ANALYSIS OF THE RELIABILITY OF THE ENGLISH SUMMATIVE TEST FOR THE SEVENTH GRADE STUDENTS OF SMP N 233 JAKARTA IN THE FIRST SEMESTER OF THE 2018/2019 ACADEMIC YEAR

A PAPER

Submitted as a Partial Fulfillment
The Requirements for Degree of
Sarjana Pendidikan

BY
SARAS WATI
1401065111

THE STUDY PROGRAMME OF ENGLISH EDUCATION
THE SCHOOL OF TEACHERS' TRAINING AND EDUCATION
THE UNIVERSITY OF MUHAMMADIYAH PROF. DR. HAMKA
JAKARTA
2020

VALIDATION SHEET

Title : **ANALYSIS OF THE RELIABILITY OF THE ENGLISH SUMMATIVE TEST FOR THE SEVENTH GRADE STUDENTS OF SMP N 233 JAKARTA IN THE FIRST SEMESTER OF THE 2018/2019 ACADEMIC YEAR**

Name : SARAS WATI

NIM : 1401065111

This paper has been presented, examined, and revised based on the advisors and examiners' suggestions

The Study Programme   : English Education

The Faculty   : The School of Teacher Training and Education

The University   : Muhammadiyah Prof. DR. HAMKA

The Day and Date   : Friday, 28th of August 2020

Validated by,		Signature	Date
Chairperson	: Drs. Zuhad Ahmad, M.Pd.		06/11/20
Secretary	: Silih Warni, Ph.D.		21/11/20
Advisor I	: Dr. Roslaini, M.Hum.		11/11/20
Advisor II	: Drs. Zuhad Ahmad, M.Pd.		06/11/20
Examiner I	: Drs. Susilo Wardoyo, M.Pd.		03/09/20
Examiner II	: Martriwati, M.Pd.		20/10/20
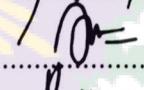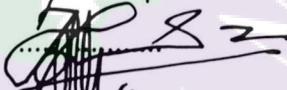
Dean,

Dr. Desvian Bandarsyah, M.Pd.

NIDN. 0317126903

i

ABSTRACT

SARAS WATI, NIM: 1401065111. *Analysis of the Reliability of the English Summative Test for the Seventh Grade Students of SMP N 233 Jakarta on the First Semester in the 2018/2019 Academic Year*. Paper: Jakarta: The Study Programme of English Education, The School of Teacher Training and Education, The University of Muhammadiyah Prof. DR. HAMKA, Jakarta 2020.

The objective of the research was to find the answer of the question of the problem: Does the English summative test for the seventh grade students of SMP N 233 JAKARTA in the First Semester of the 2018/2019 academic year have a good reliability?". This study was held in February 2019 with all of the seventh-grade students as the population and the 36 students in VII-D as the sample. This research used quantitative method, and the writer applied Pearson's Product Moment correlation analysis of the English summative test made by the English teacher and the Parallel test made by the writer. In order to collect the data, the writer took a set of English Summative test made by the English teacher. There were 50 items of the test.

The result of the analysis, the writer got r_o (r-observed) was 0.96 was higher than r_t (r-table) = 0.32. It can be concluded $r_o > r_t$ that H_o was rejected and H_i was accepted. It means that the English Summative Test for the Seventh Grade Students of SMP N 233 Jakarta in the First Semester of 2018/2019 Academic Year was **reliable**.

Key words : *Summative test, Reliability, and Parallel Test*

ii

TABLE OF CONTENTS

vi

CHAPTER I

INTRODUCTION

A. **The Background of the Problem**

Evaluation gives contribution directly to the learning and teaching process, especially in making decision. When teachers want to know about students' ability, about what they need to do, the teachers should apply an evaluation to get the data to analyze. According to Kember and Ginns, "Evaluation is introduced as the collection and analysis of data for the purposes of informing decision-making".[4] It means that evaluation is the key for making decision especially for learning and teaching process.

