

**EVALUATING ENGLISH SUMMATIVE TEST FOR NINTH GRADE
STUDENTS OF SMPS LETJEN S PARMAN MEDAN
BASED ON ITEM ANALYSIS**

AN ARTICLE

**Submitted in Partial Fulfillment of the Requirements
for the Degree of Sarjana Pendidikan**

By:

LASTRI PRIMA SARI SIBUEA

Registration Number: 2133321055



**ENGLISH AND LITERATURE DEPARTMENT
FACULTY OF LANGUAGES AND ARTS
STATE UNIVERSITY OF MEDAN**

2020

ARTIKEL

**EVALUATING ENGLISH SUMMATIVE TEST FOR NINTH GRADE
STUDENTS OF SMPS LETJEN S PARMAN MEDAN
BASED ON ITEM ANALYSIS**

Disusun dan Diajukan oleh:

Lastri Prima Sari Sibuea

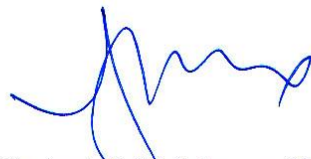
NIM. 2133321055

Telah diverifikasi dan dinyatakan memenuhi syarat
untuk diunggah pada jurnal online

Medan, Februari 2020

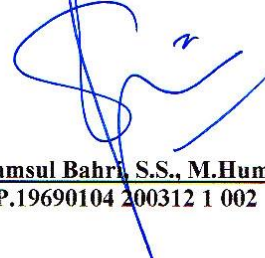
Menyetujui

Dosen Pembimbing I



Dr. Anni Holila Pulungan, M.Hum.
NIP. 19700522 200112 2 001

Dosen Pembimbing II



Syamsul Bahri, S.S., M.Hum.
NIP. 19690104 200312 1 002

Ka. Program Studi
Pendidikan Bahasa Inggris



Nora Ronita Dewi, S.Pd., S.S., M.Hum.
NIP. 19800522 200812 2 003

**EVALUATING ENGLISH SUMMATIVE TEST FOR NINTH GRADE
STUDENTS OF SMPS LETJEN S PARMAN MEDAN
BASED ON ITEM ANALYSIS**

* Lastri Prima Sari Sibuea

** Anni Holila Pulungan

** Syamsul Bahri

ABSTRACT

This research was aimed to explain the indices in evaluating English Summative test based on item analysis. In addition, this research was also aimed to find out the quality of the English Summative test for Ninth Grade Students of SMPS Letjen S Parman Medan based on item analysis, which covers item facility, item discrimination, and distractors efficiency. This research was descriptive-quantitative research. The data were taken from the English Summative Test, answer key, and the students' answer sheets. The data were collected by documentary technique. There were three indices in evaluating English test based on item analysis, they were: item facility, item discrimination, and distractor efficiency. There was 0% very difficult item, 76% moderate item, and 24% very easy item. Item with negative discrimination were 2.22%, zero discrimination were 6.67%, low discrimination were 20%, moderate discrimination were 60%, and high discrimination were 11.11%. Items with functioned distractors were 42% and items with non-functioned distractors were 58%. The quality of English summative test for ninth grade students of SMPS S Parman was moderate. There were 11.11% good items, 57.78% moderate items, and 31.11% bad items.

Keywords: *Test, Item Analysis, Item Facility, Item Discrimination, Distractors Efficiency.*

* Graduate

** Lecturer

I. Introduction

Students' ability as the result of the teaching-learning process is determined by a test. Test is used to determine how far the ability of the students is. A good test can determine the students' ability accurately. The quality of the test items influences strongly the adequacy of the overview of students' condition and the accuracy of the remedial actions needed to obtain maximum preparedness in the future. The quality of the test items can be identified by conducting evaluation. Evaluation has an important role in education as well as in teaching and learning process because by evaluation the outcome of the learning activity can be known, and from the outcome teacher can decide the follow-up that should be done. From evaluation teacher knows what things that need to be improved, so in the future, the teaching-learning process could be better. Teacher's ability in evaluation will be very beneficial when the teacher doing evaluation of students' learning output in the form of test.

