

Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## CHAPTER II

### REVIEW OF RELATED LITERATURE

#### A. Theoretical Concept

##### 1. The Language Assessment and Testing

###### a. Definition of Language Assessment

Assessment is one of the parts which include the learning and teaching process. The assessment also became the fundamental of teaching where assessment can measure the capability, performance, manner and also all of the thing that teacher needs to teach the students to get the objectives of the learning process. In according to Mihai (2010, p. 24) stated that Assessment is a combination of all formal and informal judgments and findings that occur inside and outside a classroom. Accordance with Brown (2003, p. 4) explained that the assessment is happening in the ongoing process where the process is measuring the ability or performance of students. The assessment is also using the tools to find the answer or result about the students that happen when they learn and the teacher teaches them. Related to the explanation above Mihai (2010, p. 22) said that Bachman, 1990 explained that three fundamentals concept of assessing students performance in social and scientific context, they are measurement, tests, and evaluation. Measurement is talking about the process of

## Hak Cipta Diindungi Undang-Undang

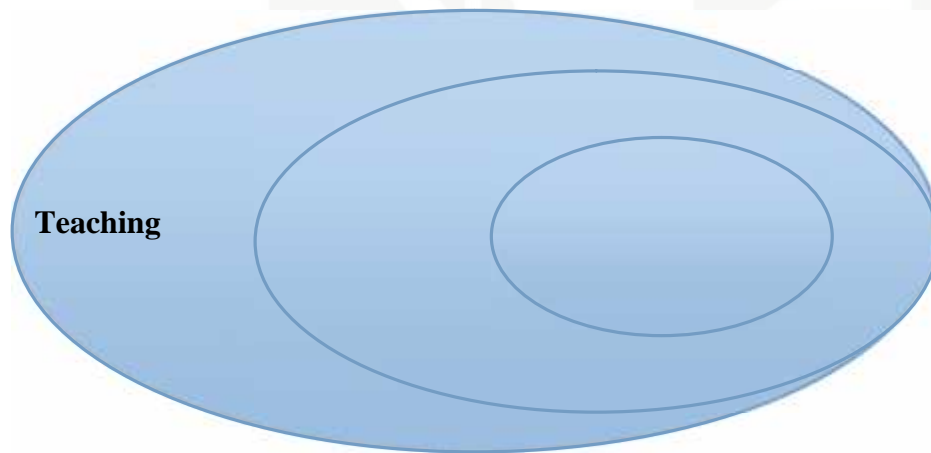
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

quantifying the characteristics of persons according to explicit procedures and rules. The test is a measurement instrument designed to elicit a specific sample of an individual behavior. And also evaluation is reliable and relevant information.

This case, the point of this is clearly discussed by the expert. The assessment is the to assess the students' performance in the class. How to assess and measure, it can be by using test. Brown (2003, p. 5) imaged that the relation of teaching, assessing and testing as in figure II.1:

**Figure II.1.**

**Teaching, Assessment, and Tests.**



Based on this figure, the relationship is very clearly explained by brown. Teaching is the way to give the information when the students

#### Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

learn in the classroom. The assessment we use to find and measure the ability of the students in the process and also the ending of the learning process. And the tools to measure the ability is the tests.

#### b. Types of Assessment

Brown (2003, p. 5-7) discussed the types of assessment, there is three kinds of the assessment:

##### 1) Formal and Informal Assessment.

Formal assessments are exercises or procedures specifically designed to tap into a storehouse of skills and knowledge. They are systematic, planned sampling techniques constructed to give teacher and student an appraisal of student achievement. To extend the tennis analogy, formal assessments are the tournament games that occur periodically in the course of a regimen of practice. Informal assessment can Like a number of forms, starting with incidental, unplanned comments and responses, along with coaching and other impromptu feedback to the student. A systematic set of observations of a student's frequency of oral participation in class is certainly a formal assessment, but it too is hardly what anyone would call a test. Tests are usually relatively time-constrained (usually spanning a class period or at most several hours) and draw on a limited sample of behavior.

#### Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Informal assessment is embedded in classroom tasks designed to elicit performance without recording results and making fixed judgments about a student's competence.

