

DESIGNING MULTIPLE CHOICE TEST OF VOCABULARY FOR THE FIRST SEMESTER STUDENTS AT ENGLISH EDUCATION DEPARTMENT OF ALAUDDIN STATE ISLAMIC UNIVERSITY OF MAKASSAR

Raidah Mahirah, Djuwairiah Ahmad & Sukirman
English Education Department of UIN Alauddin Makassar
raidahmahirah@gmail.com

ABSTRACT: This study aims to develop and design Vocabulary Test in the first semester students at English Education Department of Alauddin State Islamic University of Makassar. The research design was Research and Development (R&D). It totally applied ADDIE Model. The steps of the model are Analysis, Design, Develop, Implementation and Evaluation. The type of data of this research was quantitative data. The research instrument was a rubric dealing the quality of the test produced. The findings showed that the content of material, language, and layout of the product were totally clear and understandable. The product was valid to be implemented in testing the students' vocabulary mastery. It can be seen from the difficulty level, discrimination power, validity, and reliability of the product obtained from the score of the students' answers.

Keywords: *Vocabulary Multiple Choice Test, Difficulty Level, Discrimination Power, Validity and Reliability*

INTRODUCTION

Testing is very important in learning because it can measure and collect the information about the students' ability. English test can also benefit students in measuring their language mastery. Besides that, testing given by the lecturers or teachers aims at knowing whether the objectives of the course were achieved significantly or not and know how effective their learning process the lecturers conducted was in the last session. Based on the preliminary study conducted on April 2015 at English Education Department of Allauddin state Islamic University of Makassar, the problems faced by the lecturers were the practical constrain in measuring vocabulary ability of the students. Then, the lecturers were lack of understanding about designing test.

The problems stated previously occur because of many factors. *First*, The lecturers did not pay much attention to test vocabulary when they designed the test, they also did not create the test based on the characteristic of a good test such as difficult level, discrimination level, validity and reliability. *Second*, the test created by the lecturers were not acceptable with the materials because they designed the test only based on the ability of them and they not pay attention to make a blue print before design test. The lecturers were designing tests without based on syllabus and materials. *Third*, the lecturers only developed method how to master vocabulary.

The reason why the lecturers lack of understanding about language testing assessment; there was no more information about how to design test based on the good characteristic of test, the lecturers thought if the mastering material was more important than evaluation. There were many steps in analyzing a test. It made the lecturers lazy to do every step in designing test whereas it helped the lecturers in process on teaching.

Consequently, the lecturers cannot measure the student vocabulary level. Besides that, there was no motivation in learning, because the students did not know how high or how low their level in vocabulary ability. In this case, assessing vocabulary helped greatly the lecturer and students to know the students' vocabulary ability. Then, the lecturer cannot know whether the question was acceptable for the students or not when they design a test. Also, the lecturer cannot know whether the goals and objective of the course achieved or not.

After identifying and analyzing the factors, the researcher became aware that in order to solve the problems, the researcher has to design and develop multiple choice tests which can be acceptable and appropriate with the materials. The researcher also created the test in accordance with the characteristic of a good test; difficulty level, discrimination power, validity and reliability.

Considering the factors affected the problems above, the researcher viewed that the test which was appropriate to measure vocabulary ability was multiple choices. For the reason, the researcher viewed that it was very easy and quick for the examiner to correct this test because he or she just put ticks or crosses. On the other hand, we do not have to worry about subjectivity because only one answer should be correct (Pavlů 2009:19).

In others word, multiple choice is one of the tests for making testing that be simple but may serve as a vocabulary check (Brown 2004:194). Hopefully, the researcher can design multiple choice tests based on a good characteristic of the test. So, the researcher designed multiple choice test of vocabulary and made the testing more interesting. Besides that, this research would be information source for lecturer who will design test based on a good characteristic of test. Moreover, the other goal was to use the vocabulary more in practice and more intensively so that the students would remember the vocabulary better.

Based on the problem stated previously, the researcher conduct a research entitled *“Designing Multiple Choice Test of Vocabulary at English Education Department of Alauddin State Islamic University of Makassar”*.

