**IJHE: Indonesian Journal of History Education, 6 (2), 2018: p.172-179**
e-ISSN: 2549-0354; p-ISSN: 2252-6641
*Terakreditasi oleh Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia*
*Nomor: 21/E/KPT/2018, tanggal 19 Juli 2018*
Tersedia online https://journal.unnes.ac.id/sju/index.php/ijhe/index

# Analysis Of Instrument Test Of Historical Thinking Skills In Senior High School History Learning With Quest Programs

**Ofianto**

Departemen Pendidikan Sejarah, Universitas Negeri Padang

## ABSTRAK

Penelitian ini merupakan penelitian deskriptif kuantitatif. Tujuan penelitian ini adalah mendeskripsikan kualitas soal ujian akhir semester ganjil mata pelajaran sejarah se Sumatera Barat yang terdiri dari: validitas, tingkat kesukaran butir soal, daya beda, fungsi pengecoh dan reliabilitas. Objek penelitian adalah 1) perangkat soal; 2) lembar jawaban sebanyak 443 lembar yang diperoleh dari MAN 3 Padang, MAN Pariaman dan MAN 3 Padang Panjang; 3) kunci jawaban soal; 4) kisi-kisi penulisan soal. Analisis menggunakan program *quest*. Hasil penelitian menunjukkan bahwa*:* validitas isi, perangkat soal yang termasuk kategori valid terdapat 28 butir soal (56%) dan tidak valid terdapat 22 butir soal (44%). Validitas item, perangkat soal yang valid sebanyak 43 soal (86%) dan tidak valid sebanyak 7 soal (14%). Reliabilitas soal menggunakan formula Kuder-Ricardson 20 menunjukkan ko-efisien reliabilitas 0,53 (cukup reliabel). Tingkat kesukaran soal menunjukkan bahwa soal yang sukar sebanyak 40 soal (80%), soal yang sedang (tidak sukar) sebanyak 9 soal (18%), soal mudah sebanyak 1 soal (2%). Daya beda soal yang termasuk kategori lemah terdapat 26 soal (52%), cukup lemah terdapat 18 soal (36%), baik terdapat 4 soal (8%), baik sekali terdapat 1 soal (2%) dan soal yang buruk terdapat 1 soal (2%). Keberfungsian pengecoh soal yang berfungsi dengan baik terdapat 47 soal (94%) dan yang tidak berfungsi dengan baik terdapat 3 soal (6%).

**Kata kunci:** *historical thinking*; validitas; reliabilitas; tingkat kesukaran; daya beda; fungsi pengecoh

## ABSTRACT

This research is descriptive quantitative research and aimed to described quality of questions history subject in West Sumatra District including validity, item difficulty, item discrimination, distractor efficiency and reliability. Objects in this study are questions test and 443 sheets of answer sheet including MAN 3 Padang, MAN Pariaman and MAN 3 Padang Panjang, key answer test, a test blue print and then it's analysis with *quest* program. The result of this research shows that the content validity have category of valid was 28 questions (56%) and un-valid was 22 questions (44%). Item difficulty questions are easy category was 1 question (2%), immediate category was 9 questions (18%) and difficult category was 40 questions (80%). Item discrimination is weak category was 26 questions (52%), enough category was 18 questions (36%), good category was 4 questions (8%), very good questions was 1 question (2%) and bad category was 1 question (2%). Distractor efficiency was 47 questions (94%) and not function was 3 questions (6%). The average Kuder-Richarson formula 20 reability coefficient was 0,53 means it has enough reliability.

**Key words:** historical thinking; validity; item difficulty; item discrimination; distractor of efficiency; reliability

Diterima: November 2018, Disetujui: November 2018, Diterbitkan: Desember 2018


Korespondensi:
Email: ofianto.anto@yahoo.com

Alamat redaksi:
Gedung C5, Lt. 1 FIS-Unnes, Kampus Sekarang,
Gunungpati, Semarang, Jawa Tengah 5029
Email: sejarah@mail.unnes.ac.id

## INTRODUCTTION

One of the important aspects in education is evaluation. An evaluation is a continuous process of collecting information on the student's learning progress and result in order to make a decision based on certain criteria (Arifin, 2012:4). Evaluation is necessary since it may drive teachers to teach better and for the students to be motivated to learn more seriously in the future. Therefore, in conducting an evaluation, an instrument is needed such as a test or a non-test.

