EVALUATING STUDENTS' ABILITY IN THE MAKING OF ENGLISH WRITING TEST AT SEVENTH SEMESTER OF ENGLISH DEPARTEMENT ACADEMIC YEAR 2017 IN IAIN PALOPO

A THESIS

Submitted to the English Study language of S1 Tarbiyah and Teacher Training Faculty of State Institute for Islamic Studies of Palopo in Partian Fulfillment of Requirement for S.Pd Degree in English Education

By

AHLIS

REG. NUM. 13.16.3.0042

ENGLISH STUDY PROGRAM TARBIYAH AND TEACHER TRAINING

FACULTY STATE INSTITUTE FOR ISLAMIC STUDIES

(IAIN) PALOPO

2017/2018

EVALUATING STUDENTS' ABILITY IN THE MAKING OF ENGLISH WRITING TEST AT SEVENTH SEMESTER OF ENGLISH DEPARTEMENT ACADEMIC YEAR 2017 IN IAIN PALOPO

A THESIS

Submitted to the English Study language of S1 Tarbiyah and Teacher Training Faculty of State Institute for Islamic Studies of Palopo in Partian Fulfillment of Requirement for S.Pd Degree in English Education

By:

AHLIS

REG. NUM. 13.16.3.0042

Supervised By:

1. **Wisran, S.S.,M.Hum**
2. **Amalia Yahya, S.E.,M.Hum**

ENGLISH STUDY PROGRAM TARBIYAH AND TEACHER TRAINING

FACULTY STATE INSTITUTE FOR ISLAMIC STUDIES

(IAIN) PALOPO

2017

The researcher hopes this thesis can give some values to the students of English department, English teachers and the readers especially improveSpeakingin teaching-learning. The researcher admits that this thesis is not perfect, so that the writer will accept suggestions from the readers in order to make better. the researcher hopes that this thesis would be useful and give positive contribution for the reaaders.

Palopo, March16th 2018

Researcher

AHLIS
13..16.3.0042

THESIS APPROVAL

This thesis, entitled "**Evaluating Students' Ability in Making of English Writing Test at Seventh Semester of English Departement Academic Year 2017 in IAIN Palopo**" written by **Ahlis, Reg. Number 13.16.3.0042**, English S1 Study Program of Tarbiyah and Teacher Training Faculty at State Institute for Islamic Studies (IAIN) Palopo, has been examined and defended in **MUNAQASYAH** session which is carried out on **Monday, April 2nd 2018 M**, coincided with **Rajab 16th 1439 H**. It is authorized and acceptable as partial fulfillment of requirement for S.Pd, degree in English language teaching.

Palopo, <u>April 2nd, 2018 M</u>
Rajab 16t 1439 H

COMMITTEE OF EXAMINATION

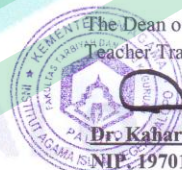|			
1.	Amalia Yahya, S.E., M.Hum.	Chairman	(.................................)
2.	Syamsudarni, S.Pd.I., M.Ed.	Secretary	(.................................)
3.	Dr. H. Dahlan, M.Hum	Examiner I	(.................................)
4.	Dewi Furwana, S.Pd.I., M.Pd.	Examiner II	(.................................)
5.	Wisran,.S.S.,M.Hum	Consultant I	(.................................)
6.	Amalia Yahya, S.E., M.Hum.	Consultant II	(.................................)

Approved by;

The Rector of IAIN Palopo

Dr. Abdul Pirol, M.Ag.
NIP. 19691104 199403 1 004

The Dean of Tarbiyah and Teacher Training Faculty

Dr. Kaharuddin, M.Pd.I.
NIP. 19701030 199903 1 003

CONSULTANT APPROVAL

Thesis Entitle : **"IMPROVING STUDENS' SPEAKING SKILL BY USING PBL (PROBLEM BASED LEARNING) METHOD AT THE TENTH YEARS OF PMDS PUTRI PALOPO".**

Written By :

Name : Restian

Reg. Number : 13.16.3.0070

Faculty : Tarbiyah and Teacher Training

Study Program :English Language Teaching

Has been corrected and approved to be examined.

Palopo, February28th 2018

Consultant I   **Consultant II**

Wisran,S.S.,M.Pd   **Amalia Yahya, SE.,M.H**
NIP. 19800613 20051 1 005   **NIP. 19720727 200504 1 003**

NOTA DINAS PEMBIMBING

Hal   : Skripsi   Palopo, February 28th 2018

Lamp   : -

Kepada Yth. Dekan Fakultas Tarbiyah IAIN Palopo

Di,-

     Palopo

Assalamu 'Alaikum Warohmatullahi Wabarakatu

     Setelah melakukan bimbingan skripssi mahasiswa tersebut di bawah ini:

Nama   : Ahlis

Nim   :13.16.3.0042

Program Study   : Tadri Bahasa Inggris

Judul   : **EVALUATING STUDENTS' ABILITY IN THE MAKING OF ENGLISH WRITING TEST AT SEVENTH SEMESTER OF ENGLISH DEPARTEMENT ACADEMIC YEAR 2017 IN IAIN PALOPO**

Menyatakan bahwa skripsi tersebut sudah layak untuk diujikan, demikian untuk di proses selanjutnya.

     Assalamu 'Alaikum Warohmatullahi Wabarakatu

                       Pembimbing II

                       Amalia Yahya, S.E, M.Hum
                       NIP. 19720727 200504 1 003

NOTA DINAS PEMBIMBING

Hal   : Skripsi   Palopo, October 16th 2017

Lamp   : -

Kepada Yth. Dekan Fakultas Tarbiyah IAIN Palopo

Di,-

        Palopo

Assalamu 'Alaikum Warohmatullahi Wabarakatu

        Setelah melakukan bimbingan skripssi mahasiswa tersebut di bawah ini:

        Nama   : Ahlis

        Nim   :13.16.3.0042

        Program Study   : Tadri Bahasa Inggris

        Judul   : **Evaluating Students' Ability in The Making of English Writing Test at Seventh Semester of English Departement Academic Year 2017 in IAIN Palopo.**

Menyatakan bahwa skripsi tersebut sudah layak untuk diujikan, demikian untuk di proses selanjutnya.

        Assalamu 'Alaikum Warohmatullahi Wabarakatu

                                Pembimbing II

                                Wahibah, Ag., M.Hum
                                NIP. 19690504 200312 2002

PRONOUNCEMENT

I have been at signature below:

Name   : Ahlis

Reg. Number   : 13.16.3.0042

Faculty   : Tarbiyah and Teacher Training

Study program   : Tadris English

With all awarness and consciousness, the researcher who signs below, pronounces that this thesis is originally written by herself except the listed sources. If somebody proves that this is duplicated, copied or made by other people as whole or partially, so this thesis will be the responsibility of the writer for law.

Palopo, on March 16th 2018

Writer,

AHLIS
Reg. Num 13.16.3.0042

CONSULTANT APPROVAL

Thesis Entitled   : "**Improving Students' Skill in Writing Passages By**

Using Context Clues at the Second Year Students of Smp Negeri 12 Palopo"
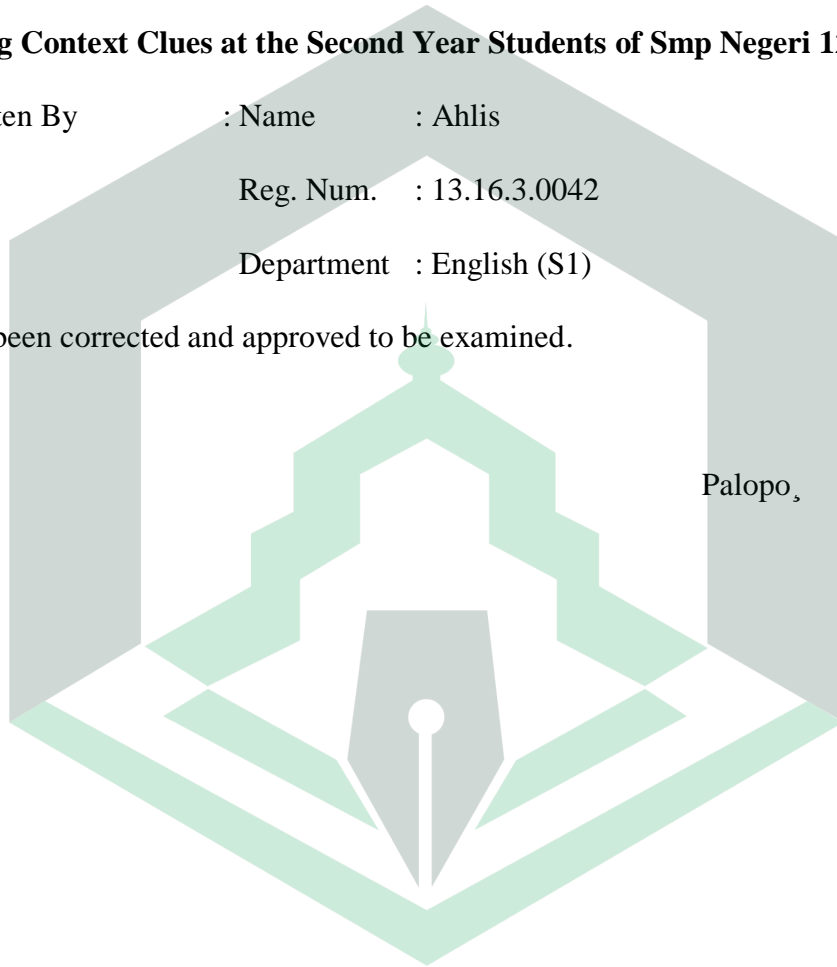
Written By   : Name   : Ahlis

Reg. Num.   : 13.16.3.0042

Department   : English (S1)

Has been corrected and approved to be examined.

Palopo¸

ACKNOWLEDGEMENT

بِسْمِ اللهِ الرَّحْمٰنِ الرَّحِيمِ

Alhamdulillah Rabbil 'Alamin, the researcher express my highest gratitude to Allah Subhanahu Wata'ala for the blessing, love, mercy on most of all change to finish this thesis. Salawat and salam be with our prophet Muhammad sallallhu'alaihi wasallam, who has guided us from the darkness to the lightness in this world.

The researcher realizes that in writing this thesis, many people have given their helps, motivation, critics and advices to her study. Therefore the researcher would like to express her deepest thank to them :

1. Dr. Abdul Pirol, M. Ag., as the Rector of IAIN Palopo, all lecturers and their assistances also the employers of IAIN Palopo who have taught, educated, helped, adviced, and guided the writer during her study in IAIN Palopo.

2. Drs. Nurdin K, M.Pd. as the head of Tarbiyah and Teacher Training Faculty of State Institute for Islamic Studies Palopo.

3. Wisran, S.Pd., M.P.d, as the first consultant who has given the writer guidance, explanation, suggestion, some ideas, and correction to the writer in finishing this thesis.

4. Amaliah Yahya S.E.,M.Hum as the second consultant, who always gives support and correction in this thesis.

The writer hopes this thesis can give some values to the students of English department, English teachers and the readers especially improve Speaking in teaching-learning. The writer admits that this thesis is not perfect, so that the writer will accept suggestions from the readers in order to make better. the

writer hopes that this thesis would be useful and give positive contribution for the

reaaders.

Palopo, March 16th 2018

Researcher

AHLIS
13..16.3.0042

TABLE OF CONTENT

ABSTRACT

Ahlis , 2017. "Evaluating the Students' Ability in Making the English Writing Test in the Seventh Semester of the Academic Year 2017/2018 in IAIN Palopo, Consultant (1) Wisran. S.S., M.Pdand Consultant (II) Amalia Yahya, S.E.,M.Hum.

This thesis researcherEvaluating the Students' Ability in constructing the English Test in the Seventh Semester of the Academic Year 2017/2018 in IAIN Palopo. This thesis focused on a evaluating about to what extent of the student's ability inmaking writing english test in seventh semester academic year20172018 in IAIN Palopo. The population of this evaluating was 100 students in the seventh semester department English year 2017.

This research uses descriptive method by using survey. Samples were taken from entireds population by purposive sampling technique.This caseresearch took 25 students of seventh semester department English as a sample. In collecting data the researcher distributes blank papers to students then write 10 items multiple choice and 5 items essay test. And explain about the purpose and how to do it in writing the test well. In addition, student are given the opportunity to write a test or a matter of English as a good test

The objectives of the research were intended to determine the quality of English test exam by multiple choice test and essay test the following criteria of a good test; test validity,test reliability and level of difficult. The result of analysis proved that good validity because the material in criteria of a good test, fair reliability (r=0,54),by multiple choice test and essay test according to level of difficult include 10 items test multiple choice and 5 items tests made at students'the result is 5 difficult, 41 fair and 3 easy by 50 respondent follows multiple choice and also the result essay test 11 difficult, 39 fair and 1 easy

CHAPTER I
INTRODUCTION

A. Background

In the process of education, evaluation is one of the important component and plays a major role in identifying the success of an educational program. Basically, the evaluation is intended to obtain data or information on the distance between the existing situation and the expected situation using certain criteria. Using data and information, teachers can take decisions on further teaching and learning activities. So that the evaluation process can take place, then the evaluation instrument should be planned, prepared and executed. One evaluation instrument is a widely used test.

The test is a systematic process of obtaining and using information to make a judgment that is used as the basis for making keputusan. In planning and preparation of the necessary tests various things, such as the rubric. Sections that are various aspects that define what should be done by the participants in the follow tes. Other words, the test section with regard to the test procedure, which includes the organization of the test, the allocation of time and the user can assign a test tes. planner expected answer is through design tests and can try to get the test instructions, specifications and input the appropriate task. Thus, the expected response is part of the test method .

In the teaching process unlimited to the teaching interaction between teacher and student. The teaching implementation completely includes evaluation as the way to get feedback from teaching process which is done. Therefore, the development, employing, and implementation good test suitable with principles occur. It will give advantages in teaching achievement completely.

