CONTENT VALIDITY OF ENGLISH SUMMATIVE TEST ITEMS OF THE EIGHTH GRADE STUDENTS AT SMPN 4 POLLUNG IN ACADEMIC YEAR 2017/2018

*Patricia Adelina Lumban Gaol

**Willem Saragih

**Anggraini Thesisia Saragih

ABSTRACT

This study aimed to describe the content validity of English summative test items of the eighth grade students at SMPN 4 Pollung in connection with the learning indicators, to investigate the reason of describing content validity and to write the test items with good content validity.

In this study, the researcher used descriptive qualitative design. The data in this study were the test items of summative test that obtained by coming and asking the test from the English teacher at the school. This study described the conformity and inconformity of the English summative test items in the even semester of the eighth grade students at SMPN 4 Pollung with the learning indicators. Then the researcher interpreted the conformity level of the English summative test to the learning indicators in term of content validity by using Arikunto's criteria.

The findings of this study showed that the test items is 35% conform to the learning indicators and it falls into the level 21 – 40% which is called poor. It means that the content validity of the English summative test items of the eighth grade students of SMPN 4 Pollung is poor.

Keywords: content validity, summative test.

_____

*Graduate status

**Lecturer Status

I. INTRODUCTION

Evaluation is one of the important aspects in teaching learning process. According to Harlen (2007: 12), evaluation is more often used to denote the process of collecting evidence and making judgments about programs, systems, materials, procedures and process. It means that through evaluation, teachers are able to find out the effectiveness or failure of a method and also students' achievement in mastering the lesson. In other word, evaluation is used to measure and evaluate teaching learning activity.

One of the evaluation forms that teachers need to hold the evaluation is a test. Through a test, teachers are able to figure out how far the students understand the lessons in teaching learning activity. There are various kinds of test that teachers can use to measure students' achievement in teaching learning. The tests used should be appropriate with things that the teachers want to measure.

Achievement test is a kind of test that is usually used by teachers in order to know how far students have mastered the lessons (Hughes (1995:13). The achievement test is intended to establish how successful individual students, group of students or the courses themselves have been in achieving objectives of language courses. Achievement test is divided into two types, namely formative test and summative test. Formative test is a test that administered during learning process. Summative test is a test that provides information and feedback that sums up the teaching and learning process. It is usually held in the end of learning period. In order to measure students' achievement accurately, the teachers should use a good test. A good test will give the real information about students' learning result and measure the ability of the students. The teachers should attend some characteristics or requirements that must be fulfilled in constructing the good test. According to Brown (2003), he classified the characteristic of good test into 5 parts: (1) validity, (2) reliability, (3) practicality, (4) authenticity and (5) washback. The statement can be understood that when constructing the test, it must fulfill those criteria.

A test with good validity should be in line with the syllabus especially with the learning indicators suggested and the content of the test must measure what is intended to measure. That is why validity is the most important consideration in test evaluation, because it is used to figure out the quality of the test entirely. There are three types of validity, namely content validity, criterion-related validity and construct validity. According to Thoha (1991:48), content validity deals with the content of the test item whether it reflects to the curriculum or not. It can be understood that in analysis of content validity we can find how far the students understand about the materials delivered by the teacher. So the teacher has to be careful when they arrange the test, because content validity needs a sharp and systematic analysis and it can represent the content of the test that will be examined.

Since a test should fulfill the characteristics of the good test, then it is necessary to analyze the quality of the test itself. The test's quality analysis is an analysis in terms of validity, reliability, practicality, authenticity and washback. In this research, the researcher focuses on the content validity of summative test.

Based on the researcher's observation towards English summative test items of the eighth grade in the odd semester, which were designed by one of the English teachers at SMPN 4 Pollung, it had some problems related to the content validity. Some of the test items' indicators are not in line with the learning indicators that are written by the teacher for instance the learning indicator "identifying information

from a greeting card" which is found in items number 26, 27, 28, and 29. They ask students to identify information from the greeting card of birthday. While in the syllabus, the learning indicator asks to "identify the similarities and differences from some greeting cards" and "identify the expression in it". It can be said that the learning indicator is not applied in the test. The incompatibility of the test items with the learning indicators consequently makes the test less appropriate to measure students' understanding towards the materials learnt or the test does not have a good content validity.