## 2) Formative and Summative Assessment

Formative assessment is evaluating students in the process of forming their competencies and skills with the goal of helping them to continue that growth process. The key such formation is the delivery (by the teacher) and internalization (by the student) of appropriate feedback on performance, with an eye toward the future continuation (or formation) of learning. For 311 practical purposes, virtually all kinds of informal assessment are (or should be) formative. They have as their primary focus the on-going development of the learner's language. So when you give a student a comment or a suggestion or call attention to an error, that feedback is offered in order to improve the learner's language ability.

Brown (2003, p. 6) stated that Summative assessment aims to measure or summarize, what a student has grasped, and typically occurs at the end of a course or unit of instruction. A summation of what a student has learned implies looking back and taking stock of how well that student has accomplished objectives but does not necessarily point the

#### Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

way to future progress. Final exams in a course and general proficiency exams are examples of summative assessment. One of the problems with prevailing attitudes toward testing is the view that all tests (quizzes, periodic review tests, midterm exams, etc.) are summative. At various points in your past educational experiences, no doubt you've considered such tests as summative. You may have thought.

### 3) Norm-Referenced Test and Criterion-References Test

Norm-referenced tests, each test-taker's score is interpreted in relation to a mean (average score), median (middle score), standard deviation (extent of variance in scores), and/or percentile rank. The purpose of such tests is to place test takers along a mathematical continuum in rank order. Scores are usually reposed back to the test-taker in the form of a numerical score (for example, 230 out of 300) and a percentile rank (such as 84 percent, which means that the test-takers score was higher than 84 percent of the total number of test-takers but lower than 16 percent in that administration). Typical of norm-referenced tests are standardized tests, intended to be administered to large audiences, with results efficiently disseminated to test-takers. Such tests must have fixed, predetermined responses in a format that can be scored

## Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

quickly at minimum expense. Money and efficiency are primary concerns in these tests.

Criterion-referenced tests are designed to give test-takers feedback, usually in the form of grades, on a specific course or lesson objectives Classroom tests involving the students in only one class, and connected to a curriculum, are typical of criterion-referenced testing. Here, much time and effort on the part of the teacher (test administrator) are sometimes required in order to deliver useful, appropriate feedback to students In a criterion-referenced test, the distribution of students' scores across a continuum may be of little concern as long as the instrument assesses appropriate objectives. In language assessment, with an audience of classroom language teachers and teachers in training, and with its emphasis on classroom-based assessment (as opposed to standardized, large-scale testing), criterion-referenced testing is of more prominent interest than norm-referenced testing.

Mihai (2010, p. 26-36) explains that the assessment is divided into four categories, there are:

**a) Intention**

In intention, the category is an assessment can occur spontaneously or it can be planned well in advance. there

**Hak Cipta Diindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

are 2 kinds of assessment that we can use, they are formal and informal assessments. Formal assessment is Formal assessment is generally accompanied by a numerical score that reflects how well the students understood the academic content that was presented in class. Formal assessment usually consists of tests, quizzes, or term papers, content area teachers may also consider portfolios or journals as more comprehensive ways of formally assessing content objectives. Besides, when it occurs without much planning, it is an informal assessment. The definition of informal assessment is including any type of ad hoc reactive feedback to students' performance in the classroom. It does not contribute directly to the final grade that students earn.

**b) Purpose**

Purpose category is an assessment can focus on the process of learning rather than on the final product, which is usually an end-of-the-semester exam. There are two types, these are formative and summative. Formative assessment is conducted throughout a course and is used to evaluate students in the process of acquiring their competencies in the academic area or in language proficiency. The goal is to provide information on how

**Hak Cipta Diindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

well the students have mastered the content so far, thus helping the process of instruction. And also, Summative assessment's purpose, on the other hand, is to measure what the students have learned at the end of a course. The focus here is not on the process of learning but rather on its product. Summative assessment can take the form of a final exam at the end of a course, or it can be a standardized test that is administered state-wide or nation-wide. The students are assigned a grade, and important educational decisions are often based on the results of these product-oriented assessment procedures.