Testing though as an integral part of the teachers business in the classroom, it has not been done properly. The way the instructional material is taught is not compatible with the way to evaluate the students learning. As a result, the student learning has not been accurately assessed yet. This condition is certainly not faithful for both the teachers teaching and student learning. Attempts need to be made immediately if consistency in teaching English and testing the students learning is considered to be of primary concern. So, the policy of educational authority should be relevant to the goals of teaching language and language testing.

In order to judge the effectiveness of any test, it is sensible to lay down criteria which the test can be measured as follow, validity and reliability. That is to say, any test that we use must be appropriate in terms of our objectives, dependable in the evidence it provides, and applicable to our particular situation.

The validity of at test is the extent to which it measure what it is supposed to measure and nothing else. Every test, whether it is a short, in formal classroom test or a public examination, should be as valid as the construction can make it.

The test must aim to provide a true measure of the particular skill which it is intended to measure.[1]

Reliability is a necessary characteristic of any good test: For it to be valid at all, a test must first be reliable as a measuring instrument.[2] By reliability is means the stability of test scores. A test can't measure anything well unless it measures consistently.

As a prospective teacher should be able to design a test that was a teacher should be able to make a test to obtain information as a basis for decision making of for the students who completed Practice Field Experience tarbiyah particular faculty and science teacher training courses English language has been able to make a test of English,Thus,students and teachers are able to make a test of English as a material for know the result of his study.During this effort to evaluate the outcomes of learning English in the form of the final stage of the evaluation study semester (final), and the midterm (middle test).

Generally, seventh semester students of tarbiyah and teacher training in IAIN Palopo who will plunge later in field practice will surely make the test or give the question that will be answered by the students as the material of learning evaluation. Meanwile, especially English language course are known to be quite difficult when making English language test subjects. One student who has taught at school is NurulHidayat that practice field experience in 2016 at SMK Negeri 1

---

[1] J.B Heaton, *Writing English Language test* (Longman: t.p, t.t), 153
[2] *Ibid...*p. 155

Palopo, her made the test when the student faces midterm exam. In this test, the type of question used is the type of multiple choice question, which will be answered student are taught to provide to student learning outcomes.

Furthermore, in terms of the level of student ability to make tests vary, there are low, medium and high. From the results of the evaluation make the test, has been known level of ability of students who make tests. Then from the result that have been known to students who have understood how to make the test well so get the maximum evaluation results. However, tests or problems that are made only aimed at upper secondary students only. From the statement the researcher but offokusknow the level of the ability of seventh semester students when giving tests or making exam question to high school students especially students of English study program.

Finally, it is should be noted that a test can be reliable without necessarily processing validity. However, reliability is clearly in adequate by itself, in a test does not succeed in measuring what it is supposed to measure.

By those reasons above, the researcher interest to analyze "**Evaluating the Students' Ability in Make the English Writing Test in the Seventh Semester of the Academic Year 2017/2018 in IAIN Palopo.**"This research needed to discuss and look for the solution in order to make a good test based on principles of evaluation.

B. ProblemStatement

Based on the description above background that the proposed formulation of the problem is: To what extent of the student's ability inmaking writing english test in seventh semester academic year2017 in IAIN Palopo ?

C. Objective of The Study

Based on the formulation of the problem described above, the purpose of this research is to find and evaluate: Know extent of the student's ability in making the writing english test in seventh semester academic year 2017/2018 in IAIN Palopo.

D. The Significance of the Research

The signifance of the research was expected to be useful information to the lecture of english about extent of students' ability in making writing test and the result this reearch to find out the good materials to help the lecture to improve their ability in student to making english test. And for the researcher, it can add experiance and knowledge about make writing test.

E. Scope of the Research

This research the title "Evaluating the Students' Ability in Making the English Writing Test in the Seventh Semester of the Academic Year 2017/2018 in IAIN Palopo". The reseearcher explain the concept, idea about good test with criteria of test,especially English subjects to know the level criteria of good test made by seventh semester students of English department,In this research but

ofconcern forValidity test (face validity), (content validity), (constructive validity), and (predictive validity),Reliability test and Level of difficulty.

F. Definitions of Term

Operational definition in this research is but of one variable in question, that is evaluation tool. Evaluation tool is a test tool used by students and teachers to find out where the taught material can be delivered well to the students at the time of teaching and learning activities for the assessment process or ongoing tests.

1. Evaluating is a process to know the extent to which a person's knowledge results.
2. Test is a proces test given to a person
3. Evaluating test is a process of an evaluation process to determine the validity of the test and the reliability of a teacher's test
4. Making test is the process of making the exam to be given to the students
5. Making test English is the process of making exams conducted by a special student English subjects.

CHAPTER II

REVIEW OF RELATED LITERATURE

A. **Previous Related Research Findings**

Before the researher raised the title of "evaluation tests students' ability to make english writing in the first half of the academic year 2017 seventh IAIN Palopo" There are several studies that have relevance to this study are:

1. Research conducted by Andy Marvona with the title "Problem Analysis of IPS School Subject Exam Materials History SMP Negeri Se-District Ngaglik" concluded that the results showed the quality of items that meet the criteria of theoretical validity of 30 grains (100%), level of difficulty 16 grains (5.333333%), 26 item (86.67%) distinguishing power, and reliability index about 0.514 (enough category). If all three criteria are used together, then the item is a 14-item quality (46.67%), thereby overall the less-qualified item

2. Research conducted by Muhammad Maksum with the title "An Overview Validy, Realibility, Item Facility, and ItemDiscriminationEvaluation Testin SDN 1Nologaten Ponorogo" This analysis is descriptive quantitative-qualitative research which is aimed to describe: (1) validity of test item in fifth grade English lesson daily test. (2) Difficulties level of test item fifth grade English lesson daily test. (3) reliability of test item in English lesson daily test at fifth grade. The subject of this research is 39 students in fifth grade at SDN 1 NOLOGATEN. The method to collect the data used is method of

collecting data of test item and result of daily test in fifth grade students at SDN 1 NOLOGATEN. Data analysis descriptively-qualitatively to determine validity, reliability, and item difficulty. To determine reliability used formula KR-20. The result of research about test item in daily test at SDN 1 NOLOGATEN shown that: (1) contain validity which covered basic competence and indicators as follow; response with complete the blank dialogues, identifications many information in very simply written message, and understanding texts and answering question by correctly answer. Test item validity classified as good, it is 73% valid test and 27% invalid test. (2) Difficulties level of test item classified good less with 4 easy test item, 8 medium test item, and 3 difficult test item. (3) The reliability classified is low, from analysis result showed rate 47%. Considering the test item are instrument which measure the students ability.

B. Some Pertinent Ideas

1. Concept of Evaluation

Evaluation is plan, the process of acquiring, and provide information that are required to make alternative decision.NormanE.Grounlund explain evaluasion is evaluationa systematic process of determining the extent to which instructional objectives or values in the curriculum.[3]

---

[3]M NgalimPurwanto,*Prinsip-PrinsipdanTekhnikEvaluasiPembelajaran,Cet1(Bandung: PT. Remaja RosdaKarya,1984),p.3*

Evaluation is the favored term when we talk of judging a program.Another opinion. **Evaluation is the process of delineating, obtaining, and providing descriptive and judgmental information about the worth and merit of some object's goals, design, implementation, and impact in order to guide decision making, serve needs for accountability, and promote understanding of the involved phenomena.**

Ralph Tyler said the evaluation is the process of determining the extent to which educational goals can be achieved. next by Mac lcolm, Provus, originator Discrepancy Evaluation defines evaluation as the difference that exists with a standard to determine whether there is a difference. Meanwhile, According to the Joint Committee, Evaluation research is a systematic or regular on the benefits or to multiple objects.[4]

While other experts point about the evaluation, one of which Jackson, that the understanding of evaluation needs to be associated with the assessment and measurument, Jackson wrote that: "if you can not measure accurately how well people are performing in their jobs before training, you are not going to be

---

[4]Farida yusuf tayibnapis, *evaluasi program (*Jakarta : PT. Rineka cipta 2000).p.3

Able to Determine what impact the training has had or Reviews their performance.[5]

unlike the case with Webster's new collegiate dictionary states that the evaluation is to Determine or fix the value of or to' examine and judge '. Meanwhile, according to Paulson, the notion of evaluation are as follows:'evaluation is process of examining certain objects and events in the light of specific value standards for the purpose of makingadaptive decisions.[6]

Meanwhile Dr. Suharsimi Arikunto also expressed Evaluation means assessing (but done by measuring beforehand). Thus the above opinion of some researchers also gives his own opinion that the evaluation is a process conducted by assessing or measuring the ability of a program that is planned as a gauge to determine the development of an activity.

In education there is a systematic learning process and consists of many components. Each teaching component is not separate or self-directed, but must run on a regular, dependent and sustainable basis.

The learning process is basically the interaction that occurs between teachers and students to achieve educational goals. Teachers as directors and mentors, while students as people who experience and actively involved to obtain changes that occur in these students after attending the process of teaching and

---

[5]Soekartawi, *monitoring dan evaluasi : proyek pendidikan, (Jakarta : PT. Dunia pustaka jaya 1995) p. 10*

[6]Ibid

learning, by the teacher in charge of performing an activity that is assessment or evaluation of student achievement in college. In addition to having the ability to arrange teaching materials and skills to submit materials to condition students' learning activities, teachers must also have the ability to evaluate the achievement of student learning, because evaluation is one of the most important components of teaching and learning activities, especially in formal education.

Evaluation comes from the English language "Evaluation". According to Mehrens and Lehman cited by Ngalim Purwanto, evaluation in the broad sense is a process of planning, obtaining, and providing information that is needed to make an alternative decision[7]. In relation to teaching evaluation activities contain several senses, including:

According to Norman Gronlund, followed by Purwanto ngalim in the book Principles and Techniques Evaluation Teaching, that "evaluation is a process that determines the decision to what extent the goals achieved by students". And according to Wrightone and colleagues, "the evaluation of education is the interpretation of student growth and progress toward the goals or values set out in the curriculum"[8].

Furthermore, Roetiyah in his book of Teacher Science problems which was later quoted by Slameto, describes the definition of evaluation, namely:Evaluation is the process of understanding or studying the meaning,

---

[7] Drs. M. Ngalim Purwanto M.P., *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran,* (Cet. XII; Bandung: PT. Remaja Rosdakarya, 2004), p. 3.
    [8] *Ibid.*, p. 3.

obtaining and communicating an information for the key guidelines of decision makers; evaluation is the activity of collecting data as wide as possible, deeply, concerned with the capability of students, teachers know cause and effect of student learning outcomes that encourage and develop learning ability; in the context of the development of an instructional system, evaluation is an activity to assess how far the program has proceeded as planned; evaluation is a tool to determine whether the purpose of education and whether the process in the development of the sciences has been on the path that has been expected[9].

Therefore, an educator must know the extent to which his teaching success is well achieved, and to improve and direct the implementation of the teaching and learning process and to obtain the decision. Then we need an evaluation process in learning or also called learning evaluation.

Evaluation of learning is an evaluation of systematic teaching and learning process, learning evaluation is directed to learning system component which includes raw input component, yaiut student's behavior behavior. The instrumental component is the professional ability of the teacher or the education personnel, the curriculum component (study program, method, media), the administrative component (tools, time, funds), the process component is the

_____

[9] Drs. Slameto, *Evaluasi Pendidikan,* (Cet. III; Jakarta: Bumi Aksara, 2001), p. 6.

learning implementation procedure. The output component is the learning outcome that marks the achievement of the learning objectives.[10]

Judging from its function that can improve the teaching program, then the evaluation of learning is categorized in formative or formative evaluation that is evaluation conducted at the end of teaching and learning program[11].

According to Anas Sujono, formative evaluation is an evaluation carried out in the middle or at the time of the learning process, which is implemented on every unit of learning program or sub-subject that can be completed, with the aim of knowing the extent to which learners have been formed in accordance with the purpose of teaching that has been determined.[12]

From the description above, it can be understood that the purpose of the formative test is to improve the level of mastery of learners and also to improve the learning process of teachers in the field of study, so that the expected evaluation objectives can be achieved with the maximum.

In general in the field of education, evaluation aims to:

a) Obtaining evidentiary data that will serve as a guide to where the level of ability and success level of learners in achieving the objectives of the curriculum after taking the learning process within the specified time period.

---

[10] Oemar Hamalik, *Kurikulum dan Pembelajaran*, (Cet. I; Jakarta: Bumi Aksara, 1995), p. 171.

[11] Nana Sudjana, Penilaian hasil Proses Belajar Mengajar, (Cet. III; Bandung: PT Remaja Rosdakarya, 1991), p. 5.

[12] Prof. Dr. Anas Sudjono, Pengantar Evaluasi Pendidikan, (Jakarta: PT. Raja Grafindo Persada, 2006), p. 23.

b) Measure and assess where the effectiveness of teaching and teaching methods have been applied or implemented by educators, participants of learning activities undertaken by participants.[13]

The specific objectives of the evaluation activities in education are:

a) To stimulate the activities of learners in the education program.

b) To find and discover the factors that cause success of learners in following the education program, so that can be searched and found a way out or ways of improvement.[14]

Evaluation in learning is done for decision making purposes, for example about whether or not an approach, method, or technique is used. The main objectives of the evaluation of the learning process are as follows:

a. Prepare information for the purpose of decision making in the learning process.

b. Identify parts that have not been implemented in accordance with the objectives.

c. Search for follow-up alternatives, forwarded, modified or stopped.[15]

In the state of decision-making process of learning, evaluation is very important because it has provided information about the implementation of teaching and learning process, so it can serve as a helper and control the implementation of teaching and learning process. In addition, the process

---

[13]*Ibid.*, p.16
[14]*Ibid.*, p.17

[15] Drs. Ahmad Sofyan, M.Pd., dkk, *Evaluasi Pembelajaran IPA Berbasis Kompetensi*, (Cet. I; Jakarta Press, 2006) p. 31-32.

evaluation function is to provide information about the results achieved, as well as the weaknesses and needs of further program improvement, then furthermore this information as feedbackfor teachers in redirecting the deviations in the implementation of the original plan towards the goal will be achievedTheteacher functioning as an evaluator needs to be taken seriously so that the evaluation is really about the target. This is based, because almost every time teachers carry out evaluation activities to assess the success of student learning and teaching programs.