Furthermore, the finding of the research done by Lestari (2015) shows that the summative test does not have good content validity. It can be proved by some learning indicators which are not suggested but they are included in the test items. That condition automatically does not create high students' achievement.

The researcher concludes that content validity is very important in a test in order to measure what is intended to measure. Therefore, the researcher wants to analyze the content validity of the test items of the English summative test for the eighth grade students at SMPN 4 Pollung in academic year 2017/2018."

II.   THEORETICAL FRAMEWORK
1.   Test

Test is a tool that teachers used to hold evaluation. Test will help teachers to get the information about students' achievement and ability dealing with teaching learning process as well as to measure how far students comprehend about the lesson.

Longman in his book "*Language Assessment*" says, "a test, in simple terms, is a method of measuring a person's ability, knowledge, or performance in a given domain". There are four components of test based on the definition. First, test is a method, instrument, a set of techniques, procedures or items that requires performance on the part of the test-taker. It means that the method must be explicit and structured to qualify the test. Second, a test must measure. It is impossible for making a test, if the test itself cannot measure what is intended to measure. Third, a test measures an individual's ability, knowledge or performance. In this term, testers should understand who the test-takers are and how their background and previous experiences. If testers have understood theirs, they would found is the test appropriate with their abilities and then test-takers can interpret their scores. Last, a test measures a given domain.

Based on definitions above, the researcher concludes that a test is a method or tool which is used to measure students' ability, knowledge and intelligence by some questions and exercises.

There are a lot of types of test that can be used to measure students' achievement and ability. But, before designing the test, teachers as the test-makers should determine the purpose of the test and choose the right kind of test in order to help teachers to focus on the specific objectives of the test.

As Brown written in his book, there are five kinds of test based on its purpose; they are: language aptitude tests, proficiency test, placement test, diagnostic test and achievement test.

2. Summative test

Summative test typically comes at the end of a course or unit of instruction. It is designed to determine the extent to which the instructional goals have been achieved and is used primarily for assigning course grades or for certifying student mastery of the intended learning outcomes. The techniques used in summative test are determined by the instructional goals, but they typically include teacher-made achievement tests, rating on various types of performance and assessments of products. Although the main purpose of summative test is grading, or the certification of student achievement, it also provides information for judging the appropriateness of the course objectives and the effectiveness of the instruction.

3. Characteristics of a good test

Teachers have to understand and comprehend on how to make or construct a good test since test used to measure a person's ability. Hughes (20013) argued that such tests are mostly of poor quality leaving harmful effects on teaching and learning. The impact of test on teaching and learning, which may positive and negative, or rather beneficial or harmful, is linguistically known as wash back. In other words, a good test or assessment is the one that has validity, reliability, practicality, authenticity and beneficial wash back in a balanced way.

Brown in his book also stated that a good test has five characteristics that will be elaborated as follows:

a) Reliability

Reliability means the stability of test scores. A reliable test is consistent and dependable. If the same test given to the same student or matched students on two different occasions, the test should yield similar results. A test cannot measure anything well unless it measures consistently. Then, it can be concluded that reliability refers to the extent to which a test score is precise, consistent or stable.

b) Validity

Validity is the most important quality to consider when constructing or selecting an assessment procedure. Validity refers to the adequacy and appropriateness of the interpretation made from assessments, with regard to a particular use. If an assessment is to be used to describe students' achievement, we should like to be able to interpret the scores as a relevant and representative sample of the achievement domain to be measured.

Validity is the extent to which inferences made from assessment results are appropriate, meaningful and useful in terms of the purpose of the assessment. Lado (1975:30) stated that a test cannot be a good test unless it is valid. Validity is divided into 5 types namely content validity, criterion-related validity, construct validity, consequential validity and face validity.

c) Practicality

An effective test is practical. It means that the test: is not excessively expensive; it stays within appropriate time constrains; it is relatively easy to administer, and; it has a scoring/evaluation procedure that is specific and time-efficient.

d) Authenticity

Bachman and Palmer (1996,p.23) define authenticity as the degree of correspondence of the characteristics of a given language test task to the features of a

target language task and then suggest an agenda for identifying those target language tasks and for transforming them into valid test items.

e) Wash back

Wash back is used to mean the impact the test is going to make on both the teacher and the learner. The wash back can be positive to yield good influence on both of them. It can also be negative effect and yield bad results on the teaching and learning processes.

