = \frac{H_c + L_c}{n} = \frac{9+7}{18} = 89\%$$

**Find the words (Part C, table 5.3):**

$$\text{item a : } DI = \frac{H_c + L_c}{n} = \frac{8+1}{18} = 50\%$$

$$\text{item b : } DI = \frac{H_c + L_c}{n} = \frac{3+0}{18} = 16\%$$

$$\text{item c : } DI = \frac{H_c + L_c}{n} = \frac{9+4}{18} = 72\%$$

$$\text{item d : } DI = \frac{H_c + L_c}{n} = \frac{8+4}{18} = 67\%$$

$$\text{item e : } DI = \frac{H_c + L_c}{n} = \frac{9+7}{18} = 89\%$$

$$\text{item f : } DI = \frac{H_c + L_c}{n} = \frac{3+4}{18} = 39\%$$

From 18 items above, 17 items are fall between 35 and 90 so it is classify into "about right". There is only one item, item b (find the words, part c) which is considered too easy. At this point the lecturer can be said has given an appropriate criterion.

### 2.1.3 The low level of difficulty

Score: 1

**Yes/No questions (Part D, table 5.4):**

$$\text{item a : } DI = \frac{H_c + L_c}{n} = \frac{8+4}{18} = 67\%$$

$$\text{item b : } DI = \frac{H_c + L_c}{n} = \frac{9+6}{18} = 83\%$$

$$\text{item c : } DI = \frac{Hc+Lc}{n} = \frac{9+6}{18} = 83\%$$

$$\text{item d : } DI = \frac{Hc+Lc}{n} = \frac{9+7}{18} = 89\%$$

$$\text{item e : } DI = \frac{Hc+Lc}{n} = \frac{8+4}{18} = 67\%$$

$$\text{item f : } DI = \frac{Hc+Lc}{n} = \frac{9+7}{18} = 89\%$$

$$\text{item g : } DI = \frac{Hc+Lc}{n} = \frac{9+4}{18} = 72\%$$

$$\text{item h : } DI = \frac{Hc+Lc}{n} = \frac{9+6}{18} = 83\%$$

$$\text{item i : } DI = \frac{Hc+Lc}{n} = \frac{9+7}{18} = 89\%$$

$$\text{item j : } DI = \frac{Hc+Lc}{n} = \frac{9+7}{18} = 89\%$$

From each item analysis, it can be seen that the items range from 67-89, in other word the percentage fall between 35% and 90% .Hence these items classify in "about right". None of the item percentage is above 90% which can be said too easy but the variety of percentage is still lower than the middle level. Therefore, it meets the criterion of its level of difficulty.

## 2.2 Students' Score

To present the students' achievement, the following sections describe how the test scores are analyzed by using some statistics formula which taken from Vallete (1977:50-52),