2. Concept of Test

The test is one of the evaluation process in which there are questions that must be answered. According to experts one of them F.L. Geodenough, test is a series of tasks given to an individual or group of individuals for the purpose of comparing the skill with each other.[16] The next Anastasi, said test is a measuring device that has an objective standard that can be used widely, and can actually be used and compare the mental state or behavior of individuals.[17]

Besides other opinions testing is a way of gathering data to provide a test to the object under study. Furthermore, the test can also be referred to a series of questions or exercises and other tools used to measure the skills, knowledge, intelligence, ability or talent possessed by individuals or groups.[18] , meanwhile

---

[16]http://selvy-pgri.blogspot.co.id/2012/10/pengertian-definisi-tes-menurut-para.html

[17]Ibid

[18]http://selvy-pgri.blogspot.co.id/2012/10/pengertian-definisi-tes-menurut-para.html

According to Allen Philips A test is commonly difined as a tool or instrument of measurement that is used to Obtain the data about a specific trait or characteristic of an individual or group.[19]

Meanwhile, According Ridwan test as an instrument of data collection is a series of questions / exercises are used to measure the skills of knowledge, intelligence, ability or aptitude of the individual or group. Than also P.Douglas Brown also stated that a test is a method of measuring a person's abilities or knowledge in given domain, with an emphasis on the concepts of method and measuring.[20] From some of the above opinion researchers also noted that the test is an instrument of questions to be answered either orally or in writing to know Capacity of someone.

1. **Types of Tests**

   a. Function

1. Achievement test

An achievement test is related directly to classroom lessons, units, or even a total curriculum. Achievement test are (or should be) limited to particular material addressed in a curriculum within a particular time frame and are offered after a course has focused on the objectives in question.[21] According to Donald Ary. "Achievement test are widely used in educational research, as well as in

---

[19]Suharsini Arikunto, *Prosedur Penelitian: Suatu Pendekatan Praktik. (Jakarta: PT. Rineka Cipta, 2006), p. 150*

[20]Tanzeh, *Metodologi Penelitian Praktis,* (Yogyakarta: Teras, 2011), p.91

[21]Ibid, 47

school systems. They measure mastery and proficiency in different areas of knowledge by presenting subject with achievement standard set of questions involving completion of cognitive task."[22]

Achievement test, thought similar in achievement number of ways to progress test, are far more formal test and are intended to measure achievement on achievement large scale. Most annual school examinations take the form of achievement test all public tests which are intended to show mastery of achievement particular syllabus are also achievement test.

Achievement test are often summative because they are administered at the end of unit or term of study, they also play an important formative role. An effective achievement test will offer wash back about the quality of a learner's performance in subset of the unit a course. This wash back contributes to the formative nature of such tests.

The specifications for an achievement test should be determined by:

1. The objectives of the unit or course being assessed.

2. The relative importance (or weight) assigned to each objective,

3. The task employed in classroom lessons during the unit of time

4. Practicality issues, such as the time frame the test and turnaround time, and

5. The extent to which the test structure lends itself to formative wash back.

---

[22]Donald Ary, Introduction to research in education. Third Edition. (New York : CBS College Publishing, t.t), 216

Achievement test measure the current status of individuals with respect in given areas of knowledge or skill. They including subtest which measure test in areas such as reading, math, spelling, social studies, science and listening, comprehension.

b. Proficiency test

Proficiency test is not limited to any one curriculum, or single skill in the language; other, it test overall ability. Proficiency tests have traditionally consisted of standardized multiple-choice items on grammar, vocabulary, reading comprehension and aural comprehension.[23]

The proficiency tests is concerned simply with measuring a student's control of the language in the light of what he or she will be expected to do with in the future performance of a particular task.[24] The proficiency test is thus concerned with measuring not general attainment but specific skills the language demands made later on the students by a future course of study or job.

An early advocate of proficiency test, there are:

1) A reason for the test

2) A particular performance to be evaluated

3) Exercise that elicit the performance

4) Systematic rating procedures.[25]

---

[23] Douglas Brown, *Language Assessment* ,p.44
[24] J.B Heaton, *Writing English*.p,173
[25] Donal Ary, *Introduction to Research*,p.220

When the proficiency test is desired, the tester should follow three basic steps in constructing the test, namely.[26]

1) A set of test specifications listing the critical dimensions to be tested will lead to a more comprehensive coverage of the domain. State weather there will be time limits, whether reference books will be available and so on.

2) Provide a problem or an exercise that gives students an opportunity to perform either a simulation or an actual task. All individuals should be asked to perform the same task.

3) Develop an instrument that lists the relevant criteria to use in evaluating the performance.

Proficiency test are almost always summative and referenced. They provide result in the form of a single score (or at best two or three sub scores, one) each section of a test, which is sufficient result for the gate-keeping role they play of accepting or denying someone passage into the next stage of a journey. A typical example of a standardized proficiency test is the test of English as a foreign language (TOEFL) produced by the educational testing service.

c. Aptitude test

A language aptitude test (or prognostic test) is designed to measure the student's probable performance in a foreign language which he or she has not started to learn: i. e it assesses aptitude for learning a language. Aptitude test generally seeks to predict the student's probable strengths and weakness in

---

[26]*Ibid*

learning a foreign language by measuring performance in an artificial language. The ability to learn new phonemic distinctions and also to use language pattern in an unfamiliar but systematic manner is tested by means of the artificial language.

A language aptitude test is designed to measure capacity or general ability to learn a foreign language and ultimate success in that under language aptitude tests are ostensibly designed to apply to the classroom learning of any language.[27]

Aptitude test attempt to measure the subject's ability to perceive relationships, solve problem and apply knowledge in a variety of contexts.

There are two variety of aptitude test:

1) General aptitude test

General aptitude test may be group test or individually administered test. Group tests are more convenient to administer, save considerable time and provide an estimate of academic potential that is educate for most educational studies. A serious disadvantage of group tests is that they require a great deal of reading.

2) Specific aptitude test

Specific aptitude tests attempt to predict the level of performance that can be expected of an individual following future instruction or training in a specific area. As with general aptitude, they are used by teachers, counselors and administrators for the same reason.

d. Diagnostic test

---

[27]Douglas Brown, *Language Assessment* ,p. 43

A diagnostic test, which is sometimes called a formative or progress test, cheeks on student's progress in learning particular element of the course. It is used for example at the end of a unit in the course book or after a lesson designed to teach one particular point.

The term diagnostic test is widely used, few tests are constructed solely. In this case, the teacher should always be alert to every facet of achievement revealed in a class progress test. If only one or two students make particular error, the teacher will not pay too much attention. However, if several students in the group make a certain error, the teacher will not the error and plan appropriate remedial teaching. The teacher may to be clinician and diagnostician.

Diagnostic test should elicit information on what students need to work on in the future. Therefore, a diagnostic test will typically offer more detailed subcategorized information on the learner.

A typical diagnostic test of oral production of English pronunciation. Test-takers are directed to read a 150 word passage while they are tape recorder. The test administrator then refers to an inventory of phonological items for analyzing a learner's production. After multiple listening, the administrator produces a checklist of errors in five separate categories, each on which has several subcategories. The main categories include:

1) Stress and rhythm

2) Intonation

3) Vowels

4) Consonants, and

5) Other factors.[28]

e. Placement test.

Certain proficiency test can act in the role of placement test, the purpose of which is to place a student into a particular level or section of language curriculum or school. A placement test usually, but not always includes a sampling of the material to be covered in the various courses in a curriculum. Placement tests come in many varieties assessing comprehension and production, responding through written and oral performance, open-ended and limited responses, selection (e. g. multiple-choice) and gap-filling formats, depending on the nature of a program and its needs.

The ultimate objective of placement test is, of course, to correctly place a student into a course or level. Secondary benefit to consider includes face validity, diagnostic information on student's performance, and authenticity. A placement test is designed to sort new students into teaching group, so that they can start a course at approximately the same level as the other students in the class.

b. Form

a. Subjective test

It is the test need response may be whatever students wish to say. Some answers will be better than other, thus perhaps causing a problem in the scoring of the item. Usually a test might present students with the stimulus more. For

---

[28]*Ibid* ,p.47

example, from the stimulus " you went to live in Cairo two years ago. Someone asks you, how long have lived there. What would you say?"[29]

So, the response to be marked as follows:

Answer I   : I have been living in Cairo since 1986.

Answer II   : I did not leave Cairo since 1986.

Answer III : I have lived in the Cairo city for above two years.

Answer IV : from 1986.

b. Objective test

Objective test are frequently criticized on the grounds that they are simple to answer than subjective test. Objective tests need not be confined to any one particular skill or element. Objective tests usually have only one correct answer (or, at, least, a limited number of correct answers), they can be scored mechanically.

An objective test will be a very poor test if:[30]

1. The test items are poorly written.

2. Irrelevant areas and skill are emphasized in the test simply because they are "testable", and

3. It is confined to language-based usage and neglects the communicative skills involved.

---

[29]J.B Heaton, *Writing English*,p.25
[30]*Ibid*, p.27

A good classroom test will usually contain both subjective and objective test item.

2. Content

a. Physic Tests

It is the test to assess our understanding of physics at the college preparatory level.

b. IQ Test

IQ test which display statistics provide a great way of comparing your skills to other people.

c. Vocabulary Test

That is should avoid grammatical structures which the students may find difficult to comprehend. It is contain only those lexical items which present no difficult to the student.

2. Criteria of a Good Test

Alt good tests process two qualities. Validity and reliability. That is to say, any test that we use must be appropriate in terms of our objectives dependable in the evidence it provides, and applicable to our particular situation.

1. Validity

The validity of a test is the extent to which it measure what it is supposed to measure and nothing else. Every test, whether it be a short, in formal classroom test or a public examination, should be as valid as the construction can make it.

The test must aim to provide a true measure of the particular skill which it is intended to measure.[31]

Validity in language test depends on the linguistic content of the test and on the situation or technique used to test this content. Validity can be achieved and verified indirectly by correlating the scores on a test with those of another test or criterion which is valid. If the two sets of scores correlate highly, that is, if students who make high scores on the valid criterion test also score high on the experimental test and if those who score low on one also score low on the other, we say that the test is valid.[32]

Validity is relatively easy to define but difficult to put into practice validity refers to whether a test measures what it purports to measure it also defines the usefulness of that test for making predictions and decision about individuals.

The validity of a test is the extent to which it measures what it is supposed to measure and anything else.[33] Every test, whether it be short, informal classroom test or a public examination should be as valid as the constructor can make it. The test must aim to provide a true measure of the particular skill which it is intended to measure, it will be a valid test.

The following list is the types of validity test, namely:

a. Face Validity

---

[31]Materi Pelatihan Guru Bahasa Inggris SLTP, *Testing and evaluation*
[32] Robert Lado, *Language* ., p.30
[33] J.B Heaton, *Writing English*,p.159

This type of validity, in fact is often referred to as face validity if a test item looks right to other testers, teacher, moderators, and testers, it can be described as having at least face validity. It is therefore, often useful for show a test to colleagues and friends. As constructors the test, we can became so involved in the test that we sometimes fail to stand back and look at the individual test item objectively.

The concept of face validity is far from new in language testing but the emphasis now placed on it is relatively new. In the past, face validity was regarded by many test writers simply as a public relation exercise. Today, however, most designers of communicative test regard face validity as the most important of all types of test validity. Indeed, many argue that a test must look valid even as far as the reproduction of the material if self is concerned.

Face validity refers to the degree to which a test looksright, and appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of the examinees who take it. The administrative personnel who decide on its use, and other psychometrically unsophisticated observers. Face validity means that the students perceive test to be valid. Face validity ask the questions "Does the test, on the face of it, appear from the learner's perspective to test what it is designed to test? Face validity will likely be high if learners encounter:

1. A well-constructed, expected formal with familiar tasks,

2. A test that is clearly double within the allotted time-limit,

3. Items that are clear and uncomplicated,

4. Directions that are crystal clear,

5. Task that relate to their course work (content validity), and

6. A difficulty level that percents a reasonable challenge.[34]

Remember, face validity is not something that can be empirically tested by a teacher or even by a testing expert for his reason, some assessment experts that face validity as a superficial factor that is dependent on the whim of the perceiver.

b. Content Validity

Since language is a highly structured andhighly complex activity we will need more than face validity to be convinced that a test measure the language element or skill it purports to measure. We will need to check the validity of the content of the items. The test should be so constructed as to contain a representative sample of the course, the relationship between the test items and the course objectives always being apparent.

Content validity generally refers to the correspondence between achievement test items and the instruction for which the test is built. An appropriate technique for checking the compliance of examination items with the rule just stated involves the use of judges. The judges you choose to look at your test items should have competence in your subject area. However, if you get 75 percent agreement is less than 50 percent, you should reexamine you item choices or discus instruction expectations with the judges.

---

[34] Douglas Brown, *Language Assessment*.,p.27

A test said has a content validity if scope and content of that test suitable with the scope and content of the curriculum taught.[35] To a certain extent a test of the content validity of a given test is a qualitative judgment. However, in judgment contain validity the test consumer can use some criterion, there are:

a) Completeness of sample.

A major component of content validity is a representative sampling of the entire domain or subject matter, not just a portion of the subject.

b) Thoroughness of the topics being sample.

In the content validity of a test, the informed test consumer should make sure that topics or subject matter areas are adequately sampled.

