

## THE ANALYSIS OF DISCRIMINATING POWER OF ENGLISH SUMMATIVE TEST

<sup>1)</sup> Acep Haryudin, <sup>2)</sup> Iman Santoso  
<sup>1)</sup>haryacep@gmail.com <sup>2)</sup>cempaka4643@ymail.com

<sup>1,2)</sup> STKIP SILIWANGI BANDUNG

### ABSTRAK

Tujuan dari penelitian ini adalah untuk menganalisis daya pembeda dari tes sumatif bahasa Inggris kelas dua SMP N 2 Padalarang. Melalui penelitian ini, diharapkan guru maupun para pembuat tes mendapatkan pengetahuan, gagasan dan pemahaman cara membuat test yang baik dan benar terutama pada bagian kualitas daya pembeda tes sumatif bahasa Inggris, sehingga guru atau para pembuat tes dapat menyuguhkan soal yang sesuai dengan standar dan membantu siswa- siswa yang mendapatkan nilai rendah. Penelitian ini dikategorikan sebagai penelitian analisis deskriptif karena penelitian ini menggambarkan kondisi objektif daya pembeda tes summative bahasa Inggris pada semester genap siswa kelas dua SMP N 2 Padalarang dengan menganalisis kemampuan butir – butir soal tes summative bahasa Inggris untuk membedakan kemampuan para siswa antara satu sama yang lainnya. Penelitian ini juga termasuk penelitian kuantitatif, karena peneliti menggunakan beberapa data penghitungan yang dianalisa dengan statistik. Hasil temuan penelitian ini menyatakan bahwa tes sumatif bahasa Inggris yang diujikan pada siswa kelas kedua SMP N 2 Padalarang memiliki daya pembeda yang baik, karena 35 item dengan kategori 0.25 sampai 0.75 atau 70% dari item tes telah memenuhi kriteria dengan daya diskriminasi positif.

**Keywords:** *Analisis, Daya Pembeda dan Tes Sumatif Bahasa Inggris*

### ABSTRACT

Evaluation gives information about how successful the efforts of education have been. It helps teachers to get the information about the progress of students' achievement of the material they have learned in order to make decision. The purpose of this research is to analyze the discriminating power of English Summative test at second grade of SMP N 2 Padalarang . Through this research, it is hoped that the teachers or test makers can get clear description about the quality of discriminating power of English summative, so they are able to help the poor students. This study is categorized as descriptive analysis; because it is intended to describe the objective condition about the discriminating power of students' English summative test at exact semester of second grade of SMP N 2 Padalarang by analyzing the quality of English summative test items in discriminating students' achievement. This research is considered as quantitative research, because the researcher used some numerical data which is analyzed statistically. The finding of this study is that the English summative test which is tested at second grade of SMP N 2 Padalarang has good discriminating power, because 35 items ranging from 0.25 until 0.75 (70%) of the test items have fulfilled the criteria of a positive discriminating power.

**Key words:** *Analysis, Discriminating Power, English Summative test*

### A. INTRODUCTION

There are three domain or components in education procedures to know the quality of teaching learning proses, they are study introductory, study of proses and study of evaluation or assessment. The indicators that education is has been succeeded we can know in out put or product of students abilities and capabilities. To know the out put we new an instruments to measure its product is evaluation and assessment. Evaluation may be defined as the systematic process of collecting, analyzing, and

interpreting information to determine the extent to which pupils are achieving instructional. Evaluation gives information about how successful the efforts of education have been. It helps teachers to get the information about the progress of students' achievement of the material they have learned in order to make decision. Evaluation is defined as the systematic gathering of information for the purpose of making decision (Lyle F. Bahman, 1990:20).

A test in plain words is "A method of measuring person' ability or knowledge in a given domain"

(H. Douglass, 2001:384) Tests are used for pedagogical purposes, either as a mean of motivating students to study or as a mean of reviewing the material taught. Students usually tend to study harder when they are going to have an examination rather than when they are not and they will emphasize in studying the material that expect to be tested.

