

## CHAPTER I

### INTRODUCTION

#### A. BACKGROUND OF STUDY

Evaluation activities has an important role in education, as well as in the learning process. It can be known the results of learning activities that have been conducted and can be determined to be done properly. Allah SWT mention the importance of the evaluation is in Al Hujurat verse 6 as follows :

نَفَعَلْتُمْ مَا عَلَىٰ تَصْبِيحُوا أَجْهَلًا قَوْمًا تُصِيبُوا أَنْ فَتَبِينُوا بِنَبِيٍّ فَاسْقُوا جَاءَ كُمْ إِنْ ءَامَنُوا الَّذِينَ يَتَأْتِيهَا

﴿نَدِيمٍ﴾

From the verse above, the God wants us to evaluate something we do not know or any news from someone or people we do not know. We should analyze it accurately in order to prevent people get hurt because of something we do not know.

Evaluation involves the interpretation of information. It is involved when the results of a test (or other assessment procedure) are used for decision making (Brown, 2010, p. 5).

Assessment is the basis to get a reversal to improve learning process and learning system as a whole that the implementation should be done on an going both to process itself and the result achieved. Assessment process is performed by observation of students activities in

learning through observation guide. While the assessment results can be done with certain techniques whether test or non test.

Success or failure of a teaching-learning process can be identified by conducted an assessment of student' learning result. Although , there is a rumor that says that the main factor why students are success in their examination is because there is a help from their teachers in answering the items. Also the test should be appropriate with the syllabus or curriculum that order by the government. So, to make the English test effective, the test should provide the good items.

The results of Regional Try Out Islamic Senior High School In Social Department Academic Year 2012-2013 At East Barito Regency,Central Borneo is very different 90 % more with the result of National Examination Academic Year 2012-2013, especially in English Subject area. There is no student of Islamic Senior High School at East Barito Regency passed the Regional Try Out, but All of them Passed the National Examination. Of Course, there is a big question, why?

Therefore, writer is interested to do research in analysis the items of Try Out. Not only because it is important to know the quality of the test items in order to get the correct and accurate data but also to check the rumor that spread widely around people. Are the test questions very difficult for students? Are the English test questions very easy? How is the quality? Because, if the test questions are very easy it is impossible for the teacher to help their students. But, if the test questions are very difficult it

is impossible. So, the writer intends to take the title: *“The Items Analysis of the Regional English Try Out Questions of Social Department of Islamic Senior High Schools at East Barito Regency, Central Kalimantan Academic Year 2012/2013”*

#### **B. Statement of Problems**

1. How is the quality of Regional English Try Out Questions of Social Department of Islamic Senior High Schools at East Barito Regency, Central Kalimantan Academic Year 2012/2013?
2. Are the Regional English Try Out Questions of Social Department of Islamic Senior High Schools at East Barito Regency, Central Kalimantan Academic Year 2012/2013 appropriate with the Syllabus or Curriculum?

#### **C. Objective of Study**

Objective of this research are :

1. To know about the quality of Regional English Try Out Questions of Social Department of Islamic Senior High Schools at East Barito Regency, Central Kalimantan Academic Year 2012/2013
2. To know about the curriculum and syllabus that used in Islamic Senior High School at Barito Timur Regency

#### **D. Reasons for title**

There are some reasons for the writer choosing the title. They are :

1. To analyze the English Test Questions Of Regional Try Out Islamic Senior High Schools In Social Department Academic Year

2012-2013 At East Barito Regency, Central Borneo. Are the Items appropriate with the Curriculum or syllabus? is the quality of the items good?

2. There is a rumor that says that the main reason why students at Senior Islamic High School at East Barito Regency pass their national examination is because they helped by their teachers.

#### **E. Significance of Research**

1. Theoretically, as an information and improving the knowledge about designing the good English items.
2. Practically, as a good experience about how to design English items well.

#### **F. Definition of Key Terms**

In order to avoid any misinterpretation toward the title adopted in this research, the writer will explain some terms based on Oxford Advanced Learner's Dictionary (2000) in the title as follow:

1. Analysis is the detailed study or examination of something in order to understand more about it.
2. Try Out is an act of testing how good or effective somebody or something is before deciding whether to use them in the future.
3. Islamic Senior High School is a Senior High School under Religion Ministry Indonesian Republic

## **CHAPTER II**

### **THEORETICAL FRAMEWORK**

Making judgments and evaluations is an integral part of everyday life (even toward the end of life if we believe the quote above). People constantly evaluate all aspects of their life and work, to such an extent that we rarely notice that we are doing so.

Gronlund (1981, cited in Nunan 1992, p.184) writes : “evaluation may be defined as a systematic process of determining the extent to which instructional objectives are achieved by pupils”. There are two aspects of this definition. First, note that evaluation implies a systematic process, which omits casual, uncontrolled observation of pupils. Second, evaluation assumes that instructional objectives have been previously identified. Without previously determined objectives, it is difficult to judge clearly the nature and extent of pupil learning.

Brown & Priyanvada (2009, p.4) states that : Measurement is the process of quantifying the observed performance of classroom learners.. Also , Wandt and Brown (1977, cited in Anas, 2011, p. 7) states that measurement means the act or process of ascertaining the extent or quantity of something in order to answer the question about “How much”

Brown & Priyanvada (2009, p. 3) Test is a method of measuring a person’s ability, knowledge, or performance in a given domain. A test is

first a method. It's an instrument – a set of techniques, procedures, or items – that requires performance on the part of the test taker. Second, a test must measure, which may be defined as a process of quantifying a test taker's performance according to explicit procedures or rules (Bachman, 1990, p. 18 – 19). Some tests measure general ability, whereas others focus on very specific competencies or objectives. Next, a test measures an individual's ability, knowledge, or performance. Testers need to understand who the test takers are. What are their previous experiences and backgrounds? Is the test appropriately matched to their abilities? How should test takers interpret their scores?. Finally, a test measures a given domain. On the other hand, there is a difference between evaluation, measurement and test, but they are connected each other.

## **A. TEST**

### **1. Definition of Test**

The term “test” is quite common in the realm of academic. In fact, there are two term relating to test. The two of them are measurement and evaluation. Therefore, sometimes it is hard to distinguish among the meaning of them. In scientific term, a test is a method of measuring a person's ability, knowledge, or performance in a given domain (Brown, 2010, p.3).

On the other hand, test are a subset of assessment, a genre of assessment techniques. They prepared administrative procedures that occur at identifiable times in a curriculum when learners muster all

their faculties to offer peak performance, knowing that their responses are being measured and evaluate (Brown & Priyavanda, 2009, p.3)

## **2. General Kinds of test**

There are many kind of test; each test has specific purpose and a particular criterion to be measured that tend to fall into a finite number of types, classified according to their purpose Anas (2010, p. 67 – 75)

### ***a. Based on its functions***

- 1) Diagnostic Test is a test usually given to a student entering an educational institution to determine specific knowledge or proficiency in various subjects for the purpose of assignment to appropriate courses or classes
- 2) Selection test is a test for selecting and choosing an object appropriately for certain purposes.
- 3) Pre-test is a carried out before the lesson was given to know how far student master the lesson that will be thought
- 4) Post test is a test given after teaching-learning process was carried out.
- 5) Formative test is a test given on certain period of teaching-learning program, for example daily examination, middle test, etc
- 6) Summative test is a test given on final period of teaching-learning program, for example semester final examination and

national final examination. Meanwhile, summative test as well as that include in these kind of this test.

