IOSR Journal of Research & Method in Education (IOSR-JRME)
e- ISSN: 2320–7388,p-ISSN: 2320–737X Volume 8, Issue 3 Ver. I (May. – June. 2018), PP 04-12
www.iosrjournals.org

Learning Assessment In Language Teaching Of Senior High School Final Test

Nila Sari Dewi
**Master of English Study Program, FKIP of Lampung University, Indonesia.*
**Foreign Language Academy of Dian Cipta Cendikia, Bandar Lampung, Indonesia.*
Corresponding Author: Nila Sari Dewi

Abstract : *The objective of this research was to describe; (1). some good items in the test; 2). some revised items in the test; 3). some good alternatives (distractors) in each items; and 4). some dropped alternatives (distractors) in each items.The methodology of this research was qualitative research that involved students of Senior High School 2 of Pangkal Pinang, for class IPS 2 of Grade XII in period 2017/2018. The subject was an English items of school final test. The instruments of research were test, document and interview. The research findings indicated that; the result of good items were two items only, those were item number 12 and 31.These items were involved in good items because they had good or high for level of difficulty, discrimination power, alternatives and reliability. There were 17 items should be revised and 21 items should be dropped. There were 200 alternatives (distractors) from this test. The result of analysis had been found that 146 distractors can be applied without revision and 54 distractors were considered to be revised or dropped.*

Keywords : *learning assessment, senior high school, school final test*

Date of Submission: 21-04-2018 Date of acceptance: 08-05-2018

I. Introduction

Assessment has always been important; it is linked to language teaching methodology, program outcomes, language teacher competencies, language standards and second language acquisition training. It is only through assessment that we can discover whether the instructional activities in which we engaged our students resulted in the intended learning. Assessment really is the bridge between teaching and learning. The term 'assessment' is used both as a general umbrella term to cover all methods of testing and assessment, and as a term to distinguish 'alternative assessment' from 'testing.' Some applied linguists use the term 'testing' to apply to the construction and administration of formal or standardized tests such as the Test of English as a Foreign Language (TOEFL) and 'assessment' to refer to more informal methods such as those listed below under the heading 'alternative assessment.' For example, Valette (1994) says that 'tests' are large-scale proficiency tests and that 'assessments' are school-based tests. Intriguingly, some testers are now using the term 'assessment' where they might in the past have used the term 'test' (Kunnan 1998). There seems, indeed, to have been a shift in many language testers' perceptions so that they, perhaps subconsciously, may be starting to think of testing solely in relation to standardized, large-scale tests. They therefore use the term 'assessment' as the wider, more acceptable term.

It is probably too early to say whether advances in the use of computer technology for testing will have apositive or a negative effect on testing and whether computer administered tests will be distrusted by 'assessors.' Until recently, computer testing tended to fossilize existing objective testing methods because objectively marked items such as multiple-choice questions and gap-filling tasks were straightforward to answer on the computer and were easy to mark mechanically. Any attempts to introduce interesting new methods of assessment and testing were foiled by limitations in the memory size and processing speed of the computers. The move of the TOEFL towards computer based testing too has at least in the short run, extended the use of multiple choice and other easy-to-mark objective items in computer tests. However, it seems that the promised testing revolution may at last be on its way.

. Computer technologies have opened up new possibilities not only for optimizing the administration of tests, but also–and especially–for test development and assessment. Computer Assisted Testing (CAT) allows for a redesign of psychological and educational tests for effective and efficient administration by interactive computers; its objective is to select, for each examinee, a set of test questions that measures that person on the given trait effectively and efficiently (VanderLinden and Glas, 2000). From this statement, we can understand that a proper test is a collection of well combined elements.

The construction of test items is a crucial step for the validity of a test. A good item construction process enhances the discrimination power, score variance, reliability, and evidence of validity for the

intended interpretation and use of scores from the overall test (Suenand McClellan, 2003). Furthermore, according to Sikolova et al. (2009), text and task authenticity, attractiveness and balance of length of texts, and relevance of topics in terms of the examinees'age, education, and common interests are other aspects that must be considered during test development. The ability to construct high quality test item requires knowledge of the principles and techniques of test construction and skills in their application.

The duties of teachers and academicians in a high school do not only stop in final test, but they must make plan to step future learning. In fact, Senior High School 2 of Pangkal Pinang has not conducted learning evaluation completely. There is only questionnaire for lecturer's performance that is given at the end of each semester. Consequently, there are many steps to assess students' learning achievement such as test, scoring, and evaluation. Stufflebeam et al (1978) stated "evaluation is as a process to get some information for having a good alternative in the test mechanism. Evaluation also defines as the activity which is to know the successful of program, its scoring, and learning evaluation which are appropriate, and this also needs a good test". Based on this statement, to reach the goal of teaching learning process, lecturers and academic must evaluate students' learning. And this must have learning evaluation. The test especially final test must be evaluated to know its validity, reliability, discrimination power, and level of difficulty.

