

**THE ITEMS ANALYSIS OF SUMMATIVE TEST
AT THE EIGHTH GRADE OF SMPN 12 LEBONG
IN THE ACADEMIC YEAR OF 2020/2021**

THESIS

Submitted as partial Requirements for Degree of Sarjana (S.Pd)
in English Study Program



BY :
RELIN KARLINA
NIM : 1711230069

**ENGLISH EDUCATION STUDY PROGRAM
TARBIYAH AND TADRIS FACULTY
STATE INSTITUTE FOR ISLAMIC STUDIES BENGKULU
2020/2021**



**KEMENTERIAN AGAMA
INSTITUT AGAMA ISLAM NEGERI (IAIN) BENGKULU
FAKULTAS TARBİYAH DAN TADRIS**

Alamat : JL.Raden Fatah Pagar Dewa Telp.(0736)15276, 51171 Fax (0736)511171 Bengkulu

ADVISORS SHEET

Subject : Relin Karlina

Nim NEE: 1711230069

**To : The Dean of Tarbiyah and Tadris Faculty
IAIN Bengkulu
In Bengkulu**

Assalamu'alaikum Wr. Wb

After reading throughly and giving necessary advices, herewith, as the advisors, we state that the thesis of :

Name : Relin Karlina

Nim NEE: 1711230069

**Title NEE: The Items Analysis Of Summative Test At The Eight Grade Of SMPN
12 Lebong In The Academic Year Of 2020/2021**

Has already fulfilled the requirements to be present before the board of examiners (munaqasah) to gain bachelor degree in english education. Thank you for the attention.

Wassalamu'alaikum Wr. Wb

Bengkulu, 17 July 2021

First Advisor

Second Advisor

Risnawati, M.pd

NIP.197405231999032002

Deti Lismayanti, M.Hum

NIP.197712222009012006



KEMENTERIAN AGAMA
INSTITUT AGAMA ISLAM NEGERI (IAIN) BENGKULU
FAKULTAS TARBIYAH DAN TADRIS
JURUSAN TADRIS

Alamat : Jl. Raden Fatah Pagar Dewa Telp (0736) 51276, Fax (0736) 5117 Bengkulu

RATIFICATION

This is certify the *Sarjana* thesis entitled "**THE ITEMS ANALYSIS OF SUMMATIVE TEST AT THE EIGHTH GRADE OF SMPN 12 LEBONG IN THE ACADEMIC YEAR OF 2020/2021**" by Relin Karlina has been approved by the Board of Thesis Examiners as the requirement for the degree of *Sarjana* in English Education Program.

Chairman
Dr. Adisel, M.Pd
NIP 197612292003121004

Secretary
Andriadi, M.A
NIP 198402212019031001

Examiner I
Deti Lismayanti, M.Hum
NIP 197712222009012006

Examiner II
Fera Zasrianita, M.Pd
NIP 197902172009122003

Bengkulu, August 2021
Approved by the Dean of Islamic and Tadris Faculty
Dr. Zibaedi, M.Ag., M.Pd
NIP 196903081996031005

PRONOUNCEMENT

Name : Relin Karlina
Nim : 1711230069
Study Program : English Program
Faculty : Tarbiyah and Tadris

I hereby sincerely state thesis entitled : **The Item Analysis of Summative Test at the Eight Grade of SMPN 12 Lebong in the Academic Year 2020/2021** is my masterpiece. All thing out my masterpiece in this thesis are signed citation and referred in the bibliography. If after proven that my theses discrepancies, I am willing to take academic sanctions in the form of repealing my thesis and academic degree.

Bengkulu, August 2021
State By,



Relin Karlina
NIM. 1711230069

ACKNOWLEDGMENTS

In the name of Allah the beneficent, the Merciful in the name of Allah the most Gracious and Merciful. Alhamdulillah, all praise be to Allah the most Gracious and Merciful, after such a hard work, finally the writer could finish research and finish the report in the form of the minor thesis (skripsi), entitled **“The Item Analysis of Summative Test at the Eighth Grade of SMPN 12 Lebong in the Academic Year 2020/2021.”**

The researcher realizes that she would not finish writing this “skripsi” without help from some people around her. Therefore, she would like to give her gratitude and appreciation to:

1. Allah SWT as the only one of my God, the researcher would like to say Alhamdulillah to Allah SWT, who has given me blessing, healthy, strongest and patience in finishing the thesis.
2. Prof. Dr. H. Siradjudin M.,M.Ag, M.H as Rector of the IAIN Bengkulu.
3. Dr. Zubaedi, M.Ag M.Pd as the Dean of Tarbiyah and Tadris faculty of IAIN Bengkulu.
4. Feny Martina, M.Pd as the Chief of English Study Program IAIN Bengkulu.
5. My advisor I Risnawati, M.Pd and my advisor II Mrs.Detti Lismayanti, M.Hum has guided me with patience.
6. My Father Bahder Winoto and My mother Ramyana, who always pray for my success. I want to express my deepest thanks for all of your love which always you give to me. I love you so much. You are my biggest motivation.
7. My beloved sister’s family. Thanks for motivation and the prayer you’ve given me.
8. I would like to dedicate this thesis to all my family who had supported me throughout the process. Thanks for my big family.

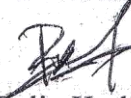
9. Special thanks for my lectures in IAIN Bengkulu

10. My best almamater, IAIN Bengkulu.

Finally the words are not enough to be expressed, except praise is to Allah the worlds for blessing and guidance, May their good deeds be accepted by Allah SWT. Aamiin..

Bengkulu, August 2021

The Researcher



Relin Karlina

1711230069

Name : **Relin Karlina**
NIM : **1711230069**
Study Program : **English Program**

ABSTRACT

This study is aimed to find out of quality summative test items on English Subject a SMPN 12 Lebong in academic year 2020/2021. This study was designed as descriptive quantitative research. Data collection techniques used documentation techniques. Data analysis in the form of item analysis was carried out using the Anates version 4.0.9 program and Marrie Tarrant's theory and Suharsimi Arikunto's theory. After analyzing the data the researcher found 40 multi-choice summative test items in English Subject a SMPN 12 Lebong in academic year 2020/2021 were reviewed from : a). The level of difficulty was 1 items (2.5%) difficult , 6 items (15%) moderate, and 33 items (82.5%) easy. b) The discrimination power was 2 items (5%) worst, 6 items (15%) poor, 8 items (20%) satisfactory, 8 items (25%) good, and 0 items (0%) excelent. c) The distractor efficiency was 44 item (22%) functioning distractor and 156 items (78%) non-functioning. Thus it can be concluded that summative test items on English Subject a SMPN 12 Lebong in the academic year 2020/2021 is not good. The implication of this study replaced with a new question, a question that is not good needs to be revised to be used again, and a good question can be entered into the question bank.

Keywords: *Analysis item, Summative Test*

Nama : Relin Karlina
NIM : 1711230069
Prodi : English Program

ABSTRAK

Penelitian ini bertujuan untuk mengetahui kualitas soal summative test pada mata pelajaran Bahasa Inggris di SMPN 12 Lebong tahun pelajaran 2020/2021. Penelitian ini dirancang sebagai penelitian deskriptif kuantitatif. Teknik pengumpulan data yang digunakan adalah teknik dokumentasi. Analisis data berupa analisis butir soal dilakukan dengan menggunakan program Anates versi 4.0.9. dan teori Marrie Tarrant dan teori Suharsimi Arikunto. Setelah menganalisis data peneliti menemukan 40 soal tes sumatif pilihan ganda pada mata pelajaran Bahasa Inggris di SMP N 12 Lebong tahun pelajaran 2020/2021 ditinjau dari : a). Tingkat kesukarannya adalah 0 item (2,5%) sulit, 6 item (15%) sedang, 33 item (82,5%) mudah, b) Daya pembedanya adalah 2 item (5%)sangat buruk, 22 item (55%) kurang, 8 item (20%) memuaskan, 8 item (20%) baik, dan item (0%) sangat baik. c) Efisiensi pengecoh adalah 44 item (22%) berfungsi dan 156 item (78%) tidak berfungsi. Dengan demikian dapat disimpulkan bahwa Soal Tes Sumatif Pada Mata Pelajaran Bahasa Inggris Di Smp N 12 Lebong Tahun Pelajaran 2020/2021 belum baik.Implikasi dari penelitian ini adalah soal yang kurang bagus harus diganti dengan soal yang baru, soal yang kurang bagus perlu direvisi untuk digunakan kembali, dan soal yang bagus bisa dimasukkan ke bank soal.