A method that usually done to evaluate test items is item analysis. The quality of the test items can be determined by doing test item analysis based on the results of the test. Item analysis is essential in improving items which will be used again in later tests. It can also be used to eliminate misleading items in a test. Kubiszyn (2003:197) states that item analysis can be used to identify items that are deficient in some way, thus paving the way to improve or eliminate them, with the result being a better overall test. There are three characteristics that usually determined for a test and it can be found by analyzing it. First is item facility or item difficulty, which indicates how difficult or easy an item is. Second, item

discrimination which tells how well the items in separating the higher students to lower students. Third, distractors efficiency that indicates how effective each option for items. Unfortunately, teachers do not usually check on the effectiveness of their tests. This is probably because teachers (1) do not always understand the importance of accurate evaluation, (2) are not aware of the methods of analyzing tests, or (3) feel that test analysis is too time-consuming (Mehrens, 1991: 160).

Based on the writer's preliminary observation, the writer found that the summative test was arranged by the major teacher by herself and the test was never been analyzed before, so the quality of the test was never known. From the short interview, the writer also found that the teacher doesn't fully understand the importance of having a good quality test. Considering that issue, researcher was interested in evaluating the English summative test. The test will be evaluated based on item analysis referring to Brown (2003).

II. Theoretical Review

Fournier (2005:140) states "Evaluation is an applied inquiry process for collecting and synthesizing evidence that culminates in conclusions about the state of affairs, value, merit, worth, significance, or quality of a program, product, person, policy, proposal, or plan."

According to Cronbach (1960: 21), a test is a systematic procedure for comparing the behaviour of two or more person". Brown (2003: 3) defines test as *a method of measuring a person's ability, knowledge, or performance in a given domain.*

According to Brown (2003: 6) formative assessment is evaluating students in the process of “forming” their competencies and skills with the goal of helping them to continue that growth process. While summative assessment, according to Brown (2003: 6), aims to measure, or summarize, what a student has grasped, and typically occurs at the end of a course or unit of instruction.

Item analysis is the systematic evaluation of the effectiveness of the individual items on a test. This is usually done for purposes of selecting the “best” items that will remain on a revised and improved version of the test. Item analysis is performed simply to investigate how well the items on a test are working with a particular group of students (Brown, 1996: 50). Brown (2003) states “The appropriate selection and arrangement of suitable multiple-choice items on a test can be accomplished by measuring items against three indices: item facility, item discrimination, and distractor efficiency.”

Item facility, also known as item difficulty, is the extent to which the item is easy or difficult for the proposed group of test-takers. Item discrimination is the extent to which an item differentiates between high-and low ability test-takers. The efficiency of distractor is the extent to which the distractors “lure” a sufficient number of test-takers, especially lower-ability ones, and those responses are somewhat evenly distributed across all distractors.

III. Research Methodology

This research was conducted by using descriptive-quantitative approach with reason that the researcher tried to describe the quality of summative test by

analyzing the test items and the analysis will be dealing with number as well percentages. In quantitative data analysis numerical data are generated to represent the social environment, and statistical methods and deductive reasoning are utilized to analyze data (Castellan, 2010).

The data was the multiple-choice items of the English summative test, the answer key and the students' answer sheets. The test consists of forty five multiple-choice questions and five essays questions, but researcher only used the forty five multiple-choice questions as data. The data in this study were collected using documentation technique. The researcher used the method as stated by Brown (2003) to conduct item analysis. There are three points of analysis covering item facility, item discrimination, and distractor efficiency.

IV. Research Findings and Discussion

The Indices in Evaluating English Test based on Item Analysis

The distribution of items of the English summative test based on item facility can be seen as follows:

Item Facility	Item Number	Total	Percentages (%)
0.0-0.15 (very difficult)	-	0	0
0.16-0.85 (moderate)	4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 33, 36, 37, 40, 41, 42, 43, 44	34	76
0.86-1.0 (very easy)	1, 2, 3, 10, 15, 31, 34, 35, 38, 39, 45	11	24

Based on the Table, it was found that there was no item (0%) that are categorized very difficult. And 34 items (76%) out of 45 has moderate difficulty, which means they have good quality in terms of its facility. These items that have moderate difficulty can be saved in the item bank to be used in further test. Then, 11 items (24%) out of 45 are very easy, these items might need some revise or be observed what caused them easily answered by most of the students, before using it in the future, or should be discarded.

