



THE RELIABILITY OF THE ENGLISH FINAL TEST FOR MANAGEMENT STUDENTS

Peni Astuti¹

¹Universitas Faletehan, Jl. Raya Cilegon KM. 06 Kramatwatu, Serang – Banten, Indonesia
Correspondent Email: peniastutiuf@gmail.com

ABSTRACT

This research aimed to determine the reliability of the final English test for the first semester of Management students of Faletehan University. The writer would like to know whether the final English test for the first semester of Management students was reliable. The writer's method in this research was a descriptive quantitative analysis method with Pearson's Product Moment Correlation technique to describe the reliability. The hypotheses that would be tested were: (1) Final English test for the first semester of Management students was reliable. (2) English final test for the first semester of Management students was unreliable. The research instrument was the parallel test of the final English test for the first semester of Management students. The respondent used was 28 students. After analyzing, the writer got $r_o = 0.87$ with the significant standard of 0.0500 and $r_t = 0.312$. Since $r_o 0,87 > r_t 0.312$, so H_a was accepted and H_o was rejected and the conclusion of this research states that the final English test for the first semester of Management students had been reliable.

Keywords: English final test, Management students, Person's Product Moment, Reliability, Validity

INTRODUCTION

Someone from the simple things always does evaluation or assessment to complicated one. Evaluation is critical to do to determine the value of an object in directing self-development. Indeed, evaluation or assessment is an activity to judge the object value, such as good-worst, effective-ineffective, successful-unsuccessful, etcetera. Evaluation can be defined as a systematic gathering of all relevant information necessary to promote the improvement of the curriculum and analyze its effectiveness for making the correct decision. There are some ways to evaluate. One of them is by giving a test. A lecturer carries out the test to get a score. She would like to use a test to provide information to help in making a decision. The lecturer has designed several tests to decide whether the teaching-learning process and the program have been successful. Ideally, to get accurate data, the test must be good and made as

well as possible. Thus, the test is one instrument of evaluation to collect the data or information. A test can be good if the items have three criteria; validity, reliability, and practicality. The test could be said poor if it does not fulfill those criteria. First, validity means that the test is officially acceptable. Second, reliability is the stability of measurement. The last one is practicality implies that the test must be practical.

From those criteria above, the writer only focuses on the reliability of this research. The reliability is related to the stability and the consistency of the test score. To know the reliability of a test, a test must be tested and analyzed, so the writer can get reliable and relevant information to make the correct decision. Based on the English Final test the writer got from the Department of Management Study Program, the final English test had already been administered to the first semester of Management students. Still,



the reliability of the test has never been analyzed before. Based on the fact above, the writer wants to make sure that the reliability of final English test for the first semester of Management students is reliable. In doing the research, the writer hopes that this study result will give the readers, teachers, institutions, and test makers valuable information about how to get the empirical evidence in determining the reliability of a test. From the research done, the writer gets an experience about how to do the reliability of a test. Thus she will be able to find out the reliability of a test better in the future.

This research aims to determine the reliability of the final English test for the first semester of Management students to answer the problem statement. Following the background of the study, it is clear that the reliability of the final English test in the first semester of Management students has not ever been analyzed yet. So, the writer is interested in finding whether it is reliable or not.

THEORETICAL FRAMEWORK

Evaluation is an intrinsic part of teaching and learning. Therefore, it has a crucial role in improving the teaching and learning activity. Evaluation has to expand importance in education as the lecturers collect data and form judgments activities designed to facilitate student learning. Based on the data, the lecturer can get information about how far the student can understand the subject learned and the success of the curriculum. Therefore, evaluation is an important activity to do by the lecturer. Evaluation is an activity that must be done accurately and effectively to make students achieve the instructional objectives. The amount or degree of student's changes in learning a subject can be known by doing an evaluation¹ According to the evaluation done, the lecturer can collect some evidence to decide the differences in student competence skill. Evaluation can be seen as the systematic process in gathering, analyzing, and presenting data to get wealth information about students and the instructional objectives, then the lecturer can create a correct decision. The lecturer usually holds a test to find out how much a student has achieved the subject matter given. The test result will provide valuable and valid information about student change after taking a course or program. Besides, the weaknesses and the strengths of the student and the instructional objectives can appear. Therefore, both teacher and student can correct the weaknesses and develop the

strength. An evaluation needs data collected through measure tools or instruments; one of them is a test. The majority of tests are used for making decisions about individuals. Thereby, an evaluation is a process, and a test is a tool or instrument.

