SMCC Higher Education Research Journal
ISSN Print: 2449-4402 · ISSN online: 2467-6322
Volume 4 · August 2017

Item Analysis of English Final Semester Test of the Third Year Students of the English Department of SMAN I Kupang

WALDETRUDIS MBEWA
http://orcid.org/ 0000-0001-7565-3987
walde.mbewa@gmail.com
STKIP Nusa Bunga Floresta Nagekeo: Nagekeo-Flores
Nusa Tenggara Timur, Indonesia

ABSTRACT

This paper describes item analysis. Item analysis is important for the test because it measures students 'development or achievement. Test items are therefore needed to be reliable for that purpose. The problems of this writing are what are the difficulty levels, discrimination power of the English Final Semester Test of the third year students of the English Department of SMAN I Kuang of the Academic Year 2011/2012? To what extent do the distracters of final test items function? With quantitative descriptive, the writer tries to answer the problem. The result of the research shows that: the difficulties level of the item given is ranged from too easy to too difficult. The items are 14, 4 items are too difficult, and 22 items are accepted. Variation is needed in this kind of test, and therefore the items in the test can be accepted as a measurement. The discrimination level of the items given is ranged from the item is not too understood to discriminating. The test given has some items contains distracter. Certain items, although not rightly answered, has no disaster.

KEYWORDS

Item Analysis, Semester Test, Testing, English, Indonesia

INTRODUCTION

Creativity Language is a means of communication which is human being needs to interact each other. Language plays an important role in building up communication skill. It is like English. English is an international language, which is taught as the first foreign language, and compulsory subject at all school from elementary school up to the university Krashen and Terrell (1983 in Nalley 2009: 15). Therefore, it needs an evaluation after teaching this subject. Evaluation is a tool to measure the students 'achievement. Evaluation and testing are related each other. Testing is a part of evaluation (Grounlund and Linn 1990:5) stated that a test is an instrument and systematic procedure for measuring a sample of behavior. It will measure each student's performance in comparison with the performance of other students. In other to see his/her teaching outcomes, a teacher has to give a test to the students. A test is also very useful for the students because it can motivate them in learning. If they know their performance, they will study hard either to maintain or to improve their level of language ability. The teacher has to choose which form of test that she or he will give to the students. Take, for example, multiple choices. There are many teachers who do not know whether their items are good or not. As stated above, the goodness of the items can be pictured by the items themselves. In this way of seeing the items, there are three criteria which are usually used: difficulty level, discrimination power, and distracters. However, the item analysis can be seen from the student's answers for the items. It is derived from an assumption that the difficulty, the power and the distracting of the items can be seen in the students' answers.

FRAMEWORK

There are some theoretical concepts related to testing. Testing is a part of evaluation, which is used to know how far the students have mastered the material taught. In this case, it can be said that a test is a tool for evaluation. This definition supported by Groundlund(1976) in Ali Imran (1996: 114) who says"... the systematic process of determining the extent to which instructional objectives are achieved by pupil". Based on the time of the test, a test consists of three parts. The first is pre-test which is given before teaching learning process. A pre-test is aimed to determine student's placement. The second is formative test administer during teaching learning process. The formative test aims to survey learning progress, detect learning errors and provide feedback to students and teacher. The third is a summative test or final test or semester test which does at the end of teaching learning process. While the aim of summative test or final

test or semester test is to determine whether a student has work to achieve instructional objectives, which has been determining or not. The result of this test is used to evaluate the effectiveness of the instructions and to determine the grade of the students at a particular time.