A test is an instrument utilized to monitor the student's ability and mastery towards questions given by the teacher. Among the types test tested to the students of Madrasah Aliyah Negeri (Islamic Public School) in West Sumatera is multiple choice questions on the History subject in the academic year of 2017/2018. A multiple choice is defined as a test which requires the students to choose a correct answer among several options provided (Arifin, 2012:135).

Based on the observation on the learning progress of the academic year 2017/2018 in some Madrasah Aliyah Negeri of West Sumatera, the History subject at this school lies in the category of poor.

**Table 1.** The Student's Average Score of National Examination of History Subject at MAN in West Sumatera

| No. | Schools | Average Exam Score |
|---|---|---|
| 1. | MAN 3 Padang | 48,7 |
| 2. | MAN Pariaman | 57,5 |
| 3. | MAN 3 Padang Panjang | 44,2 |

The problem appears since the question in the final semester has not been analyzed. Meanwhile, the analysis towards the question item may determine the questions which meet the criteria of good questions, thus the question can measure the student's comprehension well by the end of the semester.

Therefore, in order to improve the question quality, it is better to analyze the question item of a test thoroughly before administering the test to the students. The analysis of the question item is the scrutiny on the question items which aims to create a fair quality of test (Sudjana, 2009: 135). The purpose of analyzing the question item is to collect a meaningful information used to provide feedback to the betterment of the student's performance (Sudjono, 2015: 393-370).

To ease the process of question analysis, an application is needed. The program is called *quest*. This program is able to analyze the politomos as well as dichotomus data. Besides, the output of this program can analyze the question item from various perspectives in a classical theory such as reliability, the complexity, the distinction, and the distractor items. The strengths of this program are easy to use, flexible and informative output, has a flexible procedure to recover the data lose (Pisca, 2013: 3). Using this application, the teacher can analyze the question item effectively and quickly.

According to Hidayatulloh's (2009: 156) research result on the analysis of question items in the Arabic test at SMP Muhammadiyah 3 Depok Sleman Daerah Istimewa Yogyakarta, it can be concluded that the question items were in the category of fair. The analysis was a quantitative analysis with the classical theory using quest application.

Based on the result of that research, it can be learned that it is very necessary to conduct an analysis towards the question item before testing it to the students, specifically on the History subject at Islamic Public School. The analysis utilizes the *quest* program, and the factors being analyzed are validity, reliability, complexity, the distinguished factors, and the deceptive items in order to reveal the quality of question items tested to the students.

The analysis of question items is a way to improve the quality of question items, thus the test may successfully and effectively measure the student's competence. The better quality of a question items are, the better

the quality of a test will be, and the more reliable the evaluation yielded (Uno, 2012: 111).

## METHOD

This research employs a quantitative descriptive approach. The emphasize is on the characteristics of the question items tested towards the students of Islamic Public School grade X in the academic year of 2017/2018. Some aspects being analyzed are validity, reliability, complexity, the distinguished factors, and the deceptive items. The researcher interpreted and analyzed the data observation using a program called *quest*. The objects of this research are the predictions of question, question items, the answer keys, and the answer sheets of the test at the first term of 2017/2018 which have been filled in by the students of grade X MAN 3 Padang, MAN Pariaman and MAN 3 Padang Panjang as many as 443 pieces.

The data analysis used to test the validity of the data employed the content and item validity. What is defined as content validity is the relevance between the predicted question item and the real test. Meanwhile, the item validity is conducted using the relevance formula of moment product and the rough number:

$$r_{xy} = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\{N\sum x^2 - (\sum x)^2\}\{N\sum y^2 - (\sum y)^2\}}}$$

Afterwards, the significance of correlation coofisien *r* still needs to tested by comparing the $r_{table.}$ If $r_{number} > r_{table}$, it shows that the question item is invalid.

