
An Analysis of English National Final Examination for Junior High School in Terms of Validity and Reliability

Aris Sugianto

aris.sugianto@iain-palangkaraya.ac.id

State Islamic Institute of Palangka Raya

Jl. G. Obos Komplek Islamic Centre, Palangka Raya, Kalimantan Tengah, Indonesia

Received: December 30, 2015; Accepted: February 27, 2016; Published: March 25, 2016

Abstract

The purpose of this study is to analyze the English National Final Examination for Junior High School. Validity and reliability are the two kinds of characteristics of a good test which are concerned. Since the five packages of the test are the same in content and only different on the placement of the items, so it is considered to take only one package becomes sample. This study was analyzed by using the descriptive method. Content validity was analyzed by comparing the materials in the syllabus to the items of the test, and construct validity was analyzed by comparing the indicators in the syllabus to the items of the test. While the reliability was analyzed by using Kuder-Richardson Formula (KR20). The result of the study shows that the English National Final Examination for Junior High School was valid and reliable. The content validity shows 100% valid, and the construct validity shows 100% valid. While the reliability shows coefficient 0.89, and it is reliable. So, the English National Final Examination for Junior High School has fulfilled the characteristics of a good test.

Keywords: English language testing, English language evaluation, test validity, test reliability, English national final examination

How to cite this paper: Sugianto, A. (2016). An Analysis of English National Final Examination for Junior High School in Terms of Validity and Reliability. *Journal on English as a Foreign Language*, 6(1), 31-42.

Aris Sugianto

In line with the globalization era, learning a foreign language, especially English, becomes an important need for people to gain a more competitive advantage. That is why in Indonesia, English is the first foreign language learned started from Elementary School up to University.

In Indonesia, English is a foreign language. It is a major subject which functions as a tool in developing students' knowledge and skills in science, technology, culture, and art field that enables students to be more diligent. There are four major skills in learning English. They are listening, reading, speaking and writing. In order to conduct an effective Teaching Learning Process (TLP), there are some matters that should be paid attention, such as the teacher, curriculum, syllabus, method, facility, test, etc. The test is one of the matters that will be focused in this study.

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. As Arikunto (2005, p. 53) stated that test is an instrument that is used to measure a condition by the certain rule. The rule, in this case, refers to the characteristics of a good 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., depends on the kinds of a test conducted.

In this study, it focuses on National Final Examination or in Indonesia it is called *Ujian Nasional (UN)*. Until 2014 it is known that in National Final Examination, in the point of view of people generally, it is still a dilemma. People said that it is not fear for certain students and schools in remote areas if it is for educational decision passed and failed. The standard of curriculum and expectation of the government are too high for them. It is known that the standard of National Final Examination is the same in all areas whether in big cities or remote areas such small villages. In big cities, the facilities are adequate and complete such as language laboratory, a complete library, internet or easiness to find books and references. While in remote areas, it is still far from the expectation. They have limited facilities, laboratory, library, internet or books.

Rafsanjani (2007) wrote in his blog and also published at local newspaper, Galamedia, 1 November 2006, that Howard Gardner, Professor of Education at Harvard University, United States of America stated,

Student assessment has to take place locally, in class, with face to face feedback, and not performed from a long distance, assessed mechanically, and returned with some score, suggestion or rectification about what must be done.

Based on recommendation from people and experts, now based on *Peraturan Menteri Pendidikan dan Kebudayaan Indonesia* Number 57 Year 2015, National Final Examination is not for educational decision again, it is mentioned that the results of the National Exam Year are now used for mapping the quality of programs and/or the Education Unit, consideration of selection into the next education level, and for consideration in the development and provision of assistance to the Education Unit in an effort to improve the quality of education. The result of National Final Examination is now for see the achievement level of graduate competency.