The teaching-learning process has procedures. According to Gelder in Ali Imron (1996: 117), the procedures of teaching learning process are instructional design, situation analysis, teaching activity and learning aid and evaluation. An instructional objective is a definition of the working task that a worker must be competent to do after completing a course of instruction. Situation analysis is a process that examines a situation, its elements, and their relation to provide and maintain a state of situation awareness for the decision maker. While teaching learning activities involve helping students to learn more, increasing students' understanding

According to Encyclopedia of Education in Suharsini Arikunto (2001: 23) which says that test is a comprehensive assessment of an individual. So, when a teacher gives a test, he or she should find out the types of test and the criteria of a good test. This type of test is subjective test and objective test. The subjective test can be an essay test form whereas objective test is one that can be marked objectively. This test requires the students to choose the right answers, which has been prepared. Furthermore, the score will have the same judgment to that item. The objective test consists of multiple- choice test, true-false, matching, and completion. Next, the basic criteria of a good test are validity, reliability, and practicality. In addition to validity and reliability, students should also be concerned about the effect of the test, particular the extent to which the test cause undue anxiety. Where possible, one should utilize test forms that minimize the tension and stress generated by our English language test. Through item evaluation, we attempt to find out whether or not each question has functioned properly (Senaul(2004: 13).

A simple statistical way of evaluating item can be done by '' item analysis" (Madsen, 1983: 178-179). He said that test item analysis is most often used with multiple choice questions. An item analysis tells us three things. How difficult each item, whether or not the question discriminates between high and low students, and which distracters are working as they should. The discussion about item analysis proceeds with the step by step procedure that can be easily adapted by the teacher. The important is that should also be familiar to students before being used in a test. Otherwise, the students may make mistakes not so much because of a lack of understanding of the task requires.

Multiple-choice items are undoubtedly one of the most widely used types of items in objectives tests. However, it must be admitted that the usefulness of the item is limited. The chief criticism of the multiple choices item is that it does not lend itself to the testing of language as communication. The process of selecting one of the options bears little relation to the way language is used. Nevertheless, multiple-choice items can provide a useful means for testing knowledge of grammar, vocabulary, etc., rather than the ability to use language " (Mandaru, 2007: 38). Furthermore, multiple choices can be used to test items detailed understanding grammar or vocabulary; it also can be used

to test the understanding of their single extended test (Heaton 1989; 3-4). The true-false test consists of a declarative statement that should be marked true-false, right, correct or incorrect, yes or not, fact or opinion. It is commonly used in measuring the ability identifies the correctness of the statement of fact, the definition of terms, statement of principles, and the like. The matching test consists of two parallel columns, contain the items of the test. The matching exercise consists of the parallel columns with each word, number or symbols in one column being a match to a word, sentence or phrase in the other column. The item for which a match is sought is called premises, and the item in the column from which the selection is made is called responses. The matching exercise is limited for measuring factual information based on simple associations or students' ability to identify the relationship between two things (Weir; 1990: 4).

It has been previously mentioned that a test is basically by which something is measured. For the result of the measurement to be dependable, the test has to meet some basic criteria. There are three basics, namely: validity, reliability, and practicality. The validity of a test may be broadly defined as the extent to which the test does what it is intended to do. If a test of pronunciation measures pronunciation and nothing else, it is a valid test of pronunciation. It would not be a valid test grammar or vocabulary because it does not test grammar or vocabulary. Heaton (1975: 153) stated that the validity of a test is the extent to which it measures what it is supposed to measure and nothing else. He stated further that every test, whether it is a short, informal classroom test or a public examination, should be as valid as the constructor can make it. The test must aim to provide a true measure of the particular measured skills. To the extent that it measures external knowledge and other skills at the same time, it will not be a valid test. Validity has four types namely: face validity, content validity, construct validity and empirical validity. Sometimes empirical validity is further classified into concurrent validity which refers to how well the test score compares to with one or more measure; and predictive validity which is demonstrated by how well the test score correlates with a criterion measure taken at a much later date.

