ANALYZING ENGLISH SUB SUMMATIVE TEST ITEMS (MID EXAM)

A RESEARCH JOURNAL

By:
HERLINA IDA
F42108027

TEACHER TRAINING AND EDUCATION FACULTY
TANJUNGPURA UNIVERSITY
PONTIANAK
2014

ANALYZING ENGLISH SUB SUMMATIVE TEST ITEMS (MID EXAM)

A RESEARCH JOURNAL

By:
HERLINA IDA
F42108027

Approved By:

Supervisor I	Supervisor II
Drs. Bambang Wijaya	**Dra. Hj. Endang Susilawati, M.Si**
NIP.195011241979031005	**NIP. 196202261989032002**

Legalized By:

The Dean of Teacher Training and Education Faculty	The Chairman of Language and Art Department
Dr. H. Martono	**Drs. Nanang Heryana, M. Pd**
NIP. 196803161994031014	**NIP. 196107051988101001**

ANALYZING ENGLISH SUB SUMMATIVE TEST ITEMS (MID EXAM)

Herlina Ida, Bambang Wijaya, Endang Susilawati
English Education Study Program, Languages and Arts Education Department, Teacher Training Education Faculty of Tanjungpura University Pontianak
Email: line_may87@yahoo.com

Abstrak: Tujuan dari penelitian ini adalah untuk memberikan informasi tentang kualitas butir-butir soal tes sub sumatif bahasa inggris, mengukur validitas, reliabilitas, tingkat kesukaran, dan daya pembeda butir-butir tes yang dibuat oleh guru bahasa inggris SMA Kartini tahun ajaran 2013/2014. Metode penelitian ini adalah penelitian deskriptif. Data penelitian diambil dari soal-soal sub sumatif dan lembar jawaban siswa. Soal tes sumatif terdiri dari 20 butir soal dalam bentuk pilihan ganda. Berdasarkan hasil penelitian, dapat disimpulkan bahwa secara keseluruhan tingkat validitas butir soal adalah cukup dengan persentase 75%, nilai reliabilitas soal adalah 0.34% yang dikategorikan sebagai reliabilitas rendah, nilai tingkat kesukaran adalah 55% dimana kebanyakan butir tes adalah soal yang mudah, dan nilai daya pembeda menunjukkan bahwa butir tes yang lemah adalah 30%, sementara butir tes yang memuaskan adalah 70%.

Kata Kunci: analisis butir tes, tes sub sumatif

Abstract: The aim of this research is to provide information of English sub summative test items, measuring the validity, reliability, level difficulty, and discriminating power of the test items made by English teacher of SMA Kartini in academic year 2013/2014. The method of this research is descriptive research. The data were taken from sub summative test and students' answer sheets. The sub summative test consists of 20 items in the form of multiple choice. Based on the research findings, it can be concluded that as a whole the content validity of test items is sufficient with percentage is 75%, the point of test reability is 0.34 which is categorized Low (L) reliability, the point of level of difficulty is 55% which is categorized as easy items, and the point of discriminating power showed the percentage of poor items is 30 %, while the satisfactory items are 70 %.

Key Words: item analysis, sub summative test

In process of teaching learning, testing should be considered during the planning of instruction as a way for teacher to make decisions in instruction. Specifically, there are three decisions that teacher must make from the beginning of instruction to the end; decisions at the beginning of the instruction, during instruction and at the end of instruction. According to Zimmerman et al the purposes of testing can be divided into four; to set up a basis for assigning grades, to determine how well each student has achieved the course objectives, to diagnose student problems for remediation, and to determine where instruction needs improvement (1990:1).

Test is a tool to collect intended information with the aid of a numerical scale or a category system. Good test items should be made by considering some criteria such as reliability, validity and suitability for students.

According to Gronlund (1977) achievement test which is used to certify students' progress at the end of a course or unit of instruction is called summative test. It is given at the end of period of instruction. Summative test is designed to measure the broad range of learning outcome expected at the end of instruction. It usually gives at the end of a marking period and measures the 'sum' total of the material. The scopes of materials of summative test are wide because it consists of several subjects.