Nevertheless, still, National Final Examination should fulfill the characteristic of a good test. A test should be valid and reliable. Valid means that the test measures what it is intended to measure. The test should measure what the teacher wants to measure. For example, if the teacher wants to measure the speaking ability, the teacher should give the test in form of oral test, not giving the text to read or recording to listen. While reliable refers to the consistency of score. It means that if the teacher or tester gives the test repeatedly, the result should be approximately the same. Calfee and Chambliss (2005, p. 52) stated: "Validity is the extent to which a test measures what it is supposed to measure." Also, Fulcher and Davidson (2007, p. 4) stated that validity in testing and assessment has traditionally been understood to mean discovering whether test measures accurately what it is intended to measure or uncovering the appropriateness of a given test or any of its component parts as a measure of what it is purposed to measure.

Indeed, the government states that the National Final Examination is valid and reliable. Valid and reliable stated by the government are measured generally for all Junior High Schools in Indonesia without any consideration if the development of School-Based Curriculum or *Kurikulum Tingkat Satuan Pendidikan* (KTSP) of certain schools has fulfilled the standard/model of curriculum delivered by the government. So in this case, this study aimed to prove that the measurement by the government is really valid and reliable. The data of this study are taken from a State Junior High School in Palangka Raya. The data are syllabus and score of the students.

In line with this study, the writer found a previous study dealing with the analysis of English National Final Examination to criteria of a good test. It is a thesis entitled *An Analysis of English National Final Examination (UN) for Junior High Schools in Kurun Viewed from School-Based Curriculum (KTSP)*, written by Agustito (2012). The analysis is to match whether the English National Final Examination is matching with the competencies and materials in English syllabus of Junior High Schools in Kurun, and the result is matching. The strength is, by the study, it can be known about how to match the English

National Final Examination to the contents and competencies in the syllabus and curriculum of Junior High Schools in Kurun. It is useful to all Junior High Schools in Kurun. They can know how the KTSP they have been developed if it is matched with the items of the National Final Examination.

English National Final Examination as a test should be valid. The English National Final Examination should really measure what it is intended to measure. In this study, kinds of validity focused on content and construct validity.

English National Final Examination as a test, related to the Teaching Learning Process, simply, according to Sudijono (2011, p. 67), there are two functions of the 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, and (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 programme materials have been able to be reached by the students.

So, the function of the test is important either for the teachers or for the students. The former is important for the students and the latter are important for the teacher. In the case of the English National Final Examination, until the year 2014, it functions to make an educational decision. By the results, the students will be determined to be passed or failed.

The important characteristic of a good test is validity. A test should be valid. What is the purpose of measuring the validity of a test? Fulcher & Davidson (2007, p. 4) stated,

Validity presupposes that when we write a test we have the *intention* to measure something, that the 'something' is 'real', and that validity inquiry concerns finding out whether a test 'actually does measure' what is intended.

A test fulfills the content validity if it measures the materials that have been programmed and given. The programme materials are as described in the curriculum. So, to reach the content validity, it is necessary to construct the items of the test based on the materials that have been programmed in the curriculum. The details of the curriculum can be seen in the syllabus. The statements above are as Fulcher and Davidson (2007, p. 67) stated,

Content validity is defined as any attempt to show that the content of the test is a representative sample from the domain that is to be tested. In our example of the academic reading test it would be necessary to show that the texts selected for the “.

Aris Sugianto

A test also fulfills the construct validity if the items of the test constructed to measure every aspect of learning such as mentioned in the Specific Instructional Objectives or *Tujuan Instruksional Khusus*. The statement above is as Arikunto (2005, p. 67) stated that a test would have construct validity if each item of questions measures every aspect of thinking as stated in Specific Instructional Objectives.

Specific Instructional Objectives as Arikunto stated above, in School-Based Curriculum (KTSP) are equivalent to the Indicators in the Lesson Plan or curriculum or syllabus that show the success of the students in mastering the material.

Another important characteristic of a good test is reliability. Reliability refers to the consistency of a measure. A test is considered reliable if we get the same result repeatedly. For example, if a test is designed to measure English ability, the results should be approximately the same. The following is as Fulcher and Davidson (2007, p. 15) stated: "Reliability is the consistency of test scores across facets of the test".