Assessment is giving value of learning quality. Evaluation is the process to take final decision of quality learning by using information from evaluation instrument. Assessment of student learning is a major component of university curriculum, and evidently it carries substantial weight in the equation of student learning. A common belief is what gets assessed is what gets learned. Consistently, the literature acknowledges that assessment is a significant driver of student learning, an important indicator of teaching effectiveness (Daniel and King, 1998). These findings are supported by Black and William's (1998) review of studies on instructors' use of assessment for learning.

Hence, instead of motivating students to learn more, the externally initiated assessment policy works mainly on auditing student achievement and college performance. In Stiggins' (1994) terms, the assessment efforts are centered upon the assessment of learning, an indicator of curriculum-centered practice; the practice of assessment for learning is yet to be distinctly visible at institutions of higher education. These policies have not been successful to direct, create, manage, monitor, and evaluate the processes, procedures and standards of practice of assessment for learning. In addition, despite the policy interventions, not much is known about institutional support for a balanced practice of assessment of and assessment for learning in higher education (Cole, & Peterson, 1998; Peterson & Einarson, 2001). From these statement, we can conclude that the aim of the present study is to examine the effects of the institutional intervention. Specifically the purposes of the study is to examine the faculty's acceptance of the assessment policy and the effects of the policy on curriculum and assessment planning.

Detailed rules of procedure in the construction of an objective examination which would possess general utility can hardly be formulated. The type of questions must be decided on the basis of such facts as the school subject concerned, the purposes of the examination, the length and reliability of the proposed examination, preferences of teachers and pupils, the time available for the examination, whether factual knowledge or thinking is to be tested (Ruch, 1924). Furthermore, Kehoe (1995) presents a series of guidelines for creating multiple-choice test items. These guidelines are spec-like in their advice. Here are the first two, which concern the stem of a multiple-choice item (a stem is the top part of a multiple-choice item,usually a statement or question; 1) Before writing the stem, identify the one point to be tested by that item. In general, the stem should not pose more than one problem, although the solution to that problem may require more than one step, and 2) Construct the stem to be either an incomplete statement or a direct question, avoiding stereotyped phraseology, as rote responses are usually based on verbal stereotypes. Here is a sample multiple-choice item. Let us assume the test takers are school-aged children in an English-medium school setting. Our item is part of a larger test that determines whether the student needs additional English language training – for example, supplemental English courses at the end of the school day.

ITEMAN is one of the analysis programs that comprise Assessment Systems Corporation's Item and Test Analysis Package. ITEMAN analyzes test and survey item response data andprovides conventional item analysis statistics (e.g., proportion/percentage endorsieng anditem-total correlations) for each item, in order to assist in determining the extent to whichitems are contributing to the reliability of a test and which response alternatives arefunctioning well for each item. In addition to item-level statistics the ITEMAN program also provides statistical indicators on the performance of the test as a whole (e.g., mean,standard deviation, reliability, median p-value). The system is easy that the evaluator or a lecturer just input one way, that is test aswer key, so he will get eight facilities for learning evaluation. The mechanism of work system in Simpel PAS is complete, easy, and accurate.

An item analysis gets at the question of whether the test is working by asking the same question of all items—how well does it discriminate? If lots of items that didn't discriminate much if at all, the test maker or a teacher may want to replace them with better ones. In short, item analysis gives a way to exercise additional

quality control over the tests. Item analyses can give a teacher feedback on how successful for well-specified learning objectives and well-constructed items. Item analyses can also help to diagnose why some items did not work especially well, and thus suggest ways to improve them.

Before discussing the construction of such items, let's review the terminology commonly used to describe the parts of multiple-choice questions. The diagram below labels the specific components of a multiple-choice item**.**