***b. Based on psychological aspect***

- 1) Intelligence test is a test to measure the extent of somebody intelligence
- 2) Attitude test is a test to find out the predisposition to act or behave of somebody in a certain way.
- 3) Aptitude test is a test to find out the basic talent of somebody.
- 4) Personality test is a test to find out the distinctive characteristics of somebody.
- 5) Achievement test is a test to find out the extent of student achievement in teaching – learning program (Dina Hermina, 2003. p. 1-2)

***c. Based on test construction***

- 1) Teacher made test is an informal test that constructed by the teacher him/her self
- 2) Non standardized test is an informal test which is constructed by unprofessional testing services.
- 3) Standardized test is a formal test, which is prepared by professional testing services that has high extent of validity, realibility and practicability (Harris,1969, p.1)



***d. Based on the respond form***

- 1) Verbal Test is the answer of respond of the test in oral or written form
- 2) Non-verbal test is the answer of respond of the test in behavioral form

***e. Based on the form of given question and answer***

- 1) Objective test, is a question to be tested that only consist of one right answer. For examples : true false, multiple choice, matching, completion, substitution, etc
- 2) Subjective test is normally in essay form test. This test consist of a set of questions that testees answer based on his/her opinion. In deed they answer the questions by explaining, comaring, interpreting and distinguishing something (WayanNurkencana et.al,1990, p.35-36) for examples : translation, composition, scored interview, etc.

**f. Based on Score Interpretation**

There are two kinds of tests based on score interpretation. They are norm-referenced test and criterion-referenced test.

1) Norm-Referenced Test

Rudner (1994, cited inGyn, (n.d), p. 126 )Norm-referenced tests compare an individual test score to the scores of a well-defined reference group who have already taken the same test.

## 2) Criterion-Referenced Test

Anastasi, (1988, cited in Gyn (n.d),p. 126 )Criterion-referenced tests determine what test takers can do and what they know, not how they compare to others

## **B. BASIC PRINCIPLES IN TEST DESIGN**

A well-constructed test is an instrument that provides an accurate measure of the test-taker's ability in a particular domain. The definition may sound very simple, but in fact constructing a good test is a complex task involving both science and art (Jamilah, 2008, p. 1). Also, The test should be effective, appropriate, and useful.

To know the quality of item test can be done by analyze the item test. The experts says, there are some characteristics in a good items. If there is no, the items can't give the information about the results teaching-learning process clearly and accurately.

### **1. Validity**

Validity is the most important idea to consider when preparing or selecting an instrument for use. In recent years, validity has been defined as referring to the appropriateness, meaningfulness, and usefulness of the specific inferences researchers make based on the data they collect (Fraenkel, Noeman. 1993, p. 139). The meaning of validity in interpreting and using assessment results can be grasped

most easily by reviewing the following characteristics (Gronlund&Waugh, 2009, p. 46 – 47)

- a. Validity is inferred from available evidence (not measured)
- b. Validity depends on many different types of evidence
- c. Validity is expressed by degree (high, moderate, low)
- d. Validity is specific to a particular use.
- e. Validity refers to the inferences drawn, not the instrument.
- f. Validity is unitary concept.
- g. Validity is concerned with the consequences of using the assessments.

We should also look for the various types of evidence when evaluating standardized tests.

- a. *Content-related evidence of validity*; refers to the content and format of the instrument. How appropriate is the content? How comprehensive? Does it logically get at the intended variable? How adequately does the sample of items or questions represent the content to be assessed? Is the format appropriate? The content and format must be consistent with the definition of the variable and the sample of subjects to be measured. Its refers to representativeness of our measurement regarding the phenomenon about which we want information.(Gass& Mackey, 2005, p. 107).

**b.** *criterion-related evidence of validity*; refers to the degree of relationship between two sets of measures :1) the test scores and 2) the criterion to be predicted or estimated that is typically expressed by means of a correlation coefficient or an expectancy table (Gronlund & Waugh, 2009, p. 51)

**c.** *construct-related evidence of validity*; focuses on assessment results as a basis for inferring the possession of certain psychological characteristic. (Gronlund & Waugh, 2009, p. 55)

**d.** *Face Validity*,

Anastasi (1982, cited in Weir, 1990, p. 26) pointed out that face validity : is not validity in the technical sense; it refers, not to what the test actually measures, but to what it appears superficially to measure. Face validity pertains to whether the test looks valid to examinees who take it, the administrative personnel who decide on its use, and other technically untrained observers. Fundamentally, the question of face validity concerns rapport and public relations. If a test does not have face validity though, it may not be acceptable to the students taking it, or the teachers and receiving institutions who may make use of it. If the students do not accept it as valid, their adverse reaction to it may mean that they do not perform in a way which truly reflects their ability (Weir, 1990 p. 26).

Anastasi (1982 cited in Weir, 1990, p. 26) took a similar line “Certainly if test content appears irrelevant, inappropriate, silly or childish, the result will be poor co-operation, regardless of the actual validity of the test. Especially it is not sufficient for a test to be objectively valid. It also needs face validity to function effectively in practical situations.

## 2. *Reliability*

Reliability refers to the consistency of the scores obtained - how consistent they are for each individual from one administration of an instrument to another and from one set of items to another. If you give the same test to the same student or matched students on two different occasions, the test should yield similar results (Brown & Priyanvada, 2010. p.27). Phelan & Wren (2005) states that there are some types of reliability, they are :

- a. *Test-retest reliability* is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals. The scores from Time 1 and Time 2 can then be correlated in order to evaluate the test for stability over time
- b. *Parallel forms reliability* is a measure of reliability obtained by administering different versions of an assessment tool (both versions must contain items that probe the same construct,

skill, knowledge base, etc.) to the same group of individuals.

The scores from the two versions can then be correlated in order to evaluate the consistency of results across alternate versions.

c. *Inter-rater reliability* is a measure of reliability used to assess the degree to which different judges or raters agree in their assessment decisions. Inter rater reliability is useful because human observers will not necessarily interpret answers the same way; raters may disagree as to how well certain responses or material demonstrate knowledge of the construct or skill being assessed.

d. *Internal consistency reliability* is a measure of reliability used to evaluate the degree to which different test items that probe the same construct produce similar results.

1) *Average inter-item correlation* is a subtype of internal consistency reliability. It is obtained by taking all of the items on a test that probe the same construct (e.g., reading comprehension), determining the correlation coefficient for each *pair* of items, and finally taking the average of all of these correlation coefficients. This final step yields the average inter-item correlation.

2) *Split-half reliability* is another subtype of internal consistency reliability. The process of obtaining split-half

reliability is begun by “splitting in half” all items of a test that are intended to probe the same area of knowledge (e.g., World War II) in order to form two “sets” of items. The *entire* test is administered to a group of individuals, the total score for each “set” is computed, and finally the split-half reliability is obtained by determining the correlation between the two total “set” scores.

### 3. *Practicality*

Bachman & Palmer (1996, p. 36) “Practicality is the relationship between the resources that will be required in design, development and use of the test and the resources that will be available for these activities . They illustrated that this quality is unlike the others, because it focuses on how the test is conducted. Moreover, they classified the addressed resources into three types : human, material and time.

Moreover, Brown (2004) said that the test that is practical it needs to be within the means of financial limitations, appropriate time constraints, easy to administrator, score, and interpret.