The analysis of realibility on dycotomous data is done using the Kuder-Richardson 20's formula. In the *quest* program, it is marked as internal consistency. The result can be presented by the realibility coofisien index as follows:

**Table 2.** Realibility Coefisien Index

| Realibility Coefisien | Realibility Level |
|---|---|
| 0.800 - 1.000 | Very high |
| 0.600 - 0.799 | High |
| 0.400 - 0.599 | Fair |
| 0.200 - 0. 399 | Low |
| 0.00 - 0.199 | Very low |

Source: Sunarti dan Rahmawati (2014:99)

The analysis towards the question complexity in the *quest* program is yielded from the complexity index as below:

**Table 3.** Complexity Level Index

| Index | Interpretation |
|---|---|
| 0 - 0,30 | Hard |
| 0,30 – 0,70 | Medium |
| 0,70 – 1,00 | Easy |

Source: Sunarti dan Rahmawati (2014:138)

The analysis on discrimination power in *quest* program can be seen in *biserial point*. Below is the analysis of discrimination power:

**Table 4.** Discrimination Power Index

| Index | Interpretation |
|---|---|
| 0,00 – 0,20 | Poor |
| 0,20 – 0,40 | Fair |
| 0,40 – 0,70 | Good |
| 0,70 – 1,00 | Very good |
| - | Poor |

Source: Ambiyar (2014: 154)

## RESULT AND DISCUSSION

### Result

### a. Validity

Validity needs to be measured to know whether or not the instrument used in the data collection works well. One of the validity tests employed in this research is content validity. A content validity is measured through the instrument itself. The content validity can be revealed through matching the question item and the materials being tested as well as the indicators of questions written on the predicted questions arranged based on the lesson plan.

Based on the content validity analysis, it can be concluded that there are 28 valid questions, which are 1, 2, 3, 5, 8, 15, 16, 17, 23, 25, 26, 27, 29, 32, 33, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50. Moreover, there are

22 invalid questions, including question number 4, 6, 7, 9, 10, 11, 12, 13, 14, 18, 19, 20, 21, 22, 24, 28, 30, 31, 34, 35, 36, 39.

**Table 5.** The Percentage of Valid Question of Odd Semester in Madrasah Aliyah Negeri Grade X of Academic Year 2017/2018

| Category | Number | Percentage |
|---|---|---|
| Valid | 28 | 56% |
| Invalid | 22 | 44% |

Source: research result, 2018

On the other hand, for the item validity, the formula of *product moment* and $r_{table}$ is used to reveal whether the questions are valid or not. With the number if the students (N) were 443 persons, therefore the significance of $r_{table}$ 5% is 0,098. If $r_{number} > r_{table,}$ the question item can be categorized as valid. However, if $r_{number} <$ dari $r_{table,}$ it means that the question items are invalid.

Based on the item analysis, the result shows that there are 43 out of 50 questions were considered valid. Those question including the questions number 43 soal, yaitu 1, 2, 3, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 38, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 50 with the percentage 86%. The number of invalid questions were 7, including 4, 5, 6, 24, 31, 37, 44 with the percentage 14%. Below is the table of item validity analysis:

**Table 6.** The Persentage of Item Validity of Odd Semester Test Grade X of Madrasah Aliyah Negeri Academic Year of 2017/2018

| Category | Number | Percentage |
|---|---|---|
| Valid | 43 | 86% |
| Invalid | 7 | 14% |

Source: research result, 2018

**b. Reliability**

Reliability is the degree of consistency of an instrument. The result of reliability of Odd Semester test in Madrasah Aliyah Negeri Grade X of Academic Year 2017/2018 is 0,53, indicating that the test is reliable.

**c. Level of Complexity**

The level of complexity shows to what degree is a question hard, medium, or easy to students. This measurement is known by the ability of students in answering the questions given by the teacher. Regarding to the result of complexity level of questions in Odd Semester test in Madrasah Aliyah Negeri Grade X of Academic Year 2017/2018, there was found 1 question in easy category. It is the question number 50. The question with medium level of difficulty found in 9 questions, including questions number 5, 6, 9, 24, 29, 31, 32, 47, 49. Finally, the rest of the questions were categorized into hard questions (40 items), namely the question number 1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 48.

