**Jurnal Filsafat, Sains, Teknologi, dan Sosial Budaya**
Volume 28, Nomor 3, Oktober 2022

# *Validity Instrument Test Based Computer Test Higher Order Thinking (Cbt-Hot) on Intermediate Reading Course*

**Indrawati Pusparini**
IKIP Budi Utomo
puspaindra71@gmail.com

**Ike Dian Puspitasari**
IKIP Budi Utomo
ikedianps@gmail.com

**Anita Kurnia Rachman**
IKIP Budi Utomo
anita27rachman@gmail.com

***Abstract:*** *Higher Order Thinking is a skill that requires skills such as analyzing, evaluating, and creating. One way to develop high-order thinking skills in students is to get the problem solved in the form of problem solving, creative thinking and critical thinking.The good from this reasearch is to develop and create evaluation instrument of High Order Thinking (HOT) with Computer Based Test .The items developed have been validated theoretically and valid. Developed Instrument tests were tested on Students from intermediate reading class . Data Collection method used test method and questionnaire method. The data analysis technique used descriptive analysis technique which included analysis of High Order Thinking (HOT) test based on Computer Based Test (CBT ) to know the value of validity, reliability, homogenitas, and Annona . The result of question analysis have validity test with t-account > t-table (t-table =0.2573). While reliability value 0,628 for package 1, reliability value 0,651 for package 2 reliability value 0,721 for package 3, reliability value 0,582 for package 4, reliability value 0,839 for package 5 means the test reliable. From the homogeneous one-way Annona test, it is known that the significance value for each package is 0.000. Thus, P value is smaller than α = 0.05. So it can be concluded if Ho is rejected. It means, there is an effect on students' thinking from the type of test given . Based on the result from research that done, so it can be concluded that evaluation instrument of High Order Thinking (HOT) with Computer Based Test (CBT) in intermediate reading courses are valid and reliable.*

***Keywords:*** *validity; realibilit; high order thinking (hot); computer basedtest (cbt); intermediate reading.*

## INTRODUCTION

High-level thinking is a combination of several aspects, that are the ability to think critically, creatively, and problem solving. In Intermediate Reading the cognitive levels achieved are only at the C1, C2 and C3 levels. The development of questions for levels C4, C5 and C6 is very rare. The limited ability of the lecturer to make variations of that questions is because the lecturers cannot maximize the demands contained in the syllabus. Lecturers often make questions using theoretical context and there is no link between the knowledge gained in learning and the actual conditions in daily life. The current condition most students have low ability to understand complex information, understanding theory, analysis and problem solving. To overcome this problem, students need to be trained in doing higher-order thinking skills that can solve the problems.

Based on observations in learning intermediate reading course in Ikip Budi Utomo Malang so far, the reading tests have been used not oriented towards measuring students' high-level thinking skills, but they still focus on low-level learning outcomes. The test questions presented are only based on what is in the contents of the reading while learning takes place, so what happens is students answer quickly all questions because the answers are already in the reading text. There is no challenge to think at a high level, because students are less motivated in attending learning, and they experience boredom. This is indicated by the students who are still having difficulty analyzing existing information. They tend to accept what

information is obtained, passive in asking questions or answering questions from problems raised by teachers, and passive in terms of presenting ideas or problem solving ideas.

Validity also stated by Hughes ,the validity of a test. For it to be valid at all, a test must first be reliable as a measuring instrument' (Hughes, 2003). The test should measure what the teacher wants to measure. For example, if the teacher wants to measure the grammar ability, the teacher should give the text in form of written test, not giving oral form or recording to listen. Generally , there are three kinds of validity. There are content, criterion-related (concurrent and predictive) and construct. From the three kinds of validity, content validity has the important roles in interpreting the test as a tool of evaluation, so that the teacher can measure student's ability effectively. Content validity depends on careful analysis of the language being tested and of the particular course objectives. The test should be so constructed as to contain a representative sample of the course (Heaton (1998). It can be understood that the content validity needs a sharp and systematic analysis because it can represent the content of the test that will be examined. The researcher will explain the content validity in the next chapter. While reliable refers to the consistency of score. For example if the same group of students took the same test twice within two days without reflecting on the first test before they sat it again-they should get the same results on each occasion. If they took another similar test, the results should be consistent.