Test that covers several subjects is called sub summative tests. This test includes a number of specific teaching materials that have been taught in certain time, aiming to obtain a description of students' absorption to improve their achievement. The results of the test are used to improve the teaching and learning process and taken into account in determining grades. In SMA Kartini Sungai Kakap, sub summative is known as midterm test (ulangan tengah semester). Sub summative test are organize in the middle of first semester. This test only includes teaching materials that have been taught in first three months.

Item analysis is processes which examine students' responses to individual test items (questions) in order to assess the quality of those items and of the test as a whole. Based on pre observation, the researcher found that Teaching English at SMA Kartini is based on curriculum or usually called KTSP (Kurikulum Tingkat Satuan Pendidikan). With KTSP, each school can develop their own syllabus, curriculum and indicators. Although determining their own syllabus, the competence standard and the content have to be accordant with Education National Standard Board (Badan Standar Nasional Pendidikan atau BSNP).

Based on the principles of KTSP, the teacher is enabled to develop his/her own curriculum, in other words, he/she can develop his or her own materials as well as the relevant instruments for the assessment. However, the teacher of the school only use a textbook as the source of the material and also take test items from the book for her sub summative test. Considering such a condition, this research was purposed on analyzing how are the sub summative test items (mid exam) made by English teacher of SMA Kartini Sungai Kakap in Academic Year 2013/2014?"

There are three criteria for testing a test: validity, reliability, and practicality (Harris, 1969:13). In this research, the researcher only focuses on validity and reliability because these two criteria are more significant in constructing test item. Validity is the important variable of a measurement instrument. Heaton (1975:153) said "The validity of a test is the extent to which it measures what it is supposed to measure and nothing else". The validity of a test must be considered in measurement in this case there must be seen whether the test used really measures what are supposed to measure, briefly. Validity also becomes the first form of evidence relates to the content of the test (Hughes, 2003:26).

Validity is one of the important criteria of a good test. Validity has important roles to measure the quality of the test items. The validity shows that the test item will be appropriate for the students. However, only content validity is discussed in this research where it depends on a careful analysis of the language being tested and of particular course objectives. The provider of test item constructs the test to contain a representative sample of course, the relationship between the test items and the course objective is always being apparent. The test has content validity if the objectives are stated in the curriculum. The sample of activities to be included in a test is as representative of the target domain as possible. The provider of test item should compare the test with the materials which are stated in curriculum to know whether the test has content validity or not. In sum, the test has high content validity

if the test items cover the materials stated in the curriculum. Content validity can be showed by doing a careful analysis of the language being tested and of particular course objectives. In doing analysis, table specification is used to identify the achievement domains being measured and to ensure that a fair and representative sample of questions appear on the test. The use of lesson plan will add information about constructed test. From the instructional objectives in lesson plan, the researcher can ensure that the test measure what is supposed to measure.

Reliability or consistency is a necessary characteristic of any good test (Heaton, 1975:155). A test is reliable if when it has consistency in its measurement. For example, if the test is administered to the same candidates on different occasion, it will not produce different result. A test made by teacher may have potential of errors. By computing the reliability, teacher can minimize it with main purpose is to determine how much of the variability in test scores is due to measurement error and how much is due to variability in true scores. Measurement errors are essentially random: a person's test score might not reflect the true score because they were sick, anxious, in a noisy room, etc.

According to Heaton (1975:172) the difficulty level shows how easy or difficult the test item when the students take the test. It is generally expressed the percentage of the students who answer the item correctly. It can be said if the more difficult of item is the fewer will be the students who select the correct opinion. And the easier the test is the more will be the students who selected the correct one. A good test item must be appropriate to student's ability. The test must be appropriate to material that they have learned in teaching and learning activity. The values of difficulty item was being distributed in the following manner: easy items, 5%; items of medium-low difficulty, 20%; items of medium difficulty, 50%; medium-hard items, 20%; and difficult items, 5%.