**Table 7.** The Percentage of Complexity Level using *Quest* Program

| Category | Number of Question | Percentage |
|---|---|---|
| Easy | 1 | 2% |
| Medium | 9 | 18% |
| Hard | 40 | 80% |

Source: research result, 2018

**d. Discrimination Power**

Discrimination power functions to distinguish students of a higher level of cognition with those who have a lower level of cognition.

Based on the discrimination power analysis towards the students of Islamic Public School grade X in the academic year of 2017/2018 using the *Quest* Program, it can be revealed that there were 26 questions in the weak category, such as the question number 1, 4, 5, 6, 9, 10, 11, 12, 13, 14, 16, 17, 23, 24, 29, 31, 32, 33, 35, 36, 37, 38, 40, 43, 47, and 50. There were 18 questions categorized into fair questions including number 2, 3, 7, 8, 15, 18, 19, 20, 21, 25, 26, 27, 28, 30, 34, 39, 41, 42. Furthermore, 4 questions belonged to good category including question number 44, 45, 48, 49. Meanwhile, there was only one categorized into very good question, namely question number 46. Similarly, there was only one question in poor category, namely question number 22.

**Table 8.** Percentage of Discrimination Power Index (*Pt-biserial*) using *Quest* Program

| Category | Number of Questions | Persentage |
|---|---|---|
| Poor | 26 | 52% |
| Fair | 18 | 36% |
| Good | 4 | 8% |
| Very good | 1 | 2% |
| Bad | 1 | 2% |

Source: research result, 2018

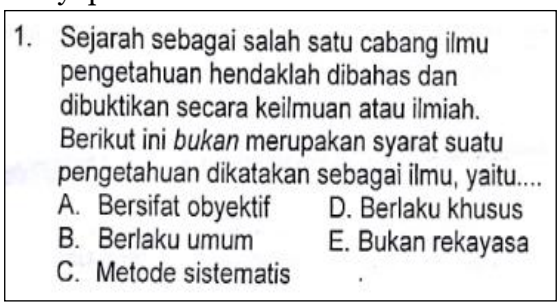
**e. The Function of Distractor**

Every question item needs to undergo a distractor analysis to reveal whether the multiplechoice questions have an effective distractor. The goal is to know the answer of the subject on the available answer choice. A distractor is categorized into good when there were 5% of test takers choose it. Based on the analysis of distractor using the Quest program, it was shown that there were 47 questions (94%) which can be categorized into good, namely question number 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48. Nevertheless, there were 3 questions (6%) fell into the category of poor, such as 47, 49, 50.

**Discussion**

**a. Validity**

Based on the result of content and item analysis towards the questions of History subject of Odd Semester test in the Islamic Public School, it was found 1 invalid question, namely question number 1.


1. Sejarah sebagai salah satu cabang ilmu pengetahuan hendaklah dibahas dan dibuktikan secara keilmuan atau ilmiah. Berikut ini *bukan* merupakan syarat suatu pengetahuan dikatakan sebagai ilmu, yaitu....
   A. Bersifat obyektif
   B. Berlaku umum
   C. Metode sistematis
   D. Berlaku khusus
   E. Bukan rekayasa

Source: Odd Semester Test 2017/2018 Document on History Subject in Madrasah Aliyah Negeri

The question item above is valid since it is based on the predicted questions of the odd

semester test. The question is relevant with the basic competence and the key answer as well. In terms of its content, the question has been clearly and firmly constructed, and the key answer is also homogenous and logic. One of the invalid questions based on the content and item analysis is the question

5. Peristiwa-peristiwa yang terjadi pada masa lalu.. akan diceritakan kembali pada generasi berikutnya. Kondisi berantai ini merupakan hakikat sejarah sebagai...
   A. Kisah
   B. Peristiwa
   C. Ilmu
   D. Seni
   E. Warisan

Source: The Document of Odd Semester Test of 2017/2018 of History Subject in Madrasah Aliyah Negeri

This invalid question item is caused by the incongruence between the question in the real test and the predicted questions. Besides, the invalidity may be caused by the student's guessing towards the question. Further actions towards the question items are as follow: (1) The invalid question items should be excluded, but if they are to be utilized in the test, they should be revised first hand. (2) The valid question item may be reused and incorporated into the test.

