ITEM ANALYSIS OF ENGLISH SUMMATIVE TEST AT SMP N 2 PURWODADI

Dewi Nurvitasari[1)]
dewinurvita06@gmail.com

Wiyaka [2)]
wiyaka@upgris.ac.id

Sukma Nur Ardini [3)]
Sukmanurandini27@gmail.com

Corresponding Author : dewinurvita06@gmail.com
Universitas PGRI Semarang

Abstract:

Evaluation is one of the important aspects of teaching and learning activities. It shows an important role, especially in terms of education. The study aimed at analyzing the item analysis English summative test at SMP N 2 Purwodadi. The objectives were to: (1) find out the level of validity of English summative test items for second-grade students of SMP N 2 Purwodadi, 2) investigate the level of reliability of English summative test for a second-grade student of SMP N 2 Purwodadi, 3) describe the level of difficulty of English summative test items for second-grade students of SMP N 2 Purwodadi. This study applied a quantitative descriptive. The population were students of SMP N 2 Purwodadi consists of 64 students. This research used purposive sampling. The sample were several students of class 8A and 8B consists of 32 students of each class. The result showed that there were sixteen valid items and nine invalid items of the English summative test. The valid items of the test were items number 2, 4, 5, 6, 7, 8, 11, 14, 16, 17, 20, 21, 22, 23, 24, and 25. On the contrary, the invalid items were items number 1, 3, 9, 10, 12, 13, 15, 18, and 19. The reliability of the test were 0,246064049 (reliable). The difficulties level of the test that showed nine items were difficult, the difficult items were the question number 1,3,9,10,12,13,15,18, and 19. Ten items were medium, the medium items were question number 4, 5,8,11,14,16,17,20,23, and 25. The last one, six items were easy, the easy items are number 2,6,7,21,22 and 24. The conclusions of the test of the research are valid, reliable, and less difficult. The test maker or teacher should recheck the test item several times before it will be given to the students. It is an action to minimize some mistakes in constructing the test.

Keywords: *English summative test, Validity, Reliability, Difficulty.*

1. Introduction

Evaluation is one of the important aspects of teaching and learning activities. It plays an important role, especially in terms of education. The information gained through the evaluation will be very useful to make improvements in the future. In the formal education system, a teacher is one of the figures who are responsible for the earning process whether it is a success or not. A good teacher not only knows how to evaluate as well as how to teach. In the teaching process, a teacher has to know how to evaluate student progress on the mastery of a lesson that has been taught in a certain period. The result of the evaluation will provide information about the quality of the teacher and the ability of the student.

Evaluation in education can be assumed as a formal and informal of examining students' achievement. The informal evaluation usually occurs at the time of the teaching and learning process taking place. Teachers can evaluate the students' achievement by observing and making a judgment based on students' performance during the process of teaching and learning. Yet, teachers cannot assume that students who never perform activities during the teaching and learning process do not understand the materials at all. It is because somehow students do not feel free to express their ideas. Thus, it needs a formal assessment to examine the students' understanding.

To evaluate students' achievement of the material which has been taught, usually, the teacher gives the students some questions in the form of a test. The teacher can conduct it after each chapter of the material is finished or at the end of the semester, the test is called the achievement test. An achievement test is a systematic procedure for determining the amount the student has learned. There are two kinds of achievement tests; formative test and summative test. In this research, the writer chooses summative test as the kind of test which is administered at the end of a unit or term semester or a year of study to measure what has been achieved both individually and by groups. The test can be in the form of an essay test in which students have to write the answer in some sentences. Besides, teachers can give the test in the form of multiple-choices to simply check students' achievement. The teacher who makes a test has to know the principles and the steps that must be done in making a good test.

This study has purposes to answer the following research questions:

1. To what degree is the validity level of English summative test items for second-grade students of SMP N 2 Purwodadi?
2. To what degree is the reliability level of English summative test items for second-grade students of SMP N 2 Purwodadi?

3. To what degree is the quality level of difficulty English summative test items for second-grade students of SMP N 2 Purwodadi?

In line with the research questions above, the writer's final project has purpose as listed below:

1. To find out the level of validity of English summative test items for second-grade students of SMP N 2 Purwodadi.
2. To investigate the level of reliability of English summative test for a second-grade student of SMP N 2 Purwodadi.
3. To describe the level of difficulty of English summative test items for second-grade students of SMP N 2 Purwodadi..