This can usually be ascertained by reading the item analysis section of the manual to see how many questions are asked on each subtopic.

c) Level of complexity of the test items

For content validity is that questions must assess learning or knowledge at various levels of complexity. It is important that a test sample a student's learning not only at the most basic levels of learning (knowledge and comprehension), but also at the higher levels (analysis, synthesis and evaluation).

d) Mastery of content

---

[35] Ngalim Purwanto, *Prinsip-Prinsip Dan Tehnik Evaluasi Pengajaran* (Bandung: Rosdakarya, 2002), 138

Test with strong content validity should assure that student posses mastery of basic skills in each subject matter area before they are allowed to progress to higher level skills.

c. Construct validity

Construct validity is a characteristic most frequently attributed to ability (for instance, intelligence) test or personality tests very generally it items that hypotheses about the relatedness of behaviors prove correct.

Perhaps the closest thing a achievement testing to the notion of construct validity relates to the use of several different types of items (true false, matching, multiple-choice, essay).

If a test has construct validity, it is capable of measuring certain specific in accordance with a theory of language behavior and learning. The type of validity assumes the existence of certain learning theories of construct underlying the acquisition of abilities and skills.

A construct is any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perceptions. Constructs may or may not be directly or empirically measured-their verification often requires inferential data. You will be tempted, perhaps, to run a quick content check and be satisfied. With the test's validity. But don't let the concept of construct validity score you. An informal construct validation of the use of virtually every classroom test is both essential and feasible.

Construct validity is a major issue in validation large-scale standardized test of proficiency. Because such test must, for economic reasons, adhere to the principle of practically, and because they must sample a limited number of domains of language. They may not be able to contain all the of a particular field or skill.

d. predictive validity

A fourth type of validity is usually referred to as statistical or empirical validity. This validity is obtained as a result of comparing the results of the test with the results of some criterion measure such as:

- An existing test, known or believed and given at the same time, or

- The teacher's ratings or any other such form of independent assessment given at the same time; or

- The subsequent performance of the tester on a certain task measured by some valid test, or

- The teacher's rating or any other such form of independent assessment given later.[36]

The test situation or the technique used is always an important factor in determining the overall validity of any test. Although an ideal test situation will by no means guarantee validity, a poor test situation will certainly detract from it.

A good test kit should be justified in terms of feasibility, validity, reliability, and practicality. That is, to assess the final outcome in the learning

---

[36] J.B. Heaton, *Writing English*,p.161

required a quality assessment tool. One of the most commonly used assessment tools is the test. The test can be known to the quality by means of a problem analysis before the problem is given to the test participants.

Theoretical analysis is the analysis of items in terms of material, construction, and language. From the material point of view, a good item item should be comprehensive and contain relevant items. Comprehensive means that the item covers all the content or lesson material that has been identified as the measuring objective, representatively and in the number of comparable questions for each section according to the weight of each part. Relevant means the points of the question really only ask about the material that has been identified and everything that has been related and considered necessary to understand the material. Theoretical analysis is done by examining each item of matter, construction, and language aspects. The material aspects studied relate to the substance of the science questioned in the test item as well as the degree of capability appropriate to the test. Theoretical validity theoretical question is the suitability between the item and the referenced kisikisi. Expert (expert judgment)

The analysis is done by matching the item with the criterion on the questionnaire, if the item corresponds between the grid with the item with the criteria on the study sheet, then it is given a check mark ($\vee$), if it is not then crossed (), and if the criterion on the sheet has matter not according to the material tested then given the dash. It is important to do theoretical validity analysis before the question is used to measure and know the student learning outcomes. This is

related to the validity and reliability of the problem, as this becomes the starting point for testing the accuracy of the measuring instrument to be used.

The following is the format of the study of multiple choice items and the descriptions in terms of material aspects, construction aspects, and language aspects.

Table 1: multiple choice Item Review Sheet

No.	Aspects reviewed	Number Question										
		1	2	3	4	5	6	7	8	9	10	…
A.	ASPECT MATERIAL											
1.	Item according to indicator											
2.	There is only one answer key or right answer											
3.	The content of the material matches the purpose of measurement											
4.	The contents of the material according to the level, type of school, and grade level											
5.	The choice actually works, if the choice is the result of the calculation, then the false is a wrong choice formula / miscalculated											
B.	CONSTRUCTION ASPECTS											
6.	The subject matter (steam) is clearly defined											
7.	The formulation of questions and choices is formulated firmly											
8.	The subject matter does not lead to the correct choice of answers											
9	The subject matter does not contain a double negative statement											
10.	When forced to use a negative											
11.												
12												
13												

	word, it must be underlined or printed otherwise												
14	Choice of homogeny answers												
15	Avoid any alternative answers: "The whole answer												
16	above true "or" not one answer above is true "and the like												
C	Alternative length / choice of												
17	answers is relatively the same, do not be very long and there is a very short												
18	Options in the form of numbers or time sorted												
19	Discourses, pictures, or graphics really work												
20	The items are independent of each other												
21	Other LANGUAGE ASPECT The formulation of a communicative sentence												
	Sentences use good and correct language, according to the language type												
	The formulation of a sentence does not lead to multiple interpretations or misunderstandings												
	Using common language or words (not local languages)												
	The formula does not contain the words												
	can offend students												

Depdiknas, 2003: 76-80)

Multiple choice That is an objective test consisting of unfinished questions

or statements, and to solve it must be selected one (or more) of several possible

answers that have been provided on each item. This form test is commonly used because it has several advantages compared with other forms, including: First, the material tested can cover most of the teaching materials that have been given. Second, learners' answers can be corrected or assessed easily and quickly. Third, the assessment will be more objective because the answer to each question is definitely right or wrong.[37]

Table 2: Essay Essential Item Review Paper

No.	Aspects reviewed	Number Question										
		1	2	3	4	5	6	7	8	9	10	…
A. **1.** **2.**	**Material** Problems in accordance with indicators (requires a written test for the Description form) Limitations of questions and											
3.	answers which is expected to be appropriate											
4. **B.** **5.**	Material in question according to competence (urgency, relevance, continuity, high daily wear)											
6. **7.** **8.**	The contents of the material in question according to the level of school or grade level											
C. **9.**	Construction											
11.	Use a question word or command that requires an											

[37]Anas Sudijono, *Pengantar* Evaluasi Pendidikan, p. 118

No.		
	answer to the description	
12.	There are clear instructions on how to do a problem	
13.	There are scroll guidelines	
	Tables, images, graphics, maps, or the like are presented clearly and legibly	
	LANGUAGE ASPECT Formulation of sentences about communicative	
	Not using words / phrases that give rise to multiple interpretations or misunderstandings	
	Not using the local / taboo language	
	The formulation of the problem does not contain	

An essay test or a description test is a question whose answer requires the testee to organize ideas or things he has learned by putting the idea into writing. The description test in the literature is also called the essay examination, is the oldest learning appraisal tool. In general, this description test is a question that requires students to answer it in the form of describing, explaining, discussing,

comparing, giving reasons, and other forms of a kind in accordance with the demands.

2. Reliability

Reliability is a necessary characteristic of any good test: for it to be valid at all, a test must first be reliable as a measuring instrument. The reliability of a set of test scores refers to the consistency with which the results place students in the same relative position to other students if the is given repeadtly. By reliability is meant the stability of test scores. A test can't measure anything well unless it measure consistently. To have confidence in a measuring measurement.

A reliable test is consistent and dependable. If you give the same test to the same student or matched students on two different occasions the test should vield similar result. The issue of reliability of a test may best be addressed by considering numbers of factors that may contribute to the unreliability of a test.[38]

---

[38]Douglas Brown, *Language Assessment*………,20

A measurement is reliable if it reflects mostly true score, relative to the error. For example, an item such as "Red foreign cars are particularly ugly" would likely provide an unreliable measurement of judies against foreign mode cars. This is because there probably are ample individual differences concerning the likes and dislikes of coolers. Thus, this item would "capture" not only a person's prejudice but also his or her color preference. Therefore, the proportion of true score (for prejudice) in subject's response to that items would be relatively small.[39]

Some types and methods of reliability have procedure about how they are assessed base.

1. Test-Retest

It is degree to which scores are consistent over time. With does this test retest, a teacher will known how far consistency of test measuring measurement. The way for determining reliability of test-retest, as follows:[40]

a) Administer the test to an appropriate group.

b) After same time has passed, say one or two week, administer the same test to the same group

c) Correlate the result from two tests.

   If the result of correlation showed high, if means that the reliability test is good. The other way, if correlation is low, it means that the test.

---

[39] http://en.wikipedia.org/wiki/reliability test, on 22 May 2010

[40] Sukardi, *Evaluasi Pendidikan Prinsip Dan Operasionalnya* (Jakarta: Bumi Aksara, 2009), 45

2. Equivalent

It is the two forms measure the same variable, have the same number of items, the same structure, the same difficulty level and the same directions, scoring and interpretation. The steps for determining reliability of equivalent test as follows:[41]

a. Choose the subject will be test

b. Administer the test to subject

c. Administration the result

d. At the same session, administer the second from of the test to the same group.

e. Correlate the two sets of scores.

If the result of equivalent coefficient is high, its means the test have a good equivalent reliability. The other way, if coefficient is low the tests have low equivalent reliability.

3. Split-Half

This type measuring the internal consistency. The split half method is based on the principle that, if an accurate measuring instrument were broken into two equal parts, the measurements obtained with one part would correspond exactly to those obtained with the other. The test is divided into two and the corresponding scores obtained, the extent to which they correlate with each other governing the reliability of the test as a whole.[42]

---

[41] *Ibid*, 47

[42] J.B. Heaton, *Writing English*,p.164

The procedure for determining reliability of split-half as follows:[43]

a. Administer the total test to a subject.

b. Divide the test into two comparable halves, generally with divide items into odd and evenness number.

c. Count subject scores each two halves group.

d. Correlate the two sets of scores using the correlation formula

The fundamental issue in all testing and measurement is to discriminate between larger and quantities for something, better and worse performances, success and failure, more or less of whatever one want to test or measure. Discrimination power it refers to the ability of an item to differentiate among students on the bases of how well they know the material being tested. Various hand calculation procedures have traditionally been used to compare item responses to total test scores using high and low scoring groups of students. [44]

There are various method of obtaining the index of discrimination:

All involve a comparison those student who performed purely on the whole test. However, while it is statistically most efficient to compare the 27,5% upper group and 27% lower group. It is enough for most purposes to divide more samples in the classroom.

There are 3 ways that can be done to determine the reliability of test scores, namely:

[43] Sukardi, *Evaluasi Pendidikan*.,p.48
[44]http://www.washington.edu/oea/score_1. htm

1) Re-measurement constancy: compatibility between the first and seconmeasurement results of a measuring instrument against the same group.

2) Constant measurement constancy: the suitability of measurement results and 2 or more measuring instruments based on the old lattice competencies.

3) Construct the halves: the fit between the first hemispherical measurements and the second half of the same measuring instrument.The use of the formula to find out the coefficients of the three types of reliability described above in detail below.

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

Notes:

r   : coefficient correlation between x and y

N   : total respondent

x   : variable x

y   : variable y

$\sum xy$: The total of x and y

$$r_{11} = \frac{2r}{1+r}$$

Notes:

r_{11} : reliabilities test

r : equivalent product moment

Reliability Test	Classification
0,90 – 1,00	More Highly
0,70 – 0,89	Highly
0,50 – 0,69	Fair
0,30 – 0,49	Low
< 0,30	Very Low

3. Level of difficulty Test

Item difficulty is the percentage of the total group that got the item correct. It is important because it reveals whether an item is to easy or to hard. In either case, the item may add to the unreliability of the test because it does not aid in differentiating between those students who know the material and those who do not. For example, and item answered correctly by everyone does nothing to aid in the assignment of grades. The same is true for items that no one answer correctly.

The item difficulty index ranges from 0 to 100, the higher the value, the easier the question. When an alternative is worth other than a single point, or

when there is more than correct alternative per question, the item difficulty is the average score on that item divided by the highest number of points for any an alternative. Item difficulty is relevant for determine whether students have learned the concept being tested. The item will have low discrimination if it is so difficult that almost everyone gets it wrong or guesses, or so easy that almost everyone get its right.[45]

Table 2.2

Level of Difficulty Classification

Index difficulty	Classification
0,00-0,30	Difficult
0,30-0,70	Fair
0,70-1,00	Easy

The following seven points are taken as criteria of a good test[46].

1. A good test must possess a very high validity.

2. It must have a high reliability.

3. It must be very objective in nature.

---

[45]http://www.washington.edu/oea/score 1. htm.
[46]*Sandfeord,Peter. The Educational Psychology on Objective*

Study (Longmans), 1939 Edition, p.318.

4. It must pick out the good students from the poor

i.e. it must possess high discriminating power.

5. It must be very comprehensive. ,

6. It must be easy to use. Its administration and scoring

   must be easy and there must be economy of time and effort.

7. Norms established on the basis of its results must be satisfactory

C. Conceptual Framework

seventh semester students have learned writing from the previous semester. In this case, researchers evaluated the ability level of the students make test writing english. So that researchers can find out how these students make test writing.