Kata kunci: Analisis butir, Tes Sumatif

TABLE OF CONTENT

COVER

ADVISORS SHEET	i
RATIFICATION.....	ii
PRONOUNCEMENT	iii
ACKNOWLEDGMENTS	iv
ABSTRACT	v
ABSTRAK	vi
TABLE OF CONTENT	vii
LIST OF TABLES	x
LIST OF CHART	xi
LIST OF APPENDICES.....	xii

CHAPTER I : INTRODUCTION

A. Background	1
B. Research Question.....	5
C. Objectives and Significances of the Research	5

CHAPTER II: LITERATURE REVIEW

A. Thereotical Framework	
1. Test	
a. Definition of Test	7
b. Kinds of Test	7
2. Summative Test	
a. Definition of Summative Test.....	8
b. Criteria of Good Test in Summative Test	8
a. Validity.....	9
b. Realibility	10
c. Practicality.....	10
d. Authenticity.....	11
e. Washback	11
c. Types of Test Item in Summative Test	12
a. Objective Test	12
1) True-False	12
2) Multiple Choice.....	13
3) Completion.....	14
4) Matching	14

b. Subjective Test.....	15
1) Essay Test.....	15
2) Types of the Essay Test.....	16
a. Extended Response Type	16
b. Restricted Response Type	16
3. Item Analysis	
a. Definition of Item Analysis	16
b. Kinds of Item Analysis.....	17
1. Level of Difficulty.....	17
2. Discrimination Power.....	18
3. Distractor Efficiency	19
c. The Importance of Item Analysis.....	20
4. Related Previous Study	21

CHAPTER III: RESEARCH METHODOLOGY

A. Research Design.....	24
B. Research Subject	24
C. Population and Sample.....	25
D. Research Variables and Indicators.....	25
E. Data Collecting Technique.....	25
F. Data Analyzing Technique.....	26

CHAPTER IV : RESULT AND DISCUSSION

A. Data Description.....	34
B. Data Analysis	39
C. Limitations of the Research	64

CHAPTER V : CONCLUSION AND SUGGESTION

A. Conclusion.....	65
B. Suggestion	65

REFERENCES

APPENDICES

LIST OF TABLES

Table 2.1 Classification Criteria Level of Difficulty	18
Table 2.2 Classification Criteria Discriminating Power	19
Table 3.1 Classification Criteria Level of Difficulty	27
Table 3.2 Classification Criteria Discriminating Power	28
Table 4.1 The Result of Item Difficulty	35
Table 4.2 The Result of Discriminating Power.....	37

LIST OF CHART

Chart 4.1 The Result of item difficulty	35
Chart 4.2 The Result of discrimination power	36
Chart 4.3 The Result of distractor efficiency	38

LIST OF APPENDICES

- Appendix 1 Soal tes sumatif Bahasa Inggris SMPN 12 Lebong
- Appendix 2 Lembar Jawaban Siswa
- Appendix 3 Kunci jawaban
- Appendix 4 The Result items difficulty di program anates
- Appendix 5 The Result discrimination power in program anates
- Appendix 6 The Data of Distractor efficiency

CHAPTER I INTRODUCTION

A. Background

The role of the school and the main teachers is to provide and facilitate students learning. The teacher must be able to generate activities that help students improve the way and results of learning. One effort is improving learning processes and results as part of improving the quality of education which can be done through a system of evaluation. Evaluation is an intrinsic part of teaching and learning. It is important for teachers because it can provide a wealth of information to use for the future direction of classroom practice, for planning of courses, and for the management of learning tasks and students.

The teacher carries out evaluation activities aimed at knowing the success level of the teaching program. However, besides that sometimes the teacher feels that evaluation is something that is contrary to teaching. This arises because it is often seen that the existence of evaluation activities is actually troubling and decreasing the arousal of learning in students. Indeed, evaluations carried out incorrectly can turn off students' enthusiasm for learning. Conversely, a well-conducted and correct evaluation should be able to improve the quality and results of learning because evaluation activities help teachers to improve the way they teach and help students improve their learning methods.

A good evaluation must help the students reach the goal as the core of the teaching and learning process. The evaluation given must pay attention to the quality of the test (evaluation tool) which is good. Because the impact of poor quality tests can affect the quality of education, even though the evaluation conducted in an effort to control the quality of education to make it better.

In addition, it will be an impact on trust in educational institutions and the credibility of teachers as evaluators. Assessment of learning outcomes by educators must be carried out continuously to monitor the

process, progress, and improvement of results in the form of daily exams, monthly exams, midterms, to final school examinations. It is used to assess the achievement of student competencies, the material for preparing progress reports on learning outcomes and improving the learning process.

Evaluation is one of method for a test. The test is a simple term a method of measuring a person's ability knowledge or performance in a given domain (Brown, 2003:3) . The test is to be valid if it measures accurately what is intended to be measured.

Tests made must be able to measure the ability of each student so that the test is made according to the criteria for making test questions. A good test as a measuring tool, if it meets the requirements of the test, which is to have: Practicality, Reliability, Validity, Authenticity, and Washback.

The results of tests obtained by students will be a mirror of whether or not the tests used. The description of the merits of a test can also be seen from the characteristics of the questions used. A good quality test will have good items. Items are the smallest unit that produces distinctive and meaningful information on a test or rating scale (Brown, 1996:49).

Therefore, after the tes process is carried out, a teacher needs to follow up by analyzing the results that have been obtained. In this case, an item analysis is carried out, which is a review of test questions in order to obtain a set of questions that have adequate quality. By analyzing these items, information can be obtained about the ugliness of a question and instructions for making improvements.

There are several items analysis, namely the analysis of the questionable item difficulty, discrimination, and distractor efficiency. of the question. Item facility (IF) is a statistical index used to examine the percentage of students who correctly answer a given item. Item discrimination is the extent to which item differentiates between high and low ability test takers (Brown, 2003:59). Distractor efficiency is a more

important measure of a multiple choice item's value in a test and one that is related to item discrimination (Brown, 2003:60).

In the research conducted by Khoshaim (2016) with the title "Assessment of the Assessment Tool: Analysis of Items in a Non-MCQ Mathematics Exam". The results of the study are as follows. The action research examines the effectiveness of an assessment process and inspects the validity of exam questions used for the assessment purpose. Using the data from 206 students, the researchers analyzed 54 exam questions with regard to the complexity level, the difficulty coefficient and the discrimination coefficient. Findings indicated that the complexity level correlated with the difficulty coefficient for only one of three semesters.

In addition, the correlation between the discrimination coefficient and the difficulty coefficient was found to be statistically significant in all three semesters. The results suggest that all three exams were acceptable; however, further attention should be given to the complexity level of questions used in mathematical tests and that moderate difficulty level questions are better classifying students' performance.

In addition, the correlation between the discrimination coefficient and the difficulty coefficient was found to be statistically significant in all three semesters. The results all three exams were acceptable; however, further attention should be given to the complexity level of questions used in mathematical tests and that moderate difficulty level questions are better classifying students' performance. Research conducted by Khoshaim and researcher have similarities, namely both of them analyze items at the junior high school level. Meanwhile, the difference in this study analyzes the items about English.

Based on a preliminary study conducted by researcher at SMP N 12 Lebong, the researchers conducted an interview English teacher. He explained that In the process of evaluating learning outcomes, in SMP N 12 Lebong uses the Summative Test as an evaluation tool to find out and measure the level of student learning outcomes, where summative test

questions are prepared by MGMP (Musyawarah Guru Mata Pelajaran) which is consisted of 40 items of multiple choice questions.

Evaluation during this was carried out not paying attention to the assessment of the item so that the quality of the items tested was not known whether it included items that met the requirements as a good measure or not. the analysis of these items needs to be done to obtain information about the quality of the test so that the quality of the test can be improved.

Whereas the results of the evaluations conducted also turned out that there were still quite a number of students who received grades below the assessment standard 7,5. Therefore, this result is not good enough to influence the research of the items that have been prepared by the MGMP. Suspicion of the unsatisfactory results of the test on the item is very important to find out whether the test item has been entered in the test items that qualify as a good measuring instrument or not.

The importance of evaluation in improving the quality of the process and learning outcomes, a teacher as part of the implementation of the teaching and learning process is required to carry out a good evaluation and automatically required to make a good test or evaluation tool. In this case, educators are required analysis of items to find out the quality level of the items used.

The purpose of analyzing the items to identify the good, average, and bad questions. good item can be saved and reused, poor items should be revised, and items that are not good should be discarded. Analysis of item questions is done so that the questions made by the teacher are a matter of quality, so they can measure the learning outcomes of students. If the question used to measure the learning outcomes of students is of doubtful quality, the results of the test may also not be of quality.

Test results from non-quality questions cannot reflect the learning outcomes of students because the test will provide less precise information about student learning outcomes. Analysis of items was carried out by

calculating aspects of item difficulty, discrimination, and distractor efficiency. Analysis of items can be done using the help of a computer device with software or manually. In this study, researchers used the ANATES Version 4.9 program.

Based on these problems, researchers are interested in conducting research with the title "**The Items Analysis of Summative Test at the Eighth Grade of SMPN 12 Lebong in the Academic Year 2020/2021**".

B. Research Question

Based on the background described above, the main problem in this study at the eight grade students at SMP Negeri 12 Lebong in academic year 2020/2021 are formulated as follows:

1. How is the difficulty level of English summative test?
2. How is the discriminating power of English summative test?
3. How is the distractor efficiency of English summative test?

C. Objective and Significances of the Research

a. Objective

Based on the research question above, the research objectives of this research are :

1. To find out the item difficulty of English summative test at SMPN 12 Lebong in academic 2020/2021.
2. To find Out the discrimination power of English summative test at SMPN 12 Lebong in academic 2020/2021.
3. To find Out the distractor efficiency of English summative test at SMPN 12 Lebong in academic 2020/2021.

b. Significances

The first of all, the results of this study can be used by the teacher as a tool to determine whether or not learning objectives have been achieved and to improve the next evaluation tool. Form the input to the teacher so that the questions are more qualified and the teacher is able

to carry out item analysis to improve quality test and the researcher hopes that the result of the item analysis can be benefit for English teachers or the test makers to improve their competences to make a good test items and also to improve their teaching and learning in the classroom. Finally, the study can be used as a reference and consideration related to the same topic for the next researchers.