Teachers usually use a test for collecting data. Test is a kind of tools to measure how far the students have understood the materials that the teachers give and the students discuss in the classroom. In addition, there are several types of test, depending on : what is the purposes, when it is applied, or who makes the test. Based on the purposes, achievement test can be used to measure the students learning progress. Achievement test is divided into formative test and summative test. Formative test is a test given by the teachers to the students in the middle of semester or when the learning progress is still on-going. Summative test is to measure accumulative student's achievement and usually given at the end of semester.

As it has been explained above, when teachers need the accurate data, they need to make right decision. Therefore, the teachers should make sure that the test given has the standard of good test, such as validity, reliability and practicality. This idea is parallel with opinion of Harris who stated, "All good test posses three qualities:

---

[4]David Kember and Paul Ginns. 2012. *Evaluating Teaching and Learning.* New York: Routledge. p.1.

1

validity, reliability and practicality."[5] It means that a test called good test when they have three qualities in it.

Of the three main characteristics of a good test mention above, it seems that the reliability influences the validity. Although the test is valid, but it is not reliable, the validity will decrease. According to McMillan and Sally, "Reliability is the consistency of measurement the extent to which the result are similar over different forms of the instrument or occasions of data collection."[6] It means that a test is reliable if students are given the same test on two different occasions, and the test can show similar results of both.

Ideally the teacher should analyze and make a good test before it is given to the students. However according to writer's pre-survey in SMP N 233 JAKARTA, she found the English teacher of this school said that the summative test administered 2018/2019, including the test of the seventh grade students, has never been analyzed yet.

Based on the result of writer interview with one of the English teacher at that school, she assumed that the test has not been in good qualities. Therefore, the writer is interested in giving a little contribution to this school by analyzing reliability of the test. The writer limits the area in the matter of reliability. Then she takes the tittle *"Analysis the Reliability of the English Summative Test for the Seventh Grade Students of SMP N 233 JAKARTA in the First Semester of the 2018/ 2019 Academic Year."*

B. The Limitation of the Problem

---

[5]David P. Harris. 1969. *Testing English as a second Language.* New York: Mc Graw-Hill book Company. p. 13.

[6]James H. McMillan and Schumarcher Sally. 2006. *Research in Education* 6th edition. New York: Pearson Education Inc., p. 183.

3

There are 3 grades on SMP N 233 JAKARTA: the seventh grade, the eighth grade, and the ninth grade. In this research, the writer only focuses on seventh grade students of SMP N 233 JAKARTA the English Summative test to be analyzed in the test for the first semester 2018/2019 academic year.

C. The Question of the Problem

Based on the background of the study, the writer formulates the problem of the research as follow "Does the English summative test for the seventh grade students of SMP N 233 JAKARTA in the First Semester of the 2018/2019 academic year have a good reliability?"

D. The Objective of the Problem

The objective of the study is to get the answer based on the formulation research of problem whether the English summative test for the Seventh grade students of SMP N 233 JAKARTA in the first semester of the 2018/2019 academic year is reliable or not.

E. The Significance of the Problem

The writer hopes that this study can give the important information to the teacher, other readers, and especially for the writer. For teachers, to motivate them to make a good test. For the writer, to enrich the knowledge about the reliability of the test and how to measure the consistency of the exam scores which is always stable.

CHAPTER II

THEORETICAL FRAMEWORK

A. The Understanding of Evaluation

Evaluation is a major aspect in the school that can be used by teacher to judge their students achievement of teaching and learning process. There are some definitions about evaluation stated by some experts. The writer quotes three of them. According to Rogers and Badham, "Evaluation is the process of systematically collecting and analysing information in order to form value judgements based on firm evidence." [7] The key term in the definition is *collecting and analyzing information.* It means that evaluation is used for collecting data and analyzing the value of students' achievement in every semester, and for measuring a quality of teaching program in the classroom what is done.