Reliability test is a test of the stability and consistency of the respondent in answering matters relating to question constructs which are the dimensions of a variable and arranged in a questionnaire form. After the validity test has been carried out, it must be continued by using the data reliability test. A reliable measuring instrument must consist of valid measuring items. So, every reliable must be valid, but every valid is not necessarily reliable. The formulas that are often used for reliability tests are Alpha Cronbach, Spearman Brown, Kristoff, Angoff, and Rullon.

According to Brown (2003), a test, in simple terms, is a method of measuring a person ability, knowledge, or performance in a given domain. A test is first a method. It is an instrument-a set of techniques, procedures, or items that requires performance on the part of the test-taker. To qualify as a test, the method must be explicit and structured: multiple-choice questions with prescribed correct answers; a writing prompt with ;t scoring rubric; an oral interview based on a question script and a checklist of expected responses to be filled in by the administrator. Second, a test must measure. Some tests measure general ability, while others focus on very specific competencies or objectives. A multi-skill proficiency test determines a general ability level; a quiz on recognizing correct use of defmite articles measures specific knowledge. Next, a test measures an individual's ability, knowledge, or performance. Testes need to understand who the test-takers are. What is their previous experience and background? Is the test appropriately matched to their abilities? How should test . takers interpret their scores? A test measures performance, but the results imply the test-taker's ability, or, to use a concept common in the field of linguistics competence. Finally, a test measures a given domain. In the case of a proficiency test, even though the actual performance on the test involves only a sampling of skills, that domain is overall proficiency in a language-general competence in all skills of a language. Other tests may have more specific criteria.

. Higher Order of Thinking Skills (HOTS) is the ability to think critically, logically, reflective, metacognitive, and creative thinking which is the ability to think at a higher level. According to Heong, et al (2011) higher order thinking is using the thinking widely to find new challenges. Higher order thinking demands someone to

apply new information or knowledge that he has got and manipulates the information to reach the possibility of answer in the new situation. In HOTS students use thinking widely to find new challenges. Higher-order thinking requires a person to apply the information or new knowledge he gets and manipulate information to reach possible answers in new situations. Several studies have written the definition of HOTS that HOTS is an internal process that occurs within a person which is characterized by several characteristics as follows: (1) Involving more than one correct answer; (2) Talk about level of understanding; (3) Characterized by complex tasks; and (4) Free content and at the same time content-related, Astutik, (tth, p. 343); Zaini (2015).

In HOTS, besides containing high-level thinking ability, it also contains creative thinking. The importance of high-level thinking ability was expressed by Fensham (2012) in order to compete in the world of work and personal life, students must have the ability to think critically, creatively and the ability to solve problems. HOTS based on Bloom's Taxonomy is included in the five highest levels, namely analysis (C4), evaluation (C5), synthesis (C6), imagination (C7), and creation (C8). Bloom's revised taxonomy is divided into two thought processes, namely higher order thinking skills or often referred to as Higher Order Thinking Skills (HOTS), and lower order thinking skills Lower Order Thinking Skills (LOTS). Low-level thinking skills involve the ability to remember (C1), understand (C2) and apply (C3) while high-level thinking skills involve analysis and synthesis (C4), evaluating (C5), and creating or creativity (C6) (Krathworl and Anderson , 2001).

Making High Order Thinking (HOT) items can be presented by utilizing information and communication technology. The items that have been made previously have been validated theoretically and declared valid. The use of information and communication technology is increasing in line with the use of the internet as a supporting aspect. This has an impact on many variations in the learning process. One variation that can be given is by making items. The items made are based on the internet with computer media to support. Utilization of information and communication technology is utilized in education, including computer-based assessment using computer tests (De-Siqueira, 2009). According to Hosseini, et.al (2014) the emergence of new technology, computer-based exams became widespread and their tests were carried out on a large scale.