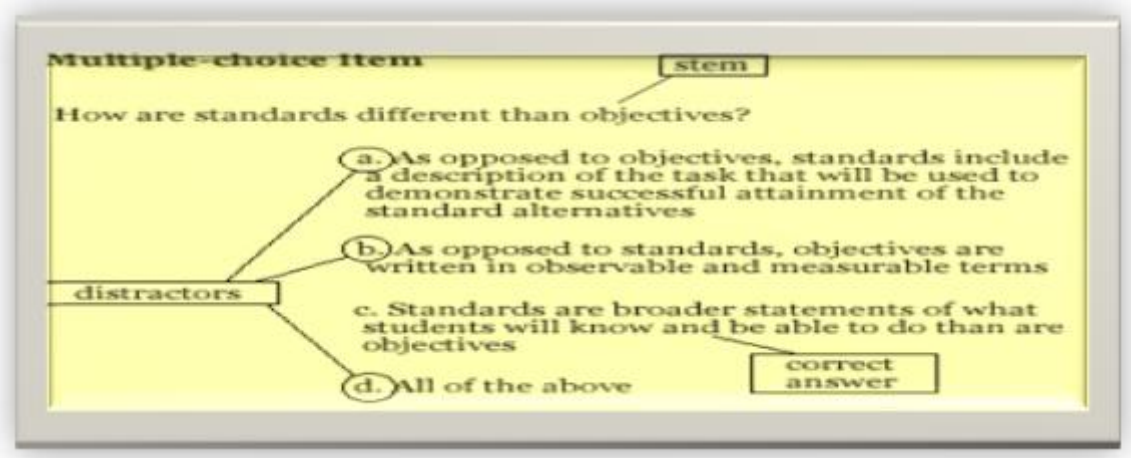

Fig.. 1.1 Components of Multiple Choice Items

From the picture above, an item of multiple choice test model consist of the followings;

- **a.** Stem : A question or statement followed by a number of choices or alternatives that answer or complete the question or statement
- **b.** Alternatives : All the possible choices or responses to the stem
- **c.** Distractors (foils) : Incorrect alternatives
- **d.** Correct answer : The correct alternative

There are several reasons to use multiple choice questions in tests. In terms of administration, multiple choice questions usually requires less time for test takers to answer, are easy to score and grade, provide greater coverage of material, allows for a wide range of difficulty, and can easily diagnose a test taker's difficulty with certain concepts. As an educational tool, multiple choice items test many levels of learning as well as a test taker's ability to integrate information, and it provides feedback to the test taker about why distractors were wrong and why correct answers were right. Nevertheless, there are difficulties associated with the use of multiple choice questions. In administrative terms, multiple choice items that are effective usually take a great time to construct. As an educational tool, multiple choice items do not allow test takers to demonstrate knowledge beyond the choices provided and may even encourage guessing or approximation due to the presence of at least one correct answer.

Multiple choice is a form of an objective assessment in which respondents are asked to select the only correct answer out of the choices from a list. Multiple choice items consist of a stem, the correct answer, keyed alternative, and distractors. The *stem* is the beginning part of the item that presents the item as a problem to be solved, a question asked of the respondent, or an incomplete statement to be completed, as well as any other relevant information. The options are the possible answers that the examiner can choose from, with the correct answer called the *key* and the incorrect answers called *distractors*. Only one answer can be keyed as correct.
Some research questions of this research were as; (1). How are some good items in the test?; 2). How are some revised items in the test?; 3). How are some good alternatives (distractors) in each items?; and 4).How are some dropped alternatives (distractors) in each items?.

II. Methodology

Research design is as a logical progression of stages or tasks, from problem formulation to the generation of conclusions or theory, that are necessary in planning or carrying out a study (e.g., Creswell, 1997; Marshall & Rossman, 1999). Qualitative design will be implemented in this research. Qualitative is a term with varying meanings in educational research. Qualitative research studies typically include case studies, and generally descriptive studies. There writer analyzed the school final test of class IPS 2 for grade XII in the SMAN 2 Pangkal Pinang. There are 40 items of multiple choice test model in English subject. The istrument used in this research was final school test, documents and interviews. The evaluation program used Iteman analysis software. The assessment was conducted after students doing school final test progam.

III. Result and Discussion

3.1 Good Items

There writer analyzed the school final test of class IPS 2 for grade XII in the SMAN 2 Pangkal Pinang. There are 40 items of multiple choice test model. After analyzing, the result of good items are two items only, those are item number 12 and 31.These items are involved in good items because they have good or high for level of difficulty, discrimination power, alternatives and reliability.

The item number 12 can be shown from the test item below;

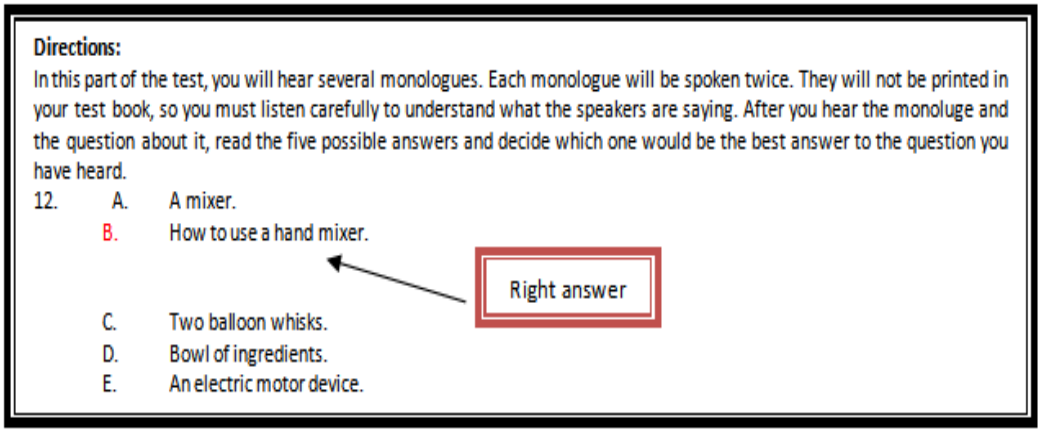

Fig 3.1 Item number 12th

From the item above, it can be described that the level of difficulty of this item is 0.28, and this means average and the discrimination power is 0.41, and this means high. The alternatives is also sufficient.