**c) Interpretation**

Interpretation category is the assessment that we use to every students' score is compared to the scores of a given group. Consist of two types, they are first is norm-referenced assessments are first administered to the norm group, and the norms, or the results of this group's performance, are used as reference points for interpreting the performance of the students who take the test in the future. The norms are the mean (the average score), the median (the middle score), the mode (the most common score), the standard deviation



**Hak Cipta Diindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

(how varied and spread out the group scores are), and the percentile rank (an indication of how well a particular test taker has performed compared to other group members. Second is a criterion-referenced assessment. The criterion-referenced assessment takes a completely different approach in terms of interpretation and design. Criterion-referenced tests and measurements are designed with a focus on the mastery of the course content. Thus, whereas norm-referenced tests evaluate students in terms of their ranking to one another, criterion-referenced tests evaluate students in terms of their mastery of course content.

**d) Administration**

In terms of administration, assessment can take place in the classroom or can be administered in the form of state or national examinations. Consisted of two types, those are Classroom-based assessment is created and given primarily by teachers; it is formative and occurs on a regular basis. It is designed and scored by teachers who shape the assessment to the context of their classroom in order to gather information on the effectiveness of their instruction. Classroom-based assessment can be formal

or informal, and it employs a variety of formats for student response. Its instruments include, but are not limited to, tests that are administered at the end of a unit, student portfolios, teacher checklists and narratives, class projects, and student (self- and peer) evaluations. And second is large-scale assessments. Large-scale assessments are usually created by external agencies and take the form of standardized tests. These instruments are summative in nature and aim to measure students' performance at the end of an instructional cycle. Because they are administered to a large number of students, the format is limited to tests in which a large majority of the questions are multiple choice and short answer. Large-scale assessments include state and national tests, standard portfolios, tests administered at the end of the course, and district norm-referenced tests.

## 2. Multiple-choice Test

Multiple-choice Question test is the simplest test technique commonly used by test-makers. It can be used any condition and situation, in any level or degree of education. Actually, its simplicity relies on its scoring and answering. It is supported by Hughes (2005, p. 75) who states the most obvious advantage of multiple-choice is that scoring can be perfectly reliable. In line with Hughes, scoring in multiple-choice techniques is

### Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## Hak Cipta Diindungi Undang-Undang

1. Diarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

rapid and economical. And it is designed to elicit specific responses from the student. Yet, designing multiple-choice question is more complicated than essay items. According to Brown (2003, p. 55) stated that multiple-choice items which may appear to be the simplest kind of item to construct are extremely difficult to design correctly. Hughes (2005, pp.75-78) said that multiple-choice items take many forms, but their basic structure is that it has stems or the question itself, and a number of options one which is correct, the others being distractors. In another case, Hughes states a number of weaknesses of multiple-choice items.

Multiple-choice questions only recognition of knowledge. They make test takers can only guess to come with a correct answer, and cheat easily. The technique severely restricts what can be tested. It is very difficult to write successful items and the answer is restricted by the optional answer. In this case, test-takers cannot elaborate their answer and understanding of the material because the answer is limited only by an optional answer. Multiple-choice comes to be the first part of test packs faced by test-takers. When we want to analyze this item we can use statistical analysis as stated in the next chapter. Since there is only one right answer, the score can very rapidly mark an item as correct and incorrect. Thus, we can use simple codes to present the answer of test-takers. Score 1 presents correct answer chosen by students, and 0 presents the wrong answer. If students choose a correct answer, we can note it by 1. And vice versa, if test-takers answer with the wrong answer we note it with number 0.

## Hak Cipta Diindungi Undang-Undang

1. Diarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 3. Validity

Mihai (2010, p. 37) said that The American Psychological Association at 1985 defines validity as the extent to which the inferences or educational decisions that are made based on test scores are meaningful, appropriate, and useful. Test scores themselves thus have to be meaningful. That is, they must reflect a student's individual ability in a particular area and make sure that the test does not measure anything else. Therefore, when determining how meaningful test scores are, test designers should be able to demonstrate that the results that are obtained by the students who are taking the test are not affected by factors external to the testing instrument itself. When test scores are affected by errors of measurement, they are not valid and cannot provide any solid evidence of a student's ability in a certain academic area.

In according to Bachman (2004, p.259) explains that validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and action based on test scores or other modes of assessment. In according to Brown (2003, pp. 23-24) said that The expert should look into whether the test content is representative of the skills that are supposed to be measured. This involves looking into the consistency between the syllabus content, the test objective, and the test contents. If the test contents cover the test objectives, which in turn are representatives of the syllabus, it could be said that the test possesses content validity.