2. Literature Review

2.1 Validity

Definitionof Validity

A test can be valid if the test really measures the ability of the test-taker. Gronlund (as cited in Brown 2004, p. 22) defined validity as 'the extent to which inferences made from assessment are appropriate, meaningful, and useful in term of purpose of the assessment.

Types of Validity

Brown also explains four types of evidence to measure the validity of the test which follows:

2.0.1 Content validity

Content validity is the degree to which a test measures and intended content area. Content validity requires both item validity and sampling validity. Item validity is concerned with whether the test items represent measurement in the intended content area, and sampling validity is concerned with how well the test samples the total content area. If a test actually samples the subject matter about which conclusions are to be drawn, and if requires the test-takers to perform the behavior that is being measured, it can claim content-related evidence validity, often popularly referred to as content validity (Mousavi, 2002; Hughas, 2003). For instance, if a test is to measure foreign students' mastery of English sentence structure, an analysis must first be made of the language itself and decisions made on which matters need to be tested, and in what

proportions. To the extent that the analysis accords with the views of recognized authorities in the skills area and the test then reflects such an analysis, it may be said to have content validity.

2.0.2 Criterion-related Validity

Criterion validity that meant 'criterion' of the test has been reached. For example, every school establishes KKM or minimum score that every student must reach or get higher from it. If the teacher decides that 80 % of students could reach or take higher than the passing grade, the test result of should reach the criteria to make it become valid. That is called criterion validity.

2.1.3 Construct Validity

Construct validity is the degree to which a test measures an intended hypothetical construct. Brown (2004: 25) stated that a construct is any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perceptions. For example, a teacher has given a procedure for conducting an oral interview. The scoring analysis for the interview includes several factors in the final score: pronunciation, fluency, grammatical accuracy, vocabulary use, and sociolinguistic appropriateness. So if you were asked to conduct an oral proficiency interview that evaluated only pronunciation and grammar, the construct validity would be inadequate.

2.1.4 Consequential Validity

Messick (1989), Gronlund (1998), McNamara (2000), and Brindley (2001), among others, underscore the potential importance of the consequences of using an assessment. Consequential validity encompasses all the consequences of a test, including such considerations as its accuracy in measuring intended criteria, its impact on the preparation of test-takers, its effect on the learner, and the social consequences of a test's interpretation and use.

2.1.5 Face Validity

An important face of consequential validity is the extent to which "students view the assessment as fair, relevant,and useful for improving learning"(Gronloud, 1998, p. 210), or what is popularly known as face validity. "face validity refers to the degree to which a test looks right, and appears to measure the knowledge or abilities it claims to measure,

based on the subjective judgment of the eaxaminees who take it, the administrative personnel who decide on its use,and other psychometrically unsophisticated observers" (Mousavi, 2002: 244 in Brown's book).

2.1 Reliability

According Brown (2004: 20) pointed out that a reliable test is consistent and dependable. As the example, if we give the same test to the same student or matched students on two different occasions, the test should yield similar results. The issue of reliability of a test may best be addressed by considering a number of factors that may contribute to the unreliability of a test. Consider the following possibilities: fluctuations in the student, in scoring, in test administration, and the test itself (Mousavi, 2002: 804 in Brown's book).

2.2 Difficulty Level

We cannot randomly argue that the test is difficult by guessing. There is something to be used as the evidence of our perception which called level of difficulty. Level of difficulty also has been known as Item Facility is the extent to which an item is easy or difficult for the proposed group of test-taker (Brown, 2004, p. 58). In another term, Brown (2005, p. 66) also defines IF as statistic used to examine the percentage of students who correctly answer a given item.

According to Mahren and Lehman (1984, p.31) the concept of difficulty or the decision of how difficult the test should be depending on a variety factor. There are:
1. The purpose of the text
2. The ability level of the students
3. The age or grade level of the student

3. Research Methodology

3.1 Object of the Study

The object of the study was an English summative test made in terms of teaching of SMP N 2 Purwodadi in the academic year of 2019/2020 for eighth grade students which consists of 25 multiple choice through online tests. This study applied a quantitative descriptive. The population were students of SMP N 2 Purwodadi consists of 64 students. This research used purposive sampling. The sample were several students of class 8A and 8B consists of 32 students of each class.