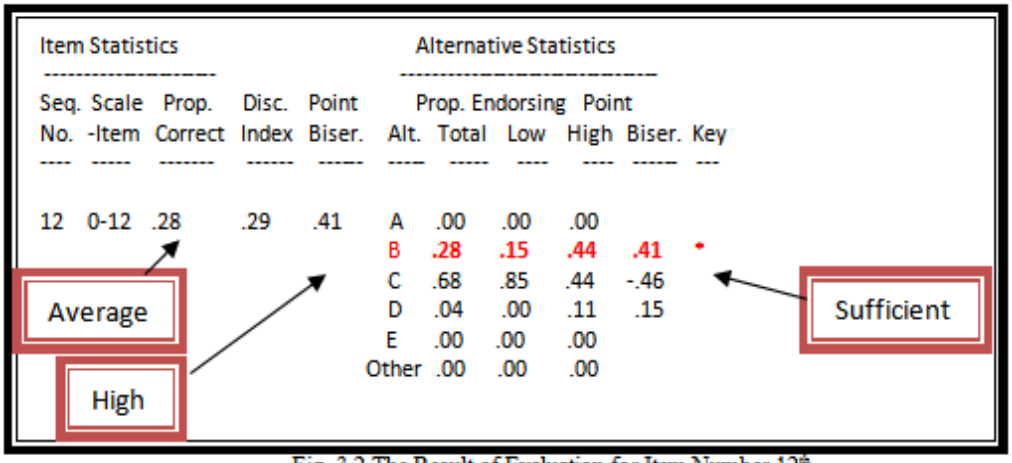

Fig. 3.2 The Result of Evaluation for Item Number 12th

And the item number 31 can be shown from the test item below;

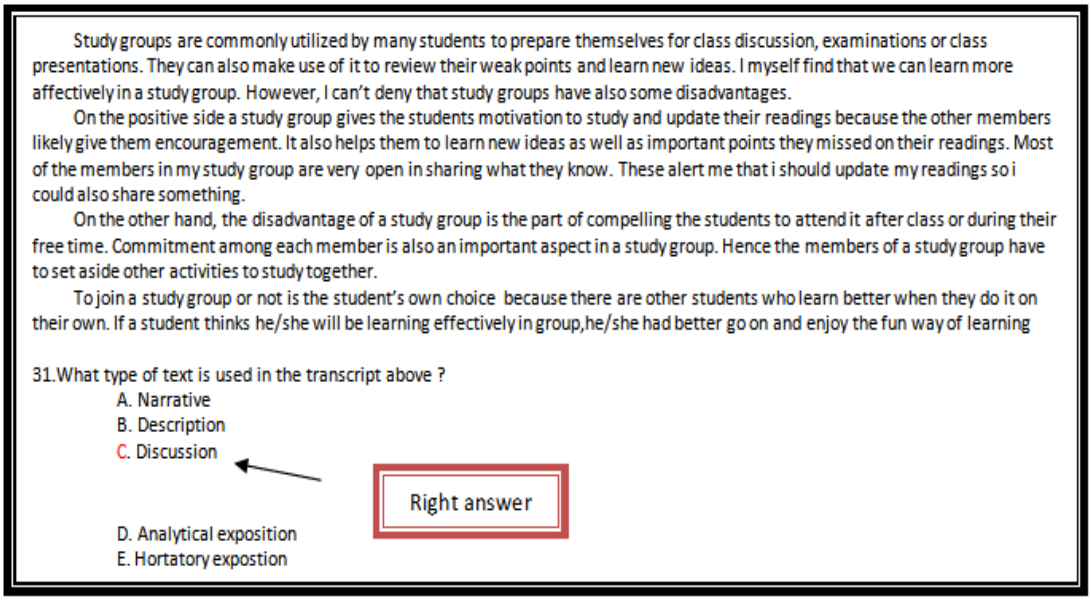

Fig 3.3 Item number 31st

From the item above, it can be described that the level of difficulty of this item is 0.72, and this means average and the discrimination power is 0.32, and this means average too. The alternatives is also sufficient.