LITERATURE REVIEW

Some researchers have conducted researches related to “designing test” and what they have found are shown such as Zhongshannvgao (2007) conducted a study on Designing and Revising a Multiple Choice Vocabulary Test. He found out that multiple choice testing appeals to many people for its high reliability and efficiency in terms of scoring, but the

construction of a good item requires a tremendous amount of time and effort. In vocabulary assessment, the decision on whether to attempt this format and how to design a test depends on the context, the needs of the taster, the test purpose and, above all, the selected construct to measure. As long as a test is proved to be valid and can bring benefits to both students and teachers”.

Another research come from Öztürk (2007) conducted a research on the designing test faced by Multiple-Choice Test Items of Foreign Language Vocabulary. The research results reveal that the English Foreign Language teachers made much more mistakes in vocabulary section than in grammar section. The findings imply that even though the EFL teachers have been provided with the principles for constructing multiple-choice items in advance, the teachers still construct improper items. Language testing plays an important role in both teaching and learning. Well-constructed tests can enhance learning and motivate students.

On the other side, Pavlu (2009) with the research “Testing Vocabulary” dealt with options how vocabulary may be tested. The thesis was divided into theoretical and practical part. The theoretical part comprised in two big subdivisions which were testing itself and Vocabulary. In the first part he dealt with the question whether testing was important and different reasons for testing, and the next part explains two basic principles of testing which were reliability and validity. And the last was focused on techniques of testing and the examples.

The related of those research findings above with this research in designing test is how to designing test in vocabulary, especially in multiple choice. They have found that much more mistakes in designing test; the test is not reliable and valid. The mistakes can make the bad test, with the result; the test cannot measure the student ability favorably. Therefore, this research tries to design and develop strategy for designing multiple choice tests in vocabulary. So, the researcher will explain how to design multiple choice tests of vocabulary in this research.

RESEARCH METHOD

The research method used by researcher in this research was Research and Development (R&D). R&D is a name of research designs involving the classroom problems, studying recent theories of educational product development, developing the educational products, validating the product to experts, and field testing the product (Latif, 2012). The researcher adopted ADDIE model. The ADDIE model as “a colloquial term used to describe a systematic approach to instructional development, virtually synonymous with instructional systems development” Molenda (2003:34), Addie is a generic instructional design model that provides an organized process for developing instructional materials

(Shelton & saltsman 2011:566). ADDIE is acronym which stands for **A**nalysis, **D**esign, **D**evelopment, **I**mplementation, and **E**valuation.

ADDIE model is design for the learners to achieve the goals and objectives of the course or syllabus. It allows for the evaluation of the materials. It also provided simple procedures to design and develop the tests.

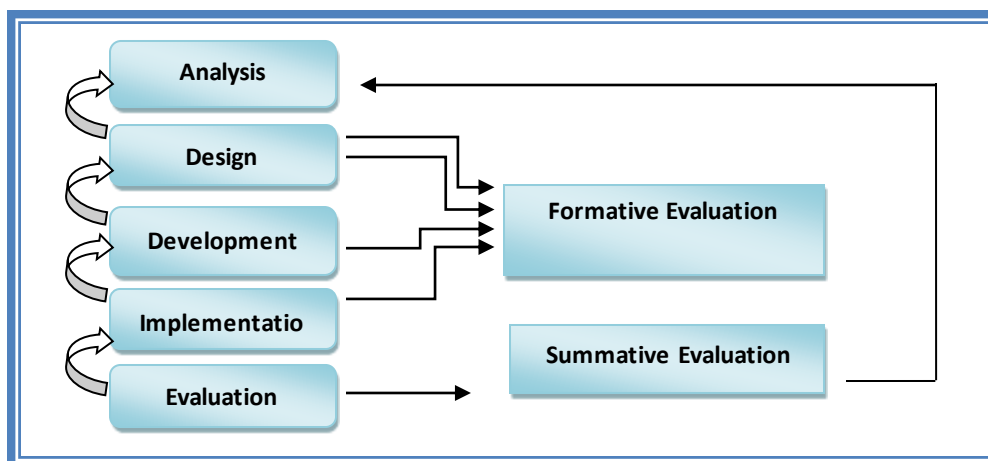


Figure 2. *ADDIE Model, Diagram by: Steven J. McGriff*

The procedures in design multiple choice test of vocabulary deals with ADDIE model which provides five phases in terms of analysis, design, development, implementation, and evaluation.