Discrimination power refers to how well the items perform in separating the better students from the poorer ones. If the good students tend to do well on an item and the poor students badly on the same item, then the item is a good one because it distinguishes the good from the bad students. To estimate the discriminating power of test item is comparing the number of students in the upper and lower group who answered the item correctly. According to Kitao (1997), discrimination power also gives more useful information about the strengths and weaknesses of an item.

METHOD

This research were descriptive research. The reason why she used this method is because descriptive method can describe the events that happened systematically and accurately. Marcsyk, DeMatteo, and Festinger (2005:209) stated that descriptive statistics allows the researcher to describe the data and examine relationships between variables. Descriptive statistics are used to describe the data collected in research studies and to accurately characterize the variables under observation within a specific sample.

The population of this research was students' answer sheets of sub summative test (mid exam) first semester of SMA Kartini (vague name) Sungai Kakap in academic year 2013/2014. From that population, twenty one students' answer sheets were taken as the sample.

In order to solve the problems objectively in this research, the researcher uses the document analysis to collect data. The researcher collected the data of related information including the result of tenth grade student's mid exam test of the first

semester. First of all, test was administered and scored by the teacher. Next, the researcher collected the data that was scored by the teacher and then analyzed based on the problems designed: validity, reliability, level of difficulty and discriminating power of the test items. The tool of data collecting used in this research was a set of test item which consist of 20 multiple choice English test items as a tool where the time given to do the test was 35 minutes. The data to be analyzed in this study is taken from the students' answer sheets of the English sub summative test items for first semester of tenth grade SMA Kartini Sungai Kakap in academic year 2013/2014. Next, the researcher put the data found into the table form and finally make the conclusion from the data analysis for each problem.

With regard to validity, the researcher only concern with the content validity. Therefore, the table of specification is analyzed to see whether test is valid or not. The purpose of a table of specifications is to identify the achievement domains being measured and to ensure that a fair and representative sample of questions appear on the test. A Table of Specifications provides the teacher with evidence that a test has content validity, that it covers what should be covered.

In order to see the percentage of content validity, the test items are studied in terms of the conformity between the sub summative test items and table of specification. Thus, to analyze the quantitative data of this research, a formula is used to analyze the content validity of English sub summative test of the tenth grade students of SMA Kartini Sungai Kakap. Then, the researcher compared the percentage with the criteria.

FINDINGS AND DISCUSSION

Findings

In this part the results of the test analysis are presented in order to answer the research questions concerning the content validity, reliability, level of difficulty, and discriminating power.

1. Analysis of content validity

   The researcher used table of specification to match the test items with table of specification.

Table 1
The Correspondence between the Sub Summative Test Items and Table of Specification

No	Learning objective	Indicators	Items Number	Correspondence	Topic of learning
1	Siswa dapat menggunakan berbagai tindak tutur dalam wacan lisan interpersonal/transaksional: berkenalan, bertemu/berpisah,	Menggunakan berbagai tindak tutur dalam wacan lisan interpersonal/transaksional: - Bertemu - Berpisah	1 2,3	 √ √	Functional text of greeting
2	Siswa dapat mengidentifikasi informasi dalam teks tulis	mengidentifikasi informasi dalam teks tulis fungsional dan	13, 14,15,16,17,18	√	Reading comprehension

	fungsional dan teks berbentuk narrative	teks berbentuk: narrative			
3		mengidentifikasi informasi dalam teks tulis fungsional dan teks berbentuk: descriptive text	4,5,6,7,8	Not correspondence	Reading comprehension
4		Menggunakan teks lisan fungsional pendek berbentuk: - surat - iklan	9,10,11,12 19,20	√ √	Functional text of learning

Based on the table above, the writer found that there are 15 number of item tests that correspondence with the table of specification. The rest of 5 items is not find in table specification. Thus, the computation of percentage is as follows:

$P = \frac{F}{N}$ x 100

Where:
P = percentage
F = frequency of conformity
N = number of sample

$P = \frac{F}{N}$ x 100 %

$P = \frac{15}{20}$ x 100 %

= 75 %

Thus, the percentage of conformity between the sub summative test items with the table of specification is 75 %. According to Arikunto, this percentage is categorized as Sufficient. It means that the content validity is good enough because the result of correspondence between the sub summative test items and table of specification, most of the test items cover the materials in the half of first semester.