Item analysis provides a quick, simple technique for appraising the effectiveness of individual test items (Norman E. Gronlund, 1981: 262) Item analysis procedures provide information for evaluating the functional effectiveness of each item and for detecting weaknesses that should be corrected. This information is useful when reviewing the test with students and it is indispensable when building a file of high quality items. There are three characteristics which are usually determined for a test item: first, item difficulty; it indicates how difficult each item was for the group. Second, discriminatory power; it tells how well the item performs in separating the better students from the poorer students. Third, item distracter; for multiple choice items, it indicates how effective each alternative was for the item. So it can be concluded that item analysis provide us the data whether the test item is too difficult or too easy, whether it can discriminate the students or not, and whether all the alternatives functioned as intended. The reason why the Researchers chooses the discriminating power is because she thinks that the discriminating power deals more with the students than the other two choices-the level on difficulty and the effectiveness of the distracter.

Based on the explanation above, the Researchers tries to limit the problem of item analysis that it will discuss, so he just focuses on the discriminating power of the test item. The test item that will be analyzed by the Researchers is a final test of Exact semester which is tested on the second grade of SMP N 2 Padalarang. So, the Researchers tries to analyze and interpret concerning “**The Analysis of Discriminating Power of English Summative Test**”. And this research will be conducted at second grade of SMPN 2 Padalarang.

### 1. Identification of Problems

Based on the background above, the Researchers identify some problems, such as:

- a. Is there an English language Expert team to make test item for Junior High School ?

- b. How is the procedure to make English Summative test?
- c. How is the English Teacher measure English Summative test well?
- d. How is the test maker decide a good discriminating power in summative test?

### 2. Research Question

Is the English Summative Test in Junior High School Class two has an good Discriminating Power ?

### 3. Research Objective

The objective of this research is to measure the quality of English summative test and to know whether the English summative test items have a good discriminating power or not.

### 4. Limitation of the Problems

To make this research is easy to be understood, the Researchers limited this research, are:

- a. The only English Summative test that have been analysed for Junior High school.
- b. This research only analyze the Discrimination Power English Summative Test.

### 5. Benefits

This research is expected can give benefits in theoretical or practical, such as:

#### a. Theoretical Benefit

To develop the knowledge about how to make a good English Summative test with a right and good.

#### b. Practical Benefits

1. To give knowledge to English Summative test makers in order that focused on best quality of English summative test
2. To enrich the literature about the characteristic and procedure in making English Summative Test

## B. LITERATURE REVIEW

### 1. The Definition of Test

Test is an important part in Education field, this is being done to know the students' knowledge improvement. As Brown (2004: 3) state that “tests as methods of measuring a person's ability, knowledge, or performance in a given domain. Most common forms of tests include fill-in-the-blanks, sentence completion, open answers, and multiple choices”. So, if we want to know the students' quality we have to give them test.

## 2. Types of Test

Test can be categorized according to the types of information they provide. This categorization will prove useful both in deciding whether an existing test is suitable for a particular purpose and in writing appropriate new tests where these are necessary (Arthur Hughes, 2003:5). Test can be classified based on its purpose and based on its test maker.

### a. Test Based on purposes

#### 1) Diagnostic Test

The diagnostic tests seek to identify those are in which a student needs further helps.

#### 2) Achievement Test

Achievement tests measure what a person has learnt during a course of instruction. It is given at the end of the course. There are two kinds of achievement test, final achievement test and progress achievement test (Desmon Allison, 1999:80).

#### 3) Proficiency test

Proficiency test are designed to measure people's ability in a language regardless of any training they may have had in that language. Rather, it is based on a specification of what candidates have to be able to do in the language in order to be considered proficient (Arthur Hughes, 1990:11.).

#### 4) Aptitude Test

Aptitude tests are often used in selecting individuals for jobs, for admission to training program, for scholarship, and for many other purposes. Sometimes aptitude tests are used for classifying individuals, as when students are assigned to different ability-grouped sections of the same course (Howard B. Lyman, Allyn & Bacon, 1998:22).