Reliability refers to the consistency of the measurement that is how consistent test scores or other evaluation results are from one measurement to another (Groundlund 1976: 105). There are three approaches to estimating the reliability of the test (Bachman 1990: 172). They are test-retest approach, equivalent forms approach, and parallel forms approach. Test-retest approach is the stability of the scores obtained by the same subjects when the same is administered to them twice with a specified time interval between the administrations. The correlation between the scores is called the coefficient of stability. Equivalent forms method indicates the consistency between subjects, scores on the test in hand and scores that would have been obtained by the same. Subjects on an equivalent form of the test might have been substituted for it on the single occasion of testing. While parallel forms approach is the consistencies of the scores obtained by the same subjects when to parallel forms of the test are administered to them. The third attribute of a good test is practicality or sometimes is called usability (Heaton (1989:

10). It means that a test must have some characteristics like administrable (the ease of administration and probability of performance required to the test), economy (a test is practical if it can be administered with minimum expenditure of time, effort, and resources), scorable and interpretability and fair (if it is not to trap the students).

OBJECTIVES OF THE STUDY

The study aimed to find out the difficulties level of final test item of the third year students of SMAN I Kupang. Besides that, this research is also aimed to find out the discrimination power of final test item and to find out whether the distracters of final test items function as they should.

METHODOLOGY

The researcher used the descriptive quantitative method which is designed to know how difficult each item, whether or not the question discriminates between high and low students, and which distracters are working as they should. In connection with the topic under discussion in this paper, the population was chosen are 50 items of the final semester of English test which done by the third year students of SMAN I Kupang. Regarding sampling, Arikunto (1992: 85) says that if the population is more than 100 only between 10% and 15% is taken as the sample, but if the population is less than 100 the whole number of population is taken as a sample. Based on the number of the population above the sample of this research is the whole number of population. In collecting the data, the instrument used by the researcher is written test especially multiple choices. The researcher prepared an item analysis using the following steps. First, the researcher scored all of the answer sheets of the tests. Second, the researcher arranged them in order from the one with the highest score to the one with the lowest score. Third, she divided the papers into three equal groups. If the class size is too small, it can also be divided into halves. The classical procedure is to choose the top 27% (High Group) and the bottom (27% of the papers to be analyzed. In conducting this research, the researcher used statistical formula. The data, especially each item with its options on each answer sheet would be computed separately from one another using statistic ways. The analysis is carried out with the used of the test formula as follows:

a. Difficulty Level or Facility Value $\quad \frac{HC+LW}{N}$

Notes:	HC	=	High Correct
	LC	=	Low Correct
	N	=	Number of Students in Both Groups

While aiming for test items with FV between 0, 4 and 0, 6, many test constructors may accept items with FV between 0, 3 and 0, 7. (Mandaru, 2007: 69).

b. Discrimination Power = $\frac{\text{HC-LC}}{\text{N}}$

Where: HC = High Correct
= Low Correct
= Number of Students in Each Group

The discrimination index is ranging from-100 to +100. The higher difference features of the problem, the more powerful or good about it. If the difference is negative (<0) means more below the group who did not understand the material correctly answer the question compare with the group (learners who understand the material)(Arikunto, 2010: 2011).

c. Distracter evaluation

It is good or weak distracter can be clearly seen in difficulty level and discrimination level. The weak distracters can cause test questions to have poor discrimination or an undesirable level of difficulty.

RESULTS AND DISCUSSION

In this discussion, the researcher did not give the lesson to the students, but she gave the test to the students directly. The test given was prepared by the teacher at that school. After giving the test, the writer collected the students' worksheets, which are 31 sheets according to the number of the students. The worksheets were scored; the number of the right answer is divided by the number of the items, multiplied by one hundred. After that, the writer arranged the students' worksheet from the highest score to the lowest score. In general, the test given was objective test, especially multiple choice. The test mostly consisted of the text-based test. Next, the researcher will give the discussion on the items of the test given. In that discussion, the three properties of an item will be explained based on the students' work. For those aims, the researcher presented as follow first, the order of the score from the highest to the lowest of the whole class, the separation of the member of the class into a high group (27%) of the lower score-having in the list) and record of the students' responses.