CHAPTER II LITERATURE REVIEW

A. Theoretical Framework

1. Test

a. Definition of Test

Assessment is an activity done by teachers to obtain information about the knowledge, skills and attitude of students. This action can involve the collection of formal assessment data as an objective test or informal assessment data as an observation checklist (Marsh, 2010:311). In doing the assessment, there is a long process in collecting, synthesizing and interpreting the information in which testing, measurement and evaluation take parts.

A test is a method to measure the ability of a person in knowledge or performance in a given program (Brown, 2003:3). A test is a formal, systematic procedure for gathering information (Airasian & Russel, 2008:9). A test also created as a device to monitor the development of the program, to diagnose the difficulties in the program and to measure the performance of the test taker in and at the end of the program (Sudaryono, 2012:102).

Based on those opinion above, the writer conclude the test as a method to monitor the development of the program in teaching-learning process, diagnose the difficulties during and at the end of the program by measuring the ability of test taker intelligence both in knowledge and or performance. Test also helps the teacher to notice which students have the same level ability in some subject.

b. Kinds of Test

There are many kinds of test ranging from a diagnostic test, achievement test, proficiency test to aptitude test. The test that mainly directed with teaching material in the classroom called as

an achievement test. Achievement tests have been created primarily to help teachers answer questions regarding the present status of pupils (Wick & Beggs, (1971:9). There are two categories of achievement test; formative and summative test.

Formative evaluation takes place when the program is still in progress that implies that the students being taught could have benefits from insights gained (Allison, 1999:65). Formative test is important since it can provide data so; the changing are possible to be made. Whereas, a summative test which at the same time is also called as terminal evaluation occurs at the end of the program when all of the experiences are being summed up. A test like summative test, which is used in a larger group of students, is suitable for use norm-referenced test because the main purpose for this measurement is to compare the ability of each student with his or her classmates.

2. Summative Test

a. Definition of Summative Test

Many experts have stated about some definitions of the test, according to Nitko (1983:6) in his book, *Educational Tests and Measurement: An Introduction*, he writes test is defined as “a systematic procedure for observing and describing one or more characteristics of a person with the aid of either a numerical scale or category system.”

Another opinion, test is a technique or way consisting of some questions, statements, or tasks that are delivered to students in term of measuring their performance or behavior (Arifin, 2009:118). To support, Tinambunan (1988:3), said that “a test is a set of questions, each of which has a correct answer, that examinees usually answer orally or in writing.”

Based on the definitions above it can be concluded that, test is a utility to collect the information about students' performances in term of a set of some items such as questions, statements in orally or writing by using category system such as a scoring.

Talking about the summative test, actually it is one of the types of the achievement test. The achievement test itself including in the test which can determine the student's competent grades in the past of the learning activities in the classroom. The type of the achievement test which often used by the teacher to evaluate the successfulness his or her teaching and learning in the classroom is summative test.

According to Wilmar Tinambunan (1988:9) in his book, "the summative test is intended to show the standard which the students have now reached in relation to other students at the same stage. Therefore it typically comes at the end of a course or unit of instruction."

To support the statement above, the summative test is given at the end of a marking period and measures the "sum" total of the material covered. On this type of a test, students are usually ranked and graded (Valette, 1977:11). The summative test is given in the end because it comes at the end of a course (or unit) of instruction. It is designed to determine the extent to which the instructional objectives have been achieved and is used primarily for assigning course grades or certifying pupil mastery of the intended learning outcomes (Groundlund, 1985:12).

It means that summative test is tested in the end of a course to know the students' competence from all the materials which have been taught by the teacher.

b. Categories of Good Test in Summative Test

Actually, the summative test has to represent all the materials which have been taught by the teacher. So, the teacher

should have some criteria of good summative test. Those are validity, reliability, practicality, authenticity, and washback.

a. Validity

- An effective test is practical. This means that it
- Is not excessively expensive,
- Stays within appropriate time constraints,
- Is relatively easy to administer, and
- Has a scoring/evaluation procedure that is specific and time-efficient.

A test that is prohibitively expensive is impractical. A test of language proficiency that takes a student five hours to complete is impractical-it consumes more time (and money) than necessary to accomplish its objective. A test that requires individual one-on-one proctoring is impractical for a group of several hundred test-takers and only a handful of examiners. A test that takes a few minutes for a student to take and several hours for an examiner too evaluate is impractical for most classroom situations.

b. Reliability

A reliable test is consistent and dependable. If you give the same test to the same student or matched students on two different occasions, the test should yield similar result. The issue of reliability of a test may best be addressed by considering a number of factors that may contribute to the unreliability of a test. Consider the following possibilities (adapted from Mousavi, 2002, p. 804): fluctuations in the student, in scoring, in test administration, and in the test itself.

c. Practicality

By far the most complex criterion of an effective test-and arguably the most important principle-is validity, “the extent to which inferences made from assessment result are appropriate,

meaningful, and useful in terms of the purpose of the assessment” (Ground, 1998:226). A valid test of reading ability actually measures reading ability-not 20/20 vision, nor previous knowledge in a subject, nor some other variable of questionable relevance. To measure writing ability, one might ask students to write as many words as they can in 15 minutes, then simply count the words for the final score. Such a test would be easy to administer (practical), and the scoring quite dependable (reliable). But it would not constitute a valid test of writing ability without some consideration of comprehensibility, rhetorical discourse elements, and the organization of ideas, among other factors.

d. Authenticity

An fourth major principle of language testing is authenticity, a concept that is a little slippery to define, especially within the art and science of evaluating and designing tests. Bachman and Palmer (1996:23) define authenticity as “the degree of correspondence of the characteristics of a given language test task to the features of a target language task,” and then suggest an agenda for identifying those target language tasks and for transforming them into valid test items.

e. Washback

A facet of consequential validity, discussed above, is “the effect of testing on teaching and learning” (Hughes, 2003, p. 1), otherwise known among language-testing specialists as washback. In large-scale assessment, wasback generally refers to the effects the test have on instruction in terms of how students prepare for the test. “Cram” courses and “teaching to the test” are examples of such washback. Another form of washback that occurs more in classroom assessment is the

information that “washes back” to students in the form of useful diagnoses of strengths and weaknesses. Washback also includes the effects of an assessment on teaching and learning prior to the assessment itself, that is, on preparation for the assessment.

c. Types of Test Item in Summative Test

Besides the categories of the summative test, in that test there are some kinds of the types of the test items. Those are:

a. Objective Test

This test can be called objective if the answers of the test have only one the correct answer as a key of the test item. In this test the students have choose one correct answer from some choices which provided by the teacher. In giving the scoring of objective items requires much less time than the scoring of essay items. With its objectivity, objective items can be accurately scored with little if any dispute about the correctness of response.

According to Zainal Arifin’s (2009:135) book, “There are many varieties of there new test, but four kinds are in most common use, true-false, multiple-choice, completion, matching.”

1) True-False

Based on Jum C. Nunnally (1964:160), “The popularity of the true-false item is probably due to the ease with which such items can be composed. It is usually easy to make up many such items in a relatively short period of time.” Besides that, James Dean Brown and Thom Hudson (2002:66), in their book “requires student to respond to the language by selecting one of two choices, for instance,

between true and false or between correct and incorrect.”

To sum up, in true-false the students are able to answer the statement with true or false by short of time. Then the function of this test is to measure the competence of the student to differentiate between the fact with the opinion. In addition, the teacher not only provides question or statement in this test but also possible to give the picture, diagram, or table.

2) Multiple-Choice

The multiple-choice is the most popular in types of test because of the multiple-choice often is used in many kinds of objective test. To support the statement above, William Wiersma and Stephan.G.Jurs (1976:169) states in his book, that “by far the most popular type of objective item is that in which the student is required to choose one alternative response to a problem or question.”

An additional, “A multiple-choice item is an item that presents a statement (called the stem) and the student is required to select one of two or more (usually more) options that correctly completes the statement or correctly answers the problem posed in the statement.” Similarly, a multiple-choice item consists of one or more introductory sentences followed by a list of two or more suggested responses from which the examinee chooses one as the correct answer (Nitko, 1983:190).

Above all, it can be concluded that in multiple-choice item consist of two parts, the question or statement it can be called as a stem and some choices which included as the correct answer, the most correct answer and the distractors,

or it can be called as option. The form of the option is possible consist of words, numbers, or statements.

3) Completion

The completion item is a written statement which requires the examinee to supply the correct word or short phrase in responses to an incomplete sentence, a question or a word association.

Actually, this item effectively to test the students' knowledge such as the definition, names of country, and so on (Tinambunan, 1988:61). To support, "Usually, completion items require the testees to supply a word or a short phrase (Heaton, 1988:209)." An example, in reading text of the summative test the teacher provides incomplete sentence or statement then the students have to fill the correct answer in the blank of that sentence or statement.