Based on the statement above, evaluation is important. It can be defined as a systematic process of determining the extent to which students achieve instructional objectives. According to Bachman and Palmer, "Evaluation is a process that involves making value judgement and decision on the basis of information, and gathering information to inform such decisions is the primary purposes gor which language assessment are used."[8] It means that evaluation is an activity to get information that can be used for making decisions and

---

[7]Glyn Rogers and Linda Badham. 1992. *Evaluation in schools.* London and New York: Routledge, p. 2.

[8]Lyle F. Bachman and Adrian Palmer. 2010. *Language Assessment in Practice.* New York: Oxford University Press, p. 3.

4

providing value. By making some decisions, the teachers need evaluation to get the accurate data. It is used to be applied in the language assessment purposes.

Tyler perceived in Glasman and Nevo, "Evaluation is the process of determining to what extent educational objectives are actually being reliazed."[9] The keywords are *process* and *determining*. Based on the Cambridge Advance Learner's Dictionary, *process* is a series of things that are done in order to achieve and *determining* is to control and influence something directly or to decide what will happen. That means evaluation is a systematic process to stage of educationalist objectives are being reliazed.

Based on the explanations above, the writer sums up that evaluation in teaching is a systematic procedure to collect data or information for making decision concerning the students' learning progress.

B. The Understanding of Test

As it has been explained before evaluation is a process of giving value to something or to make a decision based on information or data. The information or data collected by teachers are commonly about the students's progress in a teaching learning process or the data about students strengths and weaknesses in following the lessons. For getting the data, the teachers need to

---

[9]Naftaly S. Glasman and David Nevo. 1988. *Evaluation in Decision Making.* Macmillan: Kluwer Academic Publisher, p. 33.

apply tests on one of various instruments which can help to measure students' ability.

There are some explanations about test. According to Reynolds, Livingston, and Wilson, "A test is a procedure in which a sample of an individual's behaviour is obtained, evaluated, and scored using standardized procedurs,"[10] It means that a test is a procedure used to measure the qualities of someone's quality.

Brown explained, "Test is a method of measuring person's ability or knowledge in a given area.[11] The keywords are *method of measuring* means the method to measure the students' ability and *given area* means a certain time that can be used to do the test. Thus, according to Brown, a test is a method to measure the students' knowledge in a certain time.

Osterlind stated, "A test item in an examination of mental attributes is a unit of measurement with a stimulus and a prespective form for answering."[12]

From the explanation above, the writer can summarize that a test is an instrument that measure a person's ability to get data.

C. Types of Test

---

[10] Cecil Reynolds, Ronald B. Livingston and Victor Wilson. 2009. *Measurement and Assessment in Education Second Edition.* New Jersey: Pearson Education Inc., p. 3.

[11] H. Douglas Brown. 1987. *Principles of language Learning and Teaching Second Edition.* New Jersey: Prentice Hall Regents. p. 252.

[12] Steven J. Osterlind. 2012. *Constructing Test Items.* Macmillan: Kluwer Academic Publishers, p. 20.

There are five types of test that may be used to measure student's achievement such as, aptitude test, proficiency test, placement test, diagnostic test, and achievement test.

1. **Aptitude Test**

   Aptitude test is designed to measure person's capacity or general ability to learn foreign language and to be succesful in that undertaking. It means that aptitude test is design to measure students' ability to learn foreign language.

2. **Proficiency Test**

   Proficiency test is not intended to be limited to any one course, curriculum, or single skill in the language. It means that proficiency test is designed to measure the students' level of language.

3. **Placement Test**

   Placement test is used to place a student an appropriate level or section of a language curriculum or school. It means that placement test is designed to the new students entering an educational institution to determine theirs knowledge to appropriate classes.