### b. Based on the Test Maker

#### 1) Standardized Test

Standardized tests are constructed by test specialists working with curriculum experts and teachers.

#### 2) Teacher-Made Test

Standardized tests, in contrast, are used to compare students' performance in different classes or schools (Gilbert Sax, Belmont: Wadsworth, 1980:16-18). Teacher made test are constructed by teachers for use within their own classroom. Their effectiveness depends on the skill of the teacher and his or her knowledge of test construction.

## 3. The Characteristic of a Good Test

The most essential characteristic of the good test can be classified into three main aspect, they are, validity, reliability, and practicality (Norman E Gronlund, 51).

## 2. Item Analysis

To know the quality some testing or summative test, we have to analyze it in deep analysis. Item Analysis is the process to analyze the items whether its has on the right procedure in form of test or not and selection of appropriate language items is not enough by itself to ensure a good test. Each question needs to function properly; otherwise, it can weaken the exam. Fortunately, there are some rather simple statistical ways of checking individuals' item. This is done by studying the students' responses to each item. When formalized this procedure is called "item analysis ( Harold S. Madsen, 1983:180). An item analysis tells us basically three things: how difficult each item is, whether or not the question discriminate or tells the difference between high and low students, and which distracters are working as they should. An analysis like this is used with any important exam- for example, review tests and tests given at the end of a school term or course.

### 1. The Definition of Item Analysis

According to Nitko in his book he stated that, "Item analysis refers to the process of collecting, summarizing, and using information about individual test items, especially information about pupil's responses to items (Anthony J. Nitko, 1983:284).

Item analysis usually provides two kinds of information on items: item facility, which helps us decide if test items are at the right level for the target group, and item discrimination, which allows us to see if individual items are providing information on candidates' abilities consistent with that provided by the other items on the test. Here the Researchers concluded that item analysis is the process of collecting information about pupil's responses to the items, to see the quality of test items. More specific, item analysis information can tell us if an item was too easy or too hard, how well it discriminated between high and low scores on the test and whether all of the alternatives function as intended. Item analysis data also aids in detecting specific technical flaws and thus further provides information for improving the test items.

## 2. Discriminating Power

Item discriminatory power of a test is its ability to separate good students from poor students. These students groups are defined by their scores on the test as whole. The difference between the percentage of the top scoring 27% and bottom scoring 27% of students get the item right in its discrimination index (H.J.X. Fernandes, 1984: 27). As well as knowing how difficult an item is, it is important to know how it discriminates, that is how well it distinguishes between students at different levels of ability.

The discrimination index can range from -1 to +1. Items with positive values of the discrimination index are desired because those are the items that are contributing to the usefulness of the total score. When the discrimination index is near zero, it indicates that the item is contributing nothing to the discriminating power of the overall test (William Wiersma, 1990:245). When a larger proportion of students in the lower group got the item right more than those in the upper group, it discriminates negatively. And since more students in the upper group than in the lower group got the item right, it is discriminating positively (Gilbert Sax, 1999:191).

Item discriminating power can be obtained by subtracting the number of students in the lower group who got the item right (U) from the number of students in the upper group who got the item right (L) and dividing by the total number of students in one group included in the item analysis (N). It summarized in formula form, as below:

$$DI = \frac{U - L}{N}$$

Where:

DI = the index of discriminating power

U = the number of pupils in the upper group who answered the item correctly

L = the number of pupils in the lower group who answered the item correctly

N = number of pupils in each of the groups (Charles D. Hopkins & Richard L. Antes, 1990:279).

The classifications of the index of discriminating power (D) are:

DI = 0.70 – 1.00	= Excellent
0.40 – 0.70	= Good
0.20 – 0.40	= Satisfactory
≤ 0.20	= Poor

Negative value on D= Very poor (Anas Sudijono, 2006: 389).

