

**AN ITEM ANALYSIS ON ENGLISH FIRST SEMESTER TEST
FOR EIGHTH GRADE STUDENTS OF SMP N 2 TONJONG
IN THE ACADEMIC YEAR 2017/2018**

Tiya Sakina Putri¹

Sri Rejeki Pramudyawardhani²

The Graduate of English Education Study Program

Peradaban University Bumiayu – Brebes

E-mail: saputri21tya@gmail.com

Phone: +62 852 2902 3788

The Lecturer of English Study Program

Peradaban University Bumiayu – Brebes

Email: pramudyawardhani81@gmail.com

Phone: +62 815 663 5360

Abstract

This study was about an item analysis on English first Semester test for eighth grade students of SMP N 2 Tonjong in the academic year 2017/2018. The research aimed to find out the validity, the reliability, the difficulty level, and the discriminating power. This research studied an item analysis which used descriptive quantitative-qualitative design. In this research, it was found the validity of item 48% which were 24 valid items and 26 invalid items. It was reliable based on the reliability with coefficient reliable 0,83. In addition the difficulty level consisted of three kinds degree: 5 items or 10% were easy, 35 items or 70% were medium, and 10 items or 20% were difficult. Based on the discriminating power, there were five kinds of classification: 4 items or 8% were too poor, 7 items or 14% were poor, 15 items or 30% were fair, 19 items or 38%

were good, and 5 items or 10% were excellent. In conclusion, it was the good item test with equal items which were seen from the validity, the reliability, the difficulty level, and the discriminating power.

Keywords: Item analysis, the validity, the reliability, the difficulty level, the discriminating power

A. Introduction

Nowadays, English develops as major language that is used in the world which helps us as Asian people to communicate with people in the world. This is why English becomes a lessons in Indonesia. However, there are factors that make English difficult to master, such as students should understand the words in English fastly. Next, students should be able to speak in English but in their daily life they still use their mother tongue language. In other side, there are the factors from the test item which make it unclear for understood, such as all the answer choice are right, the item test has not the underlined sentences of expressing, and the answer choices gives hint to the right answer. From evaluation, teacher measures the students' competence and the students who have problems in English lesson to understand more than before.

It is stated, "Evaluation is a process through which a value judgement or decision is made from variety of observations and from the background and training of the evaluator" (Sax, 1980: 18), so this competence is appropriate with responsibility as a teacher.

One of the evaluation intruments is test. Test contains collection of questions related to the material on teaching and learning process. The

definition of test by Goodenough states that test is a task that gives to individual or groups which the purpose to compare their competence.

Teachers doing measurement should consider things when they arrange the test given to the students, such as the validity of test, the reliability of test, the difficulty level of test, the discriminating power, and the effectiveness of distractors. Nevertheless, the writer will not analyze the effectiveness of distractors, so the writer will restrict just from the validity of test, the reliability of test, the difficulty level of test, and the discriminating power. In addition especially the writer with that arrangement theory analyze in multiple choice item. Based on that things hopefully the teachers will make a good evaluation through English semester test.

Test is made by the teachers team called *MGMP (Musyawarah Guru mata Pelajaran)*. It is a forum for professional subject teachers and activities in SMP/MTs, SMPLB/MTsLB, SMA/MA, SMALB/MALB located in city with functions to communicate, learn, and share ideas and experiences to improve the teachers' competence as practitioners changes to reorientation of classroom learning. *MGMP* organization is under the Education Department Indonesia. Therefore, the writer conducts this research to help the teachers in analyzing the test finding out the quality of the test, because finding out whether the test is good or not is very important for the teachers. Moreover, the writer decides to collect the data from first semester test for eighth grade students of SMP N 2 Tonjong as it is one of favorite junior high schools in Brebes regency.

B. Literature Review

In this research there are some theories that the writers use in order to find appropriate data from the research.

1. The Definitions of Evaluation

According to Arikunto (2010: 1) says, "Evaluation is a process to determine the result from some activities that have been planned to get the purpose of education."

2. The Definitions of Test

According to Cronbach (1970), test is one of systematic procedures for observing and describing a person which includes behavior based on numerical level or category system in it.

3. Validity

Validity of test means that an instrument of evaluation can be useful when it is able to measure something what is measurement. Anastasi in Muri (2015: 60) states, "The validity of an instrument concern what does the instrument measure and how well it does so." Higher degree of validity, it means better instrument measurement that is used. In addition there is content validity based on Sudijono (2011: 164-165), "Content validity is one of analysis processes on evaluation after doing measurement and testing about the content on the student test item."

4. Reliability

On this method a tester is only trial in once time, it is called single-test-single-trial-method. Based on Arikunto (2009: 92-95) the steps on this

method are dividing students in two groups, and computing the result by using Spearman-Brown pattern.

5. The Difficulty Level

Based on Witherington in Sudijono (2011: 371) states that two kinds of a good test are not too difficult, or not too easy. Index difficulty level is called *proporsi* (p), then the pattern of difficulty level based on Du Bois in Sudijono (2011: 372)

6. Discriminating Power

According to Sudijono (2011: 387-390) explains the number to show discriminate index is called discriminating power, the specific index between 0,00 to 1,00. Nevertheless, discriminating index has no negative mark.