Presentation of scoring

The score of the students is gained from the formula

: Number of the right answer x100
Number of items

The table presents the score for each student from the highest to the lowest.

Table 1. Score from the highest to the lowest

Answer sheet	
No.	**Grade**
1	92,5
2	92,5
3	90
4	87,5
5	85
6	82,5
7	77,5
8	77,5
9	75
10	75
11	75
12	75
13	72,5
14	72,5
15	72,5
16	70
17	70
18	70
19	67,5
20	65
21	62,5
22	60
23	60
24	57,5
25	55
26	52,5
27	50
28	47,5
29	47,5
30	45
31	40

High group and low group

From the score, two groups were made: the high group is the 27% of the students taken from the upper list with the high score while the low group is a group of 27% students taken from the lowest list.

Table 2. High group and low group

High group		Low group	
No	Grade	No	Grade
1	92,5	1	60
2	92,5	2	57,5
3	90	3	55
4	87,5	4	52,5
5	85	5	50
6	82,5	6	47,5
7	77,5	7	47,5
8	77,5	8	45
9	75	9	40

Record of the students' responses

The researcher gives a sign (capital letter and bold) to the correct option for each number of the item.

Table 3.Record of the students' responses

Item number	Option	High group	Low group	Item number	Option	High group	Low group
1	A b c **D** E	- - 1 8 -	1 2 - 6 -	21	**A** b c d e	6 3 - - -	3 5 1 - -
2	A b **C** d e	- 1 8 - -	1 1 7 - -	22	A **B** c d e	- 8 1 - -	- 2 - - 7
3	A b **C** d e	- 1 8 - -	- 6 3 - -	23	a b c **D** e	- - - 9 -	- 1 1 6 1
4	A b **C** d e	1 1 7 - -	1 5 3 - -	24	a b **C** d e	- - 9 - -	- 7 2 - -

5	A **B** c d e	- 9 - - -	1 8 - - -	25	a b c **D** E	- - - 9 -	- 1 8 - -
6	**A** b c d e	9 - - - -	8 - 1 - -	26	a **B** c d e	- 9 - - -	1 4 - 4 -
7	A b c d **E**	- - - 1 8	- - - - 9	27	A b c d e	9 - - - -	2 6 - - 1
8	A b c **D** E	- - - 9 -	- 1 - 8 -	28	a b **C** d e	- - 9 - -	- 1 7 - 1
9	A **B** c d e	2 7 - - -	3 6 - - -	29	a b c d **E**	- 1 - 2 6	- - 1 2 6-
10	A b **C** d e	- - 7 1 1	- - 9 - -	30	**A** b c d e	9 - - - -	2 6 - 1 -
11	A b **C** d e	1 - 3 5 -	3 - 1 - 5	31	**A** b c d e	7 - 2 - -	1 - 8 - -
12	**A** b c d e	1 - 1 4 3	- - - 6 3	32	a **B** c d e	- 9 - - -	- 9 - - -
13	**A** b c d e	- - - 1 8	2 - 1 - 6	33	a **B** c d e	- 8 - 1 -	- 2 - 7 -
14	A b c **D** E	1 2 - 3 3	1 - 7 - 1	34	a b **C** d e	- - 9 - -	- - 3 - 6

15	**A** b c d e	7 1 1 - -	1 6 1 - 1	35	**A** b c d e	9 - - - -	8 - - - 1
16	A b c **D** E	- - 1 8 -	- 5 1 2 1	36	a b **C** d e	- - 9 - -	- - 9 - -
17	A b **C** d e	- - 8 1 -	- - 6 - 3	37	a **B** c d e	- 9 - - -	1 2 - - 6
18	A b c **D** E	- - - 9 -	- - - 9 -	38	a b c **D** E	- 4 - 5 -	- - - 3 6
19	A **B** c d e	- 3 3 - 3	1 - - - 8	39	a **B** c d e	- 9 - - -	- 9 - - -
20	**A** b c d e	7 - - - 1	1 5 1 - 2	40	a b **C** d e	1 - 7 1 -	- - 9 - -