## C. RESEARCH METHODOLOGY

### 1. Research Method

This research is categorized as descriptive analysis; because it is intended to describe the objective condition about the discriminating power of students' summative test a Exact semester of second grade of SMPN 2 Padalarang. Besides, this study is called analysis, because it analyze how well the items of English summative test can discriminate between the students who have achieved well and those who have achieved poorly. This study is considered as quantitative research, because the Researchers used some numerical data which is analyzed statistically.

### 2. Research Setting

The research was conducted at SMPN 2 Padalarang This is located at Jl. G. A. Manulang, West Bandung. The Researchers did the research in Maret 2016. The Researchers took the English summative test papers and the students' answer sheets of second grade period of 2015-2016 to be analyzed.

### 3. Technique of Sample Taking

The Researchers took the sample from second year students of SMPN 2 Padalarang. The total number of second year students is 238 students; those are divided into 6 classes. The Researchers took 25% of the total number of the second year students as a sample. That is 25% x 238 = 60 students. The Researchers used an ordinal sampling to get the students' answer sheet. The Researchers divides the students into three groups; they are upper, middle, and lower group. Then the Researchers takes upper and lower group only to be analyzed.

### 4. Technique of Data Collecting

To collect data connecting with the topic of discussion, the Researchers came to the school to get the permit from the headmaster to take students' answer sheet and the test question paper of English summative test of second year students of SMPN 2 Padalarang to be analyzed.

**5. Research Instrument**

a. Students' answer sheet

The students answer sheet is papers in which the students give their answer that correspond to the English summative test. The English summative test that the Researchers used is the final exact semester for the second year students of SMPN 2 Padalarang academic year 2015-2016, prepared by the teacher english team especially by the school that is has been implemented 13 curriculum (*kurtilas*).

b. Interview to the stakeholders such as teachers and the members of MGMP

**6. Technique of Data Analysis**

In this research, the researchehr used quantitative method to analyze the discriminating power of English summative test items of second year of SMPN 2 Padalarang by using a statistic formula, namely, the Discriminating Power Index:

$$DI = \frac{U - L}{N}$$

Where:

DI= the index of discriminating power

U= the number of pupils in the upper group who answered the item correctly

L= the number of pupils in the lower group who answered the item correctly

N= number of pupils in each of the groups

(Charles D. Hopkins & Richard L. Antes, 1990: 279)

The classifications of the index of discriminating power (*D*) are:

- DI = 0.70 – 1.00 = Excellent
- 0.40 – 0.70 = Good
- 0.20 – 0.40 = Satisfactory
- ≤ 0.20 = Poor

Negative value on D = Very poor ((Anas Sudijono, 2006: 389).

**D. RESEARCH FINDINGS**

**1. Description of Data**

The Researchers use the English summative test in the exact semester of second grade of SMPN 2 Padalarang as the data. This English summative test was held on Wednesday, July 13<sup>rd</sup> 2015, that must be finished in 120 minutes. The total numbers of test items are 50 questions, which all of them are multiple choice items.

The total numbers of students that took part in this analysis are 60 students. Kelley in the

book “*Classroom Measurement and Evaluation*” demonstrated that the selection of criterion groups based upon the upper 27 % and lower 27 % of the papers provide the greatest confidence that the upper group is superior in the trait measured by the test as compared to the lower group. The middle 46 percent of the papers is not used when 27 % in the upper and 27 % in the lower groups are employed in item analysis (Charles D. Hopkins, Richard L. Antes, 1990:275).

Based on that statement, the Researcher classified the students into three groups; upper, middle and lower group. The Researcher took only 27% of the lower group and 27% of the upper group for this analysis. And the rest students that belong to the middle group will not take part to this analysis. The next table is the students' scores and group position in English summative test.

