AN ANALYSIS ON ENGLISH SUMMATIVE TEST ITEMS

A JOURNAL

BERNASELA
F42107007

ENGLISH EDUCATION STUDY PROGRAM
TEACHER TRAINING AND EDUCATION FACULTY
TANJUNG PURA UNIVERSITY
PONTIANAK
2014

AN ANALYSIS ON ENGLISH SUMMATIVE TEST ITEMS

Research Journal

BERNASELA
F42107007

Approved by:

Supervisor 1	Supervisor 2
Drs. Bambang Wijaya	**Dewi Novita, S.Pd.,M.App.Ling**
NIP. 195011241979031005	**NIP. 197208172003122001**

Legalized by:

Dean of Teacher Training And Education Faculty	Head of Languages and Arts Department
Dr. Aswandi	**Drs. Nanang Heryana, M.Pd**
NIP. 195805131986031002	**NIP. 19610751988101001**

AN ANALYSIS ON ENGLISH SUMMATIVE TEST ITEMS

Bernasela, BambangWijaya, DewiNovita
English Education study Program, and Languages and Arts Education Department,
Teacher Training Education Faculty of Tanjungpura University in Pontianak
Email: Berna_sela@yahoo.co.id

Abstrak: Tujuan dari penelitian ini adalah untuk memberikan informasi tentang kualitas butir-butir soal sumatif bahasa Inggris, mengukur validitas, reliabilitas, tingkat kesukaran, daya pembeda dan distraktor butir-butir soal, mengidentifikasi butir-butir soal yang tepat untuk siswa kelas sebelas pada sekolah menengah kejuruan SMKN 2 Singkawang tahun akademik 2012/2013. Metode penelitian ini adalah penelitian deskriptif. Data penelitian di ambil dari soal-soal sumatif dan lembar jawab siswa. Soal tes summatif terdiri dari 50 butir soal dalam bentuk pilihan ganda. Berdasarkan analisis secara keseluruhan dapat di simpulkan bahwa ada 33 butir soal yang baik yang dapat di gunakan kembali pada tes sumatif berikutnya, 6 butir soal yang harus di eliminasi atau di ganti dengan soal yang lain dan 11 butir soal yang harus di revisi.
Kata kunci: Analisis butir tes, Tes Sumatif.

Abstract: The purposes of this research are to provide information about the quality of English summative test items, measuring the validity, reliability, level difficulty, discriminating power and distracters of the test items, identifying the appropriate test items for the eleventh grade student vocational school SMKN 2 Singakawang in academic years 2012/2013. The method of this research is descriptive study. The data were taken from the summative test and students' answer sheets. The summative test consists of 50 items in form of multiple choices. Based on the whole analysis of test items, it can be concluded that there are 33 good test items which still can be used to the next summative test, 6 test items should be discarded or changed by the other test item and 11 test items should be revised.
Key words: Item analysis, Summative test items.

Test is a tool of evaluation which has important role to measure teaching learning process. In the area vocational and technological education, tests are also used for many different purposes. A test can be used to information to guide the placement of students into appropriate classes or programs. Test is important parts of the teaching and learning process if they are integrated into daily classroom teaching and are constructed to be part of the learning process for the cumulating event. As stated by Hughes (2003:13) "a test is intended to measure students' achievement and the degree of success of the teaching learning program". Through testing, we can measure students' knowledge or ability, they allow students to see their own progress and allow teacher to make adjustment to their instruction on daily basis. In a simple term, a test is a method of measuring a person ability, knowledge, or performance in

a given domain therefore "test is also used to diagnose the strengths and weaknesses area of the students, it is to ascertain what learning still needs to take place" (Hughes, 2003:15).

In teaching learning activities, an achievement test is a systematic procedure for determining the amount of student has learned. There are two kinds of achievement test; formative test and summative test. In This research, the writer choose summative test as the kind of test which administered at the end of a unit or term, semester, or a year of study in order to measure what has been achieved both individual and by groups.