4. Content validity

The content of a course or curriculum maybe broadly defined to include both subject-matter content and instructional objectives. Both of these aspects of content are concern in determining content validity. We should like any achievement test we construct, or select, to provide results which are representative of the topics and behaviors we wish to measure.

This is the essence of content validity: more formally, content validity may be defined as the extent to which a test measures a representative sample of the subject-matter and the behavioral changes under consideration. The focus of content validity, then, is on the adequacy of the sample and not simply on the appearance of the test.

Content validity is the accuracy of a measuring instrument in terms of its contents. Content validity refers to how accurate an assessment or measurement tool taps into the various aspects of the specific construct in question. Kerlinger argues that content validity is representative of the content. Thus, content validity of an instrument depends on the adequacy of a specified domain of content that is sampled. Anderson (1975:27) refers content validity to the extent to which the test we are using actually measures the characteristics or dimension we intend to measure. Content validity includes any validity strategies that focus on the content of the test. To demonstrate content validity, testers investigate the degree to which a test is a representative sample of the content of objectives or specifications the test is originally designed to measure. Therefore, content validity measures the comprehensiveness and representativeness of the content of a scale.

5. Curriculum, Syllabus and Indicator

a. Curriculum

A curriculum is designed by the Department and qualified instructors are hired to teach that curriculum using their best judgment about the specific content to be included in individual courses and best pedagogical approaches for each course. Every course or school need curriculum to make their school be equal with others. Curriculum made by the government. Posner (2004:5) said that curriculum is the set of instructional strategies of teacher's plan to use. That statement means that curriculum has an important role in a school or course because curriculum is used as instructional strategy or basic reference on teaching learning process.

b. Syllabus

Syllabus is a reference for the preparation of a learning framework for every subject study material (Permendikbud No.22:2016). Functionally, a syllabus is a contract between the Departments—for which you act as proxy—and your students. The Department makes decisions about how to best educate students in its discipline and sub disciplines. Each course syllabus constitutes an agreement between the

instructor and her students as to course content, requirements, course policies including grading and the course calendar.

c.   Indicator

Indicator is the marker of achievement of basic competences which are characterized by measurable behavioral changes that include attitudes, knowledge and skills. Indicator is developed according to the characteristics of students, subjects, educational units, regional potential and formulated in measurable and / or observable operational verbs.

In developing indicators it is necessary to consider the demand for competencies that can be seen through the verbs used in basic competences; the characteristics of subjects, students and schools; and the potential and needs of students, the community and the environment. In developing learning and assessment, there are two formulation indicators namely:

1.   Indicators of competencies achievement known as indicators;
2.   Assessment indicators used in compiling the grid and writing questions known as the questions indicators.

Indicators are formulated in sentence form using operational verbs. The formulation of indicators includes at least two things, namely the level of competence and material that becomes the media for achieving competence.

The indicator functions as: guidelines for developing learning materials; guidelines for designing learning activities; guidelines for developing teaching materials; guidelines for designing and carrying out results assessments learning.

III.   RESEARCH METHODOLOGY

In conducting this research, the researcher used descriptive qualitative design. As Ary, Jacobs and Razavieh (2009:424) said, "the qualitative inquirer deals with data that are in the form of words, rather than numbers and statistics". The explanations of words are very important in the qualitative research.

This study described the conformity and inconformity of the English summative test items in the even semester of the eighth grade students at SMPN 4 Pollung with the learning indicators based on its content validity. The results of this research showed the percentage of conformity level between the test items and the learning indicators.

After collecting the data, the researcher continued the next step namely analyzing the data. The researcher analyzed the conformity level and the unconformity level of the English summative test to the learning indicators in a table qualitatively. All the data from the sources of research were analyzed by the researcher using descriptive analysis method. This method explained the data obtained with percentage analysis technique. The percentage was gotten by using the formula below:

$$P = \frac{F}{N} \times 100\%$$

IV. DATA, DATA ANALYSIS AND DISCUSSION

Data

The data in this research were the test items of English summative test. The data of this research is included into qualitative data. Qualitative data is usually in the form of words rather than number. The source of data in this research was the test which was given to the eighth grade students in the even semester. The researcher obtained the data by coming to the school and asking for the summative test from the English teacher.