3.2 Instruments

The explanation of these instruments can be seen as follows:

1. Paper Test Question

    English summative test for second grade student of SMP N 2 Purwodadi.

2. Answer sheets

    This answer sheets used to know the answer distribution. They were analyzed in order to find out the validity, reliability, and difficulty level to answer the problem statement.

3. Answer key

This answer key used as a valid guide in scoring each item.

3.2 Data Analysis Procedures

Since the data is defined as the secondary data source, the writer used method of documentation. This method is used in order to get the data in the form of documents owned by school. The data consist of list of student's names, problem sheet of the test, students' answer sheets and answer key.

The step that the researcher did to collect the data are:

1. Asking for a permission to the Headmaster of SMPN 2 Purwodadi city to do an observation

2. Asking for a permission to the English teacher of SMPN 2 Purwodadi city to take the problem sheet and the students' answer sheets as the sample study

3. Copying the problem sheet

4. Copying the students' answer sheet; and

5. Copying the answer keys.

Those five steps were used in method called one shot model. One shot model is a model of approach that uses one step to collect the data in one time (Arikunto, p. 83). Simply, the writer will do the five steps in one day.

4. Findings

The data that used by the writer in this researcher in English summative test for second grade student of SMP N 2 Purwodadi academic year 2019/2020. The total number of test items are 25 item multiple choice question. The test was held on June 10th, 2020 with the given 60 minutes.

4.1. The validity level of English summative test

The data of the finding show that sixteen items of the English summative test were valid and nine items were invalid. To be clear, the researcher provides the table thar give a brief description about the validity of each item.

Table 1. *The validity level Analysis*

Item	Correlation	Table	Status	Classification	Validity Level
1	-0,200095419	0,246064049	Invalid	-	-
2	0,550059847	0,246064049	Valid	0.20 – 0.40	Poor

To be clear, the writer describes each item as follows:

1. Item number 1 is an invalid item since the result of *r was* lower than table of Product Moment. This question cannot be categorized into a scale because the value of r -0,200095419.

2. Item number 2 is a valid item since the result of r was higher than table of Product Moment. This question can be categorized into a scale because the value of r 0,550059847 that belongs to poor item.

4.2. The reliability level of English summative test

The data of the findings shows that the English summative test for second grade student of SMP N 2 Purwodadi was reliable since the reliability index was 0,785081707. This reliability works on the standard index describe by Arikunto (2006:184) who highlights that an item is considered to be reliable if the coefficient correlation of each item is higher or equal to the table of critical value of product moment with the level of significance 95%. To be clear, the writer provides the table of reliability analysis as follow;

Table 2. *The reliability level Analysis*

Correlation	Table	Status	Classification	Reliability Level
0,785081707	0,246064049	Reliable	0.60 < r11 ≤ 0,70	High

From the table above, it can be seen that the level of reliability English summative test at SMP N 2 Purwodadi is 0,785081707 that belongs to high test.

4.3 The difficulty level of English summative test

The data shows that there were four medium items, four easy items, one too easy Item, and one difficult item of the test. To be clear, the researcher provides the table that give a brief description about the difficulty level of each item.

Table 3. *The difficulty level Analysis*

Item	P	Classification	Difficulty Level
1	0,09375	$0.00 < P \leq 0.30$	Difficult
2	0,796875	$0.70 < P \leq 1.00$	Easy
3	0,265625	$0.00 < P \leq 0.30$	Difficult
4	0,625	$0.30 < P \leq 0.70$	Medium
5	0,625	$0.30 < P \leq 0.70$	Medium
6	0,765625	$0.70 < P \leq 1.00$	Easy
7	0,765625	$0.70 < P \leq 1.00$	Easy
8	0,6875	$0.30 < P \leq 0.70$	Medium
9	0,25	$0.00 < P \leq 0.30$	Difficult
10	0,125	$0.00 < P \leq 0.30$	Difficult
11	0,609375	$0.30 < P \leq 0.70$	Medium
12	0,0625	$0.00 < P \leq 0.30$	Difficult
13	0,140625	$0.00 < P \leq 0.30$	Difficult
14	0,46875	$0.30 < P \leq 0.70$	Medium
15	0,15625	$0.00 < P \leq 0.30$	Difficult
16	0,671875	$0.30 < P \leq 0.70$	Medium
17	0,703125	$0.30 < P \leq 0.70$	Medium
18	0,140625	$0.00 < P \leq 0.30$	Difficult
19	0,28125	$0.00 < P \leq 0.30$	Difficult
20	0,375	$0.30 < P \leq 0.70$	Medium
21	0,796875	$0.70 < P \leq 1.00$	Easy
22	0,734375	$0.70 < P \leq 1.00$	Easy
23	0,578125	$0.30 < P \leq 0.70$	Medium
24	0,734375	$0.70 < P \leq 1.00$	Easy
25	0,453125	$0.30 < P \leq 0.70$	Medium