The writer chooses to analyze summative test because constructing good summative test items are more difficult and more time consuming than formative test. Comprehensiveness and representativeness of sampling are important features of summative test. It means that test objectives should be based directly on course objectives and test content is derived from specific course content. Although the emphasis of summative test is on measuring learning outcomes, it should not be implied that testing is to be done only at the end of instruction. It is essential to analyzed and tried on pupils before few best test items chosen for summative test. According to Heaton (1975) information concerning the performance of the students as a whole and of individual student is very important to teaching purposes, especially as many test result can show not only the types of error most frequently made but also the actual reasons for the error being made. After a test has been administered and scored, it is usually desirable to evaluate the effectiveness all the items. Item analysis is done by the studying the students' responses to each item. Item analysis information can tell us if a norm-referenced item was valid and reliable or not, too easy or too hard, how well it discriminated between high and low scores on the test, and whether all the alternatives functioned as intended.

According to Gronlund (1977), there are some benefits to do analysis of the test items: first, it is provide useful information for class discussion of the test. Second, it provides data that helps the students improve their learning. Third, it provides insight and skills that lead to the preparation of better test in the future. Item analysis is a process which examines student responses to individual test items (questions) in order to assess the quality of those items and of the test as a whole. It will show whether the test appropriate or not for the students. Item analysis is especially valuable in improving item which will be used again in later tests, but it can also be used to eliminate ambiguous or misleading items in a single test administration. In addition, item analysis is valuable for increasing instructors' skill in instruction, and identifying specific areas of course which need greater emphasis clarity.

Considering that English summative test items should fulfill the requirement of good test and by looking at the benefits of analysis of test items above the researcher regards that it is very important to conduct a research to report the quality of English summative test items related to the validity, reliability, level difficulty, discriminating power and the distractors of test items. It is expected can help the

teacher to identify / finalize the most appropriate test items and make sure that new test is designed to meet student need and truly reflect learning and fulfill the requirements of good test. Beside that, the researcher also expected that the information of the study can help the teacher to know the type of error most frequently made in constructing achievement tests and the actual reason for the error being made.

The result and the information of the analyses showed whether the teaching learning process is successful or not and also use as an instrument to help the teacher to make a good test items and prepares the better test in the future. Beside that the information of the analysis test items also can help the students to increase their learning by knowing their progress and their achievement after meet the truly reflect learning. Therefore, this study will analyze English summative test items for fourth semester of the eleventh grade student at SMK Negeri 2 Singkawang in academic year 2012/2013 and conclude about the result of the study.

METHOD

This research is descriptive study research. According to Best (2006: 118), "A descriptive study describe and interprets what is. It is concerned with conditions or relationships that exist, opinions that are held, processes that are going on, effects that are evident, or trends that are developing". This research has the purpose to describe the quality of summative test items for the fourth semester of the eleventh grade student vocational school SMK Negeri 2 Singkawang in academic year 2012/2013.

Population is the entire research objects which have characteristics. As stated by Sigh (2005:82), "Population or universe means, the entire mess of observations, which is the parent group from which a sample is to form." The population of this research is English Summative test items for the fourth semester of the eleventh grade students of SMK Negeri 2 Singkawang in academic year 2012/2013 and 94 students' answer sheets as the research objects which were taken from 3 classes of accounting department. As a part of population, the samples of this research are 50 English summative test items and 30 students' answer sheets which were taken by using simple random sampling. Simple random sampling refers to drawing a random sample from the population to assign each number of populations a distinct identification number (Ary et al., 2010). First, the researcher enumerated all the students' answer sheets in the population and assigned a number to each student's answer sheets. Then, the researcher wrote the students' answer sheets number on separated slips of paper, placed the pieces of paper in a container, shook the container and drew out a slip of paper, and continue the process until 30 slips of paper have been picked.

In order to conduct this research, the writer used the direct technique in collecting the data. The researcher analyzed directly the document of English summative test item and students' answer sheets of the first semester of second grade students at SMK Negeri 2 Singkawang in academic year 2012/2013.