Picture 1. Conceptual Framework

```
┌─────────────────────────────────────────┐
│  Making Test Students  Sevent semester of │
│    English Departement Year 2017/ 2018     │
└─────────────────────────────────────────┘
                    │
                    ▼
        ┌───────────────────────┐
        │    Evaluating Test     │
        └───────────────────────┘
                    │
      ┌─────────────┼─────────────┐
      ▼             ▼             ▼
┌───────────┐ ┌─────────────┐ ┌───────────────┐
│Test Validity│ │Test Reability│ │ Level difficult │
└───────────┘ └─────────────┘ └───────────────┘
```

Conceptual Framework Based on the above objectives researchers to determine the level of students' ability in English writing test making the seventh semester of 2017 Academic year in IAIN Palopo.
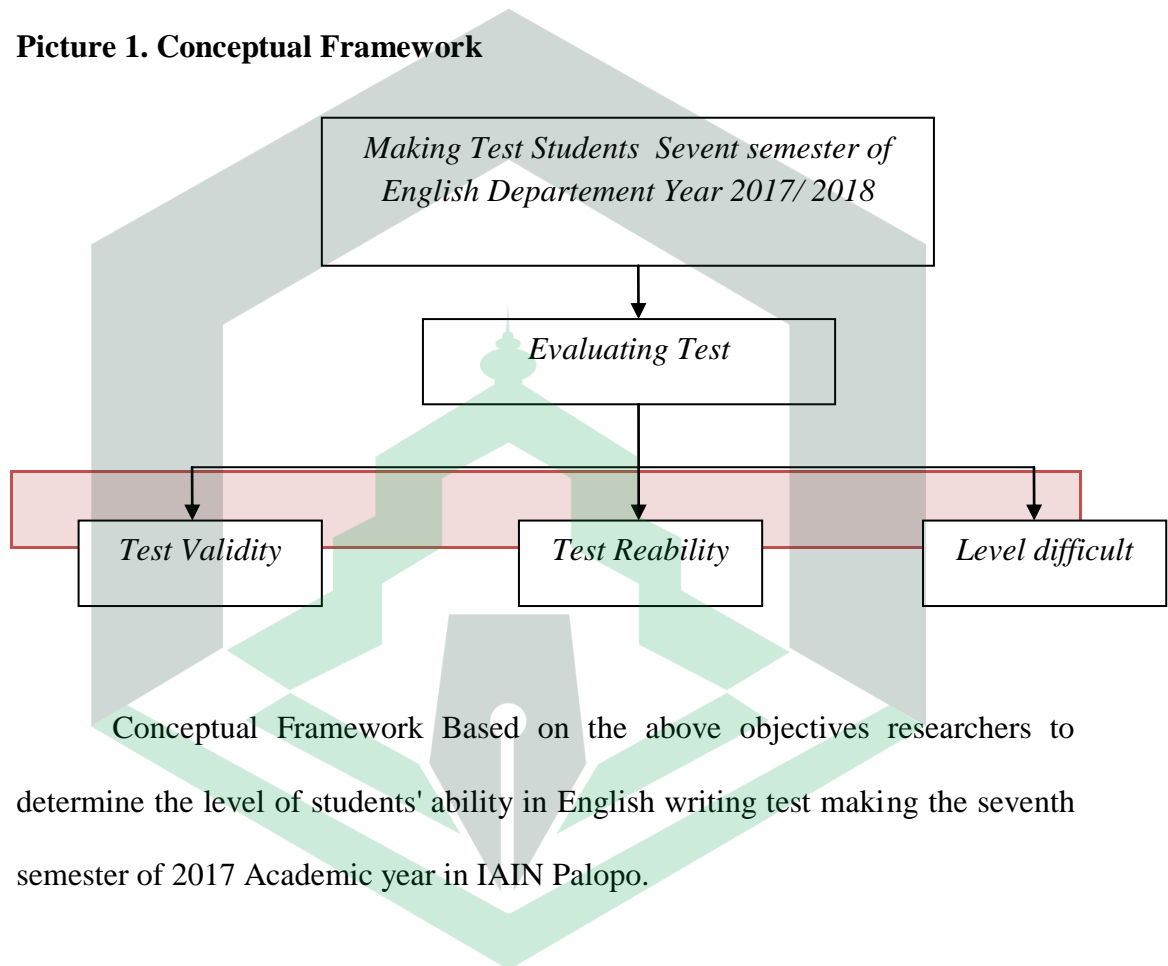
CHAPTER III

RESEARCH METHOD

A. **Research Design**

This research used descriptive method by using survey. This is to find out know the level of ability of students of Class A, B, C, D seventh semester English study program IAIN Palopo Year 2017/2018.

B. **Population and Sample**

1. Population

The population of this research is seventh semester students english study program of IAIN Palopo consisting of four classes A = 25, B = 25, C = 25, D = 25 and the number of students is 100 students

2. Sample

This research was used purpose sampling to take the sample. The researcher was take students from each class become sample. The number of sample is 25 students at the seventh English study program of IAIN Palopo.

C. **Instrument of the Research**

The research instrument is writing English test. Writing English test is used to know the level of ability to make English test at the seventh English study program of IAIN Palopo Academic Year 2017.

D. Procedure of Collecting Data

In collecting data the researcher distributes blank papers to students then write about 10 items. And explained about the purpose and how to do it in writing the test well. In addition, students are given the opportunity to write a test or a matter of English as a good test

E. Technique of Data Analysis

The results of making a student test are assessed using the formulas used by the author as follows :

1. Validity

Table 3.1

Ability in writing tests	Content validity of test
multiple choice test	30/50 x 100 = 60%
Essay test	20/50 x 100 = 40%

a. Multiple coice test

Table 3.2

resp	Test validity				Total
	Validity face	Validity content	Validity contructive	Validity predictive	
1					
2					

3					
4					
5					

b. Essay test

Table 3.3

resp	Test validity				Total
	Validity face	Validity content	Validity contructive	Validity predictive	
1					
2					
3					
4					
5					

2. Reability

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

Notes:

r   : coefficient correlation between x and y

N   : total respondent

x   : variable x

y   : variable y

$\sum xy$: The total of x and y

$$r_{11} = \frac{2r}{1+r}$$

Notes:

r_{11} : reliabilities test

r   : equivalent product moment

Table 3.4

Reliability Test	Classification
0,90 – 1,00	More Highly
0,70 – 0,89	Highly
0,50 – 0,69	Fair
0,30 – 0,49	Low
< 0,30	Very Low

3. Level of difficulty

$$Ld = \frac{R}{N}$$

Notes:

Ld   : Level of difficulty

R   : Total group that got correct answer

N   : Total respondent

Index difficulty	Classification
0,00-0,30	Difficult
0,30-0,70	Fair
0,70-1,00	Easy
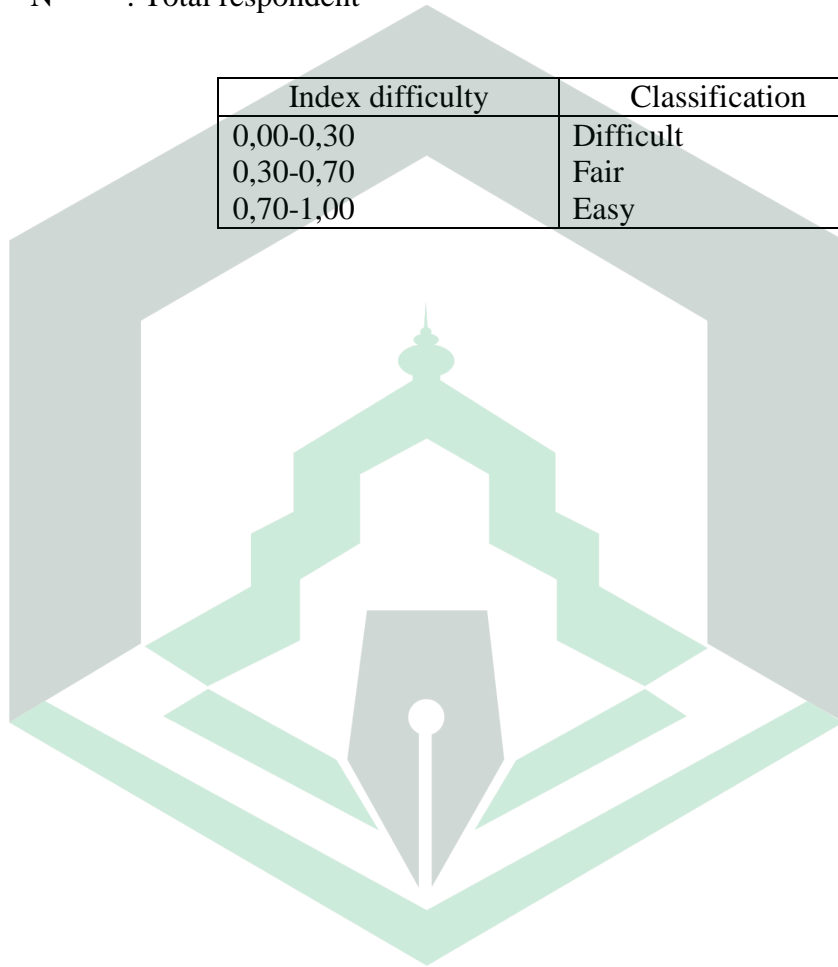
CHAPTER IV

FINDINGS AND DISCUSSIONS

This is chapter reseacrher explain about finding and discusses of the data the researcher. This chapter describes about the result of the research shows the validity test, reability test and level of difficult in multiple choice test and essay by criteria of a good test.

A. Findings

The findings of the research are showed to describe the result of the data that analayzed statistically. It comprised of the students' score in validity, reability and level of difficult in making test with criteria of a good test.

1. Validity test

To find out the validity test item made by 7th semester departement English IAIN Palopo year 2017/2018 as following presentation data (test item) used in writing test, validity test result as follows:

Table 4.1

Result Evaluating Test

Ability in writing tests	Content validity of test
multiple choice test	25/50 x 100 = 50%
Essay test	25/50 x 100 = 50%

From the table 4.1, it shows that the 50 respondents to write multiple choice test and 50 respondents write essay test. So the total respondents are 100 students.

a. Validity multiple choice of test

In this section, the researcher shows the multiple choice test aspect validity in criteria of a good test (face validity, content validity, contructive validity and predictive validity). In analyzing the items qualitatively, the use of the review form will be helpful and facilitate the implementation procedure. The problem review format is used as a basis for analyzing each item. The review format of the question is the format of the review item: description, multiple choice. In order for reviewers to easily use the problem review format, reviewers should consider the format filling instructions. The instructions are as follows.

1. Analyze each item based on all the criteria listed in the format!

2. Put a check mark (√) in the "Yes" field if the question is in accordance with the criteria!

3. Put a check mark (-) in the "No" field if the question is not in accordance with the criteria, then write the reason on the note space or on the text of the problem and the improvement
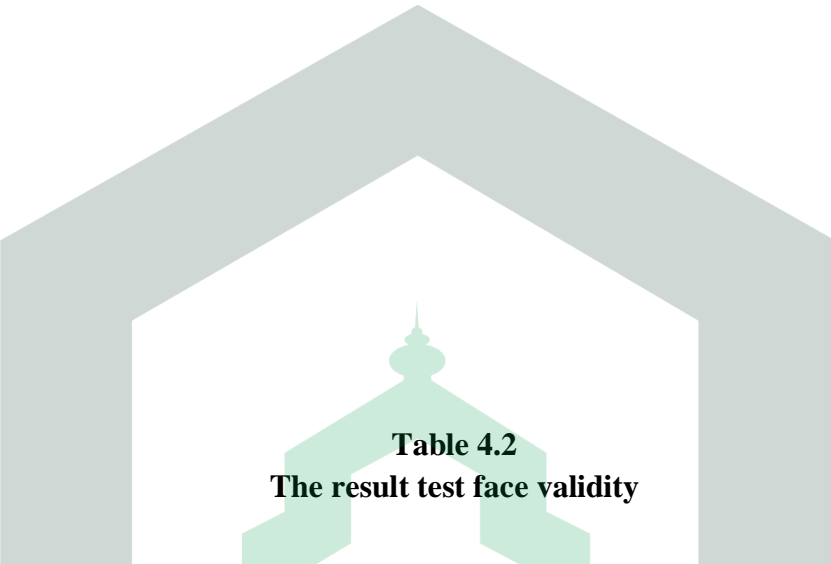
Table 4.2
The result test face validity

Respondent	No test										Total
	1	2	3	4	5	6	7	8	9	10	
1	√	√	√	-	√	√	√	√	-	√	8
2	√	√	√	√	√	-	-	√	-	√	9
3	-	-	√	√	√	-	-	√	-	-	4
4	√	√	√	√	-	√	√	√	√	√	9
5	√	√	√	√	√	-	√	-	√	-	7
6	√	√	√	√	√	-	-	-	√	√	7
7	√	√	√	√	√	-	-	√	-	-	6
8	-	√	-	√	√	√	-	-	-	√	4
9	√	√	-	√	√	√	√	-	√	√	8
10	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	√	√	-	√	√	√	√	5
12	√	√	√	√	√	√	-	√	-	-	7
13	√	√	√	√	√	√	√	√	√	√	9
14	-	√	√	√	√	√	√	√	-	-	8
15	√	√	√	√	√	√	√	√	-	-	8

No											Total
16	√	√	√	√	√	-	√	√	√	√	9
17	√	-	√	√	√	√	√	-	√	√	8
18	√	√	√	-	√	√	-	-	√	√	7
19	√	√	√	-	√	√	-	-	√	√	7
20	-	√	√	√	-	√	√	√		√	8
21	√	√	√	√	-	√	√	√	√	√	9
22	√	√	-	-	√	√	-	√	√	√	7
23	√	√	√	√	√	√	-	-	√	√	8
24	√	√	-	√	-	√	√	√	√	√	8
25	-	-	-	-	√	√	-	√	√	√	5
26	√	√	-	√	-	√	√	-	√	√	8
27	√	√	-	√	-	-	-	√	-	-	5
28	-	√	√	-	-	√	√	-	√	√	6
29	√	√	√	√	√	√	-	√	√	√	9
30	√	-	√	√	-	√	√	√	√	√	8
31	√	√	-	√	-	√	√	√	√	√	8
32	-	√	√	√	√	√	-	√	√	√	8
33	√	√	√	√	√	-	-	√	√	√	9
34	√	√	√	√	√	√	√	√	√	√	10
35	√	√	√	√	-	-	√	√	√	√	8
36	√	√	√	√	-	-	-	√	√	-	6
37	√	√	-	√	-	√	√	√	√	√	8
38	-	√	√	√	√	√	-	-	√	√	7
39	√	-	-	√	√	√	-	-	√	√	6
40	-	-	-	√	√	√	√	√	√	√	7
41	-	√	√	√	√	√	-	-	√	√	7
42	√	√	√	-	√	√	√	√	√	√	9
43	√	√	√	√	-	√	√	√	-	-	7
44	-	√	-	√	√	√	√	√	-	-	6
45	√	√	√	-	√	√	√	√	√	√	9
46	√	-	√	√	√	√	-	√	√	√	8
47	√	-	-	√	√	√	√	√	-	-	6
48	√	√	-	√	√	√	-	√	√	√	8
49	-	√	√	√	√	√	-	-	√	√	7
50	-	√	√	√	√	√	-	√	√	-	3
Total											351

From the table 4.2, it shows that the multiple choice test items made by students at sevent semester department English year 2017/2018 in IAIN Palopo according to criteria of a good test with face validity total is 351

The next result of the multiple choice test content validity. The following presents the results of validity content.

