

ANALYSIS OF VALIDITY AND RELIABILITY OF ENGLISH FORMATIVE TESTS

Aris Sugianto

Universitas Palangka Raya

sugianto_aris@yahoo.com

Abstract: This study was conducted in order to analyze the English formative tests made by the English teacher of grade VIII of SMPN-4 Mentaya Hulu based on the characteristics of a good test. The indicators to be analyzed were validity and reliability of the test. The validity covered the content and construct validity. The writer applied descriptive method in conducting this study. The population and sample of this study was the English teacher-made tests for the grade VIII of SMPN-4 Mentaya Hulu in the first semester of academic year 2008/2009. The writer analyzed 2 English formative tests as the representative of all English formative tests in the first semester of academic year 2008/2009. Based on the result of the data analysis, the writer found that the English formative test conducted on October 2008 belongs to high validity in its content validity and excellent validity in its construct validity. The English formative test conducted on November 2008 belongs to low validity in its content validity and high validity in its construct validity. While the English formative test both conducted on October 2008 and November 2008 are unreliable.

Keywords: *validity, reliability, formative test*

In line with the globalization era, learning foreign language, especially English, is becoming an important need for people to gain more competitive advantage. That is why in Indonesia, English is the first foreign language learned starting at the elementary in school level up to the university level.

In Indonesia, English is used as the first foreign language. It is included in a major subject which functions as a tool in developing the students' knowledge and skill in science, technology, culture, and art field that enables students to be diligent.

There are four major skills in learning English. They are listening, reading, speaking and writing. In

conducting an effective Teaching Learning Process (TLP), there are some matters that should be paid attention. They are the teacher, curriculum, syllabus, method, facility, test, etc. Test is one of the matters that will be focused in this study.

There are many kinds of tests. Based on Arikunto (2005:33), test is viewed from its use in measuring the students' ability. There are 3 kinds of test: (1) diagnostic test, (2) formative test, and (3) summative test.

Diagnostic test is test used to know the weaknesses of the students in learning. So by knowing their weaknesses, the teacher can give the appropriate treatment. Besides it is important for the teacher in knowing the weaknesses of the

students, through the diagnostic test, the students can also know and be conscious with their weaknesses themselves so they can improve themselves better.

Formative test is a test that is conducted after one unit or one lesson finished given by the teacher. By conducting formative test, the teachers can know how far the success of the Teaching Learning Process especially for one lesson. So they can decide what the next actions to the students are.

Summative test is a test that is conducted after all units are finished given by the teacher. This kind of test is conducted in the end of the semester. The summative test is used to get educational decision. Educational decision means the students can pass or fail in mastering the material. The summative test is interrelated to the formative test because the subjects in the summative test are including all units or lessons which are tested in the formative test.

In line with the Teaching Learning Process (TLP), simply, according to Sudijono (2005:67), there are two functions of test. They are: (1) As an instrument to measure the development or progress that has been reached by the students after they go through the teaching learning process within certain. (2) As an instrument to measure the success of the Teaching Learning Process. Through the test, the teachers will be able to know how far the programmed materials have been able to be reached by the students.

In the TLP, teachers have to conduct testing. Testing and teaching are closely interrelated to each other because the success of teaching cannot be measured and known without conducting a test. So, if it is related to the TLP, it means that test

is an instrument or procedure used to measure the students' ability, to diagnose the students' weaknesses, to get educational decision, etc., depending on the kinds of test conducted.

In line with this study, the writer is interested in analyzing formative tests made by the English teacher of Grade VIII of SMPN-4 Mentaya Hulu in the first semester. SMPN-4 Mentaya Hulu is one of the Junior High Schools in Mentaya Hulu Subdistrict, Kotawaringin Timur Regency. As the writer knows, SMPN-4 Mentaya Hulu has never been yet used as the object of research of Palangkaraya University's students.

Logically, the formative test is closely interrelated to the summative test. It is because the summative test is conducted after all units finish given by the teacher. If it is related to the meaning of formative test that is given after every one unit finishes, it means that the summative test includes some formative tests. In the other words, it means that, before conducting the summative test, the teacher gives some formative tests to the students. Based on the statements above, the result of formative test will be influential to the result of summative test. Although it can be assumed as the statements above, but it is not absolutely sure that the result of formative test will be influential to the result of summative test, so furthermore this study will try to analyze this relation whether it is influential.