The distribution of items of the English summative test based on item discrimination can be seen as follows:

Item Discrimination	Item Number	Total	Percentages (%)
<0 (negative discrimination)	43	1	2.22
0.0 (zero discrimination)	2, 38, 39	3	6.67
0.0-0.20 (low discrimination)	1, 3, 6, 10, 15, 18, 21, 22, 31,	9	20
0.21-0.79 (moderate)	4, 5, 7, 8, 9, 11, 12, 13, 14, 16, 17, 19, 20, 23, 24, 25, 26, 27, 29, 30, 34, 35, 36, 37, 41, 42, 45	27	60
0.80-1.0 (high discrimination)	28, 32, 33, 40, 44	5	11,11

Based on the Table, it was found that there is 1 item (2%) that has negative discrimination, 3 items (7%) have zero discrimination, and 9 items (20%) have low discrimination. Items in these classifications should be discarded, because they have bad quality in terms of the discrimination. There are 27 items (60%) that have moderate discrimination, these items might need some revision or

reviews before using them in further test. Then, there are only 5 items (11%) have high discrimination. These items can be used in the further test.

The distribution of items of the English summative test based on distractor efficiency can be seen as follows:

Distractor efficiency	Item Number	Total	Percentages (%)
Functioned distractors	8, 9, 12, 13, 19, 21, 22, 23, 24, 25, 26, 28, 30, 32, 33, 36, 40, 41, 44	19	42
Non-functioned distractors	1, 2, 3, 4, 5, 6, 7, 10, 11, 14, 15, 16, 17, 18, 20, 27, 29, 31, 34, 35, 37, 38, 39, 42, 43, 45	26	58

Based on the Table, it was found that there are 19 items (42%) that have functional distractors, it means that every distractor successfully attracted students. While there are 26 items (58%) which the distractors were not functioned since they did not attract anyone. These distractors might need to be reviewed, revised, or even discarded.

The Quality of the English Summative Test

The distribution of items of the English summative test based on item facility, item discrimination, and distractor efficiency can be seen as follows:

Criteria	Item Number	Total	Percentages (%)
Good Item	28, 32, 33, 40, 44	5	11.11
Moderate Item	4, 5, 7, 8, 9, 11, 12, 13, 14, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 29, 30, 36, 37, 41, 42,	26	57.78
Bad Item	1, 2, 3, 6, 10, 15, 18, 31, 34, 35, 38, 39, 43, 45	14	31.11

Based on the table, it was found that there are 5 good items (11.11%). Those items can be kept in the item bank. There are 26 moderate items (57.78%) which need some revision or reviews before using them again in further test. There are 14 bad items (31.11%) that should be discarded.

The finding of this research showed that the English test was still lack in some ways. It is important to highlight the evaluation of a test, so teacher would be more concerned in arranging the best test since it will affect the result of the students' learning.

V. Conclusion and Suggestion

Conclusion

Based on the research findings found after analyzing the data, it can be concluded that:

1. Three indices in evaluating English test based on item analysis were item facility, item discrimination, and distractor efficiency. From the point of view of item facility analysis, it can be concluded that English summative test for ninth grade students of SMPS S Parman Medan in the academic year of 2018/2019 has good quality. Based on the result of the item discrimination analysis, the test has moderate quality. While referring to the distractor efficiency, the test has bad quality.
2. The quality of the English summative test for ninth grade students of SMPS S Parman Medan in the academic year of 2018/2019 based on three indices of item analysis is moderate. The test need significant revision or reviews before using it for further test.

Suggestion

Having conducted a research about evaluating an English test, it is expected to consider these following suggestions:

1. It is suggested to the English teacher to take into consideration how to arrange a good test. Teachers are expected to recognize their students' ability, so teacher can considerate whether an item would be too easy or too difficult for the students. It is also important to avoid ambiguity and miss-typed since it would affect the students understanding to answer the items.
2. It is suggested to the English teachers as the test maker at school to have the understanding of how to analyze the test. If teacher continuously evaluate their test, it will help to achieve a better test which also help to examine the students' ability more accurately.