Also, Mousavi (2009, cited in Brown, 2010, p.26) states that practicality refers to the logistical down-to-earth, administrative issues involved in making , giving and scoring an assessment instrument.

Brown & Priyanvada (2010, p.26) also state that good test should consider the following attributes of practicality :

- a. Stays within budgetary limits
- b. Can be completed by the test-taker within appropriate time constraints
- c. Has clear direction for administration
- d. Appropriately utilizes available material resources
- e. Does not exceed available material resources
- f. Consider the time and effort involved for both design and scoring

Based on this definition, practicality can be measured by the availability of the resources required to develop and conduct the test. Therefore, our judgement of the language test is whether it is practical or impractical

#### **4. *Distractor***

One important element in the quality of a multiple choice item is the quality of the item's distractors. However, neither the item difficulty nor the item discrimination index consider the performance of the incorrect response option, or distractors.

Don McNea Fire School (n.d) states that distractors are various alternatives chosen to be as close as possible to the right answer. In good test construction, all distractors should be feasible and reasonable and should apply directly to the stem. There should be a commonality in all of the distractors. If one distractor is off base and not plausible,



then you can safely assume the person writing the test question ran out of reasonable distractors

A distractor analysis addresses the performance of these incorrect response options. Just as the key, correct response option, must be definitively correct, the distractors must be clearly incorrect. In addition to being clearly incorrect, the distractors must also be plausible. Distractors have a good function if at least chosen by 5 % of all testees (Anas, 2011, p. 411)

#### **5. *Discrimination Power***

Discrimination goes beyond determining the proportion of people who answer correctly and looks more specifically at who answers correctly. In other words, item discrimination determines whether those who did well on the entire test did well on a particular item. An item should in fact be able to discriminate between upper and lower scoring groups (Dewi&Nastiti, 2012).

Discrimination power is to determine how well each item discriminates between upper and lower level of students, for each item in a test should help to separate the proficient subjects from those who lack the tested skills or learning (Haris, 1969, p. 105)

Item discrimination or the discriminating power of a test item refers to the degree to which success or failure on an item indicates possession of the ability being measured. It determines the extent to

which the given item discriminates among examinees in the function or ability measured by the item (Maheswari, 2013)

#### **6. *Difficulty Level***

According to Wilson (2005, cited in Dewi&Nastiti, 2012), item difficulty is the most essential component of item analysis. Item difficulty is determined by the number of people who answer a particular test item correctly. It is important for a test to contain items of various difficulty levels in order to distinguish between students who are not prepared at all, students who are fairly prepared, and students who are well prepared.

A good test is not too easy and it is not too difficult if an item is very easy it will be answered by all students correctly and if an item is very complicated it will not be answered by student correctly. So, both of them are not good items. Endless a good item is the item which has the certain difficulty level (WayanNurkancana, P.P.N. Sumartana, 1986, p.25) On the other hand, The students score must be analyzed in order to know exactly the index of difficulty level of the items.

#### **7. *Wash Back***

Hughes (2003, cited in Brown & Priyanvada, 2010, p. 37) states “ good test also have the effect of testing on teaching and learning”. Alderson & Wall (1993, cited in Brown & Priyanvada, 2010, p. 37) considered washback an important enough concept to define a washback Hypothesis that essentially elaborated on how tests influence

both teaching and learning. (Brown & Priyanvada, 2010, p. 38) conclude that the following factors comprise the concept of washback test that provides beneficial washback are :

- a. Positively influences what and how teachers teach
- b. Positively influences what and how learners learn
- c. Offers learners a chance to adequately prepare
- d. Give learners feedback that enhances their language development
- e. Is more formative in nature than summative
- f. Provides conditions for peak performance by the learners.

On the other hand, wash back means the teaching-learning process should have effects to the student, teachers or educational system, etc.

### **C. OBJECTIVE TEST ARRANGEMENT PRINCIPLES**

The proper preparation and use of an achievement test that measures a broad band of learning outcomes in a balanced manner involves the following steps (Gronlund & Waugh, 2009, p.70 - 85) :

#### *1. Specifying the Instructional Objectives*

The first step in preparing instructional objectives is to list the general objectives as intended learning outcomes. When a satisfactory list of general outcomes has been identified and clearly stated, the next step is to list the specific types of student

performance that are to be accepted as evidence that the outcomes have been achieved (Gronlund & Waugh, 2009, p.70 - 71).

## 2. *Preparing the Test Specification*

When general outcomes and specific terms have been developed, the writing of test items should be guided by a carefully prepared set of test specifications. The function of the specification is to (1) describe the achievement domain being measured, (2) link the achievement domains to course subject matter, and (3) provide guidelines for obtaining a representative sample of test tasks.

- a. Building a table of specification
- b. Selecting the learning outcomes to be tested
- c. Outlining the subject matter
- d. Making the two-way chart

## 3. *Considerations in Constructing Relevant Test Items*

- a. Selecting the types of test Items to use
- b. Matching items to specific learning outcomes
- c. Improving the functioning content of items
- d. Determining the number of test items to use.

## 4. *Arranging the Items in the test*

- a. For instructional purposes, it is usually desirable to group together items that measure the same outcome
- b. Where possible, all items of the same type should be grouped together

c. The items should be arranged in terms of increasing difficulty.

5. *Preparing Directions*

The directions for an achievement test should be simple and concise and yet contain information concerning each of the following : (1) purpose of the test, (2) time allowed to complete the test, (3) how to record the answers, and (4) whether to guess when in doubt about answer.

6. *Administering and scoring the test*

Make certain that all the students know exactly what to do, and then provide them with the most favorable conditions in which to do it. Also, Give a score of each item test that have been answered by the students.

7. *Analyzing the effectiveness of test items.*

The items should be analyzed for their effectiveness, like the difficulty of the item, the discrimination power and the effectiveness of each alternative, also the reliability and validity of the item.

#### **D. ANALYZING ITEMS**

The basic unit of any test is the test item. The items should be analyzed to determine their effectiveness. It means that the test is not finished yet once the raw mark has been obtained (Heaton, 1998, p. 164). It needs further analyzed in order to get information concerning (1) the performance of the students a group, thus informing the teacher about the

effectiveness of the teaching, (2) the performance of individual student and (3) the performance of each of the items comprising the test

Analysis item is the systematic evaluation of the effectiveness of the individual items on the test (Brown, 2005. P.41) item analysis refers to the process of collecting, summarizing and using information about pupils response to items. Item analysis is an important and necessary step in the preparation of good multiple choice tests. The purpose of item analysis is to improve the validity and the reliability of a whole test in its totality, due to the validity and the reliability of a test depend on the characteristics of its items.

The item analysis is an important factor in developing of an evaluation. In this phase statistical methods are used to identify any test items that are not working well. If an item is too easy, too difficult, failing to show a difference between skilled and unskilled examinees or even scored incorrectly an item analysis will reveal it

### 1. *Validity*

According to the explanation above, an item has good validity if Scores on item in question has a fit or alignment with the direction of the total score or there is a positive correlation between score items and total score significantly. Total score as the dependent variable here is domiciled (the dependent variable), while the score of the items located as independent variable (independent variable). On

the other hand, we can say an item is valid if the item score have positive correlation with total score significantly (Anas, 2011,P. 184).