Table 4.3
The result test validity content
multiple choice

Respondent	No test										Total
	1	2	3	4	5	6	7	8	9	10	
1	√	√	√	-	√	√	√	-	√	-	7
2	√	√	√	√	√	-	√	√	√	-	8
3	-	-	√	√	√	√	√	√	√	-	7
4	√	√	√	√	-	-	√	√	√	-	7
5	√	√	√	√	√	-	-	-	√	√	7
6	√	√	√	√	√	√	√	√	√	-	9
7	√	√	√	√	√	-	√	√	√	√	9

8	-	√	-	√	√	√	√	-	√	√	7
9	√	√	-	√	√	√	-	-	√	√	7
10	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	√	√	√	√	√	√	√	7
12	√	√	√	√	-	√	√	√	√	√	9
13	√	√	√	√	√	√	-	√	√	√	9
14	-	√	√	√	√	√	-	√	√	√	8
15	√	√	√	√	-	√	√	√	√	√	9
16	√	√	√	√	√	√	√	√	-	-	8
17	√	-	√	√	√	√	√	-	√	√	8
18	√	√	√	-	√	√	√	√	√	-	8
19	√	√	√	-	√	√	-	-	√	√	7
20	-	√	√	√	√	√	-	√	√	√	8
21	√	√	√	√	-	√	√	√	√	√	9
22	√	√	-	-	√	√	√	√	-	-	6
23	√	√	√	√	-	√	√	-	-	-	6
24	√	√	-	√	-	√	√	√	√	√	8
25	-	-	-	-	√	√	-	√	√	√	5
26	√	√	-	√	√	√	-	-	√	√	7
27	√	√	-	√	-	√	√	√	√	√	8
28	-	√	√	-	√	√		-	√	√	7
29	√	√	√	√	-	√	√	√	√	√	9
30	√	-	√	√	-	√	√	√	√	√	8
31	√	√	-	√	-	√	√	√	√	√	8
32	-	√	√	√	-	√	√	√	√	√	8
33	√	√	√	√	-	√	√	√	-	-	7
34	√	√	√	√	√	√	-	√	√	√	9
35	√	√	√	√	-	√	√	√	√	√	9
36	√	√	√	√	√	√	-	√	√	√	5
37	√	√	-	√	-	√	√	√	√	√	8
38	-	√	√	√	-	√	√	√	√	-	7
39	√	-	-	√	√	√	-	-	√	√	6
40	-	-	-	√	√	√	√	√	√	√	7
41	-	√	√	√	√	√	-	-	√	√	3
42	√	√	√	-	√	√	-	√	√	√	8
43	√	√	√	√	√	√	√	√	-	-	8
44	-	√	-	√	√	√	-	√	√	√	7
45	√	√	√	-	√	√	√	√	√	√	9
46	√	-	√	√	√	√		√	√	√	9
47	√	-	-	√	√	√	√	√	√	√	8
48	√	√	-	√	-	√	√	√	-	-	6

49	-	√	√	√	√	√	√	√	√	√	9
50	-	√	√	√	√	√	-	-	√	√	7
Total											335

Based on the above presentation data, multiple choice test items made by students at sevent semester department English year 2017/2018 in IAIN Palopo by criteria of a good test with content validity is 335

The next result of the multiple choice test constsructive validity. The following presents the results of validity constsructive
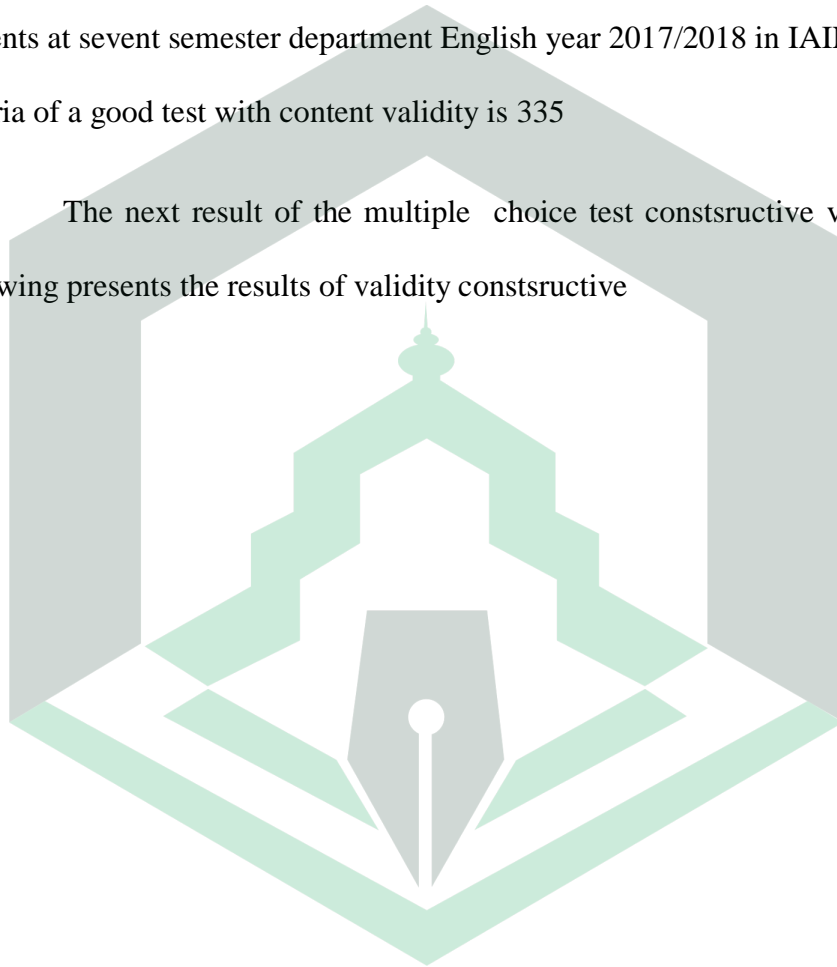
Table 4.4

The result test contsructive validity

Multiple choice

Respondent	No test										total
	1	2	3	4	5	6	7	8	9	10	
1	√	√	√	-	√	-	-	√	√	√	7
2	√	√	√	√	√	-	√	√	√	-	8
3	-	-	√	√	√	√	√	√	√	√	8
4	√	√	√	√	-	-	√	√	√	-	7
5	√	√	√	√	√	√	√	√	-	-	8
6	√	√	√	√	√	-	√	√	√	-	8
7	√	√	√	√	√	-	√	√	√	√	9
8	-	√	-	√	√	-	√	√	√	-	6
9	√	√	-	√	√	√	√	-	-	-	6
10	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	√	√	√	√	√	-	-	5
12	√	√	√	√	√	√	√	√	-	-	8
13	√	√	√	√	√	√	√	√	-	-	8
14	-	√	√	√	√	√	√	√	-	-	7
15	√	√	√	√	√	√	√	√	-	-	8
16	√	√	√	√	√	√	√	√	√	√	10
17	√	-	√	√	√	√	√	-	-	-	6
18	√	√	√	-	√	√	-	-	-	-	5
19	√	√	√	-	√	√	√	-	-	-	6
20	-	√	√	√	√	√	√	√	-	-	7
21	√	√	√	√	√	√	√	√	-	-	8
22	√	√	-	-	√	√	√	√	√	√	8
23	√	√	√	√	-	√	√	√	√	√	9
24	√	√	-	√	√	√	√	√	√	√	9
25	-	-	-	-	√	√	√	√	√	√	6
26	√	√	-	√	√	√	√	√	√	√	9
27	√	√	-	√	-	-	-	√	√	√	5
28	-	√	√	-	√	√	√	√	√	√	8

29	√	√	√	√	√	√	√	√	-	-	8
30	√	-	√	√	-	-	-	√	√	√	6
31	√	√	-	√	√	√	√	√	-	-	7
32	-	√	√	√	√	√	√	√	-	-	7
33	√	√	√	√	√	√	√	√	-	-	8
34	√	√	√	√	√	√	√	√	√	√	10
35	√	√	√	√	√	√	√	√	-	-	8
36	√	√	√	√	-	-	-	√	√	√	7
37	√	√	-	√	√	√	√	√	-	-	4
38	-	√	√	√	√	√	√	√	√	√	9
39	√	-	-	√	-	-	-	√	√	√	5
40	-	-	-	√	√	√	√	√	-	-	5
41	-	√	√	√	√	√	√	√	√	√	9
42	√	√	√	-	√	√	√	√	√	√	9
43	√	√	√	√	√	√	√	√	-	-	8
44	-	√	-	√	√	√	√	√	-	-	6
45	√	√	√	-	√	√	√	√	-	-	6
46	√	-	√	√	√	√	√	√	√	√	9
47	√	-	-	√	√	√	√	√	√	√	8
48	√	√	-	√	√	√	√	√	√	√	9
49	-	√	√	√	√	√	√	√	√	√	9
50	-	√	√	√	√	√	√	-	-	-	6
Total											377

Based on the above presentation data, multiple choice test items made by a student at sevent semester department English year 2017/2018 in IAIN Palopo according to good test criteria with constructive validity are 444

The next result of the multiple choice test predicitive validity. The following presents the results of validity predictive

Table 4.5

The result test predictive validity

Multiple choice

Respondent	No test										Total
	1	2	3	4	5	6	7	8	9	10	
1	√	√	√	-	√	√	√	√	√	√	9
2	√	√	√	√	√	-	-	√	√	-	7
3	-	-	√	√	√	√	√	√	√	-	7
4	√	√	√	√	-	√	√	√	√	-	8
5	√	√	√	√	√	√	√	-	√	√	9
6	√	√	√	√	√	√	√	√	√	-	9
7	√	√	√	√	√	-	-	-	√	√	7
8	-	√	-	√	√	-	√	√	√	√	7
9	√	√	-	√	-	-	-	-	√	√	5
10	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	√	√	√	√	√	√	√	7
12	√	√	√	√	√	√	√	√	√	√	10
13	√	√	√	√	√	√	-	√	√	√	9
14	-	√	√	√	√	√	-	√	-	-	6
15	√	√	√	√	√	√	-	√	√	√	9
16	√	√	√	√	-	-	-	√	√	√	7
17	√	-	√	√	√	√	√	√	-	-	7
18	√	√	√	-	√	√	-	-	√	√	7

19	√	√	√	-	-	√	√	-	√	√	7
20	-	√	√	√	√	√	-	√	√	√	8
21	√	√	√	√	-	-	-	√	√	√	7
22	√	√	-	-	-	-	-	√	√	√	5
23	√	√	√	√	√	√	-	-	√	√	8
24	√	√	-	√	√	√	-	√	√	√	7
25	-	-	-	-	√	√	-	√	√	√	5
26	√	√	-	√	√	√	-	-	√	√	7
27	√	√	-	√	-	√	√	√	√	√	8
28	-	√	√	-	√	√	√	√	√	√	8
29	√	√	√	√	√	√	-	√	√	√	9
30	√	-	√	√	√	√	-	√	√	√	7
31	√	√	-	√	√	√	√	√	√	√	8
32	-	√	√	√	√	√	-	√	-	-	6
33	√	√	√	√	√	√	-	√	√	√	9
34	√	√	√	√	-	√	√	√	√	√	9
35	√	√	√	√	√	√	√	√	-	-	8
36	√	√	√	√	√	√	-	√	√	√	9
37	√	√	-	√	√	√	-	√	√	√	4
38	-	√	√	√	√	√	-	-	√	√	7
39	√	-	-	√	√	√	√	√	-	-	6
40	-	-	-	√	-	√	√	√	√	√	6
41	-	√	√	√	√	√	√	√	-	-	7
42	√	√	√	-	-	√	√	√	√	√	8
43	√	√	√	√	√	√	-	√	√	√	9
44	-	√	-	√	√	√	-	√	√	√	7
45	√	√	√	-	-	√	√	√	√	√	8
46	√	-	√	√	√	√	-	√	√	√	8
47	√	-	-	√	-	√	√	-	√	√	6
48	√	√	-	√	√	√	-	√	√	√	6
49	-	√	√	√	√	√	-	-	√	√	7
50	-	√	√	√	√	√	-	√	√	-	7
Total											370

Based on the above presentation data, the essay test items made by the student at sevent semester department English year 2017/2018 in IAIN Palopo according to good test criteria with predictive validity is 370

So looking at the result score of students in making essay test' in face validity, content validity, contructive validity, and predictive validity, the researcher use table for calculates. A high correlation coefficient indicates that the test is valid. Problems tested validity feasible to use, if low then the whole test can not be used, however, to not throw all the test items it is necessary to analysis which items are valid. Calculation of grain coleration for the matter of objective form Criteria of correlation coefficient.