**b. Reliability**

Based on the result of content and item analysis towards the questions of History subject of Odd Semester test in the Islamic Public School using the Quest program, it was found that the reliability coefficient of the test was 0,53. The scale is categorized into fair based on the scale of 0,400-0,599 introduced by (Masidjo, 1995:209). The questions can be administered to the students because they are not god nor are they bad.

The factors affecting them the reliability are as follow:

1) Generally, the longer the test is, the more reliable it is.
2) The bigger the score distributions are, the more reliable they are.
3) The complexity of the test. A test which is too hard or too easy will yield a low test complexity since it causes a low score distribution.
4) Test objectivity. A highly objective test will be highly reliable as well for it does not depend on the scoring process.
5) A short test duration will cause a low reliability of the test.

**c. The Complexity Level**

The result of question item analysis using the Quest program shows that there were 40 difficult questions (80%) in the the History subject test. The questions include number 1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 48. An instance towards the difficult question is reflected in number 28 whose answer is D.

28. Proses yang menerangkan bahwa nenek moyang Indonesia berasal dari luar adalah...
   A. Mencari wilayah subur
   B. Menjelajah
   C. Naiknya permukaan laut
   D. Migrasi
   E. Perang

Source: The Document of Odd Semester Test of 2017/2018 of History Subject in Madrasah Aliyah Negeri

The complexity index of that question is 0,138. The question might become hard to the students when it has not been taught previously, thus the minimum competence has not been reached. If a question belongs to a hard category, there will be 3 possible decisions to carry out: (1) The question item shall be diminished from the test. (2) There should be an investigation on the cause of the complexity of this particular question (Sudijono, 2012: 376-377).

The questions in the medium category were found in 9 questions (18%), including questions number 5, 6, 9, 24, 29, 31, 32, 47, 49. An example of those difficult question could be seen in the question number 32.

32. Sebagian masyarakat sekarang masih ada yang meletakkan sesajian di bawah pohon besar. Hal ini merupakan warisan dari nenek moyang kita pada zaman pra aksara, yaitu...
A. Dinamisme
B. Animisme
C. Atheisme
D. Monotheisme
E. Komunisme

Source: The Document of Odd Semester Test of 2017/2018 of History Subject in Madrasah Aliyah Negeri

The question on the medium level is the best one since it is not too hard nor is it too easy for the students (Ambiyar, 2014: 150). Sudijono (2011: 376) suggested that this type of question should be recorded in the bank of question to be administered in the future. Further, the questions in the easy category were found in 1 question (2%), that is the question number 50.

50. Pengaruh Hindu Budha terhadap Indonesia di bidang pemerintahan adalah...
A. Adanya titisan Dewa
B. Seni patung
C. Ritual Keagamaan
D. Lahirnya kerajaan-kerajaan
E. Candi

Source: The Document of Odd Semester Test of 2017/2018 of History Subject in Madrasah Aliyah Negeri

It can be seen that the question above can be easily guessed by the student. It happens because the answers towards the question are matched too obviously that the students have no hesitation in choosing the right answer.

**d. The Discrimination Power**

Through the analysis of question items using the Quest program, it was found that there were 26 (52%) questions belonged to weak category. Those questions were identified in number 1, 4, 5, 6, 9, 10, 11, 12, 13, 14, 16, 17, 23, 24, 29, 31, 32, 33, 35, 36, 37, 38, 40, 43, 47, 50. There were 18 questions (36%) in the category of weak or medium such as ones in number 2, 3, 7, 8, 15, 18, 19, 20, 21, 25, 26, 27, 28, 30, 34, 39, 41, 42. On the other hand, good questions can be found in 4 numbers (8%) which are 44, 45, 48, 49. Finally, there was only one question (2%) belonged to very good category, namely question number 46. Conversely, the question in poor category found in question number 22 (2%).

Below is one example of questions in the weak category with the discrimination power index is 0,12.

33. Kebudayaan perunggu di Asia Tenggara berasal dari kebudayaan...
A. Bacson
B. Dongson
C. Hoabinh
D. Megalithik
E. Purba

Source: The Document of Odd Semester Test of 2017/2018 of History Subject in Madrasah Aliyah Negeri

The discrimination power for the question above is considered as poor since it fails to differentiate students who have understood the lesson and those who have not. Below is the question belong to fair category.