1. Analysis

In this phase, the researcher identified and developed clear understanding of materials. She also identified a set the goals and objective of the course based on materials that was given from their lecturer. Then, the researcher considered timeline and budget needed in designing the test that is also important. In Addition, this phase refer to need analysis.

Need analysis is a set of procedures used to collect information about learners, needs (Richards, 2003:51) as cited in Sukirman 2012.

2. Design

In this phase, the researcher designed multiple choice test of vocabulary considering the goals and objective of the learning process, designing blue print (*see more in appendix 2*), determining target population description, selecting delivery materials which the materials were appropriate that the signed be a test.

3. Development

This phase was done based on the two previous phases, analyze and design phase. Before phase, we have been said about blue print. In this phase, the research developed blue print in this stage. In the blue print, there are lists of materials, so the blue print guided the researcher to designing multiple choice test based on materials and syllabus.

There are some steps in doing this phase. *First*, the researcher listed what activities which can assist the learners learn the materials. Second, she selected the best way which was appropriate with learners' styles. Third, she designed, developed and produced multiple choice test of vocabulary dealing with the materials and syllabus of the course. Then, she organized the test. After that, she validated the test to experts to make sure whether the test was appropriate to materials as well as the syllabus of the course or not. Finally, the final product was ready to be implemented.

4. Implementation

This phase deals with trying-out the product. In this case, the product was implemented in the real learning/teaching. The purpose of this phase to prove whether the test was appropriate for the target learners or not. If not, the product was revised and was tried out again.

5. Evaluation

This phase was designed to measure the rate of quality of the materials as being implemented. It measured the appropriateness of the designing test. In this evaluation, one expert involved to check the quality of the product.

There were two kinds of evaluation in this phase generally, Formative and summative evaluation. *Formative evaluation* was ongoing and during between phases. The purpose to improve the quality of the content of the test before the final steps of test was implemented. Meanwhile, *summative evaluation* was the final evaluation of the process designing test.

FINDINGS AND DISCUSSION

A. Finding

The result of this research finished based on steps of R&D which have been done on the design test. There were five steps that have done to get a good product. The steps were;

1. *Analysis*

In this phase, the researcher observed about testing that teacher gave to the student in vocabulary in context course and the researcher found some problems in the item of test. There were some lecturer did not pay much attention to design test of vocabulary, therefore the lecturer did not design the test based on syllabus and materials and the lecturer also did not measure difficult level, discrimination level, validity and reliability of the testing.

2. *Design*

The researcher designed what she did in this research. The researcher designed blue print based on syllabus and materials of vocabulary in context deals with synonym, antonym, rewording, details, collocation, reference, inference, and word form.

3. *Development*

The product of this research consists of 40 items of testing. Every single number of

test developed based on syllabus and material that had been designed on blue print. (*See more in appendix 2*)

4. Implementation

This phase dealt with trying-out the product. Before trying the product, the product was analyzed by the expert. It identified the validity instrument of testing by using rubrics. It included some indicators to measure the validity of the product (*see more in appendix 5*).

a. Tried out 1

After analyzed by the expert, the product tried out. The product revised in the first based on comment expert and students answer. Based on the researcher's statically calculation, the data of the students' answers demonstrated that there were 13 valid items of the test, namely 1, 3, 5, 7, 11, 15, 17, 19, 25, 26, 27, 37 and 39. They had validated index appropriate by the indexes in the table of the critical values of product moment stated in Arikunto (2003:76). There were two items that received and repair, namely 9 and 23. They validated index gone up to the indexes in the table of the critical values of product moment. On the contrary, the other items that invalid for the data showed that their validity was not appropriated with the indexes in the table of the critical values of product moment. The researcher also analyzed the reliability of the item test. As explain that implicitly that the result of r in a test items was not appropriate with the table of product moment. It meant that the item was considered to be not reliable. To be clearer, the researcher provided the table that gave a brief description about the validity of each item.