## Hak Cipta Diindungi Undang-Undang

1. Diarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hughes (2005, p.26) also stated that a test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc which it is meant to be concerned.

Relation with Brown (2003, p.59) explained that a good test is a test which is not too easy or vice versa too difficult to students. It should give an optional answer that can be chosen by students and not too far by the key answer. Very easy items are to build in some effective feelings of “success” among lower ability students and to serve as warm-up items, and very difficult items can provide a challenge to the highest-ability students. It makes students know and record the characteristics of teacher’s test if the test is given always comes to them too easy and difficult. Thus, the test should be standard and fulfill the characteristics of a good test. Mehren and Lehmen (1984) point out that the concept of difficulty or the decision of how difficult the test should depend on various factors, notably 1) the purpose of the test, 2) ability level of the students, and 3) the age of grade.

#### 4. Item Difficulty Analysis

Item analysis refers to a statistical technique that helps instructors identify the effectiveness of their test items. In developing quality assessment and specifically effective multiple-choice test items, item analysis plays an important role in contributing to the fairness of the test along with identifying content areas that maybe problematic for students. itemanalysis can help in identifying potential mistakes in scoring, ambiguous items, and alternatives(distractors) that don’t work. When



#### Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

performing item analysis, we are analyzing the difficulty level of items in the test.

According to Shih (2010, p. 60) said that item analysis is to identify the appropriateness of item with regard to the difficulty levels, the power of discrimination and the function of distractors. Besides, item analysis helped examines particular objectives of learning which the students haven't achieved properly and seek to review and re-teach. It meant that indicated the proportion of students who got the item right. A high percentage indicates an easy item/question and a low percentage indicates a difficult item.

In according to Susanti, Tokunaga, Nishikawa, and Obari (2017, p. 4) said that item difficulty or the difficulty index is defined as the proportion of test takers who correctly answered a question item. Thus item difficulty can be estimated only after administering the item to the test takers. The item difficulty can then be used to determine whether a certain item is appropriate (not too easy nor too difficult) for a group with a certain ability. So, the writer can conclude that the function of item difficulty is to measure the level of validity the test, where the test will be examines to the students. It is because the test should be suitable to students when the test becomes the measurement tools to measure the ability of the students.

National Education Ministry of Indonesia (2008, p. 12 ) had made the Item Analysis Guidance. It explained that item difficulty or difficulty item

#### Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

analysis has 2 purpose, that are the purpose to the teacher and purpose to the examining and teaching. The purpose to the teacher is 1). As the concept introduction to the students about the remedial or additional lesson and give the evaluation to the students about the result that they learned before, 2). Take the information about the obstacles of curriculum objectives and investigate the item bias. Then, the purpose to the examining and teaching process are to give the concept introduction to the students that are necessarily to conduct the remedial class or additional lesson, to give the signals about the school's curriculum evaluation, to give the suggestion and evaluation to the students, the signals that give the information about the item bias probability, and to construct the test that has the question data accurateness (cite from Nitko, 1996, pp. 310-313).

Then, also discussed that if the items are in difficult category, predictions about the information are

- a) The items have wrong key answer, the item has 2 or more true answers.
- b) The materials that become the indicators of test are not taught by teacher in teaching and learning process. Therefore, the students cannot achieve the minimum competency.
- c) The materials are not suitable to the students.
- d) The questions and sentences that the test used are very complex and longer.

## Hak Cipta Diindungi Undang-Undang

1. Diarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

It means that if the teacher wants to know about the material that the students are still difficult. The teacher has to used the item difficulty analysis to find out the difficult items and after that analyse the items again and adjust to the components or indicators of the items' test. The components is presented the material, if the components of the difficult items, it means that the components refer to the difficult material.

To find out the difficult item, we have to used the item difficulty analysis. The way to measure the difficulty items just look the percentages of the correct answer. The number that shows difficulty or easiness of test items is known as difficulty index or level of difficulty. In according to Brown (2004,p.59) the formula used to measure it is:

$$IF = \frac{CI}{TI}$$

Where:

IF = Item Facility (Level of difficulty)

CI = number of test-takers answering the item correctly

(Correct Items)

TI = number of test-takers responding to that item (Total Items)



## Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

And also, in according to Evroro and Sylvanus (2015, p.21), uttered that to know the percentage of the difficulty level of the test we use the formula as follows:

$$P = \frac{R}{N} \times 100$$

Where:

P = the percentage of examinees who answered items correctly.