41. Banyak ahli berpendapat bahwa kebudayaan yang berkembang di Indonesia berasal dari India, hal ini disebabkan oleh faktor berikut, *kecuali*...
A. Ditemukannya candi di Indonesia
B. Bahasa sansekerta ditemukan di Indonesia
C. Indonesia pernah dijajah oleh India
D. Pernah terjalin hubungan dagang
E. Adanya kemiripan budaya

Source: The Document of Odd Semester Test of 2017/2018 of History Subject in Madrasah Aliyah Negeri

The discrimination power of the question above is 0,40, reflecting that the question can distinguish students who have mastered the lesson materials with those who have not. According to Sudijono (2012:408-409), the further actions taken towards this type of question are: (1) Fair, good, and very good category of question items should be

recorded in the question bank to enable the test makers using the question in the future. (2) Question items with a weak discrimination power can be treated by either: i) Investigate them to be revised, or ii) diminish them in order not to use them anymore in the future.

**e. The Distractor**

Based on the question analysis using the *Quest* program, it was found that there were 47 questions considered as good (94%). However, there were 3 questions fell into the poor category (6%) because some options were not chosen by the students. According to (Sudijono, 2012: 417), questions belonged to good category can be reused in the next test. Meanwhile, those who are categorized into poor must not be used in the future test.

## CONCLUSION

Based on the research explanations and elaborations above, several conclusions can be drawn that the validity of questions on History subject odd semester test academic year of 2017/2018 in Madrasah Aliyah Negeri (MAN) 3 Padang, MAN Pariaman, MAN 3 Padang Panjang shows that there are 28 questions (56%) and there are 22 invalid questions (44%). Meanwhile, according to item validity, 43 questions (86%) considered as valid and 7 questions (14%) were invalid.

The reliability of questions on History subject of odd semester test academic year of 2017/2018 in Madrasah Aliyah Negeri (MAN) 3 Padang, MAN Pariaman, MAN 3 Padang Panjang shows 0,53 as the reliability coefficient which means the questions have a fair reliability.

In terms of the complexity level of questions on History subject of odd semester test academic year of 2017/2018 in Madrasah Aliyah Negeri (MAN) 3 Padang, MAN Pariaman, MAN 3 Padang Panjang, there are 9 questions (18%) in the hard level, 1 question (2%) in easy level, and 40 questions (80%) in hard category.

In terms of the discrimination power of questions on History subject of odd semester test academic year of 2017/2018 in Madrasah Aliyah Negeri (MAN) 3 Padang, MAN Pariaman, MAN 3 Padang Panjang, the analysis shows that the test has not used a good discrimination power. It happens because from 50 questions, there are only 23 questions (46%) which have a good discrimination power (fair, medium, and good).

In terms of the distractors provided in the questions on History subject of odd semester test academic year of 2017/2018 in Madrasah Aliyah Negeri (MAN) 3 Padang, MAN Pariaman, MAN 3 Padang Panjang, the analysis reveals that the test has already given a good distractor because among 50 questions, it is found that 47 (94%) falls in good category.

## REFERENCE

Ambiyar. 2012. *Pengukuran dan Tes*. Padang: UNP Press

Anas Sudijono. 2012. *Pengantar Evaluasi Pendidikan*. Jakarta: PT. RajaGrafindo Persada

Hamzah B. Uno & Satria Koni. 2012. *Assesment Pembelajaran*. Jakarta: Bumi Aksara

Nana Sudjana. 2005. *Penilaian hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya

Sunarti & Selly Rahmawati. 2014. *Penilaian dalam Kurikulum 2013 Membantu Guru dan Calon Guru Mengetahui Langkah-langkah Penilaian Pembelajaran*. Yogyakarta: Andi OFFSET

Zainal Arifin. 2012. *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya

Muhammad Arif Hidayatullah. 2013. *Analisis Kualitas Butir Soal Ulangan Akhir Semester Genap Bahasa Arab Kelas VII Tahun Pelajaran 2012/2013 dengan Program Quest di SMP Muhammadiyah 3 Depok Sleman Daerah Istimewa Yogyakarta*. Skripsi. Yogyakarta: Universitas Islam Negeri