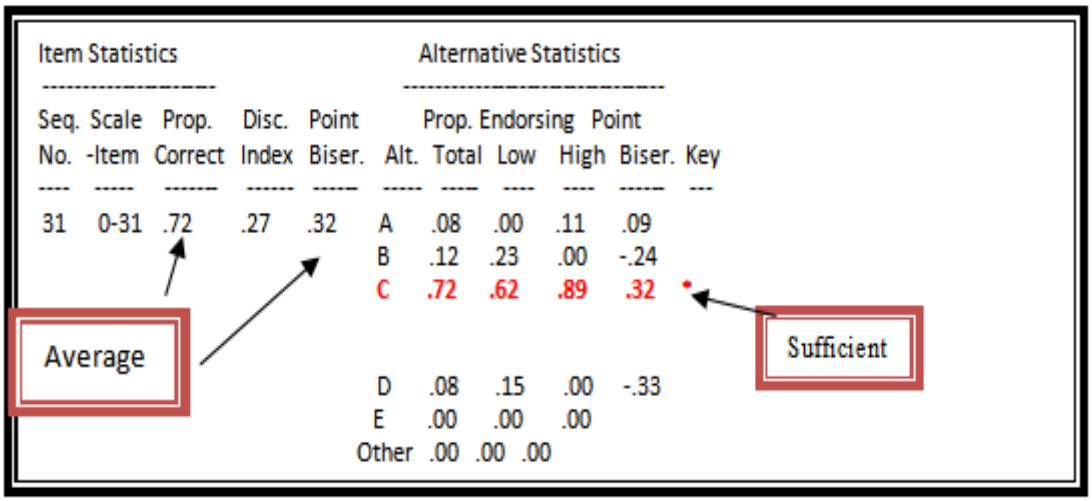

Fig. 3.4 The Result of Evaluation for Item Number 31st

From the analysis of good items above, it shows that the test is not effective because the test has only two items which are good items test, so the test maker or a teacher shoul consider and revise all items to be good test items.

3.2 Revised Items

Test items are also revised or altered because of some reason such as the item is too easy or too difficult for level of difficulty, very low of discrimination power, very low of proportion of the answer and low reliability. There are 17 items should be revised and 21 items should be dropped.Here are some examples of revised item for number 17 and 20. The question of number 17 can be shown below;

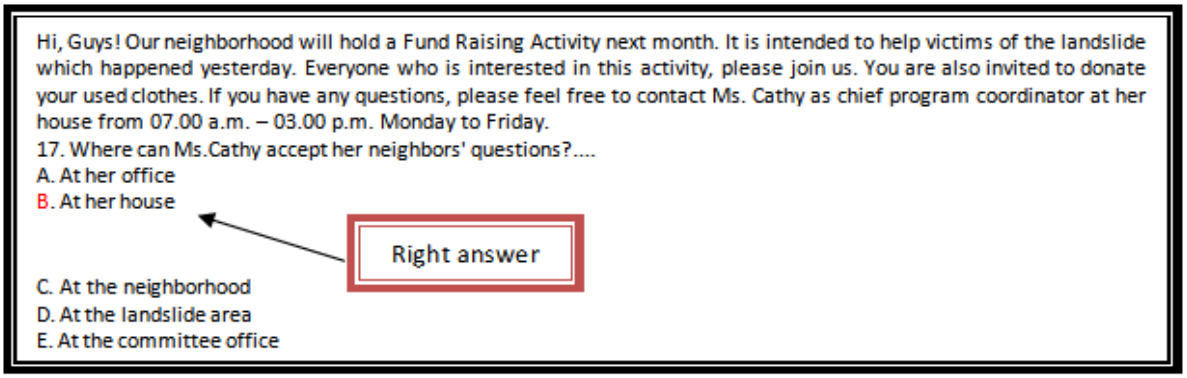

Fig. 3.5 Item number 17th

From the item above, it can be described that the level of difficulty of this item is 0.08, and this means difficult and the discrimination power is 0.64, and this means high. The alternatives is low students can not answer the right answer (0.00).