Table IV. Validity index

No. Soal	Validity	
	Index	Category
1	0.602	Valid
2	0.336	Invalid
3	0.407	Valid
4	0.122	Invalid
5	0.44	Valid
6	0.264	Invalid
7	0.502	Valid
8	0.226	Invalid
9	0.33	Invalid
10	0.169	Invalid
11	0.364	Valid
12	0.436	Valid
13	-0.272	Invalid
14	0.497	Valid
15	0.34	Valid
16	0.034	Invalid

17	0.667	Valid
18	0.052	Invalid
19	0.666	Valid
20	0.526	Valid
21	0.384	Valid
22	0.318	Invalid
23	0.5	Valid
24	0.448	Valid
25	0.502	Valid
26	0.589	Valid
27	0.353	Valid
28	-0.008	Invalid
29	0.057	Invalid
30	0.288	Invalid
31	0.189	Invalid
32	0.049	Invalid
33	0.354	Valid
34	0.125	Invalid
35	0.329	Invalid
36	0.396	Valid
37	0.613	Valid
38	0.137	Invalid
39	0.378	Valid
40	0.149	Invalid

The item analyzed the reliability. As explain implicitly that if the result of r in a test item was not appropriate with the table of product moment, it meant that the items was considered to be not reliable.

Table V. Reliability index

r-table (tarafsig 5 % & taste 36)	Reliability
0.329	-0.40 (Not reliable)

Each item of this product analyzed about difficulty levels and capacity of distinctive. The researcher provided the table that gave a brief description about the status of each item.

Table VI. Difficulty levels analysis

Easy	Average	Difficult
1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 17, 20, 21, 23, 24, 26, 29, 35	9, 11, 15, 18, 19, 22, 25, 27, 30, 31, 32, 33, 34, 36, 37, 39, 40	7, 13, 16, 28, 38

Based on difficulty level analysis the items demonstrated that there were 18 items in easy level, 17 items in average level, and 5 items in difficulty level analysis. Also the items were analyzed by the discrimination power. The researcher provided the analysis that gave brief description about the status of each item.

Tabel VII. Discrimination power analysis

Good	Receive and repair	Repair	Fail
1, 2, 3, 5, 7, 11, 15, 17, 19, 25, 26, 27, 37, 39	9, 23	8, 12, 14, 20, 22, 24, 33, 34, 40	4, 6, 10, 13, 16, 18, 21, 28, 29, 30, 31, 32, 35, 36, 38

Based on the discrimination power analysis the items demonstrated that there were 15 items in good level, 2 items in received and repaired level, 9 items in repaired level, and 15 items in failed level. As the result, there were 11 items which have to revise it and 15 items which have to change and create new items test, and there were 26 items cannot measure students' knowledge.

b. Tried out II

After the first try out, the product revised. Then the product tried out and analyzed in the second time. Based on the researcher's statically calculation, the data of the students' answer demonstrated that there were 5 invalid items of the test.

Table VIII. Validity index

No. Soal	Validity	
	Index	Category
1	0.382	Valid
2	0.439	Valid
3	0.757	Valid
4	0.574	Valid
5	0.509	Valid
6	0.462	Valid
7	0.493	Valid
8	0.547	Valid
9	0, 781	Valid
10	0.757	Valid
11	0.574	Valid
12	0.377	Valid
13	0.109	Invalid
14	0.548	Valid
15	0.582	Valid
16	0.453	Valid

17	0.462	Valid
18	-0.047	Invalid
19	0.263	Invalid
20	0.342	Valid
21	0.841	Valid
22	0.078	Invalid
23	0.459	Valid
24	0.536	Valid
25	0.645	Valid
26	0.676	Valid
27	0.648	Valid
28	0.556	Valid
29	0.811	Valid
30	0.657	Valid
31	0.251	Invalid
32	0.624	Valid
33	0.608	Valid
34	0.731	Valid
35	0.509	Valid
36	0.496	Valid
37	0.737	Valid
38	0.485	Valid
39	0.503	Valid
40	0.622	Valid

The item analyzed the reliability in the second times. As explain implicitly that if the result of r in a test item was not appropriate with the table of product moment, it meant that the items was considered to be not reliable.

Table IX. reliability index

r-table (tarafsig 5 % & taste 36)	Reliability = - 0.40 (Not Reliable)
0.329	0.094

Each item of this product analyzed about difficulty levels and capacity of distinctive in the second times. The researcher provided the table that gave a brief description about the status of each item.