To be clear, the writer describes each item as follows:

1) Item number 1 is difficult item because there are 6 students from 64 students who can answer correctly, and the difficulty level of this item is 0,09375 that belongs to difficult item.

2) Item number 2 is easy item because there are 51 students from 64 students who can answer correctly, and the difficulty level of this item is 0,796875 that belongs to easy item.

3) Item number 3 is difficult item because there are 17 students from 64 students who can answer correctly, and the difficulty level of this item is 0,265625 that belongs to difficult item.

4) Item number 4 is medium item because there are 40 students from 64 students who can answer correctly, and the difficulty level of this item is 0,625 that belongs to medium item.

5) Item number 5 is medium item because there are 40 students from 64 students who can answer correctly, and the difficulty level of this item is 0,625 that belongs to medium item.

6) Item number 6 is easy item because there are 49 students from 64 students who can answer correctly, and the difficulty level of this item is 0,765625 that belongs to easy item.

7) Item number 7 is easy item because there are 49 students from 64 students who can answer correctly, and the difficulty level of this item is 0,765625 that belongs to easy item.

8) Item number 8 is medium item because there are 44 students from 64 students who can answer correctly, and the difficulty level of this item is 0,6875 that belongs to medium item.

9) Item number 9 is difficult item because there are 16 students from 64 students who can answer correctly, and the difficulty level of this item is 0,25 that belongs to difficult item.

10) Item number 10 is difficult item because there are 8 students from 64 students who can answer correctly, and the difficulty level of this item is 0,125 that belongs to difficult item.

11) Item number 11 is medium item because there are 39 students from 64 students who can answer correctly, and the difficulty level of this item is 0,609375 that belongs to medium item.

12) Item number 12 is difficult item because there are 4 students from 64 students who can answer correctly, and the difficulty level of this item is 0,0625 that belongs to difficult item.

13) Item number 13 is difficult item because there are 9 students from 64 students who can answer correctly, and the difficulty level of this item is 0,140625 that belongs to difficult item.

14) Item number 14 is medium item because there are 30 students from 64 students who can answer correctly, and the difficulty level of this item is 0,46875 that belongs to medium item.

15) Item number 15 is difficult item because there are 10 students from 64 students who can answer correctly, and the difficulty level of this item is 0,15625 that belongs to difficult item.

16) Item number 16 is medium item because there are 43 students from 64 students who can answer correctly, and the difficulty level of this item is 0,671875 that belongs to medium item.

17) Item number 17 is medium item because there are 45 students from 64 students who can answer correctly, and the difficulty level of this item is 0,703125 that belongs to medium item.

18) Item number 18 is difficult item because there are 9 students from 64 students who can answer correctly, and the difficulty level of this item is 0,140625 that belongs to difficult item.

19) Item number 19 is difficult item because there are 18 students from 64 students who can answer correctly, and the difficulty level of this item is 0,28125 that belongs to difficult item.

20) Item number 20 is medium item because there are 24 students from 64 students who can answer correctly, and the difficulty level of this item is 0,375 that belongs to medium item.

21) Item number 21 is easy item because there are 51 students from 64 students who can answer correctly, and the difficulty level of this item is 0,796875 that belongs to easy item.

22) Item number 22 is easy item because there are 47 students from 64 students who can answer correctly, and the difficulty level of this item is 0,734375 that belongs to easy item.

23) Item number 23 is medium item because there are 37 students from 64 students who can answer correctly, and the difficulty level of this item is 0,578125 that belongs to medium item.

24) Item number 24 is easy item because there are 47 students from 64 students who can answer correctly, and the difficulty level of this item is 0,734375 that belongs to easy item.