**Table 4.1**  
**The Students' Scores and Group Position of English Summative Test In the Exact Semester**

SCORES					
NO	UPPER	NO	MIDDLE	NO	LOWER
1	82	17	64	45	52
2	80	18	62	46	50
3	78	19	62	47	50
4	76	20	60	48	50
5	76	21	60	49	48
6	74	22	60	50	46
7	74	23	60	51	46
8	74	24	60	52	44
9	74	25	58	53	44
10	72	26	58	54	42
11	72	27	58	55	40
12	72	28	58	56	38
13	72	29	58	57	34
14	64	30	58	58	34
15	64	31	58	59	34
16	64	32	56	60	32
		33	56		
		34	56		
		35	56		
		36	56		
		37	54		
		38	54		
		39	54		
		40	54		
		41	54		
		42	54		
		43	52		
		44	52		

We can see at Table 4.1, it shows that students who are taking the test are classified into 3 groups: upper group, middle group and lower group. The Researchers took 27% or 16 students from upper and lower group to be analyzed. The highest score in upper group is gained by one student in score 82. The lowest score in upper group is gained by three students in the same score 64. Meanwhile the highest score in lower group is gained by one student in the same score 52. So, the lowest score in lower group is gained by one student in score 32.

Before the Researchers analyzes the data, the Researchers have calculated the data into statistic calculation. The Researchers used Discrimination Index formula to find the discriminating power criteria of English summative test. The table is as follows:

**Table 4.2**  
**The Discriminating Power Index of the Upper and Lower Group**

Item number	Total correct answer		U- L	DI = $\frac{U-L}{N}$	Remark*
	Upper Group	Lower Group			
1	0	0	0	0	Poor
2	16	7	9	0.56	Good
3	16	16	0	0	Poor
4	12	11	1	0.06	Poor
5	15	5	10	0.62	Good
6	13	6	7	0.43	Good
7	11	2	9	0.56	Good
8	12	6	6	0.37	Satisfactory
9	12	11	1	0.06	Poor
10	15	9	6	0.37	Satisfactory
11	13	7	6	0.37	Satisfactory
12	1	1	0	0	Poor
13	14	10	4	0.25	Satisfactory
14	11	10	1	0.06	Poor
15	15	4	11	0.68	Good
16	11	2	9	0.56	Good
17	14	1	13	0.81	Excellent
18	10	9	1	0.06	Poor
19	10	6	4	0.25	Satisfactory
20	10	9	1	0.06	Poor
21	5	0	5	0.31	Satisfactory
22	3	0	3	0.18	Poor
23	16	13	3	0.18	Poor

Item number	Total correct answer		U- L	DI = $\frac{U-L}{N}$	Remark*
	Upper Group	Lower Group			
24	7	12	-5	-0.31	Very poor
25	14	7	7	0.43	Good
26	7	10	-3	-0.18	Very poor
27	9	7	2	0.13	Poor
28	14	8	6	0.37	Satisfactory
29	12	7	5	0.31	Satisfactory
30	13	4	9	0.56	Good
31	8	4	4	0.25	Satisfactory
32	1	2	-1	-0.06	Very Poor
33	12	9	3	0.18	Poor
34	13	6	7	0.43	Good
35	14	2	12	0.75	Excellent
36	16	8	8	0.50	Good
37	11	6	5	0.31	Satisfactory
38	11	7	4	0.25	Satisfactory
39	10	2	8	0.50	Good
40	16	8	8	0.50	Good
41	12	2	10	0.62	Good
42	16	16	0	0	Poor
43	15	7	8	0.50	Good
44	7	2	5	0.31	Satisfactory
45	15	12	3	0.18	Poor
46	13	4	9	0.56	Good
47	12	7	5	0.31	Satisfactory
48	15	7	8	0.50	Good
49	14	9	5	0.31	Satisfactory
50	11	7	4	0.25	Satisfactory

\* (Anas Sudijono, 2006:389).