As we known, In objective test there are just two option, right or wrong. Each item which answer correctly have 1 (one) score, and if the answer is uncorrectly have 0 (nul) score who in the world of statistics known as dichotomous or discrete data is pure. While the total score owned by each individual which is the sum of each score are owned by the individual - each item is a continuous data item. Anas (2011,p.185) concludes that according to the expert, if variable I is dicotomic data, and variable II is Continue Data, the best technique for looking the correlation is Point Biserial :

$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

*Details :*

$r_{pbi}$  : *Koefisien of Biserial Point Correlation that represent the correlation power between Variable I and variable II as Koefisien of Item validity*

$M_p$  : *Average Score of testee that answered the items correctly*

$M_t$  : *Average from Total Score.*

$SD_t$  : *Standard Deviasion from total score.*

$P$  : *proportion of testee who can answer the items correctly*

$q$  : *Proportion of testee who answer the items incorrectly*

## 2. *Reliability*

A reliable test is consistent and dependable. If you give the same test to the same student or matched student on two different occasions, the test should yield similar results. Different with Essay, in objective test in order to know the reability of the test we can use three ways :*First*, Single Test – Single Trial Method, *second*, Single Test – Double Trial Method and *third*, Double Test – Double Trial Method (Anas, 2011, p. 213).

In this case, the authors use the testing objective achievement test forms by a single test approach - single trial method using KR<sub>20</sub> Formula Kuder-Richardson (Anas, 2011, p. 252).

$$r^{11} = \left( \frac{n}{n-1} \right) \left( \frac{Sd^2 - \sum pq}{Sd^2} \right)$$

*Details :*

$r_{11}$  : Koefisien of reliability test completely

$n$  : Number of items

$l$  : Constant

$Sd^2$  : Standard Deviation

$p$  : Proportion of testee who can answer correctly

$q$  : Proportion of testee who answer incorrectly

## 3. *Discrimination Power*

Discrimination power is to determine how well each item discriminates between upper and lower level of students, for each item



in a test should help to separate the proficient subjects from those who lack the tested skills or learning (Haris, 1969,p. 105) .

Estimate the discrimination index is by comparing the member of students in the upper group and the lower group who answer the items correctly. The discrimination index can be calculated by using the formula below (Anas, 2011, p. 389 – 390) :

$$D = P_A - P_B \text{ Atau } D = P_H - P_L$$

Details :

D : Discrimination Power

$P_A$  atau  $P_H$  : Proportion of the higher group that can answer items correctly.

$P_B$  Atau  $P_L$  :Proportion of the higher group that can answer items correctly

$$P_A = P_H = \frac{B_A}{J_A}$$

Details :

$B_A$  : Number of testee of the higher group that can answer items correctly

$J_A$  : Number of testee who include in the up group

$P_B$  atau  $P_L$  : Proportion of the lower group that can answer items correctly

$$P_B = P_L = \frac{B_B}{J_B}$$

Details :

BB : Number of testee of lower group who can answer the item correctly

JB : number of testee who include in lower group

or

$$\phi = \frac{P_H - P_L}{2 \sqrt{(p)(q)}}$$

Details:

$\phi$  : Corelation Index of Phi in this section as Discrimination test index

$P_H$  : Proportion of the higher group

$P_L$  : Proportion of the lower group

2 : Bilangan constant

$P$  : Proportion of all testee who answer correctly Proporsi

$q$  : Proportion of all testee who answer discorrectly, where

$$q = (1 - p)$$

Discrimination indices can range from +1 (= an item which discriminates perfectly) through 0 (= an item which does not discriminates in any way at all) to -1 (=an item which discriminates in entirely the wrong way (Heaton, 1998, p. 180)

The statement above means that if the test discriminates perfectly with the discrimination index is +1, all the student in upper

group can answer all the items of the test correctly, while the student in the lower group cannot/. On the other hand, if the discrimination index is -1, all the students in upper group cannot answer all the items correctly, but all the students in the lower group can. This kind of test is entirely wrong and must be replaced. However, if both the students in the upper group and in the lower group can or cannot answer the items correctly, so the index of discrimination is 0. This kind of test does not discriminate in any way at all.

#### 4. *Difficulty level*

A good test is not too easy and it is not too difficult if an item is very easy it will be answered by all students correctly and if an item is very complicated it will not be answered by student correctly. So, both of them are not good items. Endless a good item iss the item which has the certain difficultylevel (Nurkancana&Sumartana, 1986, p.25) .

The students score must be analyzed in order to know exactly the index of difficulty level of the items. There are some ways to find difficulty index; one of them is du Bois formula. According to Witherington (n.d, cited in Anas, 2011, p. 371) Difficulty Level index about 0,00 to 1,00.

P	0,0	0,1	0,20	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,00
	Too Difficult			-----Medium-----				-----Too Easy			

The criteria above show that if the index of difficulty shows 1.00 the test is very easy since all of the student can answer the item. It is not

good to be given to the students. Moreover if the index of difficulty shows 0, The test is very difficult since all of the student cannot answer the items. This test is also not good being given. The test which is good to be given to the students is the test with criterion between 0.31 – 0.70

### 5. *Distractors*

One important element in the quality of a multiple choice item is the quality of the item's distractors. However, neither the item difficulty nor the item discrimination index consider the performance of the incorrect response option, or distractors.

A distractor analysis addresses the performance of these incorrect response options. Just as the key, correct response option, must be definitively correct, the distractors must be clearly incorrect. In addition to being clearly incorrect, the distractors must also be plausible. Distractors have a good function if at least chosen by 5 % of all testees (Anas, 201, p. 411).

$$D = \frac{\text{Number of student who choose the distractor}}{\text{Number of student who follow the test}} \times 100 \%$$

## **CHAPTER III**

### **RESEARCH METHOD**

#### **A. RESEARCH DESIGN**

The type of this research is field research by using descriptive quantitative approach in order to explain about an incident or certain incident also serve what happen accurately. In this research, researcher want to describe about the quality of the English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department academic year 2012/2013 at East Barito Regency, Central Borneo by analysis each of the items.

#### **B. RESEARCH SETTING**

The setting of the research is in East Barito Regency, Central Borneo Province on three specific areas:

1. MAN Dusun Timur at Tamiyang Layang City, Dusun Timur District , East Barito Regency
2. MA Ampah at Ampah City, Dusun Tengah District , East Barito Regency
3. MA Miftahul Ulum at Tuyau, Pematang Karau District , East Barito Regency

#### **C. SUBJECT AND OBJECT OF RESEARCH**

- a. Subject of the research

The subject of this research is English Test Questions scripts of English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department academic year 2012/2013 at East Barito Regency, Central Borneo and the distribution of the subject of this research is in the table below:

**Table 3.1 Names of School**

No	Name Of School	Number of Students
1.	MAN Dusun Timur	8
2.	MA Miftahul Ulum	31
3.	MA Ampah	65
TOTAL		104

b. Object of the research

The object of this research is the quality of English Try Out Questions of Social Department of Islamic Senior High Schools at East Barito Regency, Central Kalimantan Academic Year 2012/2013 in perspective validity, reliability, difficulty level, discrimination power and distractor.

**D. RESEARCH INSTRUMENT**

1. Data

The data that will be investigated in this research are primary data and the secondary data.

a. Primary data

The primary data in this research are the result of English Try Out Questions of Social Department of Islamic Senior High Schools at East Barito Regency, Central Kalimantan Academic Year 2012/2013 contains English items script, guidelines for designing the items, answer sheets and the key answer also the curriculum or syllabus.

b. Secondary data

To complete the primary data, the writer also collects the secondary data which related to the research like: the condition of English Teacher and the condition of the test-taker.