REFERENCES

- Arikunto, Suharsimi. 2016. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara
- Bachman, L. F., & Palmer, A. S. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Boopathiraj, C & Chellamani, K. 2013. Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education. *International Journal of Social Science and Interdisciplinary Research*. Vol 2:2
- Brown., Frederick G. 1981. *Measuring Classroom Achievement*. New York: Halt Rich and Winston
- Brown, H. Douglas. H. 2003. *Language Assessment: Principles and Classroom Practices*. San Francisco: Longman

- Brown, James Dean. 1996. *Testing in Language Program*. New Jersey: Prentice Hall Regents
- Castellan, C. 2010. Quantitative and Qualitative Research: A View for Clarity. *International Journal of Education*. Vol 2:2. Macrothink Institute
- Cohen, Louis. 2007. *Research Methods in Education*. New York: Routledge
- Cronbach, L. J. 1960. *Essentials of Psychological Testing*. New York: Harper & Row.
- Ebel, R.L. 1972. *Essentials of Educational Measurement (1st Edition)*. New Jersey: Prentice Hall
- Fournier, D. M. 2005. Evaluation. In S. Mathison (Ed.). *Encyclopedia of Evaluation*. Thousand Oaks, CA: Sage
- Gronlund, Norman E. 1985. *Measurement and Evaluation in Teaching, Fourth Edition*. New York: Macmillan Publishing.
- Gronlund, Norman E. 1998. *Assessment of Student Achievement*. Boston: Allyn and Bacon.
- Haladyna, Thomas. 2004. *Developing and Validating Multiple-choice Test Items*. London: Lawrence Erlbaum Associates Publisher
- Henning, Grant. 1987. *A Guide to Language Tests*. London: Longman
- Hughes, Arthur. 2003. *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Kencana, Dila Puspa. 2015. *Test Item Analysis of Final Examination on Economic Subject in Grade XI IPS SMA Negeri 10 Yogyakarta Academic Year 2014/2015*. Skripsi. Yogyakarta: FE UNY
- Kubiszyn, Tom & Borich, Gary. 2003. *Educational Testing and Measurement: Classroom Application and Practice*. USA: John Wiley & Sons, Inc.
- Kumari, Amrita & Bhattacharya, S.B. 2016. Item Analysis of Diagnostic Test in English Language Skills of Secondary School Students. *IOSR Journal of Research & Method in Education*. Volume 6, Issue 4 Ver. I
- Mehrens, William A., Lehmann, Irving J. 1991. *Measurement and Evaluation in Education and Psychology*. Belmont CA: Wadsworth Thomson Learning.
- Mousavi, S. A. 2002. *An Encyclopedic Dictionary of Language Testing*. Taiwan: Tung Hua Book Company.
- Patton, M. Q. 2008. *Utilization-Focused Evaluation (4th Ed.)*. Thousand, Oaks, CA: Sage.

- Rajhy, Hussein Ahmed Abdo. 2014. Five Characteristics of a Good Language Test. *National Journal of Extensive Education and Interdisciplinary Research*. Volume II, Issue IV, Oct.-Dec., 2014, 61-66.
- Rao C, Kishan Prasad HL, Sajitha K, Permi H, Shetty J. 2016. Item Analysis of Multiple Choice Questions: Assessing an Assessment Tool in Medical Students. *International Journal of Educational and Psychological Researches*.
- Richards, J. C., J. Platt., H. Weber. 1985. *Longman Dictionary of Applied Linguistics*. Harlow: Longman.
- Risydah, Yunita. 2014. *An Analysis of Test Items of the English First-Term Test of the Seventh Grade Students of SMP Muhammadiyah 10 Yogyakarta in The Academic Year of 2013/ 2014*. Skripsi. Yogyakarta: FBS UNY
- Setiyana, Rusma. 2016. Analysis of Summative Test for English. *English Education Journal (EEJ)*. Vol 7:4, 433-447.
- Simanullang, Rosmelati. 2016. *Item Analysis of English Teachers' Made Mid Test at the Seventh Grade of SMP N 4 Medan*. Skripsi. Medan: FBS Unimed
- Stufflebeam, D. L., et al. 1971. *Educational Evaluation and Decision Making*. Bloomington, IN: Phi Delta Kappa.
- Toksöz, Sibel & Ertunç, Ayşe. 2017. Item Analysis of a Multiple-Choice Exam. *Advances in Language and Literary Studies*.
- Worthen, B. R., Sanders, J. R. 1973. *Educational Evaluation: Theory and Practice*. Worthington Ohio: Jones Publishing Company.