Based on the data above, the percentage of discriminating power of English summative test is:

**Table 4.3**  
**The percentage of Discriminating Power**

No	Discriminating power	Total item	%	Item number
1	Excellent	2	4%	17, 35
2	Good	16	32%	2, 5, 6, 7, 15, 16, 25, 30, 34, 36, 39, 40, 41, 43, 46, 48

No	Discriminating power	Total item	%	Item number
3	Satisfactory	15	30%	8, 10, 11, 13, 19, 21, 28, 29, 31, 37, 38, 44, 47, 49, 50
4	Poor	14	28%	1, 3, 4, 9, 12, 14, 18, 20, 22, 23, 27, 33, 42, 45
5	Very Poor	3	6%	24, 26, 32

The table above showed that: there are 2 test items (4%) are categorized into excellent test item, which is showed by the test items number 17 and 35. It is categorized as excellent test item because its discriminating index is in range between 0.70 – 1.00. There are 16 test items (32%) are categorized into good items, that range from 0.40 – 0.69, they are test items number 2, 5, 6, 7, 15, 16, 25, 30, 34, 36, 39, 40, 41, 43, 46, and 48. There are 15 test items (30%) are categorized as satisfactory test items for their discriminating index are in range 0.20 – 0.39, they are test items number 8, 10, 11, 13, 19, 21, 28, 29, 31, 37, 38, 44, 47, 49, and 50.

Meanwhile, 14 test items (28%) are categorized into poor test items because their discriminating index are range in 0.00 – 0.19, they are test items number 1, 3, 4, 9, 12, 14, 18, 20, 22, 23, 27, 33, 42, and 45. At last, there are 3 test items (6%) are categorized as very poor item as their discriminating index are range in negative values.

**2. Data Analysis**

In analyzing the discriminating power of the data, the Researchers listed the students’ responses of each number of the test firstly (see appendix). Then the next step is to make a format of item analysis. This format and the result of this format labeled table 4.2. The last step is to count discriminating power of all items using this formula:

$$DI = \frac{U - L}{N}$$

Where:

- DI = the index of discriminating power
- U = the number of pupils in the lower group who answered the item correctly
- L = the number of pupils in the lower group who answered the item correctly
- N = number of pupils in each of the groups

The result of this last step can be seen also in the table 4.2 In this table, result of each item will

be in decimal then the Researchers categorized each item according to this formula: The classifications of the index of discriminating power (D) are:

- DI= 0.70 – 1.00 = Excellent
- 0.40 – 0.70 = Good
- 0.20 – 0.40 = Satisfactory
- ≤ 0.20 = Poor
- Negative value on D = Very poor

Based on the data of item analysis result in discriminating power above, the Researchers can conclude that from 50 items:

1. There are 33 test items (66%) are categorized into good test items which is range from 0.25 – 0.81.
2. There are 14 test items (28%) are categorized into poor test items because their discriminating index are range in 0.00 – 0.18.
3. There are 3 test items (6%) are categorized as very poor item as their discriminating index are range in negative values that -0.06 – -0.31

**3. Data Interpretation**

For whole items, the Researchers can interpret that the discriminating power of English summative test prepared by School which has been Applied Kurtilas tested at the second grade of SMP N 2 Padalarang belongs to good discriminating power, because there are 33 test items or 66% from 50 test items is ranging from 0.25 – 0.81.

Based on the table 4.2 on the previous page, the Researchers concluded the achievement of upper group students in their English test. From 50 multiple choice items, none of the students got the perfect score. The following description tells about the responses of each item:

1. There is no student who answered the item no 1 correctly.
2. There is 1 student who answered the item no 12 and 32 correctly.
3. There are 3 students who answered the item no 22 correctly.
4. There are 4 students who answered the item no 25 correctly.
5. There are 5 students who answered the item no 21 correctly.
6. There are 7 students who answered the item no 24, 26, and 44 correctly.
7. There are 8 students who answered the item no 31 correctly.
8. There are 9 students who answered the item no 27 correctly.