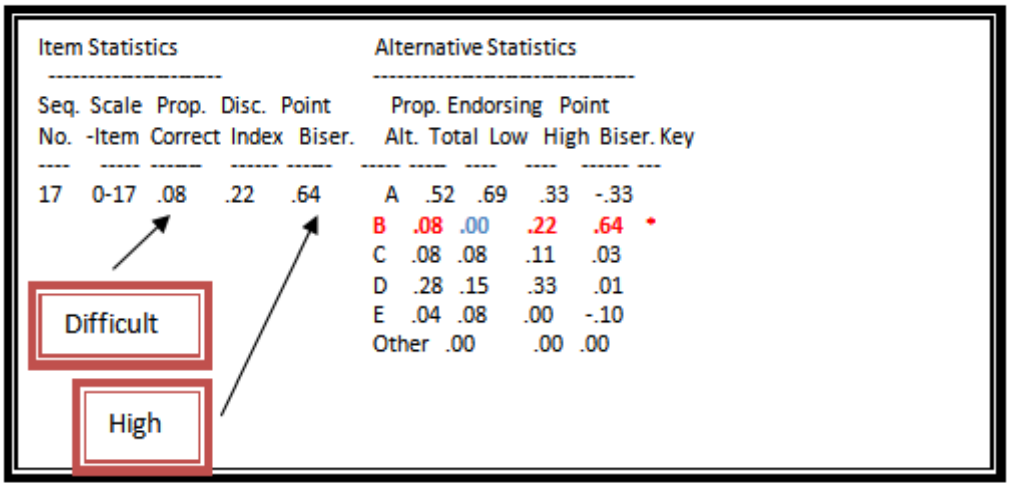

Fig. 3.6 The Result of Evaluation for Item Number 17th

The problem which is happened in this item because the low students are wrong intrepretation in doing scanning or skimming in reading comprehension. The reading text looks like an announcement at office, so they just answer *at the office* directly.
Then, the question of number 20 can be shown below;

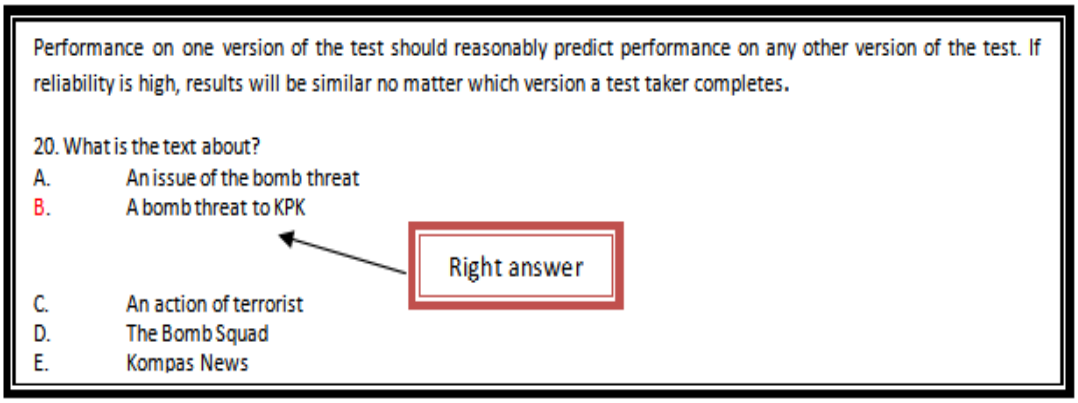

Fig. 3.7 Item number 20th

From the item above, it can be described that the level of difficulty of this item is 0.08, and this means difficult and the discrimination power is 0.70, and this means high. The alternatives is the low students can not answer the right answer (0.00), no students choose distractor (C), and most of students answer distractor (E).

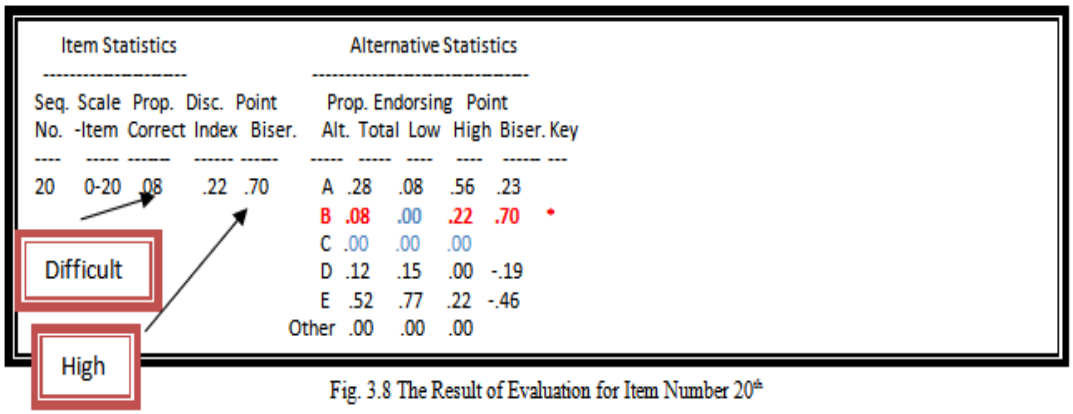

Fig. 3.8 The Result of Evaluation for Item Number 20th

The problem which is happened in this item because the low students are wrong intrepretation in understanding the questions in reading comprehension. Performance on one version of the test should reasonably predict performance on any other version of the test.

3.3 Good Alternatives (Distractors)

There are 200 alternatives (distractors) from this test. The result of analysis has been found that 146 distractors can be applied without revision and 54 distractors are considered to be revised or dropped.