2. Source of Data

For getting primary data and secondary data, this research adopt source of data as follows:

- a. Documenter, archives from *Dinas Pendidikan dan Pemuda dan Olahraga* of East Barito Regency like items script, answer sheets, and the key answer. Also the curriculum or English syllabus of each school.
- b. Informants, they are : Head master of each schools and English Teacher of each schools

## **E. DATA COLLECTING PROCEDURES**

There are some techniques that are applied in this research, they are:

1. Documentary

- a. Special document : the test items, the students' answer of Try Out and the key answer
- b. General document : the written reports of school, such as Guidelines for designing the items, number of the students, curriculum or syllabus, etc.

## 2. Interview

Interview is a conversational in specific purpose between two people or more. This technique used to ask the informant or some related people

## F. DATA ANALYSIS

### 1. Technique of Data Processing

#### a. *Editing*

The writer recheck all the collected data to make sure whether they have already completed or not in this research

#### b. *Classifying*

The writer uses this techniques to classify the collected data based on its kind in order to analyzing the data easily

#### c. *Tabulating*

The writers compiles and enters the data that has been processed into the table

### 2. Technique of Data Analyzing



To make the organizing and the analysis of data process more easily, the writer organizes the steps of measurement , as follows:

### 1. Validity

An item has good validity if Scores on item item in question has a fit or alignment with the direction of the total score or there is a positive correlation between score items and total score significantly. Total score as the dependent variable here is domiciled (the dependent variable), while the score of the items located as independent variable (independent variable).On the other hand, we can say an item is valid if the item score have positive correlation with total score significantly (Anas , 2011, P.185).

$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

### 2. Difficulty level

A good test is not too easy and it is not too difficult if an item is very easy it will be answered by all students correctly and if an item is very complicated it will not be answered by student correctly. So, both of them are not good items. Endless a good item iss the item which has the certain difficultylevel (Nurkancana&Sumartana, 1986, p.25) The students score must be

analyzed in order to know exactly the index of difficulty level of the items. There are some ways to find difficulty index; one of them is du Bois formula (Anas, 2011, p. 371 – 372):

$$P = \frac{N_p}{N} \text{ or } P = \frac{\sum B}{N}$$

### 3. Reliability

A reliable test is consistent and dependable. If you give the same test to the same student or matched student on two different occasions, the test should yield similar results (Anas, 2011, p. 252).

$$r^{11} = \left( \frac{n}{n-1} \right) \left( \frac{Sd^2 - \sum pq}{Sd^2} \right)$$

### 4. Discrimination power

Discrimination power is to determine how well each item discriminates between upper and lower level of students, for each item in a test should help to separate the proficient subjects from those who lack the tested skills or learning. Estimate the discrimination index is by comparing the member of students in the upper group and the lower group who answer the items correctly. The discrimination index can be calculated by using two formulas (Anas, 2011, p.389 – 390) :

$$D = P_A - P_B \text{ Atau } D = P_H - P_L \text{ or } P_A = P_H = \frac{B_A}{J_A} P_B = P_L = \frac{B_B}{J_B}$$

$$\text{OR } \phi = \frac{P_H - P_L}{2\sqrt{(p)(q)}}$$

## 5. Distractor

A distractor analysis addresses the performance of these incorrect response options. Just as the key, correct response option, must be definitively correct, the distractors must be clearly incorrect. In addition to being clearly incorrect, the distractors must also be plausible. Distractors have a good function if at least chosen by 5 % of all testee (Anas, 2011, p. 411)

$$D = \frac{\text{Number of student who choose the distractor} \times 100 \%}{\text{Number of student who follow the test}}$$

## G. Procedure of Research

### 1. Preliminary Step

- a. To hold a prior observation in the research object
- b. To discuss the result of the previous observation with the writer's academic counselors
- c. To make a research design proposal to be approved by the Bureau Of Thesis Writing of Education Faculty IAIN Antasari Banjarmasin

### 2. Preparatory Steps

- a. To hold a seminar on the research design proposal.
- b. To ask the Dean of Islamic Education faculty for a written mandate to conduct the research.

- c. To make the data instruments.
3. Research Steps
    - a. To contact all respondents and informants who are needed to obtain the data.
    - b. To collect all the needed data from several sources.
    - c. To process the data in procedural way and then analyze them properly
  4. Organization Step
    - a. To arrange and to write the collected data in the first draft
    - b. To consult it with the writer's advisor and his assistant to get some comments and contributions
    - c. To write the final draft after having approved by the writer's advisor
    - d. To be examined before the team of the thesis examiner of Islamic Education Faculty.

## **CHAPTER IV**

### **FINDINGS AND DISCUSSION**

#### **A. Findings**

### 1. Test Maker

English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, Central Borneo made by teachers of SMAN 1 Tamiyang Layang. They are :

**Table 4.1 The names of Test Makers**

No	Code Name	Code Number
1	001	001 – 2012
2	002	002 – 2012
3	003	003 – 2012

*Source :Dinas Pendidikan dan Olahraga Kab. Barito Timur 2013*

### 2. Test Guide Arrangement

English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, Central Borneo based on the references :

1. Guide Technic of Syllabus Development for Senior High School  
by Education Ministry of Indonesia Republic 2006
2. *Kisi – kisi soal Ujian Nasional 2013* (See Appendix)

### 3. Test Items

English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, Central Borneo consists of 50 items. But, There are

some problems when the items are being examined. *First*, All listening file for question number 1 – 15 are broken when distributed to each schools. *Second*, Multiple Choice items for questions number 47 – 50 are the wrong question. So, There are only 31 English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, Central Borneo which examines or should be answered by the testee, as shown in the table below :

**Table 4.2 Test Items Condition and Explanation**

No.	Item Number	Categories	Explanation	Number Replacement
1	1	Listening	File is Broken	-
2	2	Listening	File is Broken	-
3	3	Listening	File is Broken	-
4	4	Listening	File is Broken	-
5	5	Listening	File is Broken	-
6	6	Listening	File is Broken	-
7	7	Listening	File is Broken	-
8	8	Listening	File is Broken	-
9	9	Listening	File is Broken	-
10	10	Listening	File is Broken	-
11	11	Listening	File is Broken	-
12	12	Listening	File is Broken	-
13	13	Listening	File is Broken	-

14	14	Listening	File is Broken	-
15	15	Listening	File is Broken	-
16	16	Multiple Choice	Item is Good	1
17	17	Multiple Choice	Item is Good	2
18	18	Multiple Choice	Item is Good	3
19	19	Multiple Choice	Item is Good	4
20	20	Multiple Choice	Item is Good	5
21	21	Multiple Choice	Item is Good	6
22	22	Multiple Choice	Item is Good	7
23	23	Multiple Choice	Item is Good	8
24	24	Multiple Choice	Item is Good	9
25	25	Multiple Choice	Item is Good	10
26	26	Multiple Choice	Item is Good	11
27	27	Multiple Choice	Item is Good	12
28	28	Multiple Choice	Item is Good	13
29	29	Multiple Choice	Item is Good	14
30	30	Multiple Choice	Item is Good	15
31	31	Multiple Choice	Item is Good	16
32	32	Multiple Choice	Item is Good	17
33	33	Multiple Choice	Item is Good	18
34	34	Multiple Choice	Item is Good	19
35	35	Multiple Choice	Item is Good	20
36	36	Multiple Choice	Item is Good	21
37	37	Multiple Choice	Item is Good	22

38	38	Multiple Choice	Item is Good	23
39	39	Multiple Choice	Item is Good	24
40	40	Multiple Choice	Item is Good	25
41	41	Multiple Choice	Item is Good	26
42	42	Multiple Choice	Item is Good	27
43	43	Multiple Choice	Item is Good	28
44	44	Multiple Choice	Item is Good	29
45	45	Multiple Choice	Item is Good	30
46	46	Multiple Choice	Item is Error	-
47	47	Multiple Choice	Item is Error	-
48	48	Multiple Choice	Item is Error	-
49	49	Multiple Choice	Item is Error	-
50	50	Multiple Choice	Item is Good	31

*Source ;DinasPendidikanPemudadanOlahragaKab. Barito Timur, 2013*

#### 4. Validity

The Variables is to count correlation between dependent variable and independent variable in order to know the validity of English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, Central Borneo using Biserial Point Correlation Formula that apply in Microsoft Excel.