R = the number of examinees who answered items correctly.

N = total number of examinees who tried the items.

Classifications of level difficulty of items that are used by Evroro and Sylvanus (2015, p.21) are:

$P = 0, 00$ : test items are too difficult

$0, 00 < P \leq 0, 30$  or  $0, 00 - 0, 30$ : test items are difficult

$0, 30 < P \leq 0, 70$  or  $0, 31 - 0, 70$ : test items are medium

$0, 70 < P \leq 1, 00$  or  $0, 71 - 1, 00$ : test items are easy

$P = 1$ : test items are too easy.

## Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 5. National Examination

Based on BSNP (2015) declares that the School / Madrasah exam / Equivalency Education hereinafter referred to as S / M / PK Exam is the activity of measuring and assessing the competence of learners against the competency standard of graduates for all subjects conducted by the Education Unit. The National Examination hereinafter referred to as the UN is the activity of measuring the achievement of graduate competency in certain subjects nationally with reference to Graduates Competency Standards. Graduation criteria is a requirement of minimum achievement of graduate competency in all subjects to be stated from the education unit.

### a) General requirements of UN participants as follows.

- 1) Learners have or have been in the last year at an educational level in a particular educational unit;
- 2) Learners have a complete report on the assessment of learning outcomes at a certain level of education in a particular educational unit from the first semester of the first year to the first semester of the year;
- 3) Learners have a full report on the assessment of learning outcomes in Equivalency Education; and
- 4) Learners have not met the criteria of achieving the competence of graduates.

**Hak Cipta Diindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**b) Grid UN**

- 1) UN grid year 2015/2016 prepared based on the criteria of the achievement of the competence of graduates, content standards, and material scope of the applicable curriculum.
- 2) The UN grid contains the cognitive level and the scope of the material.

**c) Criteria for the achievement of graduate competency based on UN result.**

The value of un results is reported in the range of values 0 (zero) to 100 (one hundred), with the level of achievement of graduate competencies in the following categories:

- 1) Very good, if the value is more than 85 (eightyfive) and less than or equal to 100 (one hundred);
- 2) Well, if the value is more than 70 (seventy) and less than or equal to 85 (eightyfive);
- 3) Sufficient, if the value is more than 55 (fiftyfive) and less than or equal to 70 (seventy); and
- 4) Less, if the value is less than or equal to 55 (fiftyfive).

**d) The Components of the Listening Section of English National Examination**

**Hak Cipta Diindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

The National Examination have regulations that control and organize the distribution and significance of the test. The BSNP is the department that controls and organizes it. in every year BSNP always makes the regulation about the component of the test. In 2014/2015 the framework of the components still same in the years before. But in the 2015/2016 the BSNP make the new framework of the component of the national examination test. The new component is also used by the 2016/2017.

Accordance with BSNP (2016 and 2017) imaged that the components of English National Examination for the academic years 2015/2016 and 2016/2017 are same. So, the table of the components of English National examination imaged into two tables, there are:

Table II. 1

**The Components of the Listening Section of English National Examination for the Academic Year 2015/2016**

No	Cognitive Level	Social Function	Text Structure	Linguistic Element
1	Knowledge and understanding - identify	Students can: - identify topics / goals / backgrounds back / the overall text reason for short functional text (announcement, letter, short news (news), biography, procedure) and text essays (recount, narrative, report, analytical exposition, new items, discussion).	Students can identify steps Work / sequence / plot / plot of Events / events / actions within Text.	Students can: - identify phrases / words that Has the same meaning in Context. - determine which phrases/words Right to complete the text/dialog. - determining the grammatical form right in context
2	Application a. Classify b. Specify c. Apply	Students can: - implement the detailed function of each step / tools / events / sections / aspects which is mentioned in the text.	Students can: - describes the linkagbetween steps one with the other - determine the exact composition of the sentence - using exact phrases in text	
3	Reasoning a) Conclude b) Analyze		Students can: - concluding text read - analyze ideas - analyze the general picture	