### 2.2.1 Range

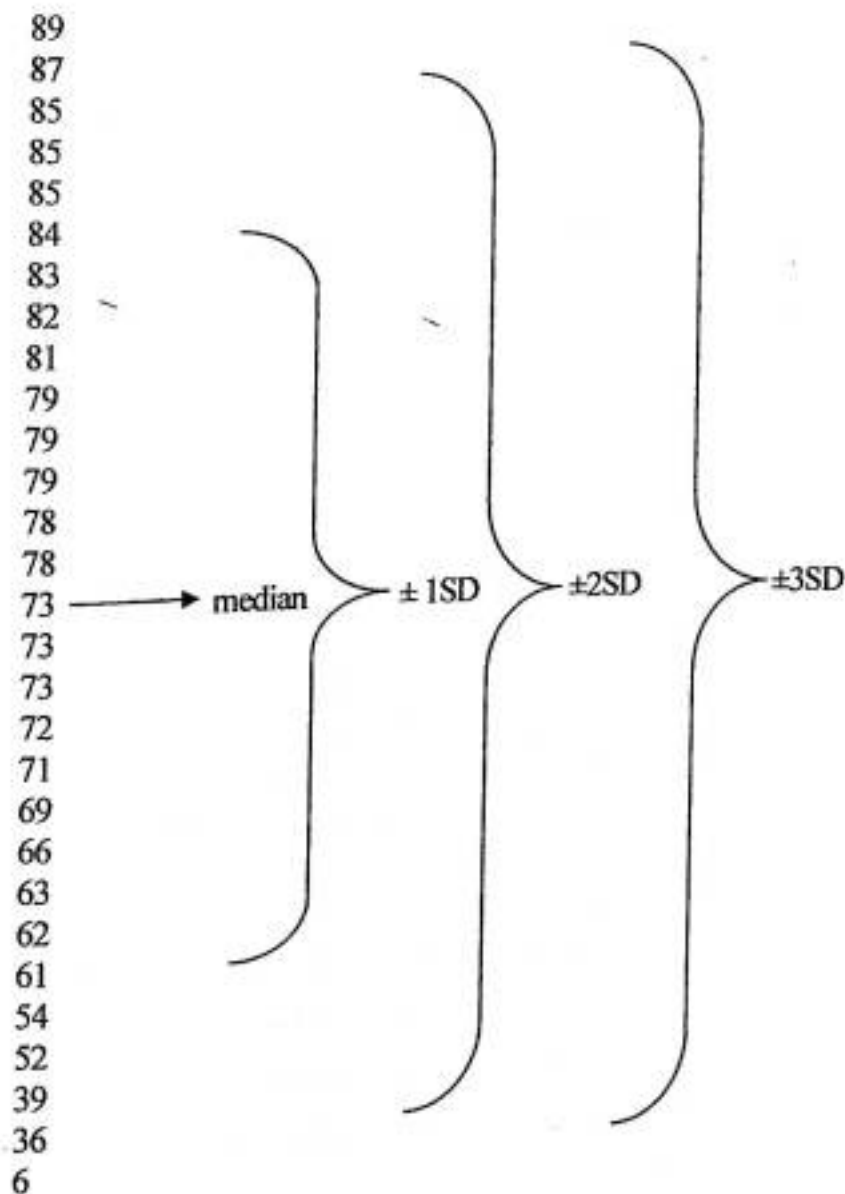
There are 54-item test are administered to 29 students. The test range is taken by finding the highest and lowest scores. From the distribution below the highest score is 89 and the lowest is 6. Therefore the test range is from 89 to 6 (or 83) and as the maximum possible score is 101.

### 2.2.2 Distribution

The score distribution of the Reading test's papers are shown in

Figure 1 as follows,

Figure 1



### 2.2.3 Median

The middle of the score after we arrange from the highest into the lowest is called median. Because of the number of papers are 29 so the median fall into score number fifteen, 73.

### 2.2.4 Mean

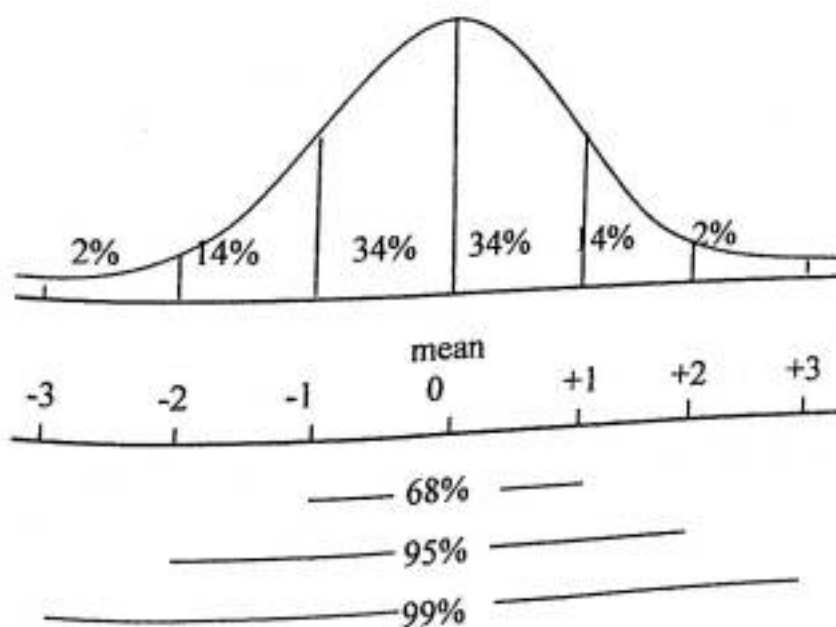
The mean score is obtained through the following formula:

$$(M) \text{ Mean} = \frac{\text{sum of the score}}{\text{number of papers}} = \frac{\sum S}{n} = \frac{2024}{29} = 69,79$$

### 2.2.5 Standard Deviation and Normal Curve

The standard deviation is a measure of the spread of the score on the test. It will be best understood in terms of normal curve. Normal curve is a theoretical distribution of the score. Thus, the writer provide the normal curve as indicated in Figure 2

Figure 2.