The use of information and telecommunications media in the process of assembling items can provide many benefits. According to Martin (2008) the advantage of computer-based exams is that in terms of test collection there is no need to collect a lot of papers and when the process of correcting the teacher does not need to take a long time. For students this can also be used to behave honestly by not cheating. The Items on the Computer Based Test (CBT) can be used with two systems, namely online systems and offline systems.

Computer Based Test (CBT) with this online system has the advantage that questions and answer options contained will automatically be randomized by the system. For lecturers with an online system can help in seeing charts in the form of students who answer a lot correctly. But besides having advantages, this online system also has disadvantages; an internet server that is used can sometime having problem with the process of working on the item. In the Computer Based Test (CBT) with this offline system it can help for schools in areas that are difficult to access the internet. The development of Computer Based Test (CBT) can reduce paper usage during exams and can minimize the time in the correction process.

Based on the explanation above, the purpose of this research is to develop a Computer Based Test (CBT) based High

Order Thinking (HOT) test instrument on Intermediate Reading courses.

## RESEARCH METHOD

This research is development research, because in this study developed items High Order Thinking (HOT) based on Computer Based Test (CBT). The target in this research is a Computer Based Test (CBT) based High Order Thinking (HOT) item which was theoretically validated and declared valid and tested on Ikip Budi Utomo third semester students who are currently taking intermediate reading course.The target of this research is the items reading test High Order Thinking (HOT) based on Computer Based Test (CBT) which had previously been theoretically validated and declared valid and was tested on 2 heterogeneous Class students of Intermediate reading.

This method uses the test method, as well as the questionnaire method. The data analysis technique uses descriptive analysis techniques which include the analysis of the results of the Computer-based High Order Thinking (HOT) Test Based Test (CBT) to determine the value of validity and reliability test item. The researcher analyzed the data using validity and reliability formula by Suharsimi Arikunto and Kuder-Richardson's. There were 20 multiple choice questions of Intermediate English test. This research used expert judgement as validator to analyzed and commented the validity and reliability of Intermediate English test

## FINDING AND DISCUSSION

### FINDING

The findings of the research were showed to describe the result of the data that analyzed statistically. It comprised of the students' score in validity and reability in making test with criteria of a good test. They were 40 students in Intermediate Reading . There were two class took as an example to analyzed the Validity and Reliability in Intermediate Reading test. To get the data, the researcher took the questions sheet, answer sheet to be analyzed. The total numbers of test items were 20 tests which consist of 20 multiple choices. The test was held online . Furthermore, this research was support by expert judgement. Based on the data, we can see the table as follow:

**Table 1 Validity Intermediate Reading Test**

| | | P1 | P2 | P3 | P4 | P5 |
|---|---|---|---|---|---|---|
| **Q1** | P Corr | 0,3887 | 0,3721 | 0,4283 | 0,5062 | 0,5013 |
| | Sig. (2-tailed) | 0,0132 | 0,0181 | 0,0058 | 0,0009 | 0,0010 |
| | N | 40 | 40 | 40 | 40 | 40 |
| **Q2** | P Corr | 0,8416 | 0,2784 | 0,2870 | 0,2578 | 0,5016 |
| | Sig. (2-tailed) | 0,6056 | 0,0820 | 0,1785 | 0,3465 | 0,0010 |
| | N | 40 | 40 | 40 | 40 | 40 |
| **Q3** | P Corr | 0,5062 | 0,5013 | 0,3887 | 0,4235 | 0,283 |
| | Sig. (2-tailed) | 0,0009 | 0,0010 | 0,0132 | 0,0065 | 0,257 |
| | N | 40 | 40 | 40 | 40 | 4 |
| **Q4** | P Corr | 0,2449 | -0,2698 | -0,2742 | -0,2632 | 0,267 |
| | Sig. (2-tailed) | 0,7834 | 0,6686 | 0,8666 | 0,4125 | 0,366 |
| | N | 40 | 40 | 40 | 40 | 4 |
| **Q5** | P Corr | 0,2588 | 0,2917 | 0,2594 | 0,2599 | 0,286 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sig. (2-tailed) | 0,3465 | 0,2361 | 0,5014 | 0,1727 | | Sig. (2-tailed) | 0,2796 | 0,1236 | 0,0776 | 0,1797 | 0,233 |
| | N | 40 | 40 | 40 | 40 | 40 | N | 40 | 40 | 40 | 40 | 4 |
| **Q6** | P Corr | 0,2582 | 0,2655 | 0,2596 | 0,2592 | **Q14** | P Corr | 0,4582 | 0,3958 | 0,3887 | 0,3721 | 0,263 |
| | Sig. (2-tailed) | 0,4392 | 0,3074 | 0,1808 | 0,3581 | 0,4867 | Sig. (2-tailed) | 0,0029 | 0,0115 | 0,0132 | 0,0181 | 0,130 |
| | N | 40 | 40 | 40 | 40 | | N | 40 | 40 | 40 | 40 | 4 |
| **Q7** | P Corr | -0,2722 | -0,2633 | -0,2645 | -0,2725 | **Q15** 0,2716 | P Corr | 0,2928 | 0,5334 | 0,2614 | 0,4283 | 0,298 |
| | Sig. (2-tailed) | 0,2880 | 0,4123 | 0,3103 | 0,2102 | | Sig. (2-tailed) | 0,2334 | 0,0004 | 0,1903 | 0,0058 | 0,219 |
| | N | 40 | 40 | 40 | 40 | 40 | N | 40 | 40 | 40 | 40 | 4 |
| **Q8** | P Corr | 0,3012 | 0,2475 | 0,2021 | 0,2165 | **Q16** | P Corr | 0,2784 | 0,2614 | 0,2668 | 0,2849 | 0,361 |
| | Sig. (2-tailed) | 0,0590 | 0,1236 | 0,2110 | 0,1797 | 0,0108 | Sig. (2-tailed) | 0,0820 | 0,1903 | 0,1247 | 0,2534 | 0,022 |
| | N | 40 | 40 | 40 | 40 | | N | 40 | 40 | 40 | 40 | 4 |
| **Q9** | P Corr | 0,2708 | 0,2721 | 0,2617 | 0,2585 | **Q17** 0,2638 | P Corr | 0,2599 | 0,3921 | 0,2822 | 0,3247 | 0,461 |
| | Sig. (2-tailed) | 0,4578 | 0,2883 | 0,7053 | 0,3528 | | Sig. (2-tailed) | 0,1199 | 0,0123 | 0,0776 | 0,0409 | 0,002 |
| | N | 40 | 40 | 40 | 40 | 40 | N | 40 | 40 | 40 | 40 | 4 |
| **Q10** | P Corr | 0,4578 | 0,4283 | 0,3865 | 0,5334 | **Q18** | P Corr | 0,4107 | 0,3744 | 0,3265 | 0,3561 | 0,344 |
| | Sig. (2-tailed) | 0,0030 | 0,0058 | 0,0138 | 0,0004 | 0,2577 | Sig. (2-tailed) | 0,0085 | 0,0173 | 0,0398 | 0,0241 | 0,029 |
| | N | 40 | 40 | 40 | 40 | | N | 40 | 40 | 40 | 40 | 4 |
| **Q11** | P Corr | 0,2597 | 0,2614 | 0,2782 | 0,2849 | **Q19** 0,3610 | P Corr | 0,8529 | -0,2998 | -0,2960 | -0,2627 | 0,287 |
| | Sig. (2-tailed) | 0,3567 | 0,1903 | 0,2711 | 0,2534 | | Sig. (2-tailed) | 0,6008 | 0,3556 | 0,5558 | 0,1875 | 0,488 |
| | N | 40 | 40 | 40 | 40 | 40 | N | 40 | 40 | 40 | 40 | 4 |
| **Q12** | P Corr | 0,2986 | 0,2655 | 0,2993 | 0,2876 | **Q20** | P Corr | 0,6443 | 0,6426 | 0,4512 | 0,6509 | 0,501 |
| | Sig. (2-tailed) | 0,2191 | 0,3074 | 0,3892 | 0,4655 | 0,1814 | Sig. (2-tailed) | 0,0000 | 0,0000 | 0,0035 | 0,0000 | 0,001 |
| | N | 40 | 40 | 40 | 40 | | N | 40 | 40 | 40 | 40 | 4 |
| **Q13** | P Corr | 0,1752 | 0,2475 | 0,2822 | 0,2165 | 0,2928 | | | | | | |