Table II. 2

## The Components of the Listening Section of English National Examination for the Academic Year 2016/2017

No	Cognitive level	Social function	Text structure	Linguistic element
1	Knowledge and understanding - identify	Students can identify aspects of social function: - topic / issue / issue - destination / function / message - background / reason - impact / impact / benefit - attitude / value carried - the role and function of the speaker / author - the role and function of the listener / reader - context of use	Students can identify meaning Between sections in the text: - views, intentions, opinions Become the main idea - details of arguments, steps, events, descriptions - plot, mind flow - reference meaning	Students can identify Related language elements With oral and written text content The following: - synonym - word order - articles, demonstrative, - Possessive pronoun - agreement and number - tense - passive voice - grammatical reference - conjunctions - prepositions
2	Application a. compare b. classify c. explain	Students can compare Classifying, explaining aspects of social functions: - topic / issue / issue - destination / function / message - background / reason - impact / impact / benefit - attitude / value carried - the role and function of the speaker / author - the role and function of the listener / reader - context of use (place a, Time, situation, etc.)	Students can compare. Classify, explain The interconnectedness of meanings between parts In the text: - views, intentions, opinions Become the main idea - details of arguments, steps, events, descriptions - plot, mind flow - reference meaning	Students can compare, Identify, explain Related language elements With the following text content: - synonym - word order - articles, demonstrative, - Possessive pronoun - agreement and number - tense - passive voice - grammatical reference - conjunctions - prepositions

## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip, sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, p  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa

			<ul style="list-style-type: none"> <li>- modal</li> <li>- conditional sentences</li> <li>- derivative construction</li> </ul>
<p>3</p> <p>Reasoning  a) summing up  b) detailing differences / Equation  c) analyze</p>	<p>Students can infer, detail  Difference / equation, analyze  Aspects of social function:  - topic / issue / issue  - destination / function / message  - background / reason  - impact / impact / benefit  - attitude / value carried  - the role and function of the speaker / author  - the role and function of the listener / reader  - context of use</p>	<p>Students can infer, detail  Differences/equations of relevance  Meaning between the inner parts  Text:  - views, intentions, opinions  M the main idea  - details of arguments, steps, events, descriptions  - plot, mind flow  - reference meaning</p>	<p>Students can compare,  Identify, explain  Related language elements  With the following text content:  - synonym  - word order  - articles, demonstrative,  - Possessive pronoun  - agreement and number  - tense  - passive voice  - grammatical reference  - conjunctions  - prepositions  - modal  - conditional sentences  - derivative construction  - sentences, phrases, phrases  -Conclusion</p>

#### Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

This component are too general, so to make it be specifically the writer used the components of English National examination in academic year 2014/2015. It caused the formation of the test items for the listening section is same. In the level of cognitive of the components are also same based on the Bloom's taxonomy. The level of listening Section of English National Examination is in the second level of cognitive. Cognitive level two in bloom is comprehension.

In according to bloom (1959, pp. 204-205) said that comprehension refers to a type of understanding or apprehension such that the individual knows what is being communicated and can make use of the material or idea being communicated without necessarily relating it to other material or seeing its fullest implications. This is divided in three, they are

- 1) Translation
- 2) Comprehension as evidenced by the care and accuracy with which the communication is paraphrased or rendered from one language or form of communication to another. Translation is judged on the basis of faithfulness and accuracy, that is on the extent to which the material in the original communication is preserved although the form of the communication has been altered. "The ability to



## Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

understand non-literal statements (metaphor, symbolism, irony, exaggeration). Skill in translating mathematical verbal material into symbolic statements and vice versa.

## 3) Interpretation

- 4) *The explanation or summarization of a communication. Whereas translation involves an objective part-for-part rendering of a communication, interpretation involves a reordering, rearrangement, or a new view of the material. The ability to grasp the thought of the work as a whole at any desired level of generality. The ability to interpret various types of social data.*

## 5) Extrapolation

- 6) *The extension of trends or tendencies beyond the given data to determine implications, consequences, corollaries, effects, etc., which are in accordance with the conditions described in the original communication.*

Therefore, the writer uses the component of Listening Section of English National Examination in academic year 2014/2015 because of the level of cognitive for the components are same. The verb that the component of Listening Section of English National Examination in academic year 2014/2015 components

## Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

used is “understanding” that refers to the cognitive level 2 and for the all of the indicators are using determine that refer to the extrapolation in the kinds of comprehension.