0,8 – 1,0 : very high

0,6 – 0,8 : high

0,4 – 0,6 : fair

0,2 – 0,4 : low

0,0 – 0,2 : very low

Table 4.6

The result test validity item multiple choice

Test validity							
respondents	Validity face	Validity content	Validity contructive	Validity predictive	total	r-items	Status
1	8	7	7	9	31	0,62	Valid
2	9	8	8	7	32	0,64	Valid

3	4	7	8	7	26	0,52	Valid
4	9	7	7	8	31	0,62	Valid
5	7	7	8	9	31	0,62	Valid
6	7	9	8	9	33	0,66	Valid
7	6	9	9	7	31	0,62	Valid
8	4	7	6	7	24	0,50	Valid
9	8	7	6	5	26	0,52	Valid
10	7	6	6	5	24	0,50	Valid
11	5	7	5	7	24	0,50	Valid
12	7	9	8	10	34	0,68	Valid
13	9	9	8	9	35	0,7	Valid
14	8	8	7	6	26	0,52	Valid
15	8	9	8	9	34	0,64	Valid
16	9	8	10	7	34	0,64	Valid
17	8	8	6	7	29	0,58	Valid
18	7	8	5	7	27	0,54	Valid
19	7	7	6	7	27	0,54	Valid
20	8	8	7	8	31	0,62	Valid
21	9	9	8	7	33	0,66	Valid
22	7	6	8	5	26	0,52	Valid
23	8	6	9	8	31	0,62	Valid
24	8	8	9	7	32	0,66	Valid
25	5	5	6	5	21	0,42	Valid
26	8	7	9	7	31	0,62	Valid
27	5	8	5	8	26	0,52	Valid
28	6	7	8	8	29	0,58	Valid
29	9	9	8	9	34	0,7	Valid
30	8	8	6	7	32	0,58	Valid
31	8	8	7	8	26	0,62	Valid
32	8	8	7	6	32	0,58	Valid
33	9	7	8	9	34	0,66	Valid
34	10	9	10	9	36	0,76	Valid
35	8	9	8	8	29	0,66	Valid
36	6	5	7	9	32	0,54	Valid
37	8	8	4	4	23	0,50	Valid
38	7	7	9	7	30	0,60	Valid

39	6	6	5	6	23	0,46	Low
40	7	7	5	6	24	0,50	Valid
41	7	3	9	7	26	0,52	Valid
42	9	8	9	8	34	0,72	Valid
43	7	8	8	9	32	0,7	Valid
44	6	7	6	7	26	0,52	Valid
45	9	9	6	8	32	0,66	Valid
46	8	9	9	8	34	0,72	Valid
47	6	8	8	6	36	0,72	Valid
48	8	6	9	6	29	0,58	Valid
49	7	9	9	7	32	0,64	Valid
50	3	7	6	7	23	0,46	Low
Total					1.345		

Based on the above presentation data, multiple choice test items made by students at sevent semester department English year 2017/2018 in IAIN Palopo fit good test criteria with face validity, content validity, contructive validity and predictive validity is 1.345

b. Validity essay of test

In this section, the researcher shows the essay test aspect validity in criteria of a good test (face validity, content validity, contructive validity and predictive validity) The first result of the multiple choice test face validity. The following presents the results of validity predictive.The first result of the essay test face validity. The following presents the results of face falidity

Table 4.7

The Result Test face Validity
Essay

Respondent	No test					Total
	1	2	3	4	5	
1	√	√	√	√	√	5
2	√	√	√	√	-	4
3	-	√	√	√	√	4
4	√	-	√	√	√	4
5	√	-	√	√	-	3
6	√	-	√	-	√	3
7	-	√	√	√	√	4
8	√	-	√	√	-	3
9	√	√	√	√	√	5
10	√	-	√	-	√	3
11	√	-	√	-	√	3
12	√	√	-	√	-	4
13	-	-	√	√	-	2
14	√	-	√	√	-	3
15	√	√	√	√	√	5
16	√	√	√	-	√	4
17	-	√	-	√	√	3
18	√	√	-	√	√	4
19	-	√	-	√	-	2
20	√	√	-	√	-	3
21	√	√	-	√	√	4
22	-	-	√	√	√	3
23	√	√	√	-	-	3
24	√	√	√	√	√	5
25	√	√	-	√	√	4
26	√	√	-	√	-	3
27	-	√	√	√	√	4
28	-	-	√	√	√	3
29	√	-	√	√	√	4
30	-	-	√	-	-	1
31	√	√	-	√	√	4
32	√	√	√	-	√	4
33	-	√	-	√	-	2
34	√	√	√	-	-	3

35	-	-	√	√	-	2
36	√	√	-	√	√	3
37	√	√	√	√	√	5
38	√	-	√	-	√	3
39	√	-	√	-	√	3
40	√	√	-	√	V	4
41	√	-	√	√	√	4
42	√	-	√	√	-	3
43	√	-	√	-	√	3
44	-	√	√	√	√	4
45	√	-	√	√	-	3
46	√	√	√	√	√	5
47	√	-	√	-	√	3
48	√	-	√	-	√	3
49	√	√	-	√	V	4
50	-	-	√	√	-	2
Total						223

Based on the above presentation data, the essay test items made by the students at sevent semester department English in 2017/2018 in IAIN Palopo according good test criteria by face validity are 223

In this section, the researcher shows the essay test aspect validity in criteria of a good test (face validity, content validity, contructive validity and predictive validity) the following result content validity essay.

Table 4.8
The result test content validity
Essay

Respondent	No test					Total
	1	2	3	4	5	
1	√	-	-	√	-	2
2	-	√	√	-	-	2
3	√	√	√	√	√	5
4	√	-	√	-	-	2
5	√	-	√	√	-	3
6	√	√	√	-	-	3
7	-	√	-	√	√	3
7	√	√	-	-	√	3
8	√	√	√	√	-	4
9	√	-	√	√	√	4
10	-	-	-	-	-	-
11	√	√	√	√	-	4
12	-	√	√	√	-	3
13	√	-	-	-	√	2
14	√	√	√	√	√	5
15	√	√	√	√	√	5
16	√	√	√	√	√	5
17	√	√	√	√	√	5
18	√	√	√	√	-	4
19	-	√	√	√	√	4
20	√	√	√	√	√	5
21	-	√	√	√	√	4
22	-	√	√	√	-	3
23	√	-	√	-	√	3
24	√	√	√	√	√	5

25	√	√	√	√	√	5
26	-	√	√	-	√	3
27	√	√	-	√	-	3
28	√	-	√	√	-	3
29	√	√	√	√	√	5
30	-	√	√	-	√	3
31	√	√	√	√	-	4
32	-	√	-	-	-	1
33	√	√	√	√	-	4
34	√	-	√	√	-	3
35	√	√	√	√	√	5
36	√	√	√	-	-	3
37	-	-	-	√	√	2
38	√	√	-	-	-	2
39	√	√	√	√	√	5
40	√	-	√	√	-	3
41						
42	√	-	√	√	-	3
43						
44	-	√	√	√	√	4
45	√	-	√	√	-	3
46	√	√	√	√	√	5
47	√	-	√	-	√	3
48	√	-	√	-	√	3
49	√	√	-	√	-	4
50	-	-	√	√	-	2
Total						225

Based on the above presentation data, the essay test items created by the students at sevent

The next result of the essaty test constructive validity. The following presents the results o

Table 4.9
The result test construct validity
Essay

Respondent	No test					Total
	1	2	3	4	5	
1	√	√	√	-	√	4
2	√	√	√	√	√	5
3	-	√	√	-	-	2
4	√	√	√	-	√	4
5	√	√	√	√	√	5
6	√	√	√	√	-	4
7	√	√	√	√	√	5
8	-	√	√	-	√	4
9	-	-	-	-	-	-
10	√	√	√	√	√	5
11	√	√	√	√	√	5
12	√	√	√	√	√	5
13	√	√	√	√	√	5
14	-	-	√	√	√	3
15	√	√	-	√	-	3
16	√	√	√	-	-	3
17	-	√	-	√	√	3
18	√	-	-	√	√	3
19	√	√	-	-	-	2
20	√	√	√	√	-	4
21	√	√	-	√	√	4
22	-	√	-	-	-	1
23	√	-	-	-	√	2
24	√	√	√	-	√	4
25	√	√	√	√	√	5
26	√	√	√	√	√	5
27	-	√	-	√	-	2
28	√	-	√	√	√	4
29	√	√	√	√	-	4
30	√	√	√	√	√	5
31	√	√	√	√	√	5
32	-	√	√	-	√	3
33	√	-	√	-	√	3

34	√	√	√	√	√	5
35	√	√	√	√	√	5
36	√	-	√	-	√	3
37	-	√	-	√	-	2
38	√	√	√	-	√	4
39	-	√	√	√	√	4
40	√	√	√	√	-	4
41	√	-	√	√	√	4
42	√	-	√	√	-	3
43	√	-	√	-	√	3
44	-	√	√	√	√	4
45	√	-	√	√	-	3
46	√	√	√	√	√	5
47	√	-	√	-	√	3
48	√	-	√	-	√	3
49	√	√	-	√	-	4
50	-	-	√	√	-	2
Total						218

Based on the above presentation data, the essay test items made by the student at sevent se[...]

The next result of the essay test predictive validitas. The following

presents the results of validity predictive

Table 4.10

The result test predictive validity
Essay

Respondent	No test					Total
	1	2	3	4	5	
1	√	√	√	√	√	5
2	-	√	√	-	√	3
3	√	-	√	√	√	4
4	-	√	√	√	√	4
5	√	√	√	-	√	4
6	-	√	√	√	√	4
7	-	√	-	-	√	2
7	√	√	√	√	-	4
8	√	-	√	√	-	3
9	√	√	√	-	√	4
10	-	-	-	-	-	-
11	√	-	-	√	√	3
12	√	√	√	√	√	5
13	√	√	√	√	√	5
14	√	√	√	√	√	5
15	-	√	√	-	√	3
16	-	√	-	-	√	2
17	√	-	-	-	√	2
18	√	-	-	-	-	1
19	√	√	√	√	√	5
20	-	√	√	√	-	3
21	√	√	-	√	-	3
22	√	√	-	√	√	4
23	-	√	√	√	-	3
24	√	√	√	-	√	4
25	√	-	√	√	√	4
26	-	√	√	√	-	3
27	√	√	√	-	√	4
28	√	√	√	-	√	4
29	-	√	√	√	√	4
30	√	-	√	-	√	3
31	√	√	√	-	√	4

32	-	√	-	√	-	2
33	√	-	-	-	-	1
34	√	√	-	√	√	4
35	-	√	√	-	-	2
36	√	-	√	-	√	3
37	√	√	-	√	-	3
38	√	√	√	√	√	5
39	-	√	-	√	-	2
40	√	-	√	√	-	3
41	√	-	√	√	√	4
42	√	-	√	√	-	3
43	√	-	√	-	√	3
44	-	√	√	√	√	4
45	√	-	√	√	-	3
46	√	√	√	√	√	5
47	√	-	√	-	√	3
48	√	-	√	-	√	3
49	√	√	-	√	-	4
50	-	-	√	√	-	2
Total						218

Based on the above presentation data, the essay test items made by the students at sevent semester department English in 2017/2018 in IAIN Palopo according to good test criteria with predictive validity are 218

So looking at the result score of students in making essay test' in face validity, content validity, contructive validity, and predictive validity, the researcher use table for calculates. The result is presents into table

Table 4.11

The result test validity item Essay

Resp	Test validity				Total
	Validity face	Validity content	Validity contructive	Validity predictive	
1	4	2	4	5	15
2	5	2	5	3	15
3	3	5	2	4	14
4	4	2	4	4	14
5	5	3	5	4	17
6	5	3	4	4	16
7	5	3	2	2	12
8	3	3	5	4	15
9	3	4	4	3	14
10	-	-	-	-	-
11	2	4	5	3	14
12	5	4	5	3	17
13	5	3	5	5	18
14	4	2	5	5	16
15	5	5	3	5	18
16	5	5	3	3	16
17	3	5	3	2	13
18	3	5	3	2	13
19	3	4	3	1	11
20	4	4	2	5	15
21	5	5	4	3	17
22	3	4	4	3	14
23	4	3	1	4	13
24	4	3	2	3	12

25	1	5	4	4	14
26	3	5	5	4	17
27	4	3	5	3	15
28	2	3	2	4	11
29	5	3	4	4	16
30	4	5	4	4	17
31	4	3	5	3	15
32	4	4	5	4	17
33	5	1	3	2	11
34	5	4	3	1	13
35	5	3	5	4	17
36	5	5	5	2	17
37	4	3	3	3	13
38	3	2	2	3	10
39	2	2	4	5	13
40	2	5	4	2	13
41	7	3	9	7	26
42	9	8	9	8	34
43	7	8	8	9	32
44	6	7	6	7	26
45	9	9	6	8	32
46	8	9	9	8	34
47	6	8	8	6	36
48	8	6	9	6	29
49	7	9	9	7	32
50	3	7	6	7	23
Total					838

By totaling the score of the students' the essay test in aspect of validity test (face validity, content validity, constructive validity and predictive validity). The essay test items made by the students at sevent semester department English year 2017/2018 in IAIN Palopo by criteria of a good test, it concludes that score total is 838

2. Reliability

Evaluation The second analysis, the authors analyze the reliability of items. Reliability refers to the stability of multiple choice test values and test description. In this study, the authors used the split-half method to determine the reliability coefficient. Of the 10 multiple choice test items and 5 essay test items that test students.

In this study, to interpret the reliability test results, the authors used the statistical formula product moment correlation product Mearson to analyze the significant between scores. Preferred test items and essays are used to represent variables X and Variable Y.Before analyzing the data, the authors present a description on the understanding table, as follows

TABLE 4.12

The Classification of Test Reability

Reliability Test	Classification
0,90 – 1,00	More Highly
0,70 – 0,89	Highly
0,50 – 0,69	Fair
0,30 – 0,49	Low
< 0,30	Very Low

Soenardi, Djiwandono, TesBahasaDalamPengajaran (Bandung : ITB, 1996), 154.

a. Reability item test multiple choice

The correlation between table for calculating the variable X = the result of the total aspects of the validity multiple choice and the Y variable the results of the total aspect of validity essay test .