Reliability indicates the degree to which a test is consistent in measuring whatever it does measure the degree to which the test measures the same thing time after time and item after item. Consistency over time items is basic to the concept of reliability. So, if it is related to the test, especially the English National Final Examination for Junior High Schools, that reliability is the consistency of the score in whenever the English National Final Examination is conducted.

This study is significant for Junior High School and the teachers as they can know if the English National Final Examination conducted by the government is valid and reliable or not. It is also as the material to improve the knowledge of theory and practice especially about the validity and the reliability of a test. Also, the result is useful for the English teachers so they can improve the syllabus and the way of teaching to be better. While the significance for the government and also the people is, through this study they can know how valid and reliable the test they constructed if it is analyzed and measured to be proved.

The test should be based on the curriculum. It is known that now the government targeted the school should use School-Based Curriculum or *Kurikulum Tingkat Satuan Pendidikan* (KTSP). School-Based Curriculum or also mentioned as the Genre-based curriculum is a curriculum targeted by the government started from 2006 replacing the *Kurikulum Berbasis Kompetensi* (KBK) targeted by the government in 2004. In KTSP, the government allows the teachers or the members of the committee of each school to arrange and improve the curriculum or syllabus by their selves under the coordination from

the regency. To help the school and the regency, the government gives guidance to approve the curriculum and the syllabus. In this study, this kind of curriculum that is being data of analysis.

METHOD

This study belongs to the descriptive research approach. Williams (2007, p. 66) stated,

The descriptive research approach is a basic research method that examines the situation, as it exists in its current state. Descriptive research involves identification of attributes of a particular phenomenon based on an observational basis or the exploration of correlation between two or more phenomena.

The writer describes the English National Final Examinations in terms of validity and reliability. The data were collected from the questions sheet, answers sheet resulted from the English National Final Examination and syllabus of Junior High School.

Since the writer used descriptive research approach, the writer analyzed the data by comparing the items of test to the syllabus to find the validity both content and construct. Content validity is comparing the materials in the syllabus to the content of items of English national Final Examination and Construct validity is comparing indicators in the syllabus to the content of items of English national Final Examination. The result was in percentage to show the level of validity. While the answer sheets were analyzed in order to find the reliability. The writer did a correction to the answer sheets the tabulated the score into the table. By the score, the reliability is measured or calculated by Kuder-Richardson Formula (KR-20).

FINDINGS AND DISCUSSION

The results of the data analysis are shown as follows:

Validity of the Test

The validity both content and construct were analyzed by tabulating in tables. The percentages were calculated based on the following formula:

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

Where:

nm : Number of item(s)

N : Total question items

Each kind of the validity was analyzed as follows:

Content Validity

The analysis of the content validity of English National Final Examination was done by comparing the materials in the syllabus to the content of each item of the test, and then calculating the percentage of the learning material in the content of each item included in the test.

It is found that all items of the English National Final Examination represent the materials in the syllabus. So, if it is calculated in percentage, the English National Final Examination contains 100% of the materials in the syllabus. Based on the description above, it means the English National Final Examination is valid in term of content validity.

As additional information, from the analysis, it is found that the materials included in the test are: grade VII is 28%, grade VIII is 34% and grade IX is 38%. So the materials were mostly taken from grade IX.

Related to the KTSP which is also called as genre-based curriculum, if it is related to genre or text type, descriptive text is dominant 38% (reading) of all materials, followed by narrative text is 14% (reading), procedure text is 10% (reading), recount text is 6% (reading) + 2% (writing), report text is 6% (reading) and exposition is 0%. While the other kinds of material (out of genres) such as instruction, greeting card, announcement and personal letter covers 22%.

Construct Validity

The analysis of the construct validity of the English National Final Examination for Junior High School was done by comparing the indicator in the syllabus to the content of each item of the test, and then calculating the percentage of the learning indicator in the content of each item included in the test.

It is found that all items represent the indicators in the syllabus. In other words, the English National Final Examination contains 100% of the indicator in the syllabus. It means that the English National Final Examination is valid in term of construct validity.