Table X. Difficulty levels analysis

Easy	Average	Difficult
10, 14, 18, 26, 27, 29, 30, 31, 32, 34, 37	2, 3, 4, 6, 7, 8, 9, 11, 16, 13, 15, 17, 19, 20, 21, 22, 23, 24, 25, 28, 33, 35, 36, 38, 39, 40	1, 5, 12

Based on difficulty level analysis the items demonstrated that there were 3 items in easy level, 26 items in average level, and 11 items in difficulty level analysis. Also the items analyzed by the discrimination power. The researcher provided the analysis that gave brief description about the status of each item.

Tabel XI. Discrimination power analysis

Good	Receive and repair	Repair	Fail
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 20, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40	19	-	13, 18, 22, 31

Based on the discrimination power analysis the items demonstrated that there were 35 items in good level, 1 item in received and repaired level, there was no item in repaired level, and 4 items in failed level. As the result, there were 5 items which have to revise. The items cannot measure students' knowledge.

c. Tried out III

In this phase, the researcher revised the items that repaired and filed in the second tried out. As the result of the second tried out, there were 5 items which have to revised. In the last tried out, the researcher added 5 items to the items that repaired and filed as selection to find good items. With the result that there were 10 items tried out and analyzed in the third tried out times as the final tried out. Based on researcher's statically calculation, the data of the students' answer demonstrated that there were 5 invalid items of the test.

Table XII. Validity index

No. Soal	Validity	
	Index	Category
1	0.707	Valid
2	0.719	Valid
3	0.284	Invalid
4	0.184	Invalid
5	0.46	Invalid
6	0.005	Invalid
7	0.409	Invalid
8	0.42	Valid
9	0.402	Valid
10	0.347	Valid

The item was analyzed the reliability in the third times. As explain implicitly that if the result of r in 10 items test were not appropriate with the table of product moment, it meant that the items was considered to be not reliable.

Table. XIII Reliability index

r-table (taraf sig 5 % & taste 36)	Reliability
0.329	- 0.16 (Not Reliable)

Each item of this product analyzed about difficulty levels and capacity of distinctive in the third times. The researcher provided the table that gave a brief description about the status of each item.

Table X. Difficulty levels analysis

Easy	Average	Difficult
1, 3, 6, 7	2, 4, 5, 8, 9, 10	-

Based on difficulty level analysis, the items demonstrated that there were 4 items in easy level, 6 items in average level, and there was no item in difficulty level analysis. Also the items were analyzed by the discrimination power. The researcher provides the analysis that gave brief description about the status of each item.

Tabel XI. Discrimination power analysis

Good	Receive and repair	Repair	Fail
1, 2, 8, 9, 10	5, 7	-	3, 4, 6

Based on the discrimination power analysis of the items demonstrated that there were 5 items in good level, 2 item were received and repaired level, there was no item were repaired level, and 3 items were failed level. As the result, there were 5 items that had good criteria of the test. The items fulfill the items' needed in the second tried out. As the result the good items in the second and third tried out can measure the students' knowledge.

The aim of final product revision was to determine whether the product was ready to use. Where, the product completed some of the criteria of good test. In this step the product was analyzed from the expert, such as; the validity, reliability, difficulty level, discrimination power that the researcher revised the product. But after the researcher analyzed the product and was commented by the expert, there was a character that not fulfill be a good test. It was reliability. It will be explained in discussion part.

5. Evaluation

There were two kinds of evaluation in this phase generally, Formative and summative evaluation. *Formative evaluation* was ongoing and during between phases. The purpose was to improve the quality of the content of the test before the final steps of test was implemented. Meanwhile, *summative evaluation* was the final evaluation of the process designing test.

Discussion

This part presents the result of the data analysis. The data were found in five steps of R&D which was adopted from ADDIE models (Shelton & saltsman 2011:566). The researcher analyzed about validity, reliability, and difficulty level and discrimination power.

The results of all items were valid. It was analyzed in three try out steps. The product was appropriate with phopam:83 that validity is, hands down, the most significant concept is in assessment. He revealed that validity is the important step of analysis of the test that helped the lecturer to make suitability between the test and materials. Therefore, the lecturer should consider the validity in the test that they made to create a good test.