The items' items validity count by using Biserial Point

CorxrelationFormula :



$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

We know number of students who follow the test (N) = 104 in significance 1% and 5%. If  $r_{pbi} > r_{tabel}$ , the item is valid. Based on the data on appendix, Validity for item number 1 is :

$M_p$  : 9,06

$M_t$  : 8,06

$SD$  : 2,39

$p$  : 0,29

$q$  : 0,71

$$r_{pbi} = \frac{9,06 - 8,10}{2,39} \sqrt{\frac{0,29}{0,71}}$$

$$r_{pbi} = \frac{1,0}{2,39} \sqrt{\frac{0,29}{0,71}}$$

$$r_{pbi} = \frac{1,0}{2,39} \sqrt{0,408}$$

$$r_{pbi} = 0,42 = 0,64$$

$$r_{pbi} = 0,2688 = 0,269$$

On the other hand, Correlation Biserial Point for item number 1 is, 0,269.

On the table of "r" product moment in significance 1% and 5%, there is no data for N=104. So, In order to know  $r_{tabel}$  for N=104 which exist between 100 – 125 by using Interpolation Linier Formula :

$$\frac{y - y^1}{y^2 - y^1} \frac{x - x^1}{x^2 - x^1}$$

1%	:	0,256	1%	:	0,176
5%	:	0,195	5%	:	0,230
		N = 100			N = 125

From the data above, we know :

$$X = 100 = 0,256 \qquad Y = 100$$

$$X_1 = 104 = ? \qquad Y_1 = 104$$

$$X_2 = 125 = 0,176 \qquad Y_2 = 125$$

$$\frac{100 - 104}{125 - 104} = \frac{0,256 - x^1}{0,176 - x^1}$$

$$\frac{-4}{21} = \frac{0,256 - x^1}{0,176 - x^1}$$

$$(-4)(0,176 - X_1) = (0,256 - X_1)(21)$$

$$- 0,704 + 4X_1 = 5,376 - 21$$

$$25 X_1 = 6,08$$

$$X_1 = 6,08/25$$

$$X_1 = 0,243$$

So, “r” product moment in significance 1% for N = 104 is, 0, 243

Next, “r” product moment in significance 5% for N = 104 using the same formula is, 0,201

Biserial Point Correlation for item number 1 is  $0,269 > r_{\text{tabel}}$  for significance 1% and 5%. So, item number 1 is valid. Next, The results of next items (Item number 2 to item number 31) use the same formula as shown in the following table below :

**Table 4.3 Validity Analysis Results**

Items Number	$r_{\text{pbi}}$	$r_{\text{tabel}}$	Interpretation
1	0,269	1 % = 0,243	Valid
		5 % = 0,201	Valid
2	0,109	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
3	0,571	1 % = 0,243	Valid
		5 % = 0,201	Valid
4	0,183	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
5	0,111	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
6	0,152	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
7	0,232	1 % = 0,243	Invalid
		5 % = 0,201	Valid
8	0,147	1 % = 0,243	Invalid
		5 % = 0,201	Valid

9	0,178	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
10	0,254	1 % = 0,243	Valid
		5 % = 0,201	Valid
11	0,078	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
12	0,294	1 % = 0,243	Valid
		5 % = 0,201	Valid
13	0,356	1 % = 0,243	Valid
		5 % = 0,201	Valid
14	0,238	1 % = 0,243	Invalid
		5 % = 0,201	Valid
15	0,299	1 % = 0,243	Valid
		5 % = 0,201	Valid
16	0,178	1 % = 0,243	Valid
		5 % = 0,201	Valid
17	0,263	1 % = 0,243	Valid
		5 % = 0,201	Valid
18	0,237	1 % = 0,243	Invalid
		5 % = 0,201	Valid
19	0,299	1 % = 0,243	Valid
		5 % = 0,201	Valid

20	0,281	1 % = 0,243	Valid
		5 % = 0,201	Valid
21	0,127	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
22	0,225	1 % = 0,243	Invalid
		5 % = 0,201	Valid
23	0,100	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
24	-0,008	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
25	0,187	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
26	0,291	1 % = 0,243	Valid
		5 % = 0,201	Valid
27	0,062	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
28	-0,206	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
29	0,030	1 % = 0,243	Invalid
		5 % = 0,201	Invalid
30	0,111	1 % = 0,243	Invalid
		5 % = 0,201	Invalid

31	0,197	1 % = 0,243	Invalid
		5 % = 0,201	Invalid

*Note :  $r_{tabel}$  for  $N=104$  with significance 1% is 0,243 and  $r_{tabel}$  for significance 5% is 0,201*

Based on the table above, from English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, Central Borneo there are 11 items (1,3,10,12, 13, 15, 16, 17, 19, 20 and 26) that valid in significance 1% and there are 20 items are invalid (2, 4, 5, 6, 7, 8,9, 11,14, 18, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31). While in Significance 5% , there are 14 items (1, 3, 7, 8, 10, 13,14, 15, 16, 17, 18, 19,20, and 22) are valid and there are 17 items (2,4,5,6,9,11,12,21,23,24,25,26,27,28,29,30 and 31) are invalid.

**Table 4.4 Validity Percentation of 1% and 5% Significance**

No	Categories	1%	Percentage	5%	Percentage
1	Valid	11 Items	35 %	14 Items	45%
2	Invalid	20 Items	65%	17 Items	55%
Total		31 Items	100%	31 Items	100%

According to the data shows on the table above, in significance 1% there are 11 items (35%) are valid and 20 items (65%) are invalid. Meanwhile, in significance 5%, there are 14 items (45%) are valid and 17 items (55%) are in valid. In the other hand, English Test

Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency. So, they should be revised, because there are too many items are invalid than items are valid.

### 5. Reliability

The Variables is to count correlation between dependent variable and independent variable in order to know the validity of English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency using Kuder-Richardson Formula that apply in Microsoft Excel. English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency are Valid if  $r_{11} \geq 0,70$ .

The reliability of of EEnglish Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency as shown on the table below :

**Table 4.5 Reliability Analysis Results'**

<b>N</b>	<b>SD<sup>2</sup></b>	<b>Σpq</b>
<b>31</b>	<b>5,7082</b>	<b>4,8635</b>

$$r^{11} = \left( \frac{n}{n-1} \right) \left( \frac{Sd^2 - \sum pq}{Sd^2} \right)$$

$$r^{11} = \left( \frac{31}{31-1} \right) \left( \frac{5,7082 - 4,8635}{5,7082} \right)$$

$$r^{11} = (1,033333)(0,14798)$$

$$r^{11} = 0,152913$$

**Interpretation,  $r_{11} = 0,15 < 0,70$  (Unreliable)**

Based on the data shows on the table above, the reliability of English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency are  $0,15 < 0,70$ . In the other hand, the items are unreliable and they should be revised.