C. Method of Invertigation

In this research applies descriptive qualitative research method from Berg adapted by Djama'an in Muhammad (2011: 30) which is qualitative research. This research focuses on descriptions of the charactericts and it is only describing something because the aims to present the validity, the reliability, the difficulty level, and the discriminating power of English first semester test.

According to Sugiyono (2017: 7-8) Qualitative research called with the new one postpositivistic method includes natural setting and interpretation the findings data which is the phenomena based on range. The writer describes the result of an item analysis on English first semester test for eighth grade

students of SMP N 2 Tonjong in the academic year 2017/2018 especially in Brebes regency.

The data consist of the item test, students answer sheet, and interview to the teacher. In other side technique of data collection uses documentation and interview method. The writer acts as observer.

Technique of data analysis uses four kinds of procedure. Based on Sudijono (2011: 185-395)

D. Findings and Discussion

Item analysis is purposed to make identification of good test, fair, or poor. There are five kinds of determining the result in item analysis as follows:

1. Determining Validity

After collecting the data, for the next step the writer makes the students answer table to counts validity on each item number and students. The writer counted the data which is consist of 50 multiple choice item by using validity formula based on Arikunto (2009: 69). The validity result item is valid when it has consultation table degree formula explained that the validity result $>$ from r_{tt} or r_{table} 0,344 (n-2) with signification level taken for education program 5% to get interpretation on each number item of the validity result. In addition, to get the content validity result, the writer analyzes the data which is consist of 50 multiple choice item based on Bloom taxonomy specification divided to classify the level skills on each item and matches the item material based on curriculum or syllabus to know the result of content validity and finally to get the test quality.

Moreover, it is shown that:

There are 24 items which are valid, it clarifies the item result $>$ from r_{table} 0,344. Therefore, the interpretation of item is valid: the details number 1, 2, 5, 11, 12, 16, 17, 18, 19, 20, 21, 22, 25, 29, 33, 35, 38, 39, 42, 43, 44, 45, 48, and 50. Furthermore, 26 items are invalid, item is invalid: the details number 3, 4, 6, 7, 8, 9, 10, 13, 14, 15, 23, 24, 26, 27, 28, 30, 31, 32, 34, 36, 37, 40, 41, 46, 47, and 49.

On the other side the analysis of content validity shows that the items only cover the material in four aspects from six aspects on Taxonomy Bloom level. There are knowledge aspect, comprehension aspect, analysis aspect, and synthesis aspect. It is clarified in detail: knowledge aspect covers items number 2, 3, 14, 20, 26, and 37. Comprehension aspect covers items number 4, 5, 6, 7, 8, 11, 19, 22, 24, 25, 27, 29, 36, 38, 39, 40, 41, and 42. Analysis aspect covers items number 23, 28, and 43, and synthesis aspect covers items number 1, 9, 10, 12, 13, 15, 16, 17, 18, 21, 30, 31, 32, 33, 34, 35, 44, 45, 46, 47, 48, 49, and 50.

The test item has followed the arrangement rules of item which is consist of material aspect, construction aspect, and languages aspect. It describes in details, based on the material: the item has followed the indicator, each item has one right answer. The construction: the item does not give hint to the right answer. The languages: the item uses communicative diction, simple, and understandable. It makes the test item has good quality.

2. Determining Reliability

In counting the reliability, the writer divides the students answer on two table groups analysis, between the uneven table for the uneven item number result and the even table for the even item number. It counts correlation product moment formula based on Spearman-Brown with split half method or $r_{\frac{1}{2} \frac{1}{2}}$ then substitution the correlation product momen result with R_{11} formula, the counting is trial in once time.

The result is reliable when it has compared with table degree formula explained that R_{11} result $>$ from r_{tt} or r_{table} 0,344 (n-2) to get interpretation the reliability result. The reliability result $r_{\frac{1}{2} \frac{1}{2}}$ 0,72 then substitution with R_{11} shows the counted 0,83 $>$ 0,344 the interpretation is reliable, it means that the item test has high reliability. The test item has reliable, because the test item have been made based on the indicator and the rules of item arrangement which is consist of element material aspect, construction aspect, and also languages aspect. It is not easy to make the item still have high consistency unit of measure, unreliable factor shows that sometimes the length of answer choices relatively unsimiliar, and it is related to ability in accuracy of instument on this part of test.

3. Difficulty Level

The result counts with total right answer divided by total all students in class. It is described: Item categorized as difficult are 10 items or 20%, it clarifies the item result has between 0,00 – 0,30. It shows details in number 3, 21, 24, 27, 30, 32, 34, 35, 49, and 50. Item categorized as medium are 35

items or 70%, it clarifies the item result has between 0,30 – 0,70. It shows details in number 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 25, 26, 28, 31, 33, 37, 39, 40, 41, 42, 43, 44, 45, 46, 47, and 48. Item categorized as easy are 5 items or 10%, it clarifies the item result has between 0,70 – 1,00. It shows details in number 1, 5, 29, 36, and 38.