First of all, the tests were administered and scored by the teacher. Next, the researcher collected the data and then analyzed the data based on the problem designed: validity, reliability, difficulty level, discriminating power and the distracter of the test items. The data was collected by using direct technique through documentary analysis. The data were taken from 50 multiple choices English test items and 30 students' answer sheets. In analyzing the data, the researcher used scoring matrix and table of student's answer as tools of data collecting.

FINDINGS AND DISCUSSION

a. Findings

1. Analysis of Content Validity

   The researcher used table of specification to analysis the content validity of the test items. From the analysis, the researcher found English summative test items for fourth semester of the eleventh (XI) grade students vocational school at SMK Negeri 2 Singkawang in academic year 2012/2013 covers almost all the materials has learned by the students. The item number 22 and 34 are not appropriate for the students in fourth semester because those items were not related to the material. In other word, 96% of English summative test items are valid. However, the content validity of the test is showed as valid test items, some of the test items could not categorized as good items because they still have problems in constructing multiple choice item.

2. Analysis of Reliability

   To get the result of analysis of the reliability the researcher used Kuder-Richardson formula (KR-21).

Table 1: The Calculation of Reliability

X	fx	χ	$\chi 2$
20	1	-15	225
21	1	-14	196
25	2	-10	100
26	1	-9	81
28	2	-7	49
30	3	-5	25
33	1	-2	4
34	1	-1	1
35	2	0	0
36	2	1	1
37	2	2	4

38	1	3	9
39	1	4	16
41	3	6	36
42	1	7	49
44	2	9	81
45	2	10	100
46	2	11	121

M= 35
$\sum f\ \chi 2$= 1676

$$KR21 = 1 - \frac{M\ (K - M)}{\mathrm{K}(\mathrm{S}^2)}$$

Where: $s = \frac{\sqrt{\sum_X 2}}{N}$

s =$\frac{\sqrt{1,676}}{30}$

s = $\sqrt{55.8}$

s = 7.4

$$r = 1 - \frac{35\ (50 - 35)}{50(7.4^2)}$$

$$r = 1 - \frac{35\ (15)}{50(54.76)}$$

$$r = 1 - \frac{525}{2{,}738}$$

$$r = 1 - 0.19 = 0.81$$

From the calculation it is found the coefficient of test item reliability is 0.81. Based on the classified coefficient of the test item reliability it means that the test items considers to high reliability.

3. Analysis of Difficulty Level

From the computation by using Heaton's formula, the result of data analysis of difficulty level of the item shows there are 53 moderate (M) test items, 7 difficult (D) test items. From the calculation of level difficulty the items which are belonging to revised (too difficult), moderate and easy as follow:

a. The items that belong to the moderate test items are the item number 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 32, 33, 35, 36, 37, 38, 39, 42, 43, 45, 46, 47, 48, 49, and 50.
b. The items were categorized in difficult test items are the item number 17, 25, 31, 34, 40, 41, and 44.

The formula for computing the mean of index difficulty level is as follows:

$$M = \frac{\sum \text{LD}}{\text{n}}$$

$$M = \frac{23.5}{50}$$

LD = 0.47

4. Analysis of Discriminating power

The researcher used Gronlund's formula to get the criteria of the discriminating power of the items. From the calculation, the researcher founds there are 10 excellent test items, 8 good test item, 10 moderate test items, 16 moderate test items and 16 revised test items in discriminating upper and lower group students (appendix 3). From the calculation of discriminating power the items which are belong to revised, moderate, good and easy as follow:

a. The items that belong to the excellent items are item number 4, 18, 21, 25, 29, 31, 33, 35, 38 and 42.
b. The items that were categorized in good test items are item number 15, 16, 20, 28, 32, 37, 43 and 45.
c. The item that classified as moderate test items in discriminating higher and lower student are the item number 3, 6, 8, 11, 12, 19, 23, 26, 27, 30, 36, 39, 46, 47, 48 and 49.
d. The item number 1, 2, 5, 7,9,10, 13, 14,17, 22, 24, 34, 40, 41, 44 and 50 were classified as poor test items and should be revised.