The following formula is used in obtaining the standard deviation.

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

one sixth of 29 students is 4,8

$$SD = \frac{(89+87+85+85+ (0,8 \times 85)) - (6+36+39+52+ (0,8 \times 54))}{14,5}$$

$$SD = \frac{414-17,2}{14,5} = \frac{273,8}{14,5} = 16,4$$

The score of the classroom follow a normal distribution because it can be seen from the various number of score in the Figure 1 then based on the theory  $\frac{1}{3}$  of the papers will fall between  $\pm 1SD$ , about 95 percent fall between  $\pm 2SD$ , and over 99 percent  $\pm 3SD$ . It can be seen from the calculation below,  $\frac{1}{3}$  of papers should fall between  $-1SD$  (69,7-16,4, or 53,3) and  $+1SD$  (69,7+16,4, or 86,1) so there will be  $\frac{1}{3} \times 29 = 19$  students between 53,3 and 86,1

95% of papers between  $-2SD$  (69,7-32,8, or 36,9) and  $+2SD$  (69,7+32,8, or 102,5, which is off the top of the scale). Therefore, there will be 27 students' papers should fall between 36,9 and 102,5.

99% of papers should fall between  $-3SD$  (69,7-49,2, or 20,5) and  $+3SD$  (69,7+49,2, or 189, which is off the top of the scale).

The result about students' score above showed that among 29 students there are 19 students has score between 53,3 and 86,1.

## CHAPTER IV

### CONCLUSION AND SUGGESTION

#### 4.1 Conclusion

Based on the analysis of the content of the Reading test and the students' Reading score, the writer concludes that:

1. The content of the reading test only covered 4 units from 12 units, therefore the test content can not be said valid.
2. The content of the Reading test only evaluated 4 reading skills from 12 reading skills required; summarizing the Reading text, guessing the difficult words, filling the summary with appropriate words and giving at least 5 new words' definition
3. The test writer's criteria in giving scores based on the items difficulty can be said inappropriate for the following reasons;
  - one of the criterion, the most difficult item should be scored as in the middle level of difficulty
  - one item in the middle level of difficulty should be scored as in the high level of difficulty
4. The students' achievement of the objectives of the reading course can be said in average's level where the points is 69,79.

## 4.2 Suggestion

1. It is suggested that the lecturers and the students understand about the objectives of the Reading course before designing the final achievement test.
2. It is suggested that the test writers design a test specification before designing the real test.
3. It is suggested that the test writers develop an appropriate criteria before scoring the students' answer.



## BIBLIOGRAPHY

- Alderson, Charles J. Clapham, Caroline. Wall Dianne. 1995. *Language Testing Construction and Evaluation*. Great Britain : Cambridge University Press.
- Bachman, Lyle F. & S. Palmer, Adrian. 1996. *Language Testing in Practice*. Hong Kong: Oxford University Press.
- Carol, Brendan J & J Hall, Patrick. 1985. *Make Your Own Language Tests 'A Practical Guide to Writing Language Performance Tests'*. England: Pergamon Press Ltd.
- Gilchrist, Alan. 1973. *Modern English Reading*. London: Longman
- Heaton, J.B. 1981. *Writing English Language Test*. England: Longman Group Ltd.
- Henning, Grant. 1987. *A Guide to Language Testing*. Los Angeles : Heinle & Heinle Publishers.
- Hughes, Arthur. 1989. *Testing for Language Teachers*. Great Britain: Cambridge University Press.
- Madsen, Harold S. 1983. *Techniques in Testing*. New York: Oxford University Press.
- Nunan, David. 2004. *Task-Based Language Teaching*. Great Britain: Cambridge University Press.
- Sudjana, Nana. 2006. *Tuntunan Penyusunan Karya Ilmiah*. Bandung : Sinar Baru Algesindo.
- Valette, Rebecca M. -----. *Modern Language Testing*. New York: Harcourt Brace Javanovich, Publishers.
- Yusuf, Haslinda. 2005. *Content Validity of an English Test Administered at SMP Negeri 12 Makassar ( A Case Study)*. Makassar : Thesis. Fakultas Sastra Universitas Hasanuddin.