4. Discriminating Power

The discriminating power counts with divide the students answer sheet into two groups table. As for 27% higher group, and 27% lower group, because it is based on empirical evidence which has given dependable results. It is described in details: Item categorized as too poor are 4 items or 8%, it clarifies the item result has between negative (-). The details in number 10, 24, 37, and 46. Item categorized as poor are 7 items or 14%, it clarifies the item result has between $< 0,20$. The details in number 7, 8, 23, 30, 31, 36, and 49. Item categorized as fair are 15 items or 30%, it clarifies the item result has between 0,20 – 0,40. The details in number 3, 4, 6, 9, 14, 15, 21, 26, 27, 28, 32, 34, 41, 47, and 50. Item categorized as good are 19 items or 38%, it clarifies the item result has between 0,40 – 0,70. The details in number 1, 2, 5, 11, 12, 13, 16, 17, 20, 22, 25, 33, 35, 38, 40, 43, 44, 45, and 48. Item categorized excellent are 5 items or 10%, it clarifies the item result has between 0,70 – 1,00. The details in number 18, 19, 29, 39, and 42.

E. Conclusion

1. The Result of Item Analysis

This item analysis has determining part to the validity also the content validity; the reliability of item is reliable. In addition there are the difficulty level and the discriminating power. It is an equal item which covers four determining analysis. The multiple choice item also has different characteristic on each item which depends on the students' answers spreading on item and the computing data uses the formula put on it.

2. The Result of interview

The result on interview shows many opinions on it, the item that is made by MGMP group. It has the test item based the concrete data empirical on range. Moreover, the item has been arranged based on the rule of item arrangement, therefore it is the good item test on it.

Acknowledgement

The writer would like to express the deepest gratitude to Dede Nurdiawati, M.Pd as the Head of English Education Study Program of Peradaban University.

Bibliography

- Arikunto, Suharsimi. 2007. *Dasar-Dasar Evaluasi Pendidikan (Edisi Revisi)*. Jakarta: Bumi Aksara.
- . 2009. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- . 2012. *Dasar-Dasar Evaluasi Pendidikan Edisi 2*. Jakarta: Bumi Aksara.
- Brown, H. Douglas. 2003. *Language Assessment Principles and Classroom Practices*. California: Longman.
- Daryanto, H. 2010. *Evaluasi Pendidikan*. Jakarta: PT Rineka Cipta.
- English Department. 2017. *Pedoman Penulisan Skripsi Jurusan Pendidikan Bahasa Inggris Fakultas Keguruan dan Ilmu Pendidikan*. Universitas Peradaban Bumiayu.
- <http://forumgurunusantara.blogspot.com/2016/09/download-silabus-smpmts-kurikulum-2013.html?m=1> Online Available (Accessed on July 31 , 2018 at 10.37am)
- Humaerah, Muspira. 2016. *Item Analysis of English Summative Test for Second Grade Student of MAN 1 Bulukumba*. A Thesis. English Education Study Program of Tarbiyah and Teaching Science Faculty UIN Alaudin Makassar: Online Accessed on August 3, 2018 at 1.09 pm.
- Muhammad. 2011. *Metode Penelitian Bahasa*. Jogjakarta: Ar-Ruzz Media.
- Naraswari, Nikmah. 2014. *Item Analysis of English Semester Test for Second Grade Students of Senior High School*. A Thesis. English Education Study Program of Educational Sciences and Teachers' Training Faculty Muhammadiyah University: Not Published.
- Nur, Tira Fitria. 2015. *Analysis of the English Item Test of English Final Examination for SMP/MTs in Academic Year 2011/2012*. A Thesis. English Study Post Graduate Program Muhammadiyah University of Surakarta: Not Published.
- Purwanto. 2014. *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar.

- Oktaviyanti, Sari. 2017. *The Analysis of English Materials Design Issued By the Hospitality Industry College of Purwokerto (An english for Specific Purposes Approach Study)*. A Thesis. English Education Study Program of Educational sciences and Teachers' Training Faculty. Peradaban University: Not Published.
- Stamboel, Conny Semiawan. 1990. *Prinsip dan Teknik Pengukuran dan Penilaian di dalam Dunia Pendidikan*. Jakarta: Mutiara Sumber Widya.
- Sudijono, Anas. 2011. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada.
- Sugiono. 2010. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta
- . 2012. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta
- Suadi, Muhammad. 2016. *Analisis terhadap Kualitas Butir Soal Buatan Guru IPA Kelas VIII MTs. Negeri Slawi Tahun Pelajaran 2015/2016*. Jurnal Kamil, Vol. 3, pp.1-15.
- Syamsudin. 2012. *Pengukuran Daya Pembeda, Taraf Kesukaran, dan Pola Jawaban Tes (Analisis Butir Soal)*. Jurnal At-Tajdid, Vol. 1, pp. 1-12.
- Utari, Retno. 2011. *Taxonomy Bloom Apa dan Bagaimana Menggunakannya*. Journal of Widya Iswaramadya.
- Yusuf, A. Muri. 2015. *Asesmen dan Evaluasi Pendidikan*. Jakarta: Prenadamedia Group.