## 6. Difficulty Level

The Variables to count Difficulty Level Index (P) of English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, Central Borneo using Du Bois Formula that apply in Microsoft Excel.

According to Witherington, Difficulty Level index about 0,00 to 1,00.

P	0,0	0,1	0,20	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,00
	Too Difficult	-----		-----	Medium	-----		-----		-----	Too Easy



The items' Difficulty Level Index count by using Du Bois Formula :

$$P = \frac{N_p}{N}$$

For Item Number 1 We know that N (Number of students who follow the test) are 104 testee and  $N_p$  (Student who can answer correctly ) are 30 testee. So, the Difficulty Level Index is :

$$P = \frac{30}{104} = 0,29$$

From the result, we know that for item number 1, the difficulty Level Index is Medium, where the Difficulty Level Index is  $0,29 > 0,25$ .

Next, The results of next items (Item number 2 to item number 31) use the same formula as shown in the following table below :

**Table 4.6 Difficulty Analysis Results'**

Item Number	$N_p$	N	P	Interpretation
1	30	104	0,29	Medium
2	67	104	0,64	Medium
3	16	104	0,15	Too Difficult
4	74	104	0,71	Medium
5	11	104	0,11	Too Difficult
6	61	104	0,59	Medium
7	72	104	0,69	Medium
8	15	104	0,14	Too Difficult
9	18	104	0,17	Too Difficult

10	49	104	0,47	Medium
11	6	104	0,06	Too Difficult
12	44	104	0,42	Medium
13	21	104	0,20	Too Difficult
14	36	104	0,35	Medium
15	29	104	0,28	Medium
16	11	104	0,11	Too Difficult
17	17	104	0,16	Too Difficult
18	45	104	0,43	Medium
19	29	104	0,28	Medium
20	20	104	0,19	Too Difficult
21	20	104	0,19	Too Difficult
22	17	104	0,16	Too Difficult
23	20	104	0,19	Too Difficult
24	10	104	0,10	Too Difficult
25	16	104	0,15	Too Difficult
26	14	104	0,13	Too Difficult
27	10	104	0,10	Too Difficult
28	8	104	0,08	Too Difficult
29	12	104	0,12	Too Difficult
30	13	104	0,13	Too Difficult
31	24	104	0,23	Too Difficult

The table shows that there are 20 items are Too Difficult, they are items number (3,5,8,9,11,13,16,17,20,21,22,23,24,25,26,27,28,29,30 and 31), There are 11 items are Medium, they are Items number (1,2,4,,6,7,10,12,14,15,18 and 19) and There is no Item is too easy.

**Table 4.7 Difficulty Level Interpretation and Percentation**

Interpretation	Frekuensi	Presentation
Too Difficult	20	65 %
Medium	11	35%
Too Easy	0	0%

Based on the results on the table above, from 31 English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, only 35 % (11 items) are medium and 65% (20 items) are Too Difficult. On the other hand, The items are too difficult, it is better to revise.

## **7. Distractors**

The Variables to count Distractors of English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency using Microsoft Excel.

Distractors have a good function if at least chosen by 5 % of all testee. To know the whether the distractors have good function or not, use the formula below :

$$D = \frac{\text{Number of student who choose the distractor} \times 100 \%}{\text{Number of student who follow the test}}$$

**Table 4.8 Distractors Analysis Results'**

Items Number	Key Answer	Distractors					Interpretation
		A	B	C	D	E	
1	A	<b>30,8</b>	37,5	5,8	6,7	19,2	GF
2	A	<b>62,5</b>	14,4	4,8	3,8	19,2	NGF
3	A	<b>15,4</b>	52,9	18,3	6,7	6,7	GF
4	A	<b>72,1</b>	10,6	1,9	6,7	8,7	GF
5	B	82,7	<b>10,6</b>	1,0	3,8	1,9	WF
6	A	<b>57,7</b>	10,6	6,7	9,6	15,4	GF
7	E	8,7	2,9	7,7	9,6	<b>28,8</b>	GF
8	A	<b>14,4</b>	5,8	35,6	24,0	20,2	GF
9	C	12,5	21,2	<b>19,2</b>	16,3	46,2	GF
10	C	6,7	9,6	<b>46,2</b>	9,6	27,9	GF
11	C	86,5	6,7	<b>5,8</b>	1,0	0,0	NGF
12	A	<b>40,4</b>	27,9	3,8	3,8	24,0	NGF
13	E	13,5	16,3	26,9	21,2	<b>22,1</b>	GF
14	C	25,0	11,5	<b>32,7</b>	10,6	20,2	GF

15	A	27,9	11,5	14,4	36,5	90,4	GF
16	B	17,3	10,6	8,7	10,6	52,9	\GF
17	C	23,1	9,6	17,3	24,0	26,0	GF
18	C	25,0	6,7	43,3	17,3	7,7	GF
19	A	26,0	1,0	14,4	14,4	44,2	GF
20	A	19,2	17,3	45,2	3,8	14,4	GF
21	A	17,3	9,6	26,0	25,0	22,1	GF
22	B	14,4	15,4	6,7	26,0	37,5	GF
23	B	29,8	16,3	9,6	21,2	23,1	GF
24	E	41,3	24,0	7,7	16,3	10,6	GF
25	B	30,8	15,4	13,5	15,4	25,0	GF
26	E	33,7	16,3	16,3	18,3	15,4	GF
27	E	10,6	26,0	14,4	34,6	14,4	GF
28	C	29,8	31,7	5,8	16,3	16,3	GF
29	D	26,9	31,7	14,4	10,6	16,3	GF
30	B	32,7	12,5	26,9	7,7	20,2	GF
31	A	24,0	22,1	12,5	13,5	27,9	GF

*Note : GF (Good-Functioned) NGF (Not Good-Functioned)*

Based on the data shown on the table 4.8 above, the distractors of English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency are well-functioned. There are only 3 items ( 2, 12 and 12) are unwell functioned.



For item number 1,  $B_A$  (Numbers of the higher group who can answer correctly) are 12 and  $B_A$  Number of students who follow the test are 28. So, Proportion of the higher group ( $P_A$ ) is :

$$P_A = \frac{12}{28} = 0,43$$

Then, Number of lower group who can answer correctly ( $B_B$ ) are 3 and numbers of student who follow the test ( $J_B$ ) are 28, So, Proportion of the lower group is :

$$P_B = \frac{3}{28} = 0,11$$

Then,  $D = P_A - P_B$

$$D = 0,43 - 0,11 = 0,32$$

$$D = 0,32$$

From the result, we know that for item number 1, the Discrimination Power Index is Satisfactory, where the Discrimination Power index is  $0,32 > 0,20$ . Next, The results of next items (Item number 2 to item number 31) use the same formula as shown in the following table below :

**Table 4.10 Discrimination Power Analysis Results'**

Item Number	$P_A$	$P_B$	$J_A$	$B_A$	$J_B$	$B_B$	D	Intrpretation
-------------	-------	-------	-------	-------	-------	-------	---	---------------