Furthermore, for the component of Listening Section of English National Examination in academic years 2015/2016 and 2016/2017 there are three levels of cognitive, these are comprehension, application, and analysis. They refer to the cognitive level 2, 3 and 4. Because of the form of the test items of Listening Section of English National Examination in every year are same. And the level of all of the tests is also same. So, the writer takes the component of Listening Section of English National Examination in academic year 2014/2015 as the specific indicators that become component of this test.

**Table II. 3**

**The component of Listening Section of English National Examination  
in academic year 2014/2015**

Competency	Indicators
LISTENING Understanding the meaning of interpersonal or transactional speech discourse formally or informally in the context of everyday life, especially in the form of short functional text, recount, news item, report, narrative, descriptive and review.	Determine the general description or specific/detailed information of a formal or informal interpersonal/transactional conversation.
	Determine the appropriate response to formal or informal transactional/interpersonal conversations.
	Determine the right image according to the information in formal or informal interpersonal/transactional conversations.
	Define an image that matches the monologue text that is played.
	Determine the general or specific / implied / detailed information of a monologue text that is played.

## Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## B. Relevant Research

There are two relevant researches which has relevance to this research. There are:

*a. Analysis of Difficulty Level of Physics National Examination's Questions was conducted by yusrizal in 2016.*

This study aimed to determine: (1) the difficulty level of items in physics National Exam of 2013 (2) physics materials that were difficult and very difficult. The subjects were all students of science major in third grade at SMAN Banda Aceh in the academic year of 2013/2014. The samples were 10 randomly selected senior high schools. The data were obtained through analyzing the answers of physics National Exam 2013. The results showed that (1) the high school students in Banda Aceh experienced difficult and very difficult level questions to be answer in the 2013 exam, (2) the difficult materials were: free fall, the potential energy and series of obstacles. The very difficult materials were: the rotational motion, motion and force on the pulley, effort, fluid, sound intensity, transformer, atomic theory, quantum theory, relativity, fusion and radio isotopes. Based on the explanation from Yusrizal, he has the similarity with this research, which is wanting to find the difficult items and also materials. In the other hand, the comparison between his research and writer's research is in the subject of the lesson. He is on Physics and the writer is on the English.

## Hak Cipta Diindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

*b. Combining Multiple Regression and CART to Understand Difficulty in Second Language Reading and Listening Comprehension Test Items by André A. Rupp & Paula Garcia and Joan Jamieson (2014),*

Sources of item difficulty allow test developers to better define the constructs that they are testing by providing empirical evidence. Past research has explored item difficulty of reading and listening comprehension items using multiple regression and classification and regression tree (CART) analyses (e.g., Freedle & Kostin, 1993, 1996, 1999; Sheehan, 1997); First, item difficulty was modelled as a function of 12 text and item and text interaction predictor variables in a multiple linear regression models. Seven of the 12 variables in the model accounted for about 31% of the variance in item difficulty as measured by the adjusted R<sup>2</sup>. Second, the data were used to build multiple regression tree models using a CART, a nonparametric technique that uncovers linear dependencies among predictor variables. Seven variables, some of which were not identified in the regression analysis, were relatively important across all trees. The researcher found that synthesizing results from the 2 methodological perspectives provided a richer picture of the interrelations of variables that affect item difficulty, and provided some empirical support for our construct definitions.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### C. Operational Concept

Operational concept was a concept as a guide used to avoid misunderstanding. It should be interpreted into a particular word in order to make it easy. In this research, there was focused to measure the difficulty level of the listening section on English National Examination at the Senior High School level and also to find the components that the students were not able to answer the test correctly or the students difficulty items.

To measure each main focus, the researcher is able to identify them in some component or indicators as follows:

**Table II. 4**

#### The Components of the Listening Section of English National Examination

Competency	Indicators
LISTENING Understand the meaning of interpersonal or transactional speech discourse formally or informally in the context of everyday life, especially in the form of short functional text, recount, news item, report, narrative, descriptive and review.	Determine the general description or specific/detailed information of a formal or informal interpersonal/transactional conversation.
	Determine the appropriate response to formal or informal transactional/interpersonal conversations.
	Determine the right image according to the information in formal or informal interpersonal/transactional conversations.
	Define an image that matches the monologue text that is played.
	Determine the general or specific / implied / detailed information of a monologue text that is played.