**a. Cannot be computed because at least one of the variables is constant.**

****. Correlation is significant at the 0.01 level (2-tailed).**

***. Correlation is significant at the 0.05 level (2-tailed).**

All items valid because $r_{hitung} > r_{tabel}$ ($r_{tabel}$ = 0.2573)

**Table 2. Reliability Statistics**

| | Cronbach's Alphaa | N of Items |
|---|---|---|
| **Package 1** | 0.628 | 20 |
| **Package 2** | 0.651 | 20 |
| **Package 3** | 0.721 | 20 |
| **Package 4** | 0.582 | 20 |
| **Package 5** | 0.839 | 20 |

All Packages are reliable
Package 1, package 2, and package 4 has a moderate reliability with α > 0.5
Package 3 has sufficient reliability with α > 0.7
Package 5 has perfect reliability with α > 0.8

**Table 3. Test of Homogeneity of Variances**

Score

| Levene Statistic | df1 | df2 | Sig. |
|---|---|---|---|
| .687 | 3 | 156 | .561 |

It is known that the type of data is homogeneous with a significance value greater than α =0.05

**Table4. One-Way Annova**

| Pack | (J) PACK | MEAN DIFF (I-J) | STD. ERR | SIG. | 95% CONF INTERVAL Lower Bound |
|---|---|---|---|---|---|
| **P 1** | Pack2 | -1 | 1,471525 | 0.000 | -5,17814 |
| | Pack 3 | 0,625 | 1,471525 | 0.000 | -3,55314 |
| | Pack 4 | 0 | 1,471525 | 0.000 | -4,17814 |
| | Pack 5 | -0,25 | 1,471525 | 0.000 | -4,42814 |
| **P 2** | Pack1 | 1 | 1,471525 | 0.000 | -3,17814 |
| | Pack 3 | 1,625 | 1,471525 | 0.000 | -2,55314 |
| | Pack 4 | 1 | 1,471525 | 0.000 | -3,17814 |
| | Pack 5 | 0,75 | 1,471525 | 0.000 | -3,42814 |
| **P 3** | Pack 1 | -0,625 | 1,471525 | 0.000 | -4,80314 |
| | Pack 2 | -1,625 | 1,471525 | 0.000 | -5,80314 |
| | Pack 4 | -0,625 | 1,471525 | 0.000 | -4,80314 |
| | Pack 5 | -0,875 | 1,471525 | 0.000 | -5,05314 |
| **P 4** | Pack 1 | 0 | 1,471525 | 0.000 | -4,17814 |
| | Pack 2 | -1 | 1,471525 | 0.000 | -5,17814 |
| | Pack 3 | 0,625 | 1,471525 | 0.000 | -3,55314 |
| | Pack 5 | -0,25 | 1,471525 | 0.000 | -4,42814 |
| **P 5** | Pack 1 | 0,25 | 1,471525 | 0.000 | -3,92814 |
| | Pack 2 | -0,75 | 1,471525 | 0.000 | -4,92814 |
| | Pack 3 | 0,875 | 1,471525 | 0.000 | -3,30314 |
| | Pack 4 | 0,25 | 1,471525 | 0.000 | -3,92814 |

## DISCUSSION

Validity test is testing the level of constraints and validity of the measuring instrument used. The instrument is said to be valid, meaning that the measuring instrument used to obtain the data is valid or can be used to measure what should be measured. Thus, a valid instrument is an instrument that is really appropriate to measure what you want to measure.