Table 4.13

Correlation Between variabel x and variabel y

Respondent	X	Y	X^2	Y^2	X.Y
1	31	15	961	225	456
2	32	15	1024	225	480
3	26	14	676	196	364
4	31	14	961	196	434
5	31	17	961	289	961
6	33	16	1089	256	1188
7	31	12	961	144	372
8	24	15	576	225	360
9	26	14	676	196	364

Respondent	X	Y	X^2	Y^2	X.Y
10	-	-	-	-	-
11	24	14	576	196	336
12	34	17	1156	289	578
13	35	18	1225	324	630
14	26	16	676	256	416
15	34	18	1156	324	544
16	34	16	1156	256	568
17	29	13	841	169	377
18	27	13	729	169	351
19	27	11	729	121	297
20	31	15	961	225	456
21	33	17	1086	289	561
22	26	14	676	196	364
23	31	13	961	169	403
24	32	12	1024	144	384
25	21	14	441	196	294
26	31	17	961	289	521
27	26	15	676	225	380
28	29	11	841	121	319
29	35	16	1225	256	560

Respondent	X	Y	X^2	Y^2	X.Y
30	29	17	841	289	493
31	31	15	961	225	444
32	29	17	841	289	493
33	33	11	1086	121	363
34	38	13	1444	169	494
35	32	17	1024	544	542
36	27	17	729	544	459
37	24	13	576	169	312
38	30	10	900	100	300
39	23	13	529	169	299
40	24	13	576	169	312
41	26	26	676	676	676
42	34	34	1156	1156	1156
43	32	32	676	196	266
44	26	26	196	196	196
45	32	32	1024	1024	1024
46	34	34	1156	1156	1156
47	36	36	1296	1296	1296
48	29	29	841	841	841
49	32	32	1024	1024	1024

Respondent	X	Y	X²	Y²	X.Y
50	23	23	529	529	529
Total	1.345	838	782199	18623	260733

From score table above, the correlation between the two score can be calculated by using Pearson's product moment correlation formula bellow:

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

Notes:

r   : coefficient correlation between x and y

N   : total respondent

x   : variable x

y   : variable y

$\sum xy$: The total of x and y

$$r_{11} = \frac{2r}{1+r}$$

Notes:

r_{11} : reliabilities test

r   : equivalent product moment

      The data were analyzed based on the table of correlation above. From the table above, it was found that:

r   : coefficient correlation between X and Y

N   : 50

$\sum XY$: 260733

$\sum X$: 1.345

$\sum Y$: 838

$\sum X^2$: 782199

$\sum y^2$: 18623

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

$$= \frac{50 \cdot 260733 - (1345)(838)}{\sqrt{[50 \cdot 782199 - (782199)^2][50 \cdot 18623 - (18623)^2]}}$$

$$= \frac{13036650 - 1127110}{\sqrt{[39109950 - 611835275601][931150 - 346816129]}}$$

$$= \frac{11909540}{\sqrt{[-611853275601][-345816129]}}$$

$$= \frac{11909540}{211588731}$$

$$= 0{,}562$$

r_{11} : reliability test

r   : equivalent product moment $= 0.562$

$$r_{11} = \frac{2r}{1+r}$$

$$r11 = \frac{2 \cdot 0.562}{1 + 0.562}$$

$$r_{11} = \frac{1{,}24}{1{,}562}$$

$$r_{11} = 0{,}54$$

The result showed the correlation between two tests (x and y) is 0,54 The consistency from the test is 50 respondents. It means that the reliability test items multiple choice and essay is fair

TABLE 4.14

Level of Difficulty Classification

Index difficulty	Classification
0,00-0,30	Difficult
0,30-0,70	Fair
0,70-1,00	Easy

Safari, EvaluasiPembelajaran (DepartemenPendidikanNasional :t.t.p, 2004), 55

a. Level of difficulty items test multiple choice

So looking at the results of the scores of students in making multiple choice tests' researchers took the results of the aspect validity as the value of the frequency determines the classification test, the researcher used the table for calculation. The result is present into the table.

TABLE 4.15

Presentation of Level of Difficulty

Question	Frequency	Level of difficulty	Classification
1	31	0,51	Fair
2	32	0,53	Fair

Question	Frequency	Level of difficulty	Classification
3	26	0,43	Fair
4	31	0,51	Fair
5	31	0,51	Fair
6	33	0,55	Fair
7	31	0,51	Fair
8	24	0,4	Difficult
9	26	0,43	Fair
10	26	0,43	Fair
11	24	0,4	Diffcult
12	34	0,56	Fair
13	35	0,70	Fair
14	26	0,43	Fair
15	34	0,56	Fair
16	34	0,56	Fair
17	29	0,48	Fair
18	27	0,45	Fair
19	27	0,45	Fair
20	31	0,51	Fair
21	33	0,55	Fair
22	26	43	Fair

Question	Frequency	Level of difficulty	Classification
23	31	0,51	Fair
24	32	0,52	Fair
25	21	0,35	Fair
26	31	0,51	Fair
27	26	0,43	Fair
28	29	0,48	Fair
29	35	0,70	Easy
30	29	0,48	Fair
31	31	0,51	Fair
32	29	0,48	Fair
33	33	0,55	Fair
34	38	0,63	Fair
35	32	0,52	Fair
36	27	0,45	Fair
37	24	0,4	Difficult
38	30	0,5	Difficult
39	23	0,38	Fair
40	24	0,4	Difficult
41	26	0,43	Fair
42	34	0,56	Fair

Question	Frequency	Level of difficulty	Classification
43	32	0,52	Fair
44	26	0,43	Fair
45	32	0,52	Fair
46	34	0,56	Fair
47	36	0,72	Easy
48	29	0,21	Difficult
49	32	0,52	Fair
50	23	0,38	Fair

From the table 4.15, it shows that the multiple choice test items and essay made by students at sevent semester department English year 2017/2018 in IAIN Palopo level of difficult in making test by item test multiple choice and essay is difficult = 6, fair 41 and easy = 3 by 50 respondents

b. Level of difficulty items test essay

So looking at the results of the scores of students in making essay tests researcher took the results of the aspect validity as the value of the frequency

determines the classification test, the researcher used the table for calculation. The

result is present into the table

Table 4.16

Question	Frequency	Level of difficulty	Classification
1	15	0,3	Fair
2	15	0,3	Fair
3	14	0,28	Fair
4	14	0,28	Fair
5	17	0,34	Difficult
6	16	0,32	Difficult
7	12	0,24	Difficult
8	15	0,3	Fair
9	14	0,28	Fair
10	14	0,28	Fair
11	14	0,28	Fair
12	17	0,34	Fair
13	18	0,36	Fair
14	16	0,32	Difficult
15	18	0,36	Difficult
16	16	0,32	Difficult
17	13	0,26	Fair

Question	Frequency	Level of difficulty	Classification
18	13	0,26	Fair
19	11	0,22	Difficult
20	15	0,3	Fair
21	17	0,34	Fair
22	14	0,28	Fair
23	13	0,26	Fair
24	12	0,24	Fair
25	14	0,28	Fair
26	17	0,34	Fair
27	15	0,3	Fair
28	11	0,22	Difficult
29	16	0,32	Difficult
30	17	0,34	Fair
31	15	0,3	Fair
32	17	0,34	Fair
33	11	0,22	Difficult
34	13	0,26	Fair
35	17	0,34	Fair
36	17	0,34	Fair
37	13	0,26	Fair

Question	Frequency	Level of difficulty	Classification
38	10	0,2	Difficult
39	13	0,26	Fair
40	13	0,26	Fair
41	26	0,52	Fair
42	34	0,68	Fair
43	32	0,64	Fair
44	26	0,52	Fair
45	32	0,64	Fair
46	34	0,68	Fair
47	36	0,72	Easy
48	29	0,58	Fair
49	32	0,64	Fair
50	23	0.64	Fair

From the table 4.15, it shows that the essay test items made by students at sevent semester department English year 2017/2018 in IAIN Palopo level of difficult in making test by item essay test is difficult = 11, fair 37 and easy = 1 by 50 respondents.

B. Discussion

1. Validity of items multiple test choice and Essay test

A test can be said to have a high validity or can be declared valid, if the scores on the item concerned have a match or alignment with the total score, or with statistical language: There is a significant positive correlation between the item score with the total score. Total score here is based as dependent variable, while item score is domiciled as variable independent variable. Thus, to arrive at the conclusion that the points that want to know validity is valid or not, can be used correlation techniques as a technique of analysis. A question matter can be declared valid if the corresponding grain score proves to have a significant positive correlation with its total score.[47]

Based on the results of validity of test, note that the presentation of both test items in testing is good enough. Criteria of good test in the aspect of multiple choice test and essay test can explain some face validity of aspect, content Validity, constructive validity and predective validity.

The results of the evaluation, moreover 50% on the test know the test criteria are good. Indeed, the validity, reability and difficulty levels need to be

---

[47] Anas Sudijono, Pengantar Evaluasi Pendidikan, P. 184

known for achievement. The good test criteria in this test are just the requirements of making the test, commonly used in making tests, multiple choice tests and the essay test in this test is the type of test in making tests, important parts that need to be before practice immediately.

Based on the tests written by the students, there are several validity aspects, which need to know in making the test items multiple choice and essay.

2. Reliability of items multiple test choice and Essay test

A test can be expressed reliably if it always gives the same result when tested in the same group at different occasions. In other words, if the learners are given the same test at different times, then each learner will remain in the same rank order in the group.

In this evaluation, the author analyzes the colleraboration method with the one-time test, rather than dividing the test into odd and even items. After that compare the reliability of the two parts as if they were two separate tests. The findings of the reliability test analysis show that the computational result of the correlation between X and Y score is greater than in both the parson product moment formulas. The result reability multiple choice test and essay 0,54 is fair with 50 respondents.

3. Level of difficulty items multiple test choice and Essay test

Based on the test of each item the level of difficulty, can be used a good test. Because many test items are classified in a fair test. The evaluation findings show the level of achievement on the test specimen written by the seventh semester student. Students majoring in English in 2017/2018 at IAIN Palopo multiple choice test and essay is 8 difficult, 41 fair and 1 is easy.

CHAPTER V

CONCLUSION AND SUGGESTION

A. Conclusion

1. In evaluating students ability in writing English test in semester seven academic year of english year 2017/2018 have good enough validity in content of multiple choice of item validity that is three face validity, content validity, constructive validity, predictive validity. It is shown of the findings of identification of test items performed by the students. It is known that the representation of test items in a fair test. Based on the material in the instrument, there are several aspects, taken in the goods and some materials in the test items taken from that aspect. The validity of the content in this test is almost a valid tes Reliability in the test evaluating students' ability in the making of english writing test at sevent semester of english departement academic year 2017/2018 there is reliability on the test items. This is shown from the analysis with the formula of a person's product moment that can realiabily answer students with multiple choice and essay test is 0,54 is fair. This means that the reliability of the test items is fair.

The degree of difficulty in making tests that evaluate students' abilities in the 2017/2018 academic year exam has the following percentage of questions: multiple choice test 5 difficult, 42 fair and easy 3 by 50 respondents and item essay test is difficult = 10, fair 39 and easy = 1 by 50 respondents.

B. Suggestion

1. Students

students should study the test items and make good tests that include the validity, reliability and difficulty levels that provide students with an intensive test. This is to achieve the standard of education in Indonesia for all students.

For the English mahasiswa, especially at departement english IAIN Palopo For teachers of English, Students are expected to study the criteria of good test based on English material.

2. Researchers

For the other researchers should be analysis test items more at the certain time.

BIBLIOGRAHY

Arikunto, Suharsimi, Dasar-Dasar Evaluasi Pendidikan, (Cet. IV; Jakarta; PT.Bumi Aksara, 2002.

Arikunto, Suharsimi, Prosedur Penelitian: Suatu Pendekatan Praktik. Jakarta: PT. Rineka Cipta, 2006

Ary, Donal, Introduction to research in education. New york : CBS College Publishing,1988

Brown, H.Douglas. Teaching by Principles,An Intereractive Approach to Language Pedalogy. Tihrd Edition, San Francisco State University : copyright by Pearson Education, 2007.

Dachja, Dadang. TOEFL, TWE & IELTS Preparation. Jakarta : Puspa Swara,2005

Hamalik, Oemar, Kurikulum dan Pembelajaran, Cet. I; Jakarta: Bumi Aksara, 1995.

Heaton, B J. writing English Tests.new edition .London and New York:Longman

Rakhmat, Jalaluddin. Metode Penelitian Komunikasi. cet.1. Bandung : PT Remaja Rosdakarya, 1984.

Ruslan, Rosadi. Metode penelitian: publik relations dan komunikasi. cet.4. Jakarta : Raja Grafindo Persada, 2008

Soekartawi. monitoring dan evaluasi : proyek pendidikan, Jakarta : PT. Dunia pustaka jaya, 1995

S. Nasution, Metode Research: Penelitian Ilmiah (Jakarta: Bumi Aksara, 2006), h. 113.

Sudjana, Nana, Penilaian hasil Proses Belajar Mengajar, (Cet. III; Bandung: PT Remaja Rosdakarya, 1991.

Sudjono, Anas, Pengantar Evaluasi Pendidikan, Jakarta: PT. Raja Grafindo Persada, 2006.

Sudrajat, M. Subana. Dasar Dasar Penelitian Ilmiah Bandung : Pustaka setia, 2015

Sugiyono, Metode Penelitian Kombinasi (Mixed Methods), Bandung: ALFABETA Cet. Ke-3 2013.

Usman, Husaini dan Purnomo Setiady Akbar, Metodologi Penelitian Sosial, Jakarta, Bumi Aksara, 2009.

http://selvy-pgri.blogspot.co.id/2012/10/pengertian-definisi-tes-menurut-para.html

http://penelitian tindakan kelas .blogspot.co.id/2013/04/jenis-jenis-tes.html.

http://pengertian-pengertian-info.blogspot.co.id/2015/05/pengertian-tujuan-dan-fungsi-evaluasi.html

http://selvy-pgri.blogspot.co.id/2012/10/pengertian-definisi-tes-menurut-para.html