4) Matching

The matching item commonly appears in a two-column-format although variations on this format can be used. The two columns of a matching item are commonly called the premises and responses. Matching items lend themselves well to testing a knowledge of relationships or definitions (Wiersma,1976:209).

It means, there are many kinds of matching form, it begins from the premises it can be list of definitions, antonym, or synonym then the responses consist of the list of choices of the best or the appropriate answer. Usually, the answers consist more than the questions.

Besides that, Jum C. Nunnally in his book, he states that, "Students are asked to write in the blank space the letter corresponding to the option on the right which matches the item on the left. The major advantage of the

matching item is that a considerable amount of material can be presented in a short space.”

b. Subjective Test

In subjective test, the students have to answer the question based on their knowledge which have they got using their words to their writing. Talking about giving score of the subjective test, the teacher will score the answer based on the students’ answering, whether it is simple or complex answering and of course it depends on the teacher’s subjectivity. According to Arthur (1989:19), in his book, “If judgment is called for, the scoring is said to be subjective.”

Also, the students’ answering not only focuses of true or false answer but also it depends of complete or incomplete answer. At this point, the teacher also will know the competence of their students from as far as they mastered the materials which have given by the teacher.

1) Essay Test

The type of essay item the students supply their answer rather than choose the correct answer. To support, “the essay-type question requires the examinee to read the question, formulate his response and express the response in his own words.” It means that the students are given a freedom to express their idea to answer the question.

In addition, J. Stanley Ahmann and Marvin D. Glock (1967:157), “an essay test item demands response composed by the pupil, usually in one or more sentences, of a nature that no single response or pattern of responses can be judged subjectively only by one skilled and informed in the subject, customarily the classroom teacher.”

Furthermore, the essay test usually consists of questions beginning with or including such directions as discuss,

explain, outline, evaluate, define, compare, contrast, and describe (Nol 1965:131).

2) Types of the Essay Test

Based on the amount of freedom of response, the essay test can be divided into two forms, those are:

a. Extended Response Type

In this type, the students have a freedom to express their argument which based on their competence, so they can begin from the definition first then the example and the opposite, on the other hand, it can be called deductive and inductive text. But the teacher also has the criteria of the scoring of the students' answering which based on the question.

According to Wilmar, "In the Extended Response Type Test, the pupil is given almost complete freedom in making his response."

b. Restricted Response Type

In the Restricted Response Type, the test item provides the specific problems and more structure. Although, they have a freedom to supply their ideas into their writing but the students have to answer systematically based on the question.

In addition, "The restricted response question usually limits both the content and the response. The content is usually restricted by the scope of the topic to be discussed. Limitations on the form of response are generally indicated in the question".

3. Item Analysis

a. Definition of Item Analysis

The item is the smallest unit that produces typical and meaningful information on the testing or ranking scale (Brown,

1996:49). Item analysis is one systematic procedure designed to obtain specific information about each test item. It is designed primarily for use with the objective test. In item analysis, the test conductor is concerned with item, difficulty level, the discriminative power of the item and effectiveness of the distracters (Evrero & Sylvanus, 2015:18).

Item analysis is the validation of a MCQs after it has appeared in a question paper (Kashyap, 2015). According Bajpai (2017:117) Item analysis is carried out to see if the items in the instrument belong there or not. Each item is examined for its ability to discriminate between those subjects whose total scores are high and those with low scores.

From the above definitions, it can be concluded that Analysis of the item is an analytical activity to determine the level of goodness of the items contained in a test so that the information generated can be used to improve the items and tests.

b. Kinds of Item Analysis

1. Level of Difficulty

The analysis of the difficulty level or Facility Value (FV) is the first step in analyzing the test items. Heaton (1988) stated that the FV of an item shows the difficulty of an item in a test. It shows which item is easy or difficult. The FV can be known from the ratio or percentage of students who answer the item correctly.

This research uses Heaton's (1988) formula to measure the FV which is gained by dividing the number of students from the upper group and the lower group students who answer a certain item correctly by the total number of the students who join the test.

The formula of the difficulty level of each item in large group is stated below:

$$FV = \frac{\text{Correct } U + \text{Correct } L}{2n} \quad \text{or} \quad FV = \frac{R}{N}$$

In which:

- F : The index of difficulty
- R : The number of correct answers
- N : The number of students taking the test
- U : Upper half
- L : Lower half
- n : Number of candidates in one group

The scale of the difficulty level of all test items ranged from 0.00 to 1.00 (Arikunto, 2006:208-210). It can be interpreted in the rank scale of difficulty level, as follow :

Table 2.1
The Clasification Criteria of Level of Difficulty

P	Interpretation
< 0.30	Difficult
0.30 – 0.70	Moderate
> 0.70	Easy

Source : Suharsimi Arikunto

From the rank scale above, it shows the easiness and the difficultness of test items. So, the teacher will know the difficulty level of each test item in the summative test.

2. Discriminating Power

The next step of item analysis is to determine the Discriminating Power (DP) that is whether the item can discriminate the students of the upper group from those in the lower group.

Heaton (1988:180) elaborates, “The index of discrimination (D) tells us whether those students who performed well on the whole test tended to do well or badly on each item of the test.”

This research uses Heaton’s (1988) formula of DP.

$$D = \frac{U - L}{N}$$

Source : J.B. Heaton

In which:

D: The index of item discriminating power

U: The number of pupils in the upper group who answered the item correctly

L: The number of pupils in the lower group who answered the item correctly

N: Number of pupils in each of the groups

Table 2.2

The Classification Criteria Of Discriminating Power

Index of Discriminating Power	Classifications
0.70 -1.00	Excellent
0.40 – 0.70	Good
0.20 – 0.40	Satisfactory
≤ 0.20	Poor
Negative value on D	Very Poor

3. Distractor Efficiency

To diagnose the problem shown in difficulty and discrimination index, the proportion of student choosing the distracter can also be calculated to assess how the distracters are functioning. It can be compared with the proportion of students choosing the correct response. The distracter is considered as effective if it is chosen by the most of students in a lower group. If the distracter is mostly chosen by the upper

group, it can be said that the distracter did not function as it should be.

One of the objectives of item analysis is to know about the answer distribution to a subject in alternative answers. Through distracter efficiency, teacher may know the number of students who answered correctly, which distracter is too showy and make it easier for students not to vote, the misleading distracter and the distracter who managed to attract lower group students.

Distractor is considered as an effective (functioning) if it is selected at least by 5% of examinees, and if it is chosen by less than 5% of examinees, it means that the distracter is categorized as an ineffective (non-functioning) distracter. This principle is based on Marrie Tarrant's theory and Suharsimi Arikunto's theory.

c. The Importance of Item Analysis

Talking about the importance of the item analysis, it is very important for the teacher to get much information from the result of the analyzing the items. By the analyzing, the teacher will know about the achievement of his or her teaching and learning process in the classroom. Not only the process itself but also the teacher knows the students' performances. It means that, there are so many benefits of the item analysis.

Based on Anthony J. Nitko (1983:284-286), there are six points of the importance of item analysis, those are:

1. "Determining whether an item functions as the teacher intends." In this book, there are five points about deciding the functioning items which teacher needs to consider: (a) whether it seems to be testing the intended instructional objective, (b) whether it is of the appropriate level of difficulty, (c) whether it is able to distinguish those who have command of the learning objectives from those who do not, (d) whether the

keyed answer is correct; and (e) (for response-choice items) whether the distractors are functioning.

2. “Feedback to students about their performance and as a basis for class discussion.” The teacher gets information about the performance of their students. For example in reviewing the test, the teacher will know the students’ errors.
3. “Feedback to the teacher about pupil’s difficulties.” In item analysis the teacher will know the students’ problem, it can be seen from the students’ errors of their answering and the teacher is suggested to take the score more than one item such as group and individual test in all the topics which has taught.
4. “Areas for curriculum improvement.” From this statement, item analysis will give information to conduct the revision of the curriculum.
5. “Revising the items.” By doing the item analysis, teacher can revise the item from students’ responses of each item if there are some items which are revised it just need a few time to revise than make the new item and the items can be used for the next testing.
6. “Improving item-writing skills.” The teacher should have item-writing skill in making items in the test and the way to improve the item-writing skill is to analyze the item from students’ responses.

B. Related Previous Studies

Some previous studies about the analysis of items including research written by Heba Bakr Khoshaim (2016) with the title “Assessment of the Assessment Tool: Analysis of Items in a Non-MCQ Mathematics Exam”. The results of the study are as follows The reported action research examines the effectiveness of an assessment process and inspects the validity of exam questions used for the assessment purpose. The instructors of a college-level mathematics course studied questions used in

the final exams during the academic years 2013–2014 and 2014–2015. Using the data from 206 students, the researchers analyzed 54 exam questions with regard to the complexity level, the difficulty coefficient and the discrimination coefficient. Findings indicated that the complexity level correlated with the difficulty coefficient for only one of three semesters. In addition, the correlation between the discrimination coefficient and the difficulty coefficient was found to be statistically significant in all three semesters. The results suggest that all three exams were acceptable; however, further attention should be given to the complexity level of questions used in mathematical tests and that moderate difficulty level questions are better classifying students' performance. There are some studies about the difficulty level and the discriminating power which had conducted by several researchers.