2. Analysis of Reliability

The researcher used Kuder-Richardson 20 (r11) formula to get the coefficient of reliability of the test item. In calculating the coefficient of reliability, the value of mean score and the standard deviation need to find out. Then, the researcher needs to calculate the deviation and the squared deviation of the scores so the value of mean score and the standard deviation can be calculate. The following is the table of deviation and squared deviation.

Table 2
Standard deviation and squared deviation

No.	Score	Deviation (d)	Squared (d^2)
1.	15 (deviates from 12.24 (mean)	3	9
2.	15	3	9
3.	15	3	9
4.	15	3	9
5.	14	2	4
6.	14	2	4
7.	14	2	4
8.	14	2	4
9.	14	2	4
10.	13	1	1
11.	13	1	1
12.	12	0	0
13.	12	0	0
14.	12	0	0
15.	12	0	0
16.	11	-1	1
17.	11	-1	1
18.	10	-2	4
19.	8	-4	16
20.	7	-5	25
21.	6	-6	36
Total : 257			**Total : 141**

Furthermore, the researcher will calculate the mean of scores and the standard deviation. The following is the computation:

a. Mean = 12,24

b. Standard deviation

$$S = \frac{\sqrt{\sum_x 2}}{N}$$

$$S = \frac{\sqrt{141}}{21}$$

$$S = \sqrt{7}$$

S = 2.6

The computation of reliability:

$$\text{R11} = \frac{N}{N-1}\left(1 - \frac{m(N-m)}{N\text{x}^2}\right.$$

$$\text{R11} = \frac{20}{20-1}\left(1 - \frac{12.24\ (20-12.24)}{20.7}\right)$$

$$\text{R11} = \frac{20}{19}\left(1 - \frac{12.24\ (7.76)}{140}\right)$$

$$\text{R11} = 1.05\ .\left(1 - \frac{95}{140}\right)$$

$$R11 = 1.05 . (1 - 0.67)$$
$$R11 = 1.05 . (0.33)$$
$$R11 = 0.34$$

From the calculation it is found that the coefficient of reliability test item is 0.34. based on the classified coeficient of the test items reliability it means that the test items considers to low reliability.

3. Analysis of level difficulty

From the computation by using Heaton's formula, the result of data analysis of level difficulty shows there is 1 difficult test item and there are 3 moderate test items. From the calculation of level difficulty, which belong to revised and easy items as follow: (1) test items classified as revised items are the item number 5, 6, 15, 16, and 17. It is because these items is having low level of difficulty; (2) test items classified as easy items are the item number 1, 2, 3, 4, 7, 9, 11, 12, 14, 19, and 20.

From those 20 items, most the English sub summative test items for first semester of tenth grade SMA Kartini Sungai Kakap in academic year 2013/2014 can be categorized as moderate items with the mean of level difficulty is 0.56.

4. Analysis of discriminating power

The researcher used Heaton's formula to analyze the discrimination power of the test item. It tells how well the item performs in separating the upper group and lower group students. From the calculation, the results of data analysis are as follows:

a. 8 items are categorized as poor items. They are item number 2, 4, 5, 9, 10, 15, 16, and 19.
b. 4 items are categorized as satisfactory items. They are item number 3, 6, 12, and 20.
c. 7 items are categorized as good items. They are item number 1, 7, 8, 11, 14, 17, and 18.
d. 1 item is categorized as excelent item. That is item number 13.

Discussion

1. Content validity

The purpose of this part is to answer the research questions on the basis of that evidence. Moreover, the research questions are answered using the data from the analysis to provide a better understanding of the result. It is important for a test item to adequately cover the lesson materials that are supposed to be measured.