A test is a group of assignments that students must do to get scores related to their behaviour and achievement². It can be analyzed that a test is a kind of assignment, which consist of questions or commands. Students do the test to get score about their behaviour and achievement. Test is a systematic procedure, a tool, an instrument for collecting information. Test also uses for measuring the student's ability, skill, knowledge, intelligence, and capacities. The test is an implement measured that has the shape of a series of task that must been done by the student in order to know the student result. Thus, the test has the dual function of measuring the students' capacity and measuring the success of the teaching program whether the students have mastered the program material comprehensively. The test as the instrument for measuring data must be good because it aims to know the weaknesses and strengths of students' competence in learning. A test is a method, instrument, or systematic procedure in collecting information to know a person's behaviour and describing it with the aid of a numerical scale or a category system. The test is also used for measuring a sample of behaviour, students' knowledge, skills, or abilities in a given domain. In constructing a good test, the lecturer should know the way to make a good test. A test called a good test has fulfilled the criterion of a good test, as follows;

1. Validity;

A test is valid if it tests what it is supposed to test. The test must be valid according to what ability the writer wants to measure. A test should only limit itself in measuring specific ability intended. In educational measurement, validity is often defined as the degree measuring instrument serves the purpose for which it is intended. Validity is the most essential characteristic of a measuring instrument.

2. Reliability;

A test must be reliable as a measuring instrument. A test should give the same results every time it measures. If the test is administered to the same respondents on different occasions and there is not differentiating in the result, it can be said that the test is reliable. The stability of a test score is consistent. In other words, the

measurement is not changing or firm. The test is reliable if its results are the same after being tested several times. Parallel tests are designed to know the reliability. It means that the correlation of the first test result must be significant to the second test result. After testing both tests and correlate the results, the test can be said reliable if both test results show a positive correlation and significance.

3. Practicality;

A third characteristics of a good test is its practicality. A test may be a highly reliable and valid instrument but still beyond our means of facilities. A test can be said practical if the tests are economical, has ease of administration and scoring. The lecturer can measure the reliability by several methods to measure the consistency of a measure with itself and compute a reliability coefficient. This coefficient will be correlated to measures the reliability of a test.

There are three general methods in estimating the reliability, but the writer only focuses on the parallel form method because this method using an alternative test, which is similar to the first test, which is given to the student to get the scores and correlate them. Therefore, parallel form is more effective than the other methods. The alternate test should be made as same as the first test given to the student. It should also be written in the exact specifications. Two parallel tests must have the same difficulty in passage and questions. The test must be identical in their material, length, number, difficulty, time, format, and etcetera. The correlation between the two parallel forms is the estimate of reliability. If the two sets of the tests are reliable, the two tests are equivalent or have the same characteristics and have the same functions.

Method

The test is administered at the University of Faletahan on Jl. Cilegon Raya KM.06 Pelamunan-Kramatwatu, Serang, Banten 42161 on October 2020 and re-test by using the parallel test on January 2021. The length time used for both the test lasted for 120 minutes. The respondent of this research is the first semester of Management students at Faletahan University Serang. The writer chooses all the Management students that contain 28 students as respondents. Variable X: The English final test scores for the first semester of

Management students administered on October 2020. Variable Y : The parallel test sores administered on January 2021. The research instrument is the second test or the parallel test sheet for the first semester of Management students of Faletahan University in the 2020/2021 academic year. The writer analyses the data scores by using a correlation technique, namely Pearson's Product Moment Correlation. The formula is :

$$r = \frac{n \cdot \sum XY - \sum X \cdot \sum Y}{\sqrt{\{n \cdot \sum X^2 - (\sum X)^2\} \cdot \{n \cdot \sum Y^2 - (\sum Y)^2\}}}$$

The hypotheses are; 1. The verbal hypotheses $H_i =$ final English test for the first semester of Management students is reliable. $H_0 =$ English final test for the first semester of Management students is unreliable. 2. The statistical hypotheses $H_i : \rho > 0$, $H_0 : \rho = 0$.