The second study is conducted by Marrie Tarrant, James Ware and Ahmed M Mohammed. This study was taken at Department of Nursing Studies, Li Ka Shing Faculty of Medicine at 2009. The title of the study is "An Assesment of functioning and non-functioning distractors in multiple-choice questions." The purpose of this study was to examine non-functioning distractors in teacherdeveloped tests to provide additional guidance to teachers in health-science disciplines regarding the optimal number of distractors to include in MCQs. This study uses item-analysis data, they assessed the proportion of non-functioning distractors on seven test papers administered to undergraduate nursing students. They reviewed 514 items, 1542 distractors and 514 correct responses. Nonfunctioning options were defined as ones that were chosen by fewer than 5% of examinees. The study shows a result that the proportion of items containing 0,1,2 and 3 functioning distractors was 12.3%, 34.8%, 39.1% and 13.8% respectively. Overall, items contained an average of 1.54 (SD = 0.88) functioning distractors. Only 52.2% (n=805) of all distractors were functioning efficiently and 10.2% (n=158) had a choice frequency of 0. The result showed that the majority of items developed by teacher will not

have more than two functioning distractors, and including more distractors may not be a good investment in item development. Three option-items have many advantages for both item writers and examiners, while further non-functioning distractors are not likely increase the item poor items and also show the need to review and modify the items.

Baiq Della Triastiwi Putri (2017) do research with the title “The Validity Analysis Of English Summative Test Of Junior High School” This research was aimed at analyzing the English summative test validity at SMP Darul Hamidin Padamara”. The object of the research was the English summative test given to seventh grade students in the academic year 2016/2017 used descriptive method. The instrument used to analyze the data is documentation such as the English summative test, syllabus, blueprint and students answer sheet. The researcher matched the English summative test and syllabus to find out the content validity, for construct validity the researcher use blueprint and the last is students answer sheet to find out the criterion validity, to support the data the researcher used questionnaire. The finding of this research showed that the English summative test of SMP Darul Hamidin had validity in term of content and construct validity because 96% of test matched with indicators and for criterion validity was not valid because out of 50 items only 10 items were valid.

The equation of this study with the previous one is both researching the items while the differences in this study with previous studies are research subjects, research objects, and research sites.

CHAPTER III RESEARCH METHODOLOGY

A. Research Design

This research is a descriptive quantitative research. Quantitative research is a means for testing objective theories by examining the relationship among variables (Creswell, 2009). A theory might appear in a research study as an argument, a discussion, a figure, or a rationale, and it helps to explain (or predict) phenomena that occur in the world (John, W. Creswell, 2014).

The quantitative approach used in this study is descriptive, meaning the results analyzed are in the form of a description of the observed symptoms that do not have to be in the form of numbers. The descriptive quantitative approach in this research will be used to show a description of the quality of the items summative test at SMP N 12 Lebong in academic 2020/2021 years.

The data obtained in the form of numbers will be identified to be drawn conclusions using the Anates version 4. The researcher used the Anates version 4 program because it can analyze multiple choice questions and descriptions. The number of subjects is multiple choice and the number of questions can be adjusted to the needs. In the Anates version 4 program there are facilities that are used to add, insert and delete subjects and delete items. The results of the analysis of the scores obtained by each test can also be transferred to Microsoft Excel to be calculated so that the use of the Anates version 4 program can be optimized.

B. Research Subject

The subjects of the research are Paper Test Question summative test on English subject at SMPN 12 Lebong in academic years 2020/2021. Data in this study obtained through research conducted on

date 27 and 28 May in SMP N 12 Lebong. Data collection techniques used are documentation techniques.

C. Population and Sample

The researcher took the data from eight grade students of SMPN 12 Lebong. Only took 40 students' answer sheet as a purposive sampling technique. The reason for using purposive sampling technique is the researcher only examined one class (A class) out of three 8th classes at SMPN 12, because the eight grade English teacher at SMPN 12 Lebong was just one for entire 3 classes. Then, the researcher divide the students into two groups : upper and lower by arranging the students' score from the highest till the lowest scores.

D. Research Variable and Indicators

Instrument is the process of selecting or developing measurement tools and methods that are appropriate for the problem being evaluated. The instrument serves as a tool in collecting the required data. In this research the instruments used with documentation techniques items summative test on English subjects in academic years 2020/2021 in the form of multiple choice questions as many as 40 items.

E. Data Collection Technique

Data collection techniques used in this study are documentation techniques. Documentation method, which is looking for data on things or variables in the form of notes, transcripts, books, newspapers, magazines, inscriptions, minutes of meetings, briefs, agendas, and so on (Siyoto and Sodik, 2005:66) . This method is used to get item summative test English subjects in academic year 2020/2021, along with a answer key sheets, and student answer sheets.

F. Data Analyzing Technique

Data analysis will be carried out on The third-grade Final School exam test multiple choice at SMPN 12 Lebong by looking for item difficulty, item discrimination from Anates V.4 Program and distractor analysis based on Marrie Tarrant's theory and Suharsimi Arikunto's theory.

a. Item Difficulty

The analysis of the difficulty level or Facility Value (FV) is the first step in analyzing the test items. Heaton (1988) stated that the FV of an item shows the difficulty of an item in a test. It shows which item is easy or difficult. The FV can be known from the ratio or percentage of students who answer the item correctly.

This research uses Heaton's (1988) formula to measure the FV which is gained by dividing the number of students from the upper group and the lower group students who answer a certain item correctly by the total number of the students who join the test.

$$FV = \frac{\text{Correct } U + \text{Correct } L}{2n} \quad \text{or} \quad FV = \frac{R}{N}$$

F : The index of difficulty

R : The number of correct answers

N : The number of students taking the test

U : Upper half

L : Lower half

n : Number of candidates in one group

The scale of the difficulty level of all test items ranged from 0.00 to 1.00 (Arikunto, 2006:208-210). It can be interpreted in the rank scale of difficulty level, as follow:

Table 3.1
The Clasification of Criteria Level of Difficulty

P	Interpretation
< 0.30	Difficult
0.30 – 0.70	Moderate
> 0.70	Easy

Source : Arikunto Suharsimi

b. Discriminating Power

The next step of item analysis is to determine the Discriminating Power (DP) that is whether the item can discriminate the students of the upper group from those in the lower group.

Heaton (1988:180) elaborates, “The index of discrimination (D) tells us whether those students who performed well on the whole test tended to do well or badly on each item of the test.”

This research uses Heaton’s (1988) formula of DP.

$$D = \frac{U - L}{N}$$

In which:

D: The index of item discriminating power

U: The number of pupils in the upper group who answered the item correctly

L: The number of pupils in the lower group who answered the item correctly

N: Number of pupils in each of the groups

The result of the use of Heaton's formula above is interpreted by using Arikunto's (1986) criterion of DP.

Table 3.2
The Classifications Of Criteria Discriminating Power

Index of Discriminating Power	Classifications
0.70 -1.00	Excellent
0.40 – 0.70	Good
0.20 – 0.40	Satisfactory
≤ 0.20	Poor
Negative value on D	Very Poor

Source : Arikunto Suharsimi

c. Distractors Efficiency

Checking the distracter efficiency, if the distractors is selected by 5% of the students, therefore the distractor categorized as effective distractor.

d. Revision

Revising the summative test items based on the result of data analysis using Moore's (2012:270) guideline :

1. The central issues or problem should be stated clearly in the stem, and there should be no ambiguity in terminology.
2. A stem and the alternatives should be in one page.
3. Avoid to provide grammatical or contextual clues to the correct answer.
4. Use language that even the most unskilled readers will understand. Keep the reading requirement question to a minimum. Write brief stems and accurate choices.
5. Avoid to use absolute terms (such as always, never, none) in the stem or alternatives.

6. Alternatives should be grammatically correct. The use of is or are, can help students to guess the correct response. All alternatives should be fit the stems to avoid giving clues to items that are incorrect.
7. Avoid the use of negatives (such as not, except, and least) and double negatives in the stem or alternatives. If negatives are used in the stem, put the as near the end of the stem as possible.
8. Avoid giving structural clues. This means, not to use one letter for the alternatives than the other, or make a pattern of correct response.
9. Make sure to use “all of the above” and “none of the above” with care. Because, “all of the above” is usually have poorer response than “none of the above”; because all the alternatives must be corret.
10. Avoid pulling statements directly from textbook. The objective of the test is for students understanding not memorization.
11. Alternatives should be plausible to less knowledgeable students. Write distractors with common errors, errors that likely, and erroneous common sense solution

e. Anates Program

1. Definition of Anates Program

Anates is an application program specifically used to analyze multiple choice tests and descriptions developed by Drs. Karnoto, M.Pd and Yudi Wibisono, ST. Anates version has the ability to analyze test questions such as:

1. Calculate the score (original and weighted).
2. Calculate the reliability of the test.
3. Group the subject into the upper or lower group.
4. Calculate the discriminating power.
5. Calculate the level of difficulty.
6. Calculate the item scores correlation with a total score.

7. Determine the quality of the distractor efficiency.

Anates version 4 is software specifically developed to analyze multiple choice tests and descriptions. But as the one who has been displayed anates version 4 has not been completely maximal in its use, there are still some who are still considered not good, namely in terms of processing data on the description. Anates is designed to be easily learned and easy to use. By using Anates, the test analysis process will be easier, fast, and accurate.