The formula for computing the mean of discriminating power is as follows:

$$M = \frac{\sum \text{DP}}{\text{n}}$$

$$M = \frac{11.94}{50}$$

DP = 0.23

5. Analysis of Distractors

   To determine the effectiveness of the distractors, the researcher comparing the students in the upper and lower group who selected each incorrect alternative (distractor). From the data analysis of distractor the researcher found there are 100 effective and functions well distractors, 9 poor distractors, and 41 ineffective distractors including the absence distractors from 150 distractors. The problem is the test items were not fulfilled the rules in constructing the multiple choice item. The distractors, which are considered poor can still used in an item because at least those distractors fulfilled they function although attracted more students in the upper group. The ineffective distractors in test items were not categorized as good test item should be discarded or revised to make those multiple choice items better.

   The Percentages of Distractors:

   a. Percentage of function well distracters:

      $$\frac{100}{150}\ 100\% = 66.7\%$$

   b. Percentage of poor distracters:

      $$\frac{9}{150}\ 100\% = 6\%$$

   c. Percentage of ineffective distracters:

      $$\frac{41}{150}\ 100\% = 27.3\%$$

The aim of this study is to provide information about the quality of English summative test item whether it can be classified as a good or poor item. According to Gronlund there are some criteria to determine which items that still can be used, revised or should be discarded. The researcher determined which item still can be used, need revision or discarded through some analysis of test items which were

collected in table of items classification. The researcher categorized the test items which can still be used, need revision and discarded as follow:

1. The items which are categorized in good test item and still can still be used are the item number 3, 4, 6, 8, 11, 12, 15, 16, 18, 19, 20, 21, 23, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37, 38, 39, 42, 43, 45, 46, 47, 48 and 49.
2. The items which are categorized can still be used but need revisions are the item number 1, 2, 5, 7, 9, 10, 13, 14, 24, 25, and 50.
3. The items which are categorized should be discarded are the item number 17, 22, 34, 40, 41 and 44.

b. Discussion

As the result of the analysis based on the whole analyzing and criteria of a good test and the rules of constructing multiple choice item the researcher found there are 11 test items which are need several revisions. Research finding show that some of the items were not fulfill the requirement of good test because they are too easy, too difficult and fail discriminate upper and lower group students.

1. The item fail discriminate upper and lower group affected by the distractors did not function because those distractors has verbal clues that enable students select the correct answer and eliminate those incorrect alternatives.

Example:

1. Item number 1
   Look at the picture and choose the statement that best describe what you see in the picture

   A. The bicycle is on the sidewalk
   B. There is a bicycle inside the house
   C. The woman is riding a bicycle
   D. The man is repairing a bicycle

   The distractors B and C has verbal clues (the man and The Woman) that enables students know those distractors are not the correct answer. It can be replaced by the following distractors:

   - The bicycle is facing the street
   - The bicycle is in the middle of street

2. The items fail discriminated upper and lower group affected by the ineffective distractors are not plausible and attractive to the uninformed that enable students select the correct answer and eliminate those incorrect alternatives.

Example:

1. Item number 5

   Which do you think will suit me better, the black gown or the red one?

   A. Please sit down and relax
   B. The black one looks more elegant
   C. You better buy it yourself

   The distractors A and D can be replaced by the following distractors:

   - The black gown suit me better than you
   - The red gown is shinning in the sun

3. The item which need revision because it is too easy test item and affected by the answer stated in the stem clearly.

Example:

1. Item number 10

   Man : One ticket to Solo Balapan station, please.
   Woman : Here you go. The next train will leave at 2:00 PM
   Man : Thank you

   When will the train leave?

   A. 12:00 AM
   B. 01:00 PM
   C. 12:30 PM
   D. 02:00 PM

   It is better if the question was constructed as follow:

   Man : One ticket to Solo Balapan station, please
   Woman : Which do you prefer, at 7 AM, 10 AM, 12.00 PM or 2 PM?
   Man : at noon please.
   Woman : Here you are.
   Man : Thank you

   What time the man will leave?