Here are the examples of items for alternatives which can be applied without revision for number 13 and 27 items below;

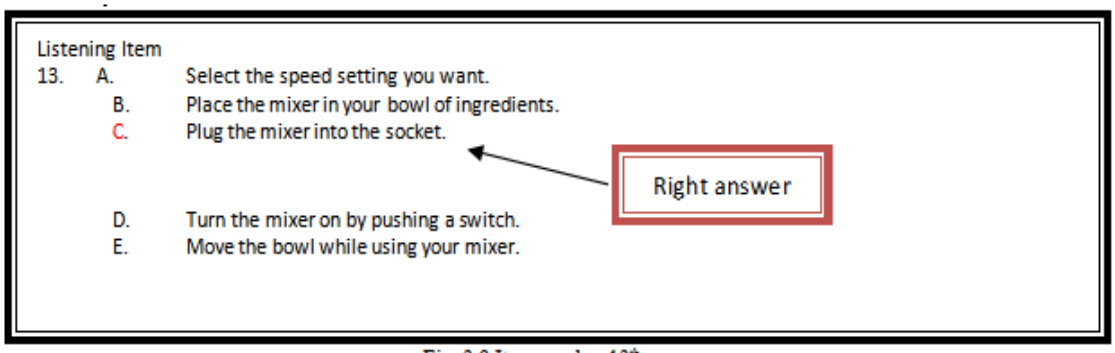

Fig. 3.9 Item number 13th

Fig. 3.10 The Result of Evaluation for Item Number 13th

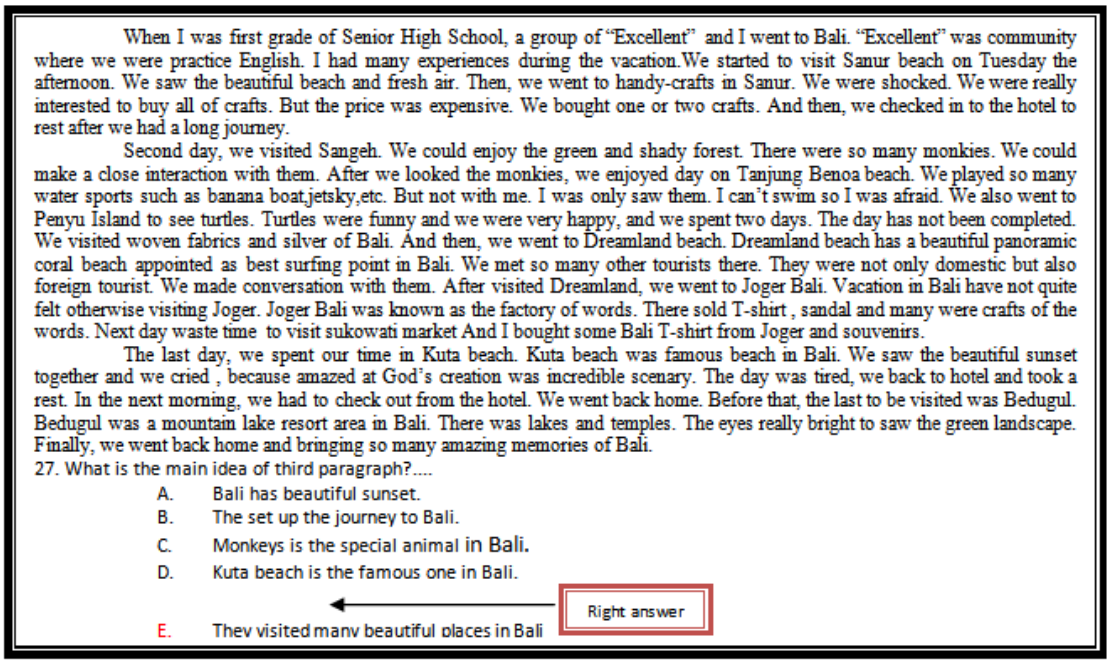

Fig. 3.11 Item number 27th

Fig. 3.12 The Result of Evaluation for Item Number 27th

2.4. Dropped Alternatives (Distractors)

Here are the examples of items for alternatives which can be dropped for number 37 and 39 items below;