Initially the anates application developed is version 4 launched operating system operating system. Due to limitations and not theory of the theory of analysis of the description, the developer creates the Anates version 4 application program can only be used to analyze the form of multiple choice questions. Only on Anates version 4 which was launched on February 22, 2004 the developer managed to create anates who can analyze the form of the description. This program is able to analyze items about multiple choice questions and descriptions easily and quickly. Another advantage is the anates program fully used Indonesian.

The advantages of this software as a matter analysis program rather than other item programs can be used for analysis of items about the description form, in addition to analysis of multiple choice forms. The use of Indonesian in this program, is also one side of the convenience in its use than other programs that use English. The results of the analysis of the score obtained by each test can also be transferred to Microsoft Excel to calculate its value, then when we explore this program at the same time can explore the Microsoft Excel application. If so, the utilization of Anates will be optimal.

2. Benefits of Anates

Basically the anatest uses are the same as other data processing items, but operating it is easier. In addition, the results have been directly analyzed by the program. So, we no longer need to struggle to analyze it back with the existing criteria.

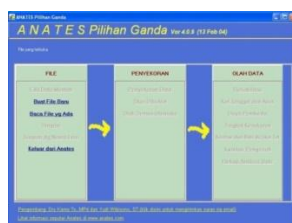
The functions and benefits of this anates are certainly to analyze the data of multiple choices, which are tested. The Anates application is very easy to use and is very helpful in analyzing multiple choices. With this version 4 of anates we can check the correct answer questions and answer questions wrong with practical and fast.

After the question is examined, with this anates we can carry out the storm and granting weight for answers to the right questions and the item is wrong. In addition, the data about the questions will be processed so that we can directly find out how to megiate multiple choice questions and descriptions, namely: 1) Reliability test, 2) superior grouping and Asor, 3) Analysis Discriminating Power, 4) The level of difficulty, 5) Correlation of Score Each item questions with a total score, 6) the quality of distractor efficiency and 7) recap of item analysis.

3. The Steps of Anates V4 Program

Multiple choice analysis steps with Anates as follows :

- 1) Click "Run Multiple Choice Anates" the following display will appear :



- 2) In the FILE column, click "Create New File" for a new analysis, "Read Existing Files" to open the saved file, "Exit Anates" to exit the program.

- 3) Click "Create New File", a dialog box will appear:

- 4) On the Number of Subjects, write down the number of test participants, the number of questions and the number of options, then click OK,

- 5) Enter the answer key for each question number, write down each name of the test participant and answer the test participants for each question, for all participants. After finishing it will look like this:

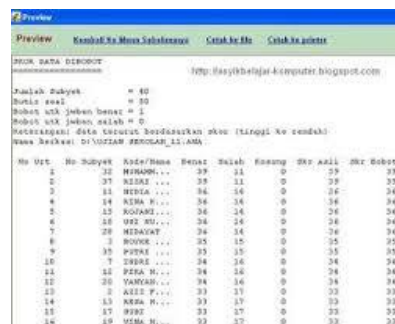
Edit Data Mentah

| Jumlah Subyek: 10 | | | Jumlah Butir Soal: 10 | | | Jumlah Pilihan Jawaban: 5 | | | | | |
| Nomor Soal | Nomor Subyek | No. Butir Baru | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
1	an		a	D	c	B	A	C	A	A	B	B
2	dian		a	D	A	B	C	E	E	A	D	D
3	rori		A	D	C	B	A	E	D	C	B	A
4	dio		A	B	C	D	E	D	C	B	A	B
5	dia		A	A	A	B	D	C	D	C	A	A
6	nan		C	A	D	A	B	E	D	A	B	A
7	rori		D	E	D	A	D	A	D	A	B	A
8	ela		E	D	C	B	A	B	A	D	A	E
9	dod		D	D	D	E	A	D	E	B	C	A
10	dovi		A	B	C	D	E	A	D	D	C	C

 </div>
 <div data-bbox="326 710 862 752" data-label="List-Group>
 <p>6) Data entry completed. Then select and click "Return to Main Menu",</p>
 </div>



7) In the CORRECTION column, select "Auto Run All"



8) The analysis process is complete, select "Print to printer" if you want to print immediately, select "Print to file" if you want to save it in Notepad.

9) Click "Return to Previous Menu", in the FILE column select "Save"

10) Click "Exit Anates" in the dialog box, click "Yes"

CHAPTER IV

RESULT AND DISCUSSION

A. Data Description

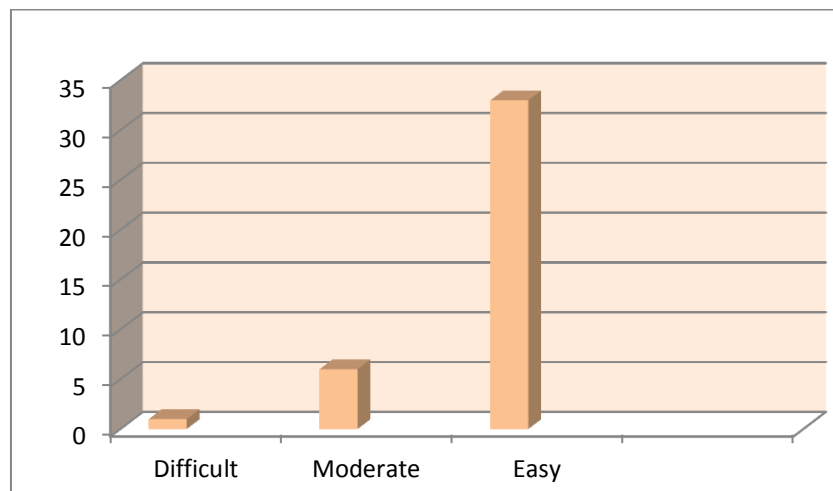
Data in this study obtained through research conducted on date 27 and 28 May in SMP N 12 Lebong. Data collection techniques used are documentation techniques. Based on the documentation technique, documents were obtained in the form of summative test questions, key answers to questions, and answer sheets of students who took the summative test in school. The items question used is multiple choice questions 40 items.

The collected data is used as a reference in analyzing the quality of summative test questions in English subjects quantitatively using the Anates version 4.9 program. Characteristics of the items produced include the level of items difficulty, discrimination power, and Distractor efficiency of the question summative test for english subjects in the 2020/2021 academic year of SMP N 12 Lebong which can be seen in appendix 3, appendix 4, and appendix 5.

1. Item difficulty of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021

The level of difficulty is the opportunity to answer a question correctly at the level of certain abilities of students. Good questions have a level of difficulty which is in a sense not too easy and not too difficult.

Description of the results of the analysis difficulty item of the question summative test English subject at SMP N 12 Lebong in academic year 2020/2021, the researcher put it into the classification score that can be seen in the following chart 4.1 :



Source : Anates Program

Chart 4.1
The Result Of Item Difficulty

Based on the data above using the Annates v.4 program, there was the result analysis of item difficulty of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021 the highest score is easy 33 (82,5%) and the lowest score is difficult (2,5%).

See more clearly the results of the analysis of item difficulty of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021 Can be seen table 4.1 below:

Table 4.1
The Result Item Difficulty

No	Category	Items	Freq	Percent(%)
1	Difficult	19	1	2,5
2	Moderate	9,15,16,22,28,31	6	15
3	Easy	1,2,3,4,5,6,7,8,10,11,12,13,14, 17,18,20,21,23,24,25,26,27,29, 30,32,33,34,35,36,37,38,39,40	33	82,5

Source : Anates Program

Based on the data above using the Annates v.4 program, there was the result of the item difficulty of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021 there

were 1 or (2,5%) items difficult, 6 or (15%) items moderate, and 33 or (82,5%) items easy.

2. Discrimination power of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021.

Discrimination power is the ability of questions to distinguish high-ability students and low-ability students. A good items must be able to distinguish the ability of students.

Description of the results of the analysis discrimination power of the question summative test English subject at SMP N 12 Lebong in academic year 2020/2021, the researcher put it into the classification score that can be seen in the following chart 4.2 below :



Source : Anates Program

Chart 4.2
The Result Of Discrimination Power

Based on the data above using the Annates v.4 program, there was the result analysis of item discrimination power of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021 the highest score is Poor 22 (55%) and the lowest score is Worst 2 (5%).