   A. 07:00 AM
   B. 10:00 AM
   C. 12:00 PM
   D. 02:00 PM

4. The test items which need revision because it is fail to discriminate the upper and lower group. These items fail discriminated upper and lower group affected by the distractors did not function because those distractor are not plausible and attractive to the uninformed. Beside that the direction of this item was wrong, it is should be questions number 14 and 15 refer to this procedure text. This problem can make the students confuse in answering the question.

Example:

1. Item number 14

   Want to make a simple dessert for your party? Try this caramel pudding. You only need to prepare four eggs, 200 cc of milk, 200 grams of sugar, and a half table spoon of vanilla. First of all, you need to melt100 grams of sugar, so it becomes caramelized. And then pour it into a Pyrex dish and stem it for about 45 minutes.

   What process did the speaker describe?

   A. How to stir the eggs
   B. How to melt sugar
   C. How to choose good eggs
   D. How to make a simple dessert

   The distractors A and B can be replaced by the following distractors:

   - How to order a dessert in a restaurant
   - How to prepare a party

5. The test item which need revision because it is too difficult test item which is affected by the intended answer is not correct or clearly best or the distractor also can be the answer.

Example

1. Item number 25

   Shop assistant : May I help you?
   Listy : Yes, the cardigan I bought yesterday was torn at the end of the right sleeve.
   Shop assistant : Let me check it. Oh we are sorry, …
   Listy : Oh, what a pity!

   A. All kinds of cardigans are sold out
   B. All items bought cannot be returned
   C. I will change it with the new one
   D. All customers will get a discount

   The distractor A also can be the answer, it can be replaced by this distractor:

- I will change it with the other color

6. The test item which need revision because it is too easy test item which is affected by the words in the stem of item number 50 are too less and there are repeating same material in each alternatives.

Example:

1. Items number 50

   Junius Alberto Lumban Tobing spends at least eight hours a day in front of his computer. The programs he likes to use more are Adobe Photoshop and Adobe Illustrator. Junius thinks with these two programs all the time to increase his ability in the field of graphic design. "My work experience of three years with the advertising company made me ready for competition", said Junius. 'Becoming a graphic design is not easy, you need diligence and must be skilled in operating computer. This is the profession so we designer must be accorded high appreciation,' he said

   What are the programs Junius likes to use?
   A. Adobe Audition and Adobe Photoshop
   B. Adobe Photoshop and Adobe Illustrator
   C. Adobe Audition and Adobe Illustrator
   D. Adobe Photoshop and photo story

   It is better if the question was constructed as follow:

   What are the programs Junius use to increase his ability?
   A. Graphic design and experience in operating computer
   B. Photoshop and illustrator
   C. Work experience and advertising
   D. Diligence and skill in operating computer

CONCLUSION

Based on the analysis of the English summative test item for the fourth semester of the eleventh grade students in academic year 2012/2013 the researcher draws the conclusion that there are 33 good test items which still can be used for the next summative test, 6 test items should be discarded or changed by the other test item and 11 test items should be revised. The ineffective distractors in test items which are need revision should be replaced since as stated in the discussion the ineffective distractors affected the test item both the difficulty level and discriminating power of the test items.

BIBLIOGRAPHY

Ary, Donald., Jacobs, Lucy C., Sorensen, Christine K. 2010. *Introduction to Research in Education*, eight edition. Canada : Nelson Education Ltd.

Best, John W & Kahn, John V. 2006. *Research in Education*, tenth edition. United Stated of America : Pearson Education Inc.

Gronlund, N.E., 1977. *Constructing Achievement Test,* 2nd Edition. New York: Prentice Hall, Inc. University of Illinois.

Heaton, J.B. 1975. *Writing English Language Test*. England: Longman Group Limited.

Hughes, A. 2003. *Testing for language teachers, second edition*. Cambridge: Cambrige University Press.

Singh, Kumar Y. 2006. *Fundamental of Research*. Ansari Road, Daryaganj. New Dehli. New Age International Publishers Ltd.