Difficulty Level (FV) and Discrimination Power (D) per item

Table 4. Difficulty Level (FV) and Discrimination Power (D) per item

Item Number	FV	Item Number	D
	0,77	1	0,22
	0,83	2	0,11
	0,61	3	0,55
	0,55	4	0,44
	0,94	5	0,11
	0,94	6	0,11
	0,94	7	-0,11
	0,94	8	0,11
	0,72	9	0,11
	0,88	10	-0,77
	0,22	11	0,22

	0,05	12	0,11
	0,77	13	0,22
	0,16	14	0,33
	0,44	15	0,66
	0,55	16	0,66
	0,77	17	0,22
	1,0	18	0
	0,16	19	0,33
	0,44	20	0,66
	0,50	21	0,33
	0,55	22	0,66
	0,83	23	0,33
	0,61	24	0,77
	0,50	25	0,99
	0,72	26	0,55
	0,61	27	0,77
	0,88	28	0,22
	0,66	29	0
	0,61	30	0,77
	0,44	31	0,66
	1,00	32	0
	0,55	33	0,66
	0,66	34	0,66
	0,94	35	0,11
	1,00	36	0
	0,61	37	0,77
	0,44	38	0,2
	1,00	39	0
	0,88	40	-0,77

With the acceptable range of Difficulty Level (FV) from 0, 3 to 0, 7(seeMandaru, 2007: 69) it can be said that only 22 items were acceptable.

- The items number 18, 32, 36 and 39 are too easy because its FV is 1, 0, and 0, 3 higher than the highest FV acceptable0, 7. The items number 5,6,7,8 and 35 are too easy because its FV is 0, 9, and 0, 2 higher than the highest FV acceptable 0, 7. The items number 2, 10, 23, 28 and 40 are also too easy because its FV is 0, 8, and 0, 1 higher than the highest FV acceptable 0, 7.

- The item number 12 is too difficult because its FV is 0, 0. 0, 3 lower than the lowest FV acceptable 0, 3. The item number 14 and 19 are too difficult because its FV is 0, 1, 0, 2 lower than the lowest FV acceptable 0, 3.
- The items number 3,24,27, 29,30, 34 and 37 are acceptable because the FV is 0,6 which is in the range of FV accepted: 0,3-0,7. The items number 4,16,21,22,25 and 33 are acceptable because the FV is 0,5 which is in the range of FV accepted: 0,3-0,7. The items number 1,9,13,17 and 26 are acceptable because the FV is 0,7 which is in the range of FV accepted: 0,3-0,7. The items number 15, 20, 38 and 31 are also acceptable because the FV is 0, 4 which is in the range of FV accepted: 0,3-0,7.

With the table of Discrimination Power above we can conclude that:

- D<0 the item is not understood by the students in the class. (See Mandaru, 2007: 69) It can be said that only 8 items were discriminating. The item number 10 and 40 are misunderstood because the value of D is -0, 77 which is the lowest than ''0". The item number 7 is also misunderstood because the value of D is -0, 11 which is the lowest than ''0".
- D=0. The item is not discriminating. The items number 18, 29, 32, 36 and 39 are not discriminating because the value of D is ''0".
- D > 0 the item is discriminating.

The item number 1, 11,13,17,38 and 28 are discriminating because the value of D is 0, 2 which is higher than ''0". The item number 2,5,6,8,9,12 and 35 are discriminating because the value of D is 0,1 which is higher than ''0". The item number 14, 19, 23 and 21 are discriminating because the value of D is 0,3 which is higher than ''0". The item number 4 is discriminating because the value of D is 0,4 which is higher than ''0". The item number 3 and 26 are discriminating because the value is 0,5 higher than ''0". The item number 15,16, 20, 22, 31, 33, and 34 are discriminating because the value of D is 0,6 higher than ''0". Item number 24, 27, 30 and 37 are discriminating because the value of D is 0,7 higher than ''0". The item number 25 is discriminating because the value of D is 0,9 higher than ''0".