The writer found that there were some previous studies dealing with analysis of teacher-made test related to criteria of a good test. They were *An Analysis of Validity of English Summative Test Constructed by the Teachers for the*

Seventh Grade Students of SLTPN-1 Pahandut by Marleni in 2006 and *An Analysis on the Summative Test made by the Teacher of SMP Katolik Palangkaraya* by Claritha in 2006. Some results of Marleni's study showed that the content validity belongs to the poor qualification, the construct validity belongs to the very low qualification, the criterion-related validity belongs to the good validity level, and the wash back validity belongs to the low qualification. While some results of Claritha's study showed that the content validity belongs to the high qualification, the construct validity belongs to the fair qualification, the criterion-related validity belongs to the very low category, the reliability of the Multiple Choice Question (MCQ) belongs to the reliable level, and the reliability of the Essay Question belongs to the unreliable level. The writer used these Sarjana Theses as references in conducting this study.

The writer is willing to conduct a study in a school that never become the object of research. It is very potential to be studied that the result of the study may help the English teachers of the school to construct a good test based on the characteristics of a good test. By giving a good test constructed by the teacher, the students would have an opportunity to get a good quality in learning. In this study, the writer uses the concept of the characteristics of a good test proposed by Arikunto (2005).

Based on Arikunto (2005), the concept of validity is a test that is given by the teacher should be valid. Valid test means the test can really measure what it is wanted to measure, not else. For example, if the teacher wants to measure speaking ability of the students, so the

teacher should give question orally and the students should answer orally too. Arikunto (2005:64) stated that "*A test is valid if it measures what it purposes to measure*". So, in other words, the test is valid if it really measures what it is intended to measure.

In this study, the writer analyzed 2 English formative tests. They were in order to be representative of all formative tests and to know whether there were some changes or differences between them in the case of the form of the English formative teacher made-tests in terms of the validity and reliability of the test. The English formative tests that had analyzed by the writer were conducted in October and November. There were 2 classrooms on the grade VIII. But only a classroom that was became the object for the need this study. It was class A. the data was collected from the result of the 2 English formative tests of this class. This class was recommended by the English teacher because this class was the best class of the 2 available classroom of the grade VIII.

As information, the school had some facilities such as sport facilities, a natural sciences laboratory, and a library. There was no language laboratory. It is known that SMPN-4 Mentaya Hulu has used KTSP. It is the same as the other schools that also have used this kind curriculum. Kurikulum Tingkat Satuan Pendidikan (KTSP) is curriculum targeted by the government started from 2006 replacing the KBK (Kurikulum Berbasis Kompetensi) targeted by the government in 2004. In KTSP, the government allows the teachers or the members of committee of each school to arrange and improve the curriculum or syllabus by themselves under the coordination from the regency.

To help the school and the regency and province, the government gives guidance to approve the curriculum and the syllabus. If it is related to the test, test is to measure the success of the material in curriculum and syllabus that has been reached.

Furthermore, in formal education, a curriculum (plural curricula) is the set of courses, and their content, offered at a school or university. While a syllabus is an outline and summary of topics to be covered in a course. It is often either set out by an exam board, or prepared by the professor who teaches the course, and is usually given to each student during the first class session. A syllabus usually contains specific information about the course, such as information on how, where and when to contact the teacher/lecturer and teaching assistants; an outline of what will be covered in the course; a schedule of test dates and the due dates for assignments; the grading policy for the course; specific classroom rules; etc.

METHOD

This study used descriptive quantitative method which are collected, arranged and classified the data, analyzed and interpreted the data. The variable of this study is the English formative tests for grade VIII students of SLTP-4 Mentaya Hulu at the first semester of academic year 2008/2009. Each English formative test had 15 items, that divided into two forms, 10 items were Multiple Choice question ((MCQ) , and 5 items were true-false questions (dichotomous scoring).

In analyzing the validity of the test, the writer analyzed the content validity and construct validity. In analyzing content validity, the writer prepared the

curriculum or syllabus of English subject for grade VIII of SLTP and compared the materials in syllabus with the content of each item. Then, the writer calculated the percentage of the material in the curriculum of English subject included in the test and put the result into a table based on Slameto (2001:115). Meanwhile, in analyzing the construct validity, the writer did the same steps in analyzing content validity, but the writer compared the indicator in the syllabus of English subject included in the test.

In analyzing the reliability, the writer calculated the right answers of each item, the total right answer, the value of p and q of each item, and tabulated the result into the table. Then, calculated the total variance, and finally, calculated the reliability of the test by using the Kuder-Richardson's formula (KR-20).

FINDINGS AND DISCUSSION

In the previous chapter, the writer has discussed that this study analyzed the validity and the reliability of the English formative tests made by the English teacher of SMPN-4 Mentaya Hulu in the first semester of academic year 2008/2009.

Actually, the English teacher of the school had some English formative tests in this semester, but for this study, the writer only studied 2 English formative tests as a drawing of the whole English formative tests conducted in grade VIII.