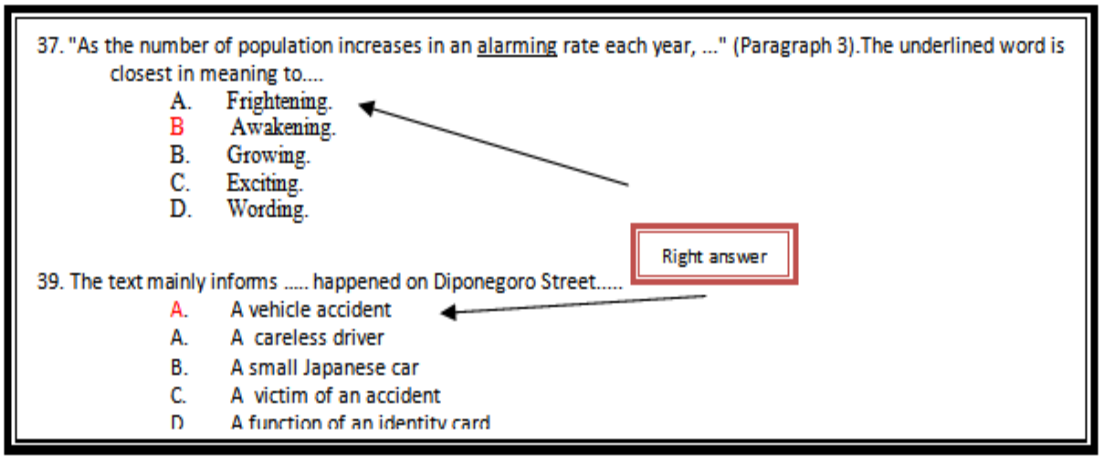

Fig. 3.13 Item number 37th and 39th

Item analysis is valuable for increasing instructors' skills in test construction, and identifying specific areas of course content which need greater emphasis or clarity. Separate item analyses can be requested for each raw scor created during a given score. Each items of test should involve good and high level of difficulty, discrimination power, proportion of alternatives or distractors and reliability.

Even after the test has been administered, statisticians and test developers review to make sure that test questions are working as intended. Before final scoring takes place, each question undergoes preliminary statistical analysis and results are reviewed question by question. If a problem is detected, such as the identification of a misleading answer to a question, corrective action, such as not scoring the question, is taken before final scoring and score reporting takes place.

IV. Conclusion

The *Iteman* program only prepared multiple choice items. In fact, teachers has priority to make kinds of test for final semester. In one test, it can be whole multiple choice items, essay items,combined both of them, moreover another kinds of items such as true false, matching, fill in the blank and the like. The researcher found difficult when she found many combination kinds of items in one test to fill in the answer key in *Iteman*. To overcome this problem, before the final test would be conducted, all lecturers must discuss together about category of test. The goal of this discussion was to find the similarity kinds of test for final test.

Researcher found that there were many revised and dropped items in the test, so, the institution should prepare test well. By this research, teachers and institution of school should make a better plan for the next test of items effeciently and accurately.

References

[1]. Valette, R. 1994. Teaching, testing and assessment: Conceptualizing the relationship. In C. Hancock (ed.). Teaching, testing and assessment: Making the connection . Lincolnwood, IL: National Textbook Company. 1–42

[2]. Kunnan, A. J. 1998. An introduction to structural equation modeling for language assessment research. Language Testing. 15.295–332.

[3]. VanderLinden, W. J. And Glas, C. A. W. (2000).Computerized Adaptive Testing: Theory and Practice. Boston: Kluwer.

[4]. Suen, H. K. And McClellan, S. (2003). Test Item Construction Technique and Principles. In: Encyclopedia ofVocational and Technological Education (Vol.1). pp.777798. Taipei: ROC Ministry of Education.

[5]. Sikolova, M., Slozilova E. And Svoboda, P. (2009). Developing an Entrance English Test at the University of Defence. Ekonomicka univerzita.

[6]. Stufflebeam, Daniel L (1978). Metaevaluation; An overview, Evaluation and the Health Profession, 1 (2) 146-163

[7]. Daniel, Larry and King, Debra A. (1998). Knowledge and Use of Testing and Measurement Literacy of Elementary and Secondary Teachers. Florida

[8]. Black, Paul and William, Dylan. (1998). Assessment in Education: Principles, Policy and Practice. Vol 5. Issue 1.

[9]. Stiggins, Richard J. (1994). Student-Centered Classroom Assessment. 2nd Edition

[10]. Cole, Agustine and Peterson. (1998). Development of Assessment Policy and Standard. Stanford University

[11]. Peterson, M. W. And Einarson, M. K. (2001). What are college doing about student assessment? Does it make a difference? Journal of Higher Education, 72, 629-669.

[12]. Ruch, G. M. (1924) **.**The Improvement of the Written Examination. Chicago: Scott, Foresman and Company

[13]. Kehoe, J. (1995). 'Writing multiple-choice test items.' Practical Assessment, Research & Evaluation, 4, 9. Retrieved 4 January 2006 from http://PAREonline.net/getvn.asp?v=4&n=9

[14]. Creswell, J. W. (1997). Qualitative inquiry and research design: Choosing among five traditions.Thousand Oaks, CA: Sage

[15]. Marshall, C., & Rossman, G. (1999). Designing qualitative research (3rd ed.). Thousand Oaks,CA: Sage

Nila Sari Dewi "Learning Assessment In Language Teaching Of Senior High School Final Test." IOSR Journal of Research & Method in Education (IOSR-JRME) , vol. 8, no. 3, 2018, pp. 04-12.