25) Item number 25 is medium item because there are 29 students from 64 students who can answer correctly, and the difficulty level of this item is 0,453125 that belongs to medium item.

5. Discussion

This part is in line with the interpretation of the findings derived from the previous quantitative analysis.

5.1 The validity level of English summative test

Based on the findings, the outcome of the existing data of the test reported that sixteen items of the test were valid and nine items were invalid. This fact simply provides us a point about the current condition of the English summative test used for the second-grade students at SMP N 2 Purwodadi.

Arikunto in Noveria (2015: 51) points out that an item is stated valid if the coefficient correlation of each item is higher or equal to the table of the critical value of product-moment with the level of significance 95 %. In line with this, Gay (1981: 110) also states that validity is the degree to which a test measures what it is supposed to measure and, consequently, permits appropriate interpretation of scores.

Therefore, the invalid items need to be eliminated or revised and the activity should be truly conducted by the teacher to be suitable with the normal validity index of a high-quality test. This information should let the test constructor master the item analysis of the validity to create the test items which work on the ability of those items to measure what is supposed to measure.

5.2 The reliability level of English summative test

Referring to the result of data, the result of the reliability of these test items by using product moment + Tuchman showed that the reliability index of the English test items used for the second-grade students at SMP N 2 Purwodadi was reliable. Referring to the result of data, the result of the reliability of these test items by using product moment + Tuchman showed that the reliability index of the English test items used for the second-grade students at SMP N 2 Purwodadi was reliable. In this study, the test-items are called reliable items when the number of R is more than R table 0,246064049. The reliable index shows the number of 0,785081707 with the number of test-takers (n) are 64 students. It means that R11 is higher than the R table.

This fact simply provides us a point about the current condition of the English summative test used for the second-grade students at SMP N 2 Purwodadi. Thus, all of the test-items are reliable. And it can be said that test has high reliability since the number of R11 is higher than the R table.

5.3 The difficulty level of English summative test

The data of the findings showed that there were ten medium items, six easy items, and nine difficult items of the test. This fact simply provides us a point about the current condition of the English summative test used for the second-grade students at SMP N 2 Purwodadi. A good test is a test that is not too easy or conversely too difficult for students. It should give an optional answer that can be chosen by students and not too far from the key answer. Very easy items are to build in some affective feelings of "success" among lower ability students and to serve as warm-up items, and very difficult items can provide a challenge to the highest-ability students (Brown, 2004:59). It makes students know and record the characteristics of the teacher's test if the test that was given always comes to them too easy and difficult. Thus, the test should be standard and fulfill the characteristics of a good test. The number that shows the level difficulty of a test can be said as a difficulty index (Arikunto, 2006:207).

6. Conclusion

According to the analysis result in the previous study, the conclusions are drawn as follows:

Based on the findings and discussion, the researcher concludes that there were sixteen valid items and nine invalid items of English summative test for a second-grade student of SMP N 2 Purwodadi. The valid items of the test were items number 2, 4, 5, 6, 7, 8, 11, 14, 16, 17, 20, 21, 22, 23, 24, and 25. On the contrary, the invalid items were items number 1, 3, 9, 10, 12, 13, 15, 18, and 19.

Then, the researcher also concludes. That English summative test for the second grade of SMP N 2 Purwodadi test was reliable because the reliability index was 0,785081707 which was higher than the table of the critical value of product-moment with level significance 95%.

And the researcher finds the difficulties level of English summative test for a second-grade student of SMP N 2 Purwodadi showed there were ten medium items, six easy items, and nine difficult items. The medium items are question number 4,5,8,11,14,16,17,20,23, and 25. The easy items are number 2,6,7,21,22 and 24. In addition, the difficult items are the question number 1,3,9,10,12,13,15,18, and 19.

References

Ajeng. 2009. *The Analysis of Validity, Reliability, Discrimination Power, and Level of Difficulty of First Mid-Term Test in the Case of the Eight Grade Students of SMP 33 Semarang*. Unpublished Final Project. Semarang: Semarang State University.

Ani, L. I. "An Item Analysis on The Difficulty Level of an English Summative Test for Second Grade of SMP Muhammadiyah 29 Cinangka-Sawangan Depok".

Arikunto, S. 2006. *Prosedur Penelitian:Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.