Data Analysis

Having the data been collected, they were analyzed to describe its content validity. In this part, the researcher described the content validity of the test by analyzing the conformity of the test items in connection with the learning indicators. there are 14 test items that conform to the learning indicators. It means that there are 26 of 40 test items that are not conform to the learning indicators.

Answering the third research problem, the researcher wrote the test items with good content validity. Before write the test, the researcher improved the learning indicators which are not appropriate to use for making a test especially, for example the learning indicator "*3.8.1 menyebutkan kegiatan yang sedang dilakukan di kelas, sekolah dan rumah pada saat diucapkan, dengan ucapan dan tekanan kata yang benar*". It means that the students are asked to mention the activity directly (speaking), while the test is in multiple choice form and they have to choose the right answer.

There are 12 basic competences which are developed in the even semester and they have to be represented by 40 items number. Dealing with it, the researcher has improved the learning indicators and wrote the test items with good content validity.

By improving the learning indicators and wrote the test items above, so the third research question has answered. After that, to answer the first research question, the researcher counts the percentage of conformity of the test items by using the following formula:

$$P = \frac{F}{N} \times 100\%$$

$$P = \frac{F}{N} \times 100\%$$

$$P = \frac{14}{40} \times 100\%$$

$$\mathbf{P = 35\%}$$

Based on the calculation above, the conformity of English summative test with the learning indicators is 35%, which is consist of 14 test items. Based on Arikunto's criteria, the percentage obviously falls into the level of 21% - 40%, and it means

poor. So, the English summative test which was administered in the eighth grade students of SMPN 4 Pollung is 35% valid in terms of its content validity.

The inconformity between the test items and the learning indicators which is consist of 26 items is explained below:

$$P = \frac{F}{N} \times 100\%$$

$P = \dfrac{F}{N} \times 100\%$

$P = \dfrac{26}{40} \times 100\%$

$P = 65\,\%$

Then, the reason why content validity of English summative test is important to describe is to know whether the test is able to measure what is intended to measure and whether the test have represented the materials which were learnt by the students and the learning objectives.

Discussion

In teaching learning process, test is used as a tool to hold the evaluation. Summative test as a test that held in the end of learning period used to measure how far the students understand the materials learnt. Then, it is necessary to analyze the test based on its content validity to find out whether the test measure what is intended to measure.

This research is aimed to describe the content validity of English summative test items of the eighth grade students at SMPN 4 Pollung. The obtained data for this study were the English summative test items which is consist of 40 items. The researcher analyzed the test items in connection with learning indicators and counts the percentage and adjusts the percentage with Arikunto's criteria.

The result of this study showed that the English summative test is 35% conform to the learning indicators. Based on Arikunto's criteria, the percentage falls into the level 21% - 40% which means the test has *poor* content validity.

Regarding to the result, this study used the similar criteria from Arikunto that consist of five criteria of conformity level; they are very good, good, fair, poor and very poor. It is similar with the previous study by Fathony (2017) that used the same criteria. The result of the research showed that the percentage of the content validity of the test is 77,8%, means that the percentage falls into the level 61 – 80% and the criteria is good.

In addition, the criterion is also used by Dari (2014) in her research. The percentage of content validity of the test is 51,11%, falls into the level 41-60% and it means that the content validity of the test is fair.

Meanwhile, Lestari (2015) in her research used Arikunto's criteria but with the different interpretation, they are good, sufficient, less good and bad. The percentage

of content validity in her research is 55% and falls into the level of 40 – 55%. Based on its percentage, the test has less good content validity.

Then, Kristiana (2014) was also used the same criteria with the research conducted by Lestari (2015). The content validity of the test in her research is sufficient, because the percentage of conformity is 58,7% and it falls into the level 56 – 75%.

Furthermore, Adha (2014) was also conducted a research by analyzing the test and interpret the percentage with the similar criteria with Lestari (2015) and Kristiana (2014). The test which was analyzed has sufficient content validity. The percentage is 65% and it falls into the level 56 – 75%.