4. **Diagnostic Test**

   Diagnostic test can be used to expose the learners difficulties, the gaps in their knowledge, and the skill deficiencies during a course.

5. **Achievement Test**

An achievement test is designes to measure learners language skill progress in relation to the syllabus program they have been following.

D. The types of achievement test

1. Formative Test

Formative test is a kind of achievement test applied when the process of teaching and learning process is still on-going.

Based on the explanation above, the test formative is a test that is conducted at the end of material. The test has a function in order the teachers know how far the students understand about the material. If the students already understand about the material, the teacher can continue the next material, but if the students do not still understand about the previous material, the teacher will repeat the explanation clearly. This kind of test is also called as daily test.

2. Summative Test

Summative test conducted in the end of teaching to find out the whole teaching and learning activities in a period (semester). According to Rebecca, "Summative test is usually given at the end o making period and measure the "sum" total o the material covered."[13]

Based on the explanations the writer chose English summative test in doing this research because English summative test is hold in the last

---

[13] Rebecca M. Vallete. 1997. *Modern Language Testing Second Edition.* New York: Harcourt Brace Joanoich Inc., p. 11.

instructional program and it is used to put the lesson or to see the students'

ability of instructional objectives.

E. The Characteristics of a Good Test

1. Validity

According to Brown, "Validity is the degree to which the test actually measures what it is intended to measure."[14] In other words, validity requires the test to reflect the contents of the instructional materials. For example, a test of reading ability does in fact meaasure reading ability, and not some characteristics.

2. Reliability

According to Brown, "Reliability is consistent and dependable."[15] It means that the score of the test has to be relatively stable and consistent. In other words, the result of measurement has to remain similar eventhough the test is administered repeatedly to the same group of students as long as there is no treatment between the repeated test administered. For example, a reliable English test will accurattely measure English knowledge for every students who take it and reliable research findings can be replicated over and over.

3. Practicality

---

[14]*Ibid.,* p.387.
[15]*Ibid.*, p.386.

According to Brown, "A good test is practical. It is within the means of financial limitations, time constraints, ease of administrations, and scoring interpretation."[16] It can be known that practicality is about financial restriction, time of the test, and the ease of administration of the test itself.

F. The Understanding of Reliability

Reliability is one of the characteristics of a good test. According to Thorndike and Christ, "Reliability refers to the accuracy or precision of a measurement procedure."[17] It means that reliability is the accuracy of test scores in the procedure of easurement.

Wrightstone stated, "Reliability is an estimate of the degree of consistency or constancy among repeated measurement of individuals with the same instrument."[18] It means that reliability is the stability and consistently of test scores eventhough the test issued to measure repeatedly by teacher for students.

Meyer stated, "Reliability is the extent to which test scores are consistent with another set of test scores produced from a similar process."[19] The keywords are *test scores, consistent, scores of another tests.* It means that reliability is the test scores are consistent from a similar process.

From the explanation of the experts above, the writer sums up that reliability is instruments measured the relevant content which test scores are consistent. Thus,

---

[16]*Ibid.*, p. 386

[17]Robert Thorndike and Tracy T. Christ. 2010. *Measurement and Evaluation in Psychology and Education* 8th edition. Boston Pearson Education Inc., p. 118.

[18]Wrightstone, et al. 1996. *Evaluation in Modern Education.* New York : American Book Company, p. 46-47.

[19]J. Patrick Meyer. 2010. *Reliability.* New York: Oxford University Press., Inc., p. 9.

if the test is administered for the same students and different occasion, the score will be relatively similar.