The procedures of doing the research are as follows:

1. Asking the head of the Management department permission to research the reliability of the final English test for the first semester of Management students.
2. Mapping the final English test.
3. Constructing the parallel test, this still holds on the standardized English final test that has been mapped.
4. Administering the parallel test.
5. Checking and giving a score to the parallel test.
6. Gathering and tabulating the variable data of X for the standardized test score and Y for the parallel test score.
7. Analyzing the pre-requirements: analyzing the normality, analyzing the linearity.
8. Analyzing the collected data using Pearson's Product Moment Correlation.
9. Testing the hypotheses.
10. Making conclusion or suggestions.

Result

The writer gives the report concerning the data description; the X is the English final test group of scores. The Y is the parallel-test group of scores.



1. The Standard Deviation of the X and Y group scores

NO.	COMPONENTS	SYMBOLS	FORMULAS	X	Y
1.	Number of Respondents	N	-	28	28
2.	Mean	\bar{X}	$\frac{\sum X_i}{N}$	68.4	72.45
3.	Varian	S^2	$\frac{n\sum X^2 - (\sum X)^2}{n(n-1)}$	132.553	119.484
4.	Deviation Standard	S	S^2	11.52	10.94

2. The frequency distribution of the X and Y group of scores

NO.	COMPONENTS	SYMBOLS	FORMULAS	X	Y
1.	Range	R	$H_s - L_s$	28	28
2.	Number of Class	Bk	$1 + 3.3 \log n$	6.29	6.29
3.	Class Interval	CI	$\frac{R}{Bk}$	7	7

Based on pre-requirement test;

1. Normality test using chi-square (χ^2), χ^2 count the opportunity $\beta = 0.95$ is 7.81. since χ^2 countless, so H_0 accepted. It can be concluded that distribution data of the English final test score for the first semester of Management students are normal.
2. The Parallel test score using chi-square (χ^2) count with opportunity $\beta = 0.95$ is 7.81. since χ^2 countless, so H_0 accepted. It can be concluded that the data of the parallel test score is normal.

The pre-requirement test; 1. Normality test: from the chi-square table (χ^2), χ^2 count the opportunity $\beta = 0.95$ is 7.81. Since χ^2 count less than χ^2 table, so H_0 accepted χ^2 account χ^2 table (χ^2 count $7.19 < \chi^2$ table 7.81). it can be concluded that distribution data of the final English test score for the first semester of Management students at Faletahan University is NORMAL. 2. Parallel-test score: from the chi-square table (χ^2), χ^2 count with opportunity $\beta = 0.95$ is 7.81. Since $\beta = 0.95$ count less than χ^2 table, so H_0 accepted χ^2 table (χ^2 count $5.01 < \chi^2$ table 7.81). it can be concluded that the distribution data of the parallel-test score is NORMAL.

Analyzing regression is a process to predict the value of variable Y (dependent) based on variable X (independent) to determine the regression similarity of the contribution of X to Y . It has been known that $n = 28$, $\sum X = 2736$, $\sum Y = 2989$, $\sum X^2 = 192312$, $\sum Y^2 = 214620$, and $\sum XY =$

202484. Constant $a = 16.07$, constant $b = 0.82$ so, $Y = a + bX = 16.07 + 0.82 X$.

Correlation analysis is done to get an index called coefficient correlation $r_o >$ from r_t so, H_0 is rejected (it means there is a relationship).

$$r = \frac{n \cdot \sum XY - \sum X \cdot \sum Y}{\sqrt{\{n \cdot \sum X^2 - (\sum X)^2\} \cdot \{n \cdot \sum Y^2 - (\sum Y)^2\}}}$$

$$= \frac{28 \cdot 202484 - 2736 \cdot 2989}{\sqrt{\{28 \cdot 192312 - (2736)^2\} \cdot \{28 \cdot 214620 - (2989)^2\}}}$$

$$= \frac{170432}{196325}$$

$$= 0.87$$

$$(0.87^2 \cdot 100)$$

$$= 75.69\%$$

Analyzing regression is a process to predict the value of variable Y (dependent) based on variable X (independent) to determine the regression similarity of the contribution of X to Y . It has been known that $n = 28$, $\sum X = 2736$, $\sum Y = 2989$, $\sum X^2 = 192312$, $\sum Y^2 = 214620$, and $\sum XY = 202484$. Constant $a = 16.07$, constant $b = 0.82$ so, $Y = a + bX = 16.07 + 0.82 X$.