From the findings above, it can be concluded that analyzing content validity of English summative test is important for teachers to obtain information about whether the test has conform to the learning indicators and has measure what is intended to measure or in other words whether the test has good content validity. Thus, after knowing the content validity of the test that has been written, the teachers are able to arrange and to decide the better test items for the next examination.

V. CONCLUSION AND SUGGESTION
Conclusion

The problem of this study tries to describe the content validity of English summative test of the eighth grade student of SMPN 4 Pollung by analyzing them in connection with the learning indicators. After analyzing the data, the researcher concludes that:

1. The English summative test which was administrated in the eighth grade students of SMPN 4 Pollung is 35% valid in terms of conformity with the learning indicators. Based on Arikunto's criteria, the percentage obviously falls into the level of 21-40% and the interpretation is *poor*. It means that the English summative test for the eighth grade students of SMPN 4 Pollung has *poor* content validity.
2. The reason why content validity of English summative test is important to describe is to know whether the test is able to measure what is intended to measure and whether the test have represented the materials which were learnt by the students and the learning objectives.
3. The researcher has done the correction in writing learning indicators and the test items.

Suggestion

The following are several suggestions which are hopefully useful to others who want to conduct a research related to this study:

1. To English teachers

The English teachers as well as the test designer have to understand how to write a good test in term of content validity before making a test. Therefore, it will be better if the test designer/teachers make the test items by considering the conformity between the test items and the learning indicators. In constructing the test items, teachers also have to consider about the content. They have to distribute all the learning indicators into the test item fairly.

2. To other researchers

   The result of this research can be used as additional reference on the related study for the next research. Moreover the next researcher can research deeper with the other aspect.

REFERENCES

Arthur, Hughes.1989. *Testing for Language Teachers Second Edition.* Cambridge: Cambridge University Press.

Arthur, Hughes.1995. *Testing for Language Teachers, Great Britain:* Cambridge University Press. p.10

Ary, D., L.C., and Razavieh, A. 2009. *Introduction to Research in education (8th Ed.).* Belmont: Wadsworth Group.

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

Cresswell, J. 2014. Reseacrh Design: Qualitative, Quantitative and Mixed Methods Approaches. California: Sage Publication

Dawson, Debra and Atkinson, Michael. 2006. *Assessment: Designing Effective Multiple-choice and Short Answer Questions.* The University of Western Ontario

Eid Alrfooh, Dr. Atif. 2012. *Types of Achievement Tests which are Preferred by Outstanding Students at Al-Hussein bin Talal University.* Jordan: IISTE

F. Yaghmaie PhD. 2003. *Content Validity and Its Estimation.* Shahid Beheshti University: Spring

Fitzpatrick, anne R. 1983. *The Meaning of Content Validity.* Amherst: Psychological Measurement Inc.

Grounlund, Norman .E. 1977. *Constructing Achievement Test Second Edition.* New Jersey: Prentice-Hall

Grounlund, Norman .E. 1920. *Measurement and Evaluation in Teaching Third Edition.* New York: Macmillan

Guma Siddiek, Dr. Ahmed. 2010. *The Impact of Test Content Validity on Language Teaching and Learning.* Canadian Center of Science and Education

Hanna, Gerald S., and Peggy Dettmer. (2004). *Assessment for Effective Teaching. Using Context-Adaptive Planning.* Boston, MA: Pearson A&B

Harlen, W. 2007. *Assessment of Learning.* London: SAGE Publications.

Harris, David P.1969. *Testing English as a Second Language* .USA.McGraw-Hill Book Company

Nitko, Anthony J., and Susan M. Brookhart. (2005). *Education Assessment of Students.* New Jersey: Person Education Ltd

Shaoqian, L. (2003). *The Research of English Classroom Teaching.* Beijing: Foreign Language Teaching and Research Press

Suter, Christoph. 2001. *Describing and evaluating a syllabus in a context of compulsory secondary schooling.* University of Birmingham

Taherdoost, Hamed. 2016. *Validity and Reliability of the Reseacrh Instrument; How to Test the Validation of a Questionnaire/Survey in a Research.* Switzerland. Helvetic Editions LTD