The analysis shows positive result. The content of English sub summative test items for the first semester of the tenth grade SMA Kartini Sungai Kakap in academic year 2013/2014 mostly cover the lesson materials showed by table of specification. However, the content validity of the test that is showed as valid test items can be categorized as good items. Moreover, based on the analysis of the items that conform to the table of specification, it clearly describes the content validity of the test.

2. Reliability

The reliability for the English sub summative test items for the first semester of the tenth grade SMA Kartini Sungai Kakap in academic year 2013/2014

is estimated with the Kuder-Richardson (KR 20) reliability coefficient. The reliability value of the test is 0.34 which is categorized as low reliable test items.

3. Level Difficulty

Based on the analysis level of difficulty is found that 11 items are categorized as easy. The easy items can be improved to be used in next the test. 3 items are categorized as moderate, 5 items are categorized as revised, and only one item is categorized as difficult. The moderate and difficult items can be used in the next sub summative test. The too difficult items can be revised or rejected in the next sub summative test.

4. Discriminating Power

The discriminating power of the test is estimated with the Heaton's formula. Most of items test is categorized as poor. The discriminating power coefficient is lower than expected. The poor items should be rejected or improved in the next test. Moreover, the mean of analysis does not clarify the entire test items have the poor coefficient of discriminating power. Based on the analysis, there are 8 items which categorized as poor items. They are item number 2, 4, 5, 9, 10, 15, 16, and 19. These test items need the improvement to be the good test items. Next, 4 test items number 3, 6, 12, and 20 are categorized as satisfactory items. These items usually needing and being subject to improvement. 7 items are categorized as good items. They are item number 1, 7, 8, 11, 14, 17, and 18. These items subject to improvement. And only 1 item is categorized as excelent item, that is item number 13.

CONCLUSION AND SUGGESTION

Conclusions

Based on the result of data analysis of English sub summative test items for first semester of tenth grade SMA Kartini Sungai Kakap in academic year 2013/2014 the researcher draws conclusions that the English sub summative test items is categorized as sufficient items which most of the test items cover the materials in the first semester. The percentage of content validity is 75%, the point of test reliability is 0.34 which is categorized Low (L) reliability, the percentages of easy items are 55%, the moderate items are 15 %, revised items are 25 %, and the difficult items are 5%. It can be concluded that the test constructor should improve the easy items if they want to use in the next test. The moderate and difficult items can be accepted and used in the next test, the analysis of test showed that the percentage of poor items is 40 %, the satisfactory items are 20 %, the good items are 35 %, and the very good item is only 1 %.

Based on the conclusion above, English sub summative test items for first semester of tenth grade SMA Kartini Sungai Kakap in academic year 2013/2014 has some test items that need to revise because they are categorized as easy and poor items. However, the problems above should be resolved to construct the better English sub summative test in the next test.

Suggestions

After describing the conclusion above, the researcher advances suggestion as follows: (1) test constructor should hold the training of how to construct test based on criteria of good test which includes validity, reliability, discriminating power, and level difficulty. (2) Teachers should develop the materials in the syllabus with the

suitable materials for students in SMA Kartini Sungai Kakap, and do not only depend on the text book but also in line with the student's need and ability, (3) the Teachers should check the test items before it given to the students to avoid error in typing that can make students confused in answering the question.

BIBLIOGRAPHY

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

Harris, David P. 1969. *Testing English As a Second Language*. New York: Mc. Grow Hill.

Heaton, J.B. 1975. *Writing English Language Test*. New York: Longman.

Hughes, Arthur. 2003. *Testing for Language Teachers, 2nd Edition*. Cambridge University Press, UK

Kitao, S. K. & Kitao, K. (1997). *Writing a good test.* http://www.ling.lancs.ac.uk/staff/visitors/kenji/kitao/design2.htm (Retrieved April 8, 2014)

Zimmerman, Beverly B, et al. 1990. *How to Prepare Better Tests: Guidelines for University Faculty.* Brigham Young University Testing Services and The Department for Instructional Science.

Zucker, Sasha. 2003. *Fundamentals of Standardized Testing.* Pearson Inc. San Antonio.