Also, the all items were analyzed in order to measure their reliability, difficulty level and discrimination power. It was found that the results of the analyses were difficulty level and discrimination power showing that the items test were balance and acceptable. This statement appropriate with Madsen:181 that difficulty level is simply the percentage of student (high and low combined) who got each question right, so the researcher could find the differentiate between difficult items test or easy items test. It was in line with discrimination power that was how well it differentiates between high and low level with more advanced language skill and those with less skill. So, the researcher can measure the ability of the student through the test.

After the researcher analyzed all of the characteristics of the item test, the researcher found a different result from reliability. In rate of r-Table (significant level 5% with 36 testers) was 0.329. But, the result of the analyzing the test based on the student answer was 0,286. It meant that the reliability for all the items was not acceptable.

The reasons why the reliability was not acceptable because the data analysis was not fulfill the criteria of reliability that was consistent and dependable. The test can use in a classroom to measure the ability of the student, but the test cannot be being a part of a bank test. It because the test was not constant, it can be used in a change situation, also just for measure the ability of the student in midterm or final examination..

The other reason came from the result of the analyzing data was in an unbalance way. The diagram of the result showed the ability of the student and the difficulty level of the test was unevenness of the data (*see more in appendix 7*), and for the chart statically collection see more in *Appendix 8*. Due to the researcher was limited by the time, ultimately the researcher decided to end this research. With the consequence, this research will be developed in the future.

CONCLUSION

Based on the research findings and discussion in the previous chapter, the researcher comes to the following conclusion. *First*, the product analyzed about Difficulty level analysis the items demonstrated that is good to use it in the students, because each item on the

difficulty level was balanced between easy, average and difficult levels. *Second*, the product analyzed about Discrimination Power. The result showed that the entire items test was good and acceptable. *Third*, the product analyzed about the validity. The result showed that the product was good. *Fourth*, the product analyzed about Reliability. The result showed that the product was not reliable.

Suggestion

Concerning with the result of this research, the researcher would like to give the following suggestion:

1. The result of this research found that the validity, discrimination power, difficulty level were appropriate with the purpose of the research, except the reliability. So, for the next researcher can be more focused in reliable analysis.
2. For the lecturers, they should know how to develop Vocabulary test based on Difficulty level, Discrimination Power, validity and Reliability. Because a good test is the important thing to measure the knowledge and ability of the students.
3. The lecturer should analyze their item of testing before use it. Considering, not all the student are good at the same test. And the students have different level of ability.
4. When the lecturer took the items test from the bank test, they have to select the items that appropriate with the materials. Considering, not all the test appropriate with the material that the student have learned.

REFERENCES

- ADDIE Model. *Instructional Systems, College of Education, Penn State University*. on October 10th, 2014 from http://www.instructionaldesigncentral.com/html/IDC_instructional_designmodels.htm#addie
- Arikunto, S. *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: PT. Rineka Cipta. 2006
- Asmawati Andi, *using JEOPARDY strategy to improve the first semester students' vocabulary at PIKIH Program of UIN ALAUDDIN MAKASSAR*. Thesis of PBI UINAM. 2011.
- Athok, Fu'adi. *Sistem pengembangan evaluasi*. Ponorogo: STAIN Po Press. 2008. December 22nd, 2015 Retrieved from <http://www.kumpulanmakalah.blogspot.com/2012/06/tingkat-kesukaran-dan-daya-beda.html?m=>
- Bc. Ivana Pavlů, *Testing Vocabulary*, thesis of Department of English Language and Literature at Faculty of Education, MASARYK UNIVERSITY BRNO. 2009.
- Birjandi, P. & Rezaei, S. *Developing a Multiple-choice Discourse Completion Test of Interlanguage Pragmatics for Iranian EFL Learners. Teaching Journal (Special Issue: Proceedings of the First Conference on ELT in the Islamic World)*. Iran: 2010.
- Blanchard Karen and Christine Root. *Ready to Read More*. USA: Longman Pearson. 2006.
- Brown, H. Douglas. *Language assessment*. America. Longman.com 2004