See more clearly the results of the analysis of quality item discrimination power of summative test items on English subject at

SMP N 12 Lebong in academic 2020/2021 Can be seen table 4.2 below:

Table 4.2
The Result Of Discrimination Power

No	Category	Items	Freq	Percent(%)
1	Good	7,9,12,13,32,34,36,40	8	20
2	Satisfactory	2,4,5,6,23,27,31,33	8	20
3	Poor	1,3,8,10,11,14,15,16,18,20, 21,22,24,25,26,28,29,30,35, 37,38,39	22	55
4	Worst	17,19	2	5

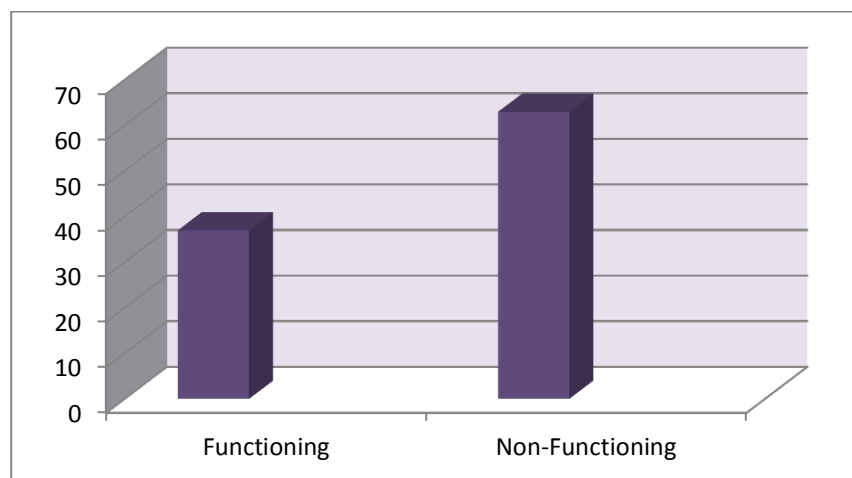
Source : Anates Program

Based on the table above using the Annates v.4 program, there was the result of the discrimination power item summative test English subject in Academic years 2020/2021 there were 8 or (20%) items good, 8 or (20%) items satisfactory, 22 or (55%) items poor, and 2 or (5%) items worst.

3. Distractor efficiency of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021.

Deception effectiveness is one of the characteristics to determine the quality of a question. The effectiveness of the option referred to here is to find out the trickster on a function or not. Distractor works well if chosen by a minimum of 5% of all participants.

Description of the results of the analysis distractor efficiency of the question summative test English subject at SMP N 12 Lebong in academic year 2020/2021, the researcher put it into the classification score that can be seen in the following chart 4.3 below :



Source : Marrie Tarrant's Theory and Suharsimi Arikunto's Theory

Chart 4.3
The Result of distractor efficiency

Based on the data above based on Marrie Tarrant's theory and Suharsimi Arikunto's theory. For the distractor effectiveness, the researchers found there are 44 (22%) distractors with effective distractor and 156 (78%) distractors with in-effective distractor.

See more clearly the results of the analysis of item distractor efficiency of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021 Can be seen appendices no.6.

The results of the analysis of the level of difficulty, discrimination power, and distractor efficiency of the tests that have been obtained are then analyzed to determine the quality of the questions between the questions that are of good quality, good enough, and not good based on the following considerations :

1. Item questions have good quality, if the question fulfills the three criteria, namely the level of difficulty, distinguishing power, and deceptive effectiveness.
2. Item questions have quite good quality, if the question only fulfills two of the three criteria.
3. Item questions have poor quality, if the question does not meet two or all criteria

B. Data Analysis

1. Level of Difficulty

The comparison between easy-medium-difficult questions can be made 3-4-3. This means 30% easy items, 40% moderate items, and 30% difficult items. While the summative test in the eighth grade of SMPN 12 Lebong does not yet have a good comparison of questions. So, in the future, the eighth grade English teacher of SMPN 12 Lebong must be more careful in making questions so that they are in accordance with the comparison of questions with good criteria.

The results of the analysis obtained level of difficulty in this study that item summative test English subject in Academic years 2020/2021: there were 1 items difficult, 6 items moderate, and 33 item easy.

2. Discriminating Power

The results of the analysis obtained in this study that the differentiation power item summative test English subject in Academic years 2020/2021: there were 10 items very good, 8 items good, 8 items satisfactory, 22 items poor and 2 items worst.

Item discrimination (ID) indicates the degree to which an item separates the students who performed well from those who performed poorly (Brown, 1996:66). Questions that have enough discrimination power should be revised, and the question with the difference between a bad and very ugly differentiator must be replaced in order to be able to distinguish the level of students' abilities.

Items number 7,9,12,13,32,34,36 and 40 have good discriminating power. This shows that the items of this item are able to distinguish the level of ability of students well so that the items that have good distinguishing abilities can be immediately recorded in the question bank book so that the items can be reused as a test of results learn in the future.

Items number 2,4,5,6,23,27,31 and 33 have satisfactory discriminating power. This shows that the items of this question are quite capable of distinguishing the abilities of students, but these items should be examined again, tracked and traced so that the factors that cause these items have not been able to distinguish the ability level of students. The items of this question should also be revised to have a good discrimination power so that they can be reused in the test of learning outcomes in the future.

Items number 1,3,8,10,11,14,15,16,18,20,21,22,24,25,26,28,29, 30,35,37,38 and 39 This shows that these items are not able to distinguish the level of ability of the students, so the items that have very poor and poor discrimination power abilities should be shed and not used again in future learning outcomes tests.

3. Distractor Efficiency

The results of the analysis obtained in this study that the seen there was the result of the distractor efficiency item summative test English subject in Academic years 2020/2021: there were 22% distractors with effective distractors and 78% distractors with in-effective distractors.

Distractor efficiency is one more important measure of a multiplechoice item's value in a test, and one that is related to item discrimination (Brown, 1996:60). So it can be said that one of the requirements of a question is said to be good is to have good quality deception.

4. Revision

Based on item difficulty (ID) and discriminating power (D) of the test, there are 20 items that did not meet the criteria both in ID and DP. Here the writer tries to revise the items. They are item number 3, 5, 6, 10, 11, 14, 19, 20, 21, 23, 25, 29, 30, 34, 35, 36, 37, 38, 39 and the last is 40.

1. Weak item number 3:
 - a. The copy of original item:

The generic structure of the text is .. .

- a. Orientation - Events - Reorientation
- b. Orientation - Complication – Resolution *
- c. General Classification - Description
- d. Identification – Description
- e. Orientation - Description – Reorientation

- b. The analysis and prediction

This item has ID 0.95 and DP 0.09. It discriminates students poorly and makes the item categorized as too easy. The writer found out that the stem of the item is not clear, because it did not refer to specific text. Therefore, the stem should be changed into “the generic structure from the text above is...” Also, the proportion of the student response who choose option A are 1, option B are 38, option C is 1, option D is 0, and option E is 0 pupils; it can be seen that option A, C, and E are not working as they should. As stated in D. Moore guideline, all of the alternatives should be fit the stems to avoid giving clues to items that are incorrect.

In order to distract lower group students choose the right answer (option B), all of the option (A, C, D and E) should have 3 answer and place “orientation” in the first answer. Therefore, option A. Orientation – Events – Reorientation is changed into A. Orientation – Complication – Reorientation; option C. General Classification – Description is changed into C. Orientation – Description – Complication; option D. Identification – Description is changed into D. Orientation – Identification – Resolution; and the last option E. Orientation –

Description – Reorientation is changed into E. Orientation – Complication – Description.

c. The suggested :

- The generic structure from the text above is.....
- a. Orientation – Complication – Reorientation
 - b. Orientation – Complication – Resolution *
 - c. Orientation – Description – Complication
 - d. Orientation – Identification – Resolution
 - e. Orientation – Complication – Description

2. Weak item number 5:

a. The copy of original item:

1. Which statement is TRUE according to the text?
 - a. The parrot could say Catano
 - b. catano was the name of the parrot
 - c. The man never got angry at the parrot
 - d. The man ate the parrot for dinner
 - e. At last the parrot could say catano*

b. The analysis and prediction:

This item has ID 0,90 and DP 0,12. It discriminates students poorly and the item categorized as too easy.

From distractor efficiency for number 5; the proportion of the student response who choose option A is 1, option B is 1, option C is 1, option D is 2, and option E is 35 pupils.

Moreover, the writer found out the ambiguity in option A and E. option A and E has similar meaning that is “the parrot could say the word of Catano”. A revised taxonomy of multiple-choice item-writng guidelines by haladyna stated that the test-maker should make sure there is only one choice is the right

answer.1 Therefore, one of the option should be changed while the other one remains the same.

Option A. The parrot could say Catano is changed into A. The parrot could not say Catano. Also, the writer found out that option D has ambiguous message, it did not clearly explain when the man ate the parrot for dinner, therefore it should be changed into D. The man ate the parrot for Sunday's dinner.

c. The suggested:

Which statement is TRUE according to the text?

- a. The parrot could not say Catano
- b. Catano was the name of the parrot
- c. The man always angry at the parrot
- d. The man ate the parrot for Sunday's dinner
- e. At the end the parrot could say Catano *

3. Weak item number 6:

a. The copy of original item:

What is the main character of the story?

- a. The man
- b. The parrot*
- c. Chicken
- d. Catano
- e. The dead chicken

b. The analysis and prediction:

This item has ID 0,88 and DP 0,19. It discriminates students poorly and the item can categorized as too easy.

From distractor efficiency for number 6; the writer found out that no students have selected option C due to the obviousness.

Therefore, the writer changed option C. Chicken into C. The Bird. Also, the writer found a grammar mistake in the option D. So, the writer changed option D. Catano into D. The Catano in order to clear the meaning.

Moreover, the writer found out the direction in the stem is not clear. The stem stated “What is the main character of the story?” which supposed to be “Who is the main character of the story?”.