The result of the correlation is got $r_o = 0.87$, so r_t which can be seen in the coefficient Pearson's Product Moment correlation in the appendixes is 0.312. Since r_o bigger than r_t ($r_o = 0.87 > r_t = 0.312$) so H_i is accepted and H_0 is rejected (it means there is a relationship). It can be concluded that there is a relationship between the English final test score



for the first semester of Management students in the 2020/2021 academic year with the parallel-test score.

It shows that $r_o = 0.87$, with the signification standard 0.0500, r_t is 0.312. Since r_o more than r_t ($r_o 0.87 > r_t 0.312$) so H_i is accepted and H_0 is rejected (there is a relationship). There is a relationship between the English final test score for the first semester of Management students of Faletahan University in the 2020/2021 Academic year with the parallel-test score. Therefore, it can be concluded that the final English test made by the English lecturer is reliable.

Conclusion

The writer would like to give some suggestions based on this research that might be useful (for the readers, institutions, and teachers), especially for the writer herself, and they are; 1. In measuring the reliability, the lecturer constructs a parallel test. In creating the parallel test, the lecturer should consider not only a standardized test but also parallel-test should be well prepared. 2. It is proved that the tests being analyzed are reliable. Which means that the two tests have fulfilled the characteristics of a good test. Thus, it is suggested for the English lecturer at Faletahan University of Management program study to keep both tests to be used as an exercise for the students before facing an exam.

REFERENCES:

- Alim, E. Astuti, P. 2011. The Influence of Student's Attitude toward ICT to their English Achievement. UHAMKA PRESS. Jakarta. One-day International Seminar on Policy, Practice, and Research in Technology Transformed Learning. <https://lemlit.uhamka.ac.id/index.php/berita/artikel-penelitian/88-30092011224151/the-influence-of-students-attitude-toward-ict-to-their-english-achievement>
- Airasian, P.W. 2012. Classroom Assessment; Concepts and Applications (7th ed.). New York: McGrawhills Companies.
- Arikunto, S. Jabar. C. 2004. Evaluasi Program Pendidikan. Jakarta: Bumi Aksara.
- Astuti, P. 2020. Analysis of Content Validity in English Examination Test on Public Health's Students. Journal of Industrial Engineering & Management Research, 1 (4), 100-113. <https://www.jiemar.org/index.php/jiemar/article/view/106>
<https://doi.org/10.7777/jiemar.v1i4.106>
- Brown, H.D. 2003. Language Assessment Principle and Classroom Practices. California: Longman.
- Fulcher, G. & Fred. D. 2007. Language Testing and Assessment; an Advanced Resource Book. Canada: Routledge
- Hughes, A. 2003. Testing for Language Teachers, (2nd ed.). UK: Cambridge University Press <http://www.statisticshowto.com/kuder-richardson/access> on Wednesday, 18 July 2018.
- Lynch, Brian. K. 2003. *Language Assessment and Programme Evaluation*. Great Britain: Edinburgh University Press. Ltd. Series Editors: Alan Davies and Keith Mitchell.
- Mc, Namara. 2000. Language Testing. Oxford: Oxford University Press.
- Reinard, John. 2001. *Introduction to Communication Research*. Singapore: McGraw Hill Companies, Inc.
- Sudjana, Nana. 2011. Penilaian Hasil Proses Belajar Mengajar. Bandung: PT. Remaja Rosdakarya.
- Sumarsono, Sigit. 2005. Statistika Pendidikan Bahasa. Jakarta: FKIP UHAMKA. EDISI V.
- Supranata, Sumarna. 2009. Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes. Bandung: PT. Remaja Rosdakarya.
- William, W., & Stephent G.J. 1990. Educational Meaurement and Testing. Boston: Allyn & Bacon.