Distracter

The following example is taken from the item number 29 of the test given, that researcher would like to present that distracted.

Charli: Sorry honey I do not know to pay the bill, I forgot my wallet at home.
Winda: So??? You always keep your wallet at home, if we are going to lunch in the restaurant!!!
Winda was .....with her boyfriend manner.

a. love *c. happy* *e. sorrow*
b. attention *d. embarrassed*

Table 5. The record of the students' answer in both groups

Option	High group	Low group
A	-	-
B	1	-
C	-	1
D	2	2
E	6	6

In the item of the test, the right answer is E. However, four(4) students from both groups, two from each make mistakes by answering the item with option D. In the item number 29 above, it seems that the D is the distracter for them. It can be understood because the option D can also be accepted to a certain degree. Perhaps, there should be more appropriate to give another option like angry, which seems to be more accepted for the situation in the conversation.

CONCLUSIONS

Item analysis is important for test because it is a best measure for students' development or achievement. Test items are therefore needed to be reliable for that purpose. After the analysis of the items of the test it can be concluded that: first, the difficulties level of the items given is ranged from too easy to too difficult. There are 14 items are too easy (2,5,6,7,8,10,18,23,28,32,35,36,39 and 40), 4 items are too difficult (11,12,14, and 19) and 22 items are accepted(1,3,4,9,13,15,16,17,20,21,22, 24,25,26,27,29,30,31,33,34,37 and 38). Second, There are 3 items are not understand (7, 10 and 40), 5 items are not discriminating (18, 29, 32, 36, and 39) and 32 items are discriminating (11,13,17,38,28,2,5,6,8,9,12,35,14,19,23,21,4,3,26,15,16,20,22,3 1,33,34,24,27,30,37 and 25) . Third, the test given has some items contains disaster. Certain items, although not rightly answered, has no disaster. It's rather the student miss understanding or lack of understanding of the item.

RECOMMENDATIONS

Based on the result of this research study, the researcher suggested the students should always be motivated to consider the importance of the test item and the students completed or corrected those sentences by selecting appropriate multiple choice item. To overcome the students in the test items, the writer would like to suggest that the teacher of English should give the items which are variation in the option. Variation

is needed in this kind of test, and therefore the items in the test can be accepted as a measurement.

LITERATURE CITED

Arikunto, S.2001. Penilaian Program Pendidikan, Jakarta: Depdikbud RI.

Arikunto, S. 2010. Dasar-DasarEvaluasiPendidikan.BumiAksara, Jakarta.

Bachman, Lyle F. 1990. Fundamental Considerations in Language Testing. Oxford University Pres.

Gronlund, N.E, and Linn (1990), Measurement and Evaluation in Teaching (6th Ed). Ny: Mac Millan.

Imron,Ali. 1996. The Teaching Learning Process. (online) (http://www.edpsycinteractive.org/papers/modeltch.html, (accesed on 8 th May 2012).

Mandaru, Language Testing and Evaluation, Kupang, FKIP Undana.

Nalley, H. 2009. Teaching English as Foreign Language.An unpublished teaching material.Kupang, FKIP UNDANA

Senaul, Sisilia. 2004. An Analysis of the teacher made test items administered to the third year students of SLTPN 8 Kupang in the Academic Year 2004/2005. Unpublished Thesis.Undana.Kupang.

Tonge, Djen, P. 2001, Analysis ButirSoal (Item Analysis), Widyaswara BPG Kupang. (online) (http://w3.gre.ac.uk/-bj61/talessi/atl.html, (Accessed on, 2nd June 2012)