The validity test question made by the English lecture in Ikip Budi Utomo Malang as following presentation data the question used in intermediate reading test. The test question in the form of multiple choice question can be done used Suharsimi and Arikunto formula. In analyze the items qualitatively, used the review form would be helpful and facilitated the implementation procedure. Based on compared in percentage with Arikunto's criteria the validity of Intermediate Reading test had a good validity Test

Reliability test is a test of the stability and consistency of the respondent in answering matters relating to question constructs which are the dimensions of a variable and arranged in a questionnaire form. After the validity test has been carried out, it must be continued by using the data reliability test. A reliable measuring instrument must consist of valid measuring items. So, every reliable must be valid, but every valid is not necessarily reliable. The formulas that are often used for reliability tests are Alpha Cronbach, Spearman Brown, Kristoff, Angoff, and Rullon.

The researcher analyzed the reliability of question. Reliability referenced to the stability of multiple choice test values. In this research, the researcher used the Kuder-Richardson's formula (KR-20) in the Arikunto's book to determine the reliability coefficient of the 20 multiple choice test items. Based on the presentation data above, the reliability of English summative test made by the researcher were 0,628 for package 1, reliability value 0,651 for package 2 reliability value 0,721 for package 3, reliability value 0,582 for package 4, reliability value 0,839 for package 5. It means that the reliability of English summative test was in classification into the high level (0.60-0.799).

The homogeneity test is a test of whether the variances of two or more distributions are equal. The variance homogeneity test (variance) is needed before we compare two or more groups, so that the differences are not caused by differences in baseline data (the inhomogeneity of the groups being compared). From the result, it is known that the type of data is homogeneous with a significance value greater than $\alpha = 0.05$

Next, to test whether the data is homogeneous or not, it can be done by clicking Analyze - Compare Means - One Way Annova. From the homogeneous one-way Annona test, it is known that the Bonferroni significance value for each package is 0.000. Thus, P value is smaller than $\alpha = 0.05$. So it can be concluded if Ho is rejected. Thus, there is an effect of the type of test given on students' thinking.

## CONCLUSION

Based on the results of the analysis and discussion, it can be concluded that the High Order Thinking (HOT) based on Computer Based Test (CBT) has valid and reliable. The result of question analysis have validity test with t-account > t-table (t-table =0.2573). While reliability value 0,628 for package 1, reliability value 0,651 for package 2 reliability value 0,721 for package 3, reliability value 0,582 for package 4, reliability value 0,839 for package 5 means the test reliable. From the homogeneous one-way Annona test, it is known that the significance value for each package is 0.000. Thus, P value is smaller

than $\alpha = 0.05$. So it can be concluded if Ho is rejected. It means, there is an effect on students' thinking from the type of test given . Based on the result from research that done, so it can be concluded that evaluation instrument of High Order Thinking (HOT) with Computer Based Test (CBT) in intermediate reading courses are valid and reliable.

This research provided expert judgment to analyzed the result validity and reliability of this research, and based on the expert judgment in the content validity it was not suitable with the researcher result. While in test reliability, expert judgment found that this research was reliable with researcher result. The researcher used Mix method to design this research. The researcher used qualitative descriptive to designed validity of Intermediate English test, and used quantitative to designed reliability of Intermediate English test.

## SUGGESTION

Computer Based Test (CBT) based High Order Thinking (HOT) evaluation instrument developed by the reseacher only in the form of multiple choices, there needs to be a new innovation so that students can include reasons in choosing the answer chosen. Further research is needed regarding the addition of features in the form of multimedia so that students can solve problems through videos and regarding the Computer Based Test (CBT) server so that it can be used in an offline state that uses many computers.

The second recommendation goes to further studies. Relevant to the finding, suggestions for further research are made as follows: 1) the present study uses only students of Ikip Budi Utomo who learn Intermediate Reading as the target population. Further studies are suggested to be conducted to students at university who learn other subject course, 2)the present study is limited only to use test intruments in Intermediate Reading . It is suggested that future study be conducted to include the evaluation level other comprehension, such as: speaking, writing and listening

## REFERENCES

Anderson, L.W., and Krathwohl, D.R. (2001). A Taxonomy of Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York Longman. P.