Reliability of the Test

The reliability of the test was measured by using the Kuder Richardson's formula (KR 20). The writer chose the formula (KR20) because the English National Final Examination for Junior High School was conducted one time and in form of Multiple Choice Question (MCQ). MCQ is kind of dichotomy scoring. The formula is 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 (2011, p. 209) are as follows:

$r_{ii} \geq 0.70$: reliable

$r_{ii} < 0.70$: unreliable

Before calculating the reliability of the test, firstly it has to calculate the total variance (S^2) of the test.

$$\begin{aligned} S^2 &= \frac{\sum X^2 - \left(\frac{\sum X^2}{N}\right)}{N} \\ &= \frac{47542 - \left(\frac{1194^2}{31}\right)}{31} \\ &= \frac{47542 - \left(\frac{1425636}{31}\right)}{31} \\ &= \frac{47542 - 45988.26}{31} \\ &= \frac{1553.74}{31} \\ &= 50.12 \end{aligned}$$

The result of calculating the total variance is 50.12.

After getting the value of the total variance (S^2), the reliability of the English National Final Examination for Junior High School is calculated, and the result is 0.89.

$$\begin{aligned} r_{ii} &= \frac{n}{n-1} \left(\frac{S^2 - \sum pq}{S^2} \right) \\ &= \frac{50}{50-1} \left(\frac{50.12 - 6.58}{50.12} \right) \\ &= \frac{50}{49} \left(\frac{43.54}{50.12} \right) \\ &= 1.02(0.87) \end{aligned}$$

Aris Sugianto

= 0.89
 r_{ii} = 0.89

So, based on the interpretation of reliability coefficient ($r_{ii} \geq 0.70$: reliable, $r_{ii} < 0.70$: unreliable), it means the English National Final Examination for Junior High School is reliable.

Besides validity and reliability, many aspects that actually should be analyzed to prove that the English National Final Examination has fulfilled the characteristics of a good test such as item difficulty, item discrimination, distracter analysis, and validity of each item of the question. But on this occasion, the writer only has a chance to analyze the validity and reliability. It will be worth consideration if the next researcher will analyze the other aspects so can make the final exam really proved that it is proper to be the instrument of the educational program.

Another finding of this study is about exposition text programmed in the syllabus (Grade VIII) of Junior High Schools where the syllabus was gotten as data (Sugianto, 2011), it is redundancy and useless because actually the government through the committee of Educational National Standard or *Badan Standar Nasional Pendidikan* (BSNP) and the model of syllabus, it does not program exposition text in Junior High Schools level, it is for Senior High Schools. Fortunately, it does not influence the result of this study.

CONCLUSION

Based on the analysis, it can be concluded as follows. First, the English National Final Examination for Junior High School is valid in its content validity. The value of content validity is 100%. Second, the English National Final Examination for Junior High School is valid for its construct validity. The value of construct validity is 100%. Third, the English National Final Examination for Junior High School is reliable in its reliability. The value of reliability is 0.89.

From this result of analysis, the problem of the study is solved. The problem of the study is to see whether the English National Final Examination for Junior High School is valid and reliable. The validity (content and construct) is analyzed by comparing the test items to the syllabus. Content validity is the suitability of materials in the syllabus to the content of items of English national Final Examination and Construct validity is the suitability of indicators in the syllabus to the content of items of English national Final Examination. Then reliability is analyzed by data of answer sheet which is in form of score. By the score, the reliability is measured or calculated by Kuder-Richardson Formula (KR-20). Thus, it shows that the English National Final Examination for Junior

High School has fulfilled the characteristics of a good test in terms of validity and reliability.

In this study, the writer suggests to those who have a duty to give the test, it is expected they keep paying attention to the characteristics of a good test in constructing the test. For the government who constructs the National Final Examination, it is expected to pay attention to the schools in the remote area where it is known that the quality of syllabus development, the facilities, and access to the schools are very limited and it may affect the result of the test. In writer's opinion, although the test is actually valid and reliable if it is measured in a certain school, it does not mean it will be valid and reliable too if it is measured in other schools especially in a remote area. It is because of the quality of syllabus development, the facilities and access really influence the result. In this case, the ability of the teachers in TLP and developing material and syllabus affects the success of the TLP, whether the teachers have taught based on the syllabus or not, moreover whether the syllabus has been developed in line with the model that has been delivered by the government or not. Here the attention of the government is really needed.