- Bull Victoria (*editor*): *Oxford Learner's Pocket Dictionary*, New York: Oxford University Press. 2008.
- Butler Linda. *Password 2, A Reading and Vocabulary Text*. USA: Longman. 2004.
- Carnine, D., Kameenui, E.J., & Coyne, G. *Contextual Information*. 1984. Retrieved from <http://www.blwd.k12.pa.us/schools/es/readii/Shared%20Documents/Vocabulary/Contextual%20Definition.pdf>
- CNN, News. *december* 12th 2015 from <http://www.news.com.au/world/breaking-news/forest-fires-flare-in-indonesian-province/news-story/fdf7b53b6674b49e8e77db3307fa55ea>
- <http://www.news.com.au/world/breaking-news/forest-fires-flare-in-indonesian-province/news-story/fdf7b53b6674b49e8e77db3307fa55ea>
- <http://www.news.com.au/world/breaking-news/lifestyle-can-help-reduce-dementia-risk/news-story/26ca5a3770a75a098908c64b4e87ce9d>
- Cunningham, J.W., P.M. Cunningham, and S.V. Arthur. *Middle and Secondary School Reading*. New York: Longman. 1981
- Gay, L.R. *Educational Research: Competencies for Analysis and Application. Second Edition*. London: A. Bell & Howel Company. 1981
- Green John A. *Teacher. Made Tests 2nd adition*, New York: Harper and Row Publisher. 1975.
- I.S.P Nation. *Teaching and Learning Vocabulary*, Boston, Massachusetts. Heinle Publisher. 1990
- Keyla Shelton & George Saltsman, *Applying the Addie model to online instruction*, U.S.A. IGI GLOBAL 2011.
- Khair Nam Miftahul, *Developing English Learning Materials for Young Learners Based on Needs Analysis*. Thesis of PBI UIN Alauddin, 2014.
- Gall, M.D. & Borg, W.R. *Educational Research: An Introduction*. New York: Longman. 1983.
- Koda, K. *Insights into Second Language Teaching*. New York: Cambridge University Press. 2005
- Latief, Muhammad Adnan. *Metode Penelitian Pembelajaran Bahasa*. Malang. UM Pers. 2014.
- Lin Leiyi Zhongshannvga. *Designing and Revising a Multiple Choice Vocabulary Test*. Taiwan. (2007)
- Madsen Harold S. *Techniques in Testing*. Oxford University Press. 1983
- Morse-Cluley Elizabeth, Richard Read. *Websters' NewWorld Power Vocabulary*. New York. Prentice Hall Trade. 1988.
- Nasum Fardi. <http://wwwFardy-nasum.blogspot.in/2011/11/metode-pengumpulan-data-quantitative.html?m=1> . Access on 10th may 2015, 22:00 PM.
- Öztürk Meral. *Multiple-Choice Test Items of Foreign Language Vocabulary*. Turki: *Eğitim Fakültesi Dergisi*, 200.
- Popham W. James. *Classroom assessment (what teachers need to know) sixth edition*. Los Angeles. Pearson. 2000.

- Readence, J.E., Bean, T.W., & Baldwin, R.S. *Content Area Reading: An Integrated Approach*, Dubuque, IA: Kendall/Hunt. 1995
- Read John. *Assessing Vocabulary*. United Kindom. Cambridge University Perss.2000
- Riyanto Slamet. *1000 selected Adjective, Adverbs, Preposition used in Speaking English*. Yogyakarta. Pustaka Pelajar. 2008
- . *1000 selected Verbs used in Speaking English*. Yogyakarta. Pustaka Pelajar. 2008
- . *1000 Selected Nouns used in Speaking English*. Yogyakarta. Pustaka Pelajar. 2008
- Sheldon, Leslie E. *ELT Textbooks and materials: Problems in Evaluation and Development*. Oxford. Modern English Publication. 1987.
- Simanjuntak. *Developing Reading Skill for EFL Students*. Jakarta: Departemen Pendidikan dan Kebudayaan. 1988
- Smith Lorraine C. *Exploring Content 2 (Reading for Academic Success)*.USA: Longman. 2005.
- Sukirman. *The Ability of The Fifth Semester Students of English Language And Literature Department at Alauddin Islmic State University Makassar in Analyzing Sentence of Reading Text Using Cooperative Learning*. Thesis of PBI UIN Alauddin, Makassar. 2010
- Tierney,RJ & Readence, JE & Disher, EK. *Reading Strategies and Practices: A Compendium*. 2005
- Tomei Laurance A. *Information Science Reference*, U.S.A. IGI GLOBAL 2008.