9. There are 10 students who answered the item no 18, 19, 20, and 39 correctly.
10. There are 11 students who answered the item no 7, 14, 16, 37, 38, and 50 correctly.
11. There are 12 students who answered the item no 4, 8, 9, 29, 33, 41, and 47 correctly.
12. There are 13 students who answered the item no 6, 11, 30, 34, and 46 correctly.
13. There are 14 students who answered the item no 13, 17, 28, 35, and 49 correctly.
14. There are 15 students who answered the item no 5, 10, 15, 43, 45, and 48 correctly.
15. There are 16 students who answered the item no 2, 3, 23, 36, 40, and 42 correctly.

## E. CONCLUSION AND SUGGESTION

### 1. Conclusion

Finally the Researchers can conclude based on the analysis and the interpretation in the previous chapter, its can be categorized into five different range of discrimination power indexes. First, they are two test items (4%) that is categorized into excellent test items. Then, there are 16 test items (32%) that are categorized into good test items. Besides that, 15 test items (30%) are categorized into satisfactory test items. Fourth, there are 14 test items (28%) are categorized as poor test items. Lastly, 3 test items (6%) are categorized into very poor test items. And there are 33 test items (66%) of English summative test regarded as a good discriminating power that range from 0.25 – 0.81 and it can be used for the next test. Meanwhile, 14 test items (28%) are needed to be revised for their poor value in differentiating the ability of the upper from the lower group that range from 0.00 – 0.18. And 3 test items (6%) have to be eliminated, because those items have negative discrimination index that range from -0.06 - -0.31.

From the explanation above, the Researchers concludes that the English summative test which is tested at second grade of SMPN 2 Padalarang has good discriminating power, because 33 items (66%) of the test items have fulfilled the criteria of a positive discriminating power which range from 0.25 – 0.81.

### 2. Suggestion

The Researchers aware that this research is not perfect yet but the Researchers wants to gives some suggestions to all stakeholders, such as:

1. Teachers, the teachers should give good techniques when they prepare to make summative test items.
2. Test makers should make and give test items which have satisfactory, good and excellent criteria in order can be used by the teachers for the future evaluation
3. Test makers should revise the test items which have poor criteria and discard those which have very poor criteria, so that they can be used for the next evaluation.
4. The best test is always indicated the difficult one, but the best test item is the test based on the material that the students have been learnt before.

## BIBLIOGRAPHY

- Anderson, J. Charles, et.al. *Language Test Construction & Evaluation*, Melbourne: Cambridge University Press, 1995.
- Bahman, Lyle F. *Statistical Analysis for Language Assessment*, Cambridge: Cambridge University Press, 2004.
- Brown, H. Douglas. *Teaching by Principles, An Interactive Approach to Language Pedagogy*, White Plains: Addison Wesley, Longman, 2001.
- Brown, H. Douglas. (2004) *language Assessment Principles and Classroom Practices*. Longman: San Francisco State University.
- Brown, J. (2005) *Testing in Language Programs*. Englewood Cliffs: Prentice Hall Regents.
- Brown, James Dean, *Testing in Language Programs*, New Jersey: Prentice Hall Regent, 1996.
- Brown, J.D. and Hudson, T. (1998). *The alternatives in language assessment*. TESOL Quarterly 32, 653–75.
- Gay, L.R. *Educational Evaluation and Measurement*, New York: Macmillan, Inc.,1985.
- Gronlund, Norman E. *Measurement and Evaluation in Teaching*, New York: Macmillan Publishing Co., Inc, 1981.
- Hughes, Arthur. *Testing for Language Teachers*, Cambridge: Cambridge University Press, 2003.
- Lado, Robert. *Language Testing*, London: Longman Group Limited, 1983.
- Madsen, Harold S. *Techniques in Testing*, New York: Oxford University Press, 1983.
- Sudijono, Anas. *Pengantar Evaluasi Pendidikan*, Jakarta: PT. Raja Grafindo Persada, 2006.



Wiersma, William. *Educational Measurement and Testing*, Boston: Allyn & Bacon, 1990.