1. Factors that affect the Reliability

In order to minimize the measurement error, teachers have to pay attention to the factors that affect the reliability. According to Uno and Koni, the factors that affect the reliability have five types, they are:

1) *Secara umum jika tes semakin panjang, maka semakin tinggi reliabilitasnya.*
2) *Penyebaran skor, semakin besar penyebaran skor, maka akan semakin tinggi perkiraan reliabilitasnya.*
3) *Kesulitan tes, umumnya tes yang terlalu mudah atau berjalan terlalu sulit akan menyebabkan reliabilitas semakin rendah. Hal ini disebabkan terbatasnya penyebaran skor,*
4) *Objektivitas tes, tes yang objektivitasnya tinggi memiliki reliabilitas yang tinggi, karena hasil tesnya tidak dipengaruhi oleh prosedur penskorannya,*
5) *Interval waktu tes, tes dengan interval waktu yang pendek menyebabkan koefisien reliabiltas tes yang besar.*[20]

Based on the explanation above, the writer summarizes that the factor that affect the reliability are : (1) Format; (2) Distribution; (3) Complication; (4) Objectivities, and (5) Time Interval.

2. The Method of Estimating Reliability

There are three methods of estimating reliability, they are : (a) test-retest (b) parallel test, and (c) split half.

a. Test-retest

---

[20]Hamzah B. Uno dan Satria Koni. 2012. *Asesmen Pembelajaran.* Jakarta: Bumi Aksara, p. 155-156.

In this method, the same individuals with the same test would be to retest. The result of the two administration are highly correlated. It can be assume that the test had temporal stability-one of the concept of reliability discussed.

b. **Parallel test (Equivalen forms)**

A second method is used of alternate is parallel test, with different versions of the same test which are equivalent in length, difficulty, time limits, formal, and all other such aspects.

c. **Split half**

Split half is another method for estimating the reliability of a test. According to Hopkins and Antes, "Split half is a measure of original score consistency of a test may be estimated by comparing two score divide of a single test."[21] It means that in using split half method, the single test scores is divided into two parts, they are odd-number and even-number score.

From those three methods, the writer chose parallel test to estimate the reliability of English summative test in order to get the empirical evidence whether or not the English summative test for the seventh grade students of SMP N 233 Jakarta in the first semester of the 2018/2019 academic year is reliable.

---

[21]Charles D. Hopkins and Richard L. Antes. *Op. Cit.*, p. 297.

REFERENCES

Bachman, Lyle F. & Adrian Palmer. 2010. *Language Assesment in Practice.* New York: Oxford University Press.

Brown, H. Douglas. 1987. *Principles of Language Learning and Teaching Second Edition.* New Jersey: Prentice Hall Regents.

Glasman, Naftaly S. & David Nevo. 1988. *Evaluation in Decision Making.* Macmillan: Kluwer Academic Publisher.

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

Kember, David & Paul Ginns. 2012. *Evaluating Teaching and Learning.* New York : Routledge.

McMillan, James H. & Schumarcher Sally. 2006. *Research in Education 6th Edition.* New York: Pearson Education Inc.

Meyer, J. Patrick. 2010. *Reliability.* New York: Oxford University Press Inc.

Osterlind, Steven J. 2012. *Constructing Test Items.* Macmillan: Kluwer Academic Publishers.

Reynolds, Cecil, & Ronald B. Livingston & Victor Wilson. 2009. *Measurement and Assesment in Education 2nd Edition.* New Jersey: Pearson Education Inc.

Rogers, Glyn & Linda Badham. 1992. *Evaluation in schools.* London and New York: Routledge.

Sumarsono, Sigit. 2014. *Statistika Terapan dengan Alur Penelitian Pendidikan Bahasa.* UHAMKA Press.

Thorndike, Robert & Tracy T. Christ. 2010. *Measurement and Evaluation in Psychology and Education 8th Edition.* Boston: Pearson Education Inc.

Uno, Hamzah B. & Satria Koni. 2012. *Asesmen Pembelajaran.* Jakarta: Bumi Aksara.

Vallete, Rebecca M. 1997. *Modern Language Testing 2nd Edition.* New York: Harcourt Brace Joanoich Inc.

Wrightstone, et.al. 1996. *Evaluation in Modern Education.* New York: American Book Company.

31