Bachman, L.F. 1990. *Fundamental Considerations in Language Testing*. London: Oxford University Press.

Backhoff, E., Larrazolo, N., & Rosas, M. (2000). *The level of difficulty and discrimination power of the Basic Knowledge and Skills Examination (EXHCOBA). Revista Electrónica de Investigación Educativa, 2 (1).*

Brown, H. D. 2004. *Language Assessment: Principles and Classroom Practice 2nd Edition)*. San Francisco: Longman, Inc.

Brown¸ J. D. 2004. Performance Assesment: Existing Literature and Direction for Research. *Second Language Studies*, 22(2), pp. 91-139

Brown¸ J. D. 2004. Performance Assesment: Existing Literature and Direction for Research. *Second Language Studies*, 22(2), pp. 91-139

Burton, S. J., Sudweeks, R. R., Merril, P. F., Wood, B. (1991). *How to Prepare Better Multiple-Choice Test Items: Guidlines for University Faculty*. Provo: Brigham Young University Testing Services and The Department of Instructional Science.

Cohen, A.D. 1994. *Assessing Language Ability in the Classroom*. 2nd Ed. Boston: A Division of Wadsworth Inc.

Creswell, J. W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (4th Edition)*. Thousand Oaks: SAGE Publication, Inc

Ebel, R. L. and Frisbie, D. A. 1991. *Essentials of Educational Measurement* (5th ed.). Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Gay, L. R. 1981. *Educational Research: Competencies for Analysis & Application (2nd Ed)*. Columbus: A Bell & Howell Company.

Hidayanti, A. D. *The Analysis of Validity, Reliability, Discrimination Power, and Level of Difficulty of First Mid-Term Test in the Case of the Eight Grade Students of SMP 33 Semarang in Academic Year 2008/2009 (*Undergraduate thesis).State University of Semarang, Semarang.

Hiza, A. 2018. *"Item analysis of english first mid-term test in the case of the eight grade students of SMPN 06 MAGELANG CITY (in academic year of 2017/2018)". (A Final project UniversitasNegeri Semarang.)*

Hughes, A. 2005. *Testing for Language Teacher.* 2nd Ed. London: Cambridge University Press.

Kothari, C. R. 2004. *Research Methodology: Method and Techniques.* 2nd Ed. New Delhi: New Age International Publishers.

Larasati, R. (2015). *Item Analysis of English National Exam Trial Test Held by Kedu Residency in SMA N 1 Kebumen*(Undergraduate thesis). State University of Semarang, Semarang.

Madsen, Harold S. Techniques in Testing. Hongkong: Oxford University Press, 1983.

Muspira, H. 2016. *"Item Analysis of English Summative Test for second grade student of MAN 1 TENETE BULUKAMB".* (A Thesis UIN Alaudin Makassar)

Sadeghi, K. and Masoumi, G. A. (2017). *Does number of options in multiple choice tests affect item facility and discrimination? An examination of test taker preferences.* Journal of English Language Teaching and Learning, 19, 123-143.

Salwa, A. (2010). *The Validity, Reliability, Level of Difficulty and Appropriateness of curriculum of the English Test* (Master's thesis, Diponegoro State University, Semarang).

Tuckman, B. W. 1975. *Measuring Educational Outcomes Fundamentals of Testing.* New York: Harcourt Brace Javanovich Inc

Valette, R.M. 1967. *Modern Language Testing.* 2nd Ed. New York: Harcourt Brace Jovanovich Publishers, Inc.

Valette, R.M. 1967. *Modern Language Testing.* 2nd Ed. New York: Harcourt Brace Jovanovich Publishers, Inc.

Veirissa, A. H. (2017). *The analysis of Appropriate Selection and Agreement of Multiple Choice Items in Mid-Term Test in SMP N 1 Secang* (Undergraduate thesis). State University of Semarang, Semarang.

Versatile. 2018. In *Merriam-Webster.com.* Retrieved July 12 2018, from https://www.Merriam-webster.com/dicitonary/versatile

Weir, C.J. 1993. *Understanding and Developing Test.* London: Phoenix Elt.

Zulaiha, Rahmah. 2008. *Analisis Soal Secara Manual.* Departemen Pendidikan Nasional Badan Penelitian dan Pengembangan Pusat Penilaian Pendidikan. Jakarta: PUSPENDIK.