Description of the Data

In this study, the writer analyzed the validity (content and construct validity) and the reliability of the English formative test. The data needed in this study were 2 English formative tests made by the English teacher of SMPN-4 Mentaya Hulu

for grade VIII students in the first semester of academic year 2008/2009, the curriculum used by the English teacher, and the result sheets of the 2 English formative tests.

Description of the English formative test.

The 2 English formative tests were made by the English teacher of SMPN-4 Mentaya Hulu. Each English formative test had 15 items. It was divided into 2 forms, 10 items were Multiple Choice Questions (MCQ), and 5 items were true-false questions (dichotomous scoring). The MCQ had 4 options (a, b, c and d). The English teacher made the tests based on KTSP Model (Model Kurikulum Tingkat Satuan Pendidikan) arranged by BSNP (Badan Standar Nasional Pendidikan), National Education Department in 2007, and based on an English textbooks entitled "English on Sky 2" by Mukarto, M.Sc.

Description of Curriculum and syllabus.

The teacher of SMPN-4 Mentaya Hulu used KTSP Model to teach the students and to construct the English formative test. Materials in curriculum were needed to analyze the content validity and indicators were needed to analyze the construct validity.

Description of the Result Sheet

Result sheet define as a sheet that was arranged by the teacher by tabulating the result of the test (English formative test) in a table, item by item, as the summary of the result of English formative test that had been conducted. It should be 2 result sheets because the data were collected from the 2 English formative tests.

Description of Students' Scores

The writer tabulated the scores of the students into a table, then replaced the name of students into codes A1, B1, C1, etc. Replacing the name of the students is in order to hide the privacy of the students. It was known that the writer analyze 2 English formative tests. They were conducted on October 2008 and November 2008. The English formative test which was conducted on October 2008 done by 25 students and had 15 items. While the English formative test which was conducted on November 2008 also done by 25 students and had 15 items.

Result of the Data Analysis

The analysis of the content validity of each English formative test was done by comparing the material in the curriculum to the content of each item of the test. Then calculate the percentage of the learning material in the content of each item included in the English formative test.

After analyzed the data, it can be seen that the English formative test that was conducted on November 2008 was only available in reading and writing, while listening and speaking were not available. From 15 items of the English formative test, only 7 items were including in the materials in curriculum. So, if it was calculated in percentage, the English formative test that was conducted on November was only 46% of the materials in curriculum. Based on the description above, the percentage was less than 50% of the materials in curriculum, it means that the English formative test conducted on October 2008 was not valid.

The analysis of the construct validity of each English formative test was done

by comparing the indicator in curriculum to the content of each item of the test. Then calculate the percentage of the learning material in the content of each item included in the English formative test.

After analyzed the data, it can be seen that the English formative test that was conducted on October 2008 was only available in reading and writing, while listening and speaking were not available. The whole items were including in the indicators in curriculum. Based on the description above, the percentage was 100% of the indicators in curriculum, it means that the English formative test conducted on October 2008 was valid.

The percentages in the tables above were calculated based on Slameto (2001:115).

Where:

$$\text{Percentage}(PQ) = \frac{nm}{N} \times 100\%$$

Nm : Number of item(s)

N : Total question items

Reliability of the test

The reliability of the test measured by using the Kuder Richardson's formula (KR 20) based on Arikunto (2005). The writer chose the formula (KR-20) because the English formative tests that were taken by the writer for the need of this study were conducted one time and the form of test is kinds of dichotomy scoring. The formula was as follow:

$$r_{11} = \frac{n}{n-1} \left(\frac{S^2 - \sum pq}{S^2} \right)$$

The interpretations of reliability coefficient based on Sudijono (2003:209) were as follows:

≥ 0.70 : reliable

< 0.70 : unreliable

The result of reliability calculation of the English formative test conducted on October 2008.

Before calculating the reliability of the test, firstly it has to calculate the total variance (S^2) of the test. The total variance (S^2) of the English formative test conducted on October 2008 is calculated as follow:

$$\begin{aligned} S^2 &= \frac{\sum X^2 - \left(\frac{\sum X^2}{N} \right)}{N} \\ &= \frac{450 - \left(\frac{94^2}{25} \right)}{25} \\ &= \frac{450 - \left(\frac{8836}{25} \right)}{25} \\ &= \frac{450 - 353.44}{25} \\ &= \frac{105.560}{25} = 4.22 \end{aligned}$$