While for the teachers who have a duty to make tests, it is suggested to keep developing the materials and syllabus become better and also keep paying attention in how to make a good test. The ability of the teacher in constructing a good test has the effect of the result of English national Final Examination.

TLP cannot be measured without a test, so the test is very important to conduct. Relating to the finding of this study which there is exposition text programmed in the syllabus (grade VIII) of Junior High School where the syllabus was gotten as data, it is redundancy and useless because actually the government through the committee of Educational National Standard (BSNP) and the model of syllabus, it does not program exposition text in Junior High Schools level, it is for Senior High Schools. Fortunately, it does not influence the result of this study. Nevertheless, redundancy is wasting time. A lot of time is wasted. The wasted time should be used for materials which are appropriate to the well-arranged syllabus.

REFERENCES

- Abdullah, S. (2012). *Evaluasi pembelajaran*. Semarang: Pustaka Rizki Putra.
- Agustito. (2012). *An analysis of English national final examination (UN) for junior high schools in Kurun viewed from the school-based curriculum (KTSP)*. Unpublished Thesis. Palangka Raya: University of Palangka Raya.
- Arikunto, S. (2005). *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara.
- Arikunto, S., & Jabar, C. S. A. (2010). *Evaluasi program pendidikan*. Jakarta: Bumi Aksara.

Aris Sugianto

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Djiwandono, M. S. (2008). *Tes bahasa*. Jakarta: PT. Indeks.
- Fathoni, A. (2010). *Metode penelitian dan teknik penyusunan skripsi*. Jakarta: Rineka Cipta.
- Flood, J. (2005). *Methods of research on teaching the English language arts*. London: Lawrence Erlbaum Associates, Inc.
- Fulcher, G. (2010). *Practical language testing*. London: Hodder Education.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London and New York: Routledge.
- Osman, M. R. (2010). *Educational evaluation and testing*. Nairobi: African Virtual University.
- Rafsanjani, A. (2007). *National final examination (UN), damages the happiness of learning*, <http://heilraff.blogspot.com/2007/11/national-final-examination-un-damages.html>, (November 25, 2015).
- Ridwan. (2009). *Metode dan teknik menyusun proposal penelitian*. Bandung: Alfabeta.
- Setyosari, H. P. (2012). *Metode penelitian pendidikan dan pengembangan*. Jakarta: Kencana Prenada Media Group.
- Singh, K. Y. (2006). *Fundamental of research methodology and statistics*. New Delhi: New Age International Publishers.
- Sudijono, A. (2011). *Pengantar evaluasi pendidikan*. Jakarta: Raja Grafindo Persada
- Sugianto, A. (2011). Analysis of validity and reliability of English formative tests. *Journal on English as a Foreign Language*, 1(2), 87-94.
- Sugiyono. (2013). *Metode penelitian pendidikan*. Bandung: Alfabeta.
- Toendan, W. H. (2006). *Educational research methods*. Palangka Raya: University of Palangka Raya.
- Williams, C. (2007). Research methods. *Journal of Business & Economic Research*, 5(3), <http://www.cluteinstitute.com/ojs/index.php/JBER/article/view/253>. (30 March 2016).

Author's Brief CV

Aris Sugianto is a full-time lecturer of State Islamic Institute of Palangka Raya, Central Kalimantan, Indonesia. At 2014, he was a part-time lecturer of this institution. He has been as a full-time lecturer from 2015 until present. He earned his Undergraduate degree (2009) and Master degree (2013) in ELT from State University of Palangka Raya. His expertise is English Language Evaluation. Email: aris.sugianto@iain-palangkaraya.ac.id, HP: 085251179905.