<b>1</b>	0,43	0,11	28	12	28	3	0,32	Fair
<b>2</b>	0,71	0,68	28	20	28	19	0,03	Poor
<b>3</b>	0,29	0,07	28	8	28	2	0,22	Fair
<b>4</b>	0,93	0,64	28	26	28	18	0,29	Fair
<b>5</b>	0,14	0,07	28	4	28	2	0,07	Poor
<b>6</b>	0,61	0,39	28	17	28	11	0,22	Fair
<b>7</b>	0,86	0,54	28	24	28	15	0,32	Fair
<b>8</b>	0,18	0,07	28	5	28	2	0,11	Poor
<b>9</b>	0,29	0,07	28	8	28	2	0,22	Fair
<b>10</b>	0,68	0,32	28	19	28	9	0,36	Fair
<b>11</b>	0,11	0,04	28	3	28	1	0,07	Poor
<b>12</b>	0,61	0,21	28	17	28	6	0,40	Good
<b>13</b>	0,39	0,04	28	11	28	1	0,35	Fair
<b>14</b>	0,50	0,18	28	14	28	5	0,32	Fair
<b>15</b>	0,43	0,11	28	12	28	3	0,32	Fair
<b>16</b>	0,14	0,04	28	4	28	1	0,10	Poor
<b>17</b>	0,32	0,04	28	9	28	1	0,28	Fair
<b>18</b>	0,50	0,21	28	14	28	6	0,29	Fair
<b>19</b>	0,43	0,11	28	12	28	3	0,32	Fair
<b>20</b>	0,32	0,07	28	9	28	2	0,25	Fair
<b>21</b>	0,18	0,14	28	5	28	4	0,04	Poor
<b>22</b>	0,29	0,04	28	8	28	1	0,25	Fair
<b>23</b>	0,25	0,14	28	7	28	4	0,11	Poor
<b>24</b>	0,14	0,07	28	4	28	2	0,07	Poor
<b>25</b>	0,29	0,07	28	8	28	2	0,22	Fair
<b>26</b>	0,32	0,07	28	9	28	2	0,25	Fair
<b>27</b>	0,18	0,14	28	5	28	4	0,04	Poor



28	0,00	0,18	28	0	28	5	-0,18	Poor
29	0,07	0,07	28	2	28	2	0,00	Poor
30	0,14	0,04	28	4	28	1	0,10	Poor
31	0,36	0,14	28	10	28	4	0,22	Fair

*Note :*

*If testee more than 100 person, we just need to take 27% of them for each group (lower and higher)*

The table shows that there are 12 items are Poor, they are items number (2, 5, 8, 11, 16, 21, 23, 24,27, 28, 29 and 30), There only 1 item is Good, it is item number 12 and There are 18 items are Fair, they are item numbers (1, 3 4, 6. 7. 9, 10, 13, 14, 15, 17, 18, 19, 20, 22, 25, 26, and 31) .

**Table 4.11 Discrimination Power Percentation**

Interpretation	Frekuensi	Presentation
Poor	12	39%
Fair	18	58%
Good	1	3%
Excellent	0	0%

Based on the results on the table above, from 31 English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, there are 39 % (12 items) are Poor and 58% (18 items) are Satisfactory and

There are 3% (Only 1 item) on the other hand, The items are too difficult. So, they should be revised

## **B. DISCUSSION**

After all the findings are described, then they also need to be discussion based on the order of issue that have introduced earlier in this study.

### **1. The quality of the English Test Questions of Regional Try Out Islamic Senior High Schools In Social Department Academic Year 2012-2013 At East Barito Regency, Central Borneo?**

As having seen in findings about quality of the English Items that analyze from five principle, validity, reliability, Discrimination Power, Difficulty Level and distractor. The fact shows that the quality of the English items is not good, and it is better to revised.

#### **a. Validity**

Based on the result of validity analysis using Biserial Point Correlation, as shown on the table 4.3 and 4.4 , there are only 11 items (35%) are valid and 20 items (65%) are invalid in significance 1%. Then, there are 14 items (45%) are valid and 17 items (55%) are in valid in significance 5%. In the other hand, English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito

Timur Regency is better to revise, because there are too many items are invalid than items are valid .

The writer found the data that the reasons why the items are invalid, because of, 1) The items only make by three teachers without any discussion or sharing with other English teacher from each school who follow the examination, 2)The condition of items itself that broken, also some dictions of the question are not clear or hard to understand by the students 3) Although the items are appropriate with the curriculum that used in Barito Timur Regency, but each school have it own problem to apply all instruction in the curriculum.

**b. Reliability**

Based on the data shows on the table 4.5 above, the reliability of English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency are  $0,15 < 0,70$ . In the other hand, the items are unreliable and they should be revised with new items. Because, the reliability of the items is very low and the items can not be a good instrument in order to know the readiness of students who will follow the National Examination.

**c. Difficulty Level**

Based on the results of Difficulty Level Analysis that shown on the table 4.6 and 4.7 above, from 31 English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, only 35 % (11 items) are medium and 65% (20 items) are Too Difficult. On the other hand, The items are too difficult, it is better to revise.

The writer found the reasons why the English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency are very difficult to answer by the students because : 1) The items only made by three teacher from one school (SMAN 1 TamiyangLayang) without invite other English Teachers from each school who will follow the Try Out. 2) There are some dictions that use in the question are not clear or hard to understand by the students.

#### **d. Distractors**

Based on the data shown on the table 4.8 above, the distractors of English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency are Good-functioned. There are only 3 items ( 2, 12 and 12) are Not Good functioned.

**e. Discrimination Power**

Based on the results on the table above, from 31 English Test Questions of Regional Try Out Islamic Senior High Schools in Social Department Academic Year 2012-2013 at Barito Timur Regency, there are 39 % (12 items) are Poor and 58% (18 items) are Fair and There are 3% (Only 1 item) the other hand, The items are too difficult, it is better to revised.

**2. Appropriateness of English Test Questions Of Regional Try Out Islamic Senior High Schools In Social Department Academic Year 2012-2013 At East Barito Regency,Central Borneo with the Syllabus or Curriculum.**

Based on the data researcher gained , from the staff of the institution that teachers who made the items test are guided by 2 references are qualified as the main reference to make items.On the other hand, the English Test Questions of Regional Try Out Islamic Senior High Schools In Social Department Academic Year 2012-2013 At East Barito Regency,Central Borneo are appropriate with the Syllabus or Curriculum. But, why there is no students who follow the test are passed the examination? It can be answer by other research who concentrate to find the factors why it happens.

## CHAPTER V

### CLOSURE

#### A. Conclusion

In this chapter, the writer concludes all the research results as follows:

1. The quality of English Test Questions of Regional Try Out Islamic Senior High Schools In Social Department Academic Year 2012-2013 At East Barito Regency, Central Borneo is not good
2. English Test Questions of Regional Try Out Islamic Senior High Schools In Social Department Academic Year 2012-2013 At East Barito Regency, Central Borneo are appropriate with the Syllabus or Curriculum

#### B. Suggestion

In this framework of intensifying the developing of English competence, the writers gives suggestions or contributions as follows:

1. For the Test makers, it is better to invite each English Teacher from each school in order to arrange the item test. Because, although the test are made based on the references that use globally, each school has its own problem in order to apply the instructions from the guide book. Also, the materials or Media in each school are not the same and limited.
2. For “*Dinas Pendidikan Pemudadan Olahraga Kab. Barito Timur*” it is better to make a team which consist of one English Teacher from each school in Barito Timur Regency in order to arrange the item test. Also, it is better if Try Out not only once.