Anderson, L.W., dan Krathwohl, D.R. (2001). A Taxonomy for Learning, Teaching, and Assesing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Addison Wesley Longman, In.

Anderson, L. W., & Krathwohl, D. R. (2010). *Kerangka Landasan Untuk Pembelajaran, Pengajaran, dan Asesmen*. Yogyakarta*: Pustaka Pelajar*.

Astutik (tth) *Integrasi Penguatan Pendidikan Karakter (PPK) Dan Higer Order Thinking Skills (Hots) Dalam Pembelajaran Tematik SD,* disajikan dalam Seminar Nasional Pendidikan – Fakultas Ilmu Pendidikan Universitas Negeri Malang Sinergitas Keluarga, Sekolah, dan Masyarakat dalam Penguatan Pendidikan Karakter. Makalah.

Arikunto, S. (1999). *Dasar-dasar Evaluasi Pendidikan (Rev. ed.*). Jakarta: Bumi Aksara.

Barnett, J. E and Francis, A.L. (2012). *Using Higher Order Thinking Questions to Foster Critical Thinking: A Classroom Study*.

Educational Psychology An International Journal of Experimental Educational Psychology ISSN1469-5820

Brown, Douglas. 2003. *Language Assessment: Principles and Classroom Practice*. USA: Longman.

Gall, J.P., Gall, M.D & Walter, R.B. (2005). *Applying Educational Research: A Practical Guide*. London: Pearson

Hughes, A. (2003). Testing for Language Teachers, (2nd ed.). UK: Cambridge University Press http//www.statisticshowto.com/kuder-richardson/ access on Wednesday,18 July 2020

Heong, Y. M.,Othman, dkk. (2011). The Level of Marzano Higher Order Thinking Skills Among Technical Education Students . *International Journal of Social and humanity*, Vol. 1,No. 2, July 2011, 121- 125

Kemendikbud, 2017. Modul Penyusunan Higher Order Thinking Skill (HOTS). Jakarta: Direktort Jenderal Pendidikan Dasar dan Menengah Departeman Pendidikan dan Kebudayaan.

Lailly, N.R & Wisudawati, A.W. 2015. Analisis Soal Tipe Higher Order Thinking Skill (HOTS) dalam Soal UN Kimia SMA Rayon B Tahun 2012/2013. *Jurnal Kaunia* Vol. XI No. 1, April 2015/1436:27-39 ISSN 2301-8550

M.R. Patel and Pravin M. Jain (2008). *English Language Teaching (Methods, Tools and Techniques)* , Vaishali Nagar, Jaipur : Sunrise Publihers and Distributors

Olson, Joane P and Diller , Martha, H. *Learning to Teach Reading at Elementary School* New York : Macmillan Publishing Co, Inc

Pratiwi, U., & Fasha, E. F. (2015). Pengembangan Instrumen Penilaian HOTS Berbasis Kurikulum 2013 Terhadap Sikap Disiplin. *Jurnal Penelitian dan Pembelajaran IPA*, *1*(1), 123-142.

Sudjana. 2005. *Metoda Statistika*. Bandung: Tarsito.

Syamsudin. 2012. *Pengukuran Daya Pembeda, Taraf Kesukaran, dan Pola Jawaban Tes*.Jurnal Ilmu Tarbiyah "At-Tajdid". Volume 1 No.2.

Zein, Anizam., Muhyiatul Fadillah, dan Rahma Novianti. Hubungan Antara Validitas Butir, Reliabilitas, Tingkat Kesukaran dan Daya Pembeda Soal Ujian Semester Genap Bidang Studi Biologi Kelas XI SMA/MA Negeri di Kota Padang Tahun Ajaran 2010/2011. Prosiding Semirata FMIPA UNIVERSITAS LAMPUNG

Zaini, M. (2015). Hasil belajar dan keterampilan berpikir tingkat tinggi siswa SMA pada pembelajaran biologi menggunakan model pembelajaran berdasarkan masalah. *Jurnal Pendidikan Biologi*, *20*(207)