After getting the value of the total variance (S^2), the reliability of the English formative test conducted on October 2008 is calculated as follow:

$$\begin{aligned} r_{ii} &= \frac{n}{n-1} \left(\frac{S^2 - \sum pq}{S^2} \right) \\ &= \frac{15}{15-1} \left(\frac{4.22 - 2.32}{4.22} \right) \\ &= \frac{15}{14} \left(\frac{1.90}{4.22} \right) \\ &= 1.07 \times 0.45 \\ &= 0.48 \\ r_{ii} &= 0.48 \end{aligned}$$

The result of reliability of the English formative test conducted on November 2008

The total variance (S^2) of the English formative test conducted on November 2008 is calculated as follow:

$$S^2 = \frac{\sum X^2 - \left(\frac{\sum X^2}{N} \right)}{N}$$

$$\begin{aligned}
&= \frac{1242 - \left(\frac{162^2}{25}\right)}{25} \\
&= \frac{1242 - \left(\frac{26244}{25}\right)}{25} \\
&= \frac{1242 - 1049.76}{25} \\
&= \frac{192.24}{25} = 7.69
\end{aligned}$$

So, the reliability of the English formative test conducted on November 2008 is calculated as follow:

$$\begin{aligned}
r_{ii} &= \frac{n}{n-1} \left(\frac{S^2 - \sum pq}{S^2} \right) \\
&= \frac{15}{15-1} \left(\frac{7.69 - 2.78}{7.69} \right) \\
&= \frac{15}{14} \left(\frac{4.91}{7.69} \right) \\
&= 1.07 \times 0.64 \\
&= 0.68 \\
r_{ii} &= 0.68
\end{aligned}$$

Based on the calculation above, it can be concluded that the English formative test made by the teacher of grade VIII of SMPN-4 Mentaya Hulu in the first semester of academic year 2008/2009 conducted on October 2008 is measured 0.48 and on November 2008 is measured 0.68. So, based on the interpretation of reliability coefficient based Sudijono (2003:209), it means that English formative test conducted both on October 2008 and on November 2008 are unreliable.

CONCLUSION

Based on the result of the study, the writer concludes that: (1) the English formative test conducted on October 2008

is valid in its content validity. Since the value of content validity is 87%, (2) the English formative test conducted on November 2008 is not valid in its content validity. Since the value of content validity is 46%, (3) the English formative test conducted on October 2008 is valid in its construct validity. Since the value of construct validity is 100%, (4) the English formative test conducted on November 2008 is valid in its construct validity. Since the value of construct validity is 79%, (5) the English formative test conducted on October 2008 is unreliable in its reliability. Since the value of reliability is 0.48, and (6) the English formative test conducted on November 2008 is unreliable in its reliability. Since the value of reliability is 0.68.

Based the conclusion of this study, the writer would like to give some suggestions especially to the English formative tests made by the English teacher of SMPN-4 Mentaya Hulu and all teachers that will construct tests: (1) in constructing the test, the teacher should pay attention to the characteristics of a good test especially in content and construct validity, and reliability, (2) in constructing the test, the teacher should construct based on the materials and indicators in curriculum. So, it will belongs the validity, either content or construct validity, and (3) the teacher should teach the students by using hand books that have the same curriculum with the curriculum that is used to construct the test.

REFERENCES

- Arikunto, S. (2005). *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara.
- Baron, D. & Bernard, H. W. (1958). *Evaluation techniques for classroom teachers*. New York: McGraw-Hill Book Company, Inc.
- Brown, J. (1996). *Testing in language programs*. New Jersey: Prentice-Hall, Inc.
- Claritha, F. (2006). *An Analysis of the summative test made by the teacher of SMP katolik Palangka Raya*. Palangka Raya: Universitas Palangka Raya.
- Fakultas Keguruan dan Ilmu Pendidikan, Universitas Palangka Raya. (1997). *Pedoman penulisan proposal dan skripsi*. Palangka Raya: FKIP Unpar.
- Heaton, J. B. (1979). *Writing English test*. London: Longman.
- Hornby, A. S. (1995). *Oxford advanced learner's dictionary of current English*. Oxford: Oxford University Press.
- Marleni. (2006). *An analysis of validity of English summative test constructed by the teachers for the seventh grade students of SLTPN-1 Pahandut*. Palangka Raya: Universitas Palangka Raya.
- Slameto. (2001). *Evaluasi pendidikan*. Jakarta: Bumi Aksara.
- Sudijono, A. (2006). *Pengantar evaluasi pendidikan*. Jakarta: Raja Grafindo Persada.
- Toendan, W. H. (2006). *Educational research methods: An introduction*. Palangka Raya: University of Palangka Raya.
- Tuckman, B. W. (1975). *Measuring educational outcomes fundamentals of testing*. New York: Harcourt Brace Jovanovich, Inc.