- c. The suggested:

Who is the main character of the story?

a. The man	d. The Catano
b. The parrot*	e. The dead chicken
c. The bird	

4. Weak item number 10:

- a. The copy of original item:

The above passage is in the form of

a. Narrative	d. Hortatory Exposition
b. Descriptive	e. Analytical Exposition
c. Report *	

- b. The analysis and prediction:

This item has DI 0,97 and DP 0,07. It discriminates students poorly and it can be categorized as too easy question.

From distractor efficiency for number 10; the proportion of the student response who choose option A is 0, option B is 1, option C is 0, option D is 1, and option E are 38 pupils. The writer found out that the distractors are too obvious so they are not working at all. Also it seems like the test-maker give a

slight clue in the length of the distracters. In order to make the length of all distracters equal and distract more students in lower group, the word “text” should be add at the end of the distracters.

Option A was not chosen by the students because it is already obvious that the passage above is not categorized as narrative text, therefore option A should be changed into A. Exposition text; Option B. Descriptive is changed into B. Descriptive text; Option C. Report is changed into C. Report text; Option D. Hortatory Exposition is changed into D. Hortatory text; Option E. Analytical Exposition is changed into E. Analytical text. Moreover, the stem of this question has grammar mistake, the stem is “The above passage is in the form of...” it should be “the passage above is the form of...”

c. The suggested:

The passage above is the form of

a. Exposition text	d. Hortatory text
b. Descriptive text	e. Analytical text
c. Report text *	

5. Weak item number 11:

a. The copy of original item:

What is the text organization of the first paragraph?

a. Thesis *	d. Orientation
b. Argument	e. Reorientation
c. Reiteration	

b. The analysis and prediction:

This item has DI 1.00 and DP 0.00. It discriminates students poorly and the item categorized as too easy.

Beside the item instruction was not clear, the writer also found that this types of question has already asked in the first place after the text, therefore this kind of text is not needed anymore.

Also, the proportion of the student response who choose option A are 40, option B is 0, option C is 0, option D is 0, and option E is 0 pupil. Since the option A was selected by all of the students, it can be concluded that the distractors did not functioned at all. The item should be changed.

c. The suggested:

What is the purpose of the writer wrote this passage?

- a. To give advice on health problem
- b. To complain about having too much fat
- c. To support the reader to have a lot of fat
- d. To show the reader the bad effect of too much fat*
- e. e. To persuade the reader to not have too much fat

6. Weak item number 14:

a. The copy of original item

The generic structure of the text is ...

- a. Orientation – Events – Reorientation
- b. Orientation – Complication – Resolution
- c. General Classification – Description
- d. Thesis – Arguments – Reiteration *
- e. Orientation – Description - Reorientation

b. The analysis and prediction:

This item has ID 0.98 and DP 0.05. It discriminates students poorly and make the item categorized as too easy.

Beside the item instruction was not clear, the writer also found that types of question has already asked in the previous question. Also, the proportion of the student response who chose option A are 1, option B is 0, option C is 1, option D is 38, and option E is 0 pupil. Since option D was selected by most of the students, it can be concluded that the distractors did not function at all. Therefore, this item must be changed.

c. The suggested:

What does the text tell about?

- a. The effect on having high blood pressure
- b. The effect of having heart problem
- c. The effect on having too much fat *
- d. The studies about health problem
- e. The studies about heart attack

7. Weak item number 19:

a. The copy of original item:

Choose the correct statement based on the text above!

- a. The text is to persuade the readers to be fat
- b. More researches is not needed anymore
- c. The heart will work harder as the effect of being fat
- d. Being fat may not lead you to heart attack *
- e. Diabetics is not serious disease

b. The analysis and prediction:

This item has ID -0,02. It discriminates students worst and since this item has DP 0,01 it categorized as very difficult item.

The writer found out that the item has miskeyed item. The right key should be option “C” rather than option “D.” It clearly stated in the previous text that “if you are fat, your heart has to work harder.” Apparantly, most of the upper group and lower group students choose option C or the “wrong” answer. In fact, they choose the right answer.

c. The suggested:

The key for number 19 should be option C.

Choose the correct statement based on the text above!

- a. The text is to persuade the readers to be fat
- b. More researches is not needed anymore
- c. The heart will work harder as the effect of being fat *
- d. Being fat may not lead you to heart attack
- e. Diabetics is not serious disease

8. Weak item number 20:

a. The copy of original item:

There are over 3000 species of spiders and they belong to arachnids family. The statement tells the readers about....

- a. Behaviour of a spider
- b. Colony of spider
- c. Habitat of a spider
- d. Species of a spider *
- e. Friends of a spider

b. The analysis and prediction:

The Item has ID 0,07. It discriminates students poorly and since it has DP 0,97; the item categorized as too easy.

The writer found out the options for this item is too obvious. Some studies said, the items are easier when a keyword in the stem is also included in the answer. In this case, the writer found out that the sentence “Species of a spider” clearly stated both in the instruction and the text. That is why the students are easy to choose the right answer. Therefore, the item should be changed into: Most spiders have poor eyesight and little sense of hearing and smell. They do have a well-developed sense of touch in the hairs and spines that cover theirs body.

c. The suggested:

Most spiders have poor eyesight and little sense of hearing and smell. They do have a well-developed sense of touch in the hairs and spines that cover theirs body. The statements above tell the reader about...

- | | |
|-------------------------|----------------------|
| a. Behavior of a spider | d. Species of spider |
| b. Colony of spider | e. Habitat of spider |
| c. Anatomy of spider * | |

9. Weak item number 21:

a. The copy of original item:

From the text above, we can conclude that spiders...

- | | |
|----------------------|---------------------------------|
| a. are not insects * | d. have least number of species |
| b. eat leaves | e. are marsupials |
| c. are mammals | |

b. The analysis and prediction:

The Item has DI 0,99 and DP 0,02. It discriminates students poorly, and categorized as too easy item.

Moreover, the proportion of the student response who choose option A are 39, option B is 0, option C is 0, option D is 0, and option E is 1 pupil.

The writer found out the options for this item are too obvious since the keyword also included in the answer. In line with Burton, he stated to keep the distracters free from clues to which the response is correct.³ In this case, the writer found out that the sentence “are not insects” clearly stated in the stem (first paragraph of text). That is why the students can easily choose the right answer.

Therefore, Option A. Are not insects should be changed into A. Related to arachnids. Option B. Eat leaves should be changed due to unlogic answer, into B. Eat solid food. Option C. Are mammals also should be changed due unlogic answer, into C. Are insects. Option D. “Have least number of species” should be changed, because in the stem clearly stated that there are over 3000 species of spider; it change into D. Have undeveloped sense of touch. Option E. Are marsupials should be changed into E. Related to insect.

c. The suggested:

From the text above, we can conclude that spiders...

- | | |
|--------------------------|------------------------------------|
| a. Related to arachnids* | d. Have undeveloped sense of touch |
| b. Eat solid food | e. Related to insects |
| c. Are insects | |

10. Weak item number 23:

a. The copy of original item:

They are able to produce this from glands in the abdomen and the silk comes out through organ called spinnerets. The word abdomen has similar meaning to...

- | | |
|--------------------------------|--------------|
| a. Spinnerets | c. Gland |
| b. Organs | d. Spiders * |
| c. None of the options correct | |

b. The analysis and prediction:

This item has DP 0,91 and ID 0,19. It discriminates poorly and the item categorized as too easy.

From distractor efficiency for number 23; the proportion of the student response who choose option A are 1, option B are 1, option C are 1, option D are 36, and option E are 1 pupils. The writer found that the distractors are too obvious since most of the students in the lower group choose the right answer. It seems that the stem is too easy, so the students are easily to choose the right key. Therefore, this item should be changed

c. The suggested:

They are able to produce this from glands in the abdomen and the silk comes out through organ called spinnerets. The word abdomen has similar meaning to...

- | | |
|--------------|------------|
| a. Diaphragm | d. Belly* |
| b. Pelvis | e. Cardiac |
| c. Thorax | |

11. Weak item number 25:

a. The copy of original item:

Paragraph 4 th tells us about ...	
a. Reproduction *	d. behavior
b. Appearances	e. food
c. Family	

b. The analysis and prediction:

This item has ID 0.97 and DP 0.07. It discriminates students poorly and categorized as too easy item.

The proportion of the student response who choose option A are 39, option B is 1, option C is 0, option D is 0, and option E is 0 pupil. In line with the proportion of students, it looks like the instruction is not clear and have grammar mistake.

At the stem of this question states “paragraph 4th tells us about...” however, it should be “paragraph 4 tells the reader about...” Also, the distractors are not function properly since most of the students get the right answer. Since option D and option E has been chosen by no students, therefore option D. Behaviour should be changed into D. Types and option E. Food is changed into E. Digestion. For option A, almost all of the students in lower group choose that option, therefore option A. Reproduction should be changed into A. Mating in order to distract lower students group.

c. The suggested:

From the text above, at the paragraph 4, the writer want to tell us about spiders'	
a. Mating *	d. Types
b. Appearance	e. Disgestion
c. Habit	

12. Weak item number 29:

a. The copy of original item:

The statements below are correct, except ...

- a. Female spiders lay large numbers of tiny eggs
- b. Spiders only suck the fluids of their victims
- c. Some male spiders are eaten by the female after mating
- d. Spiders's silk is weak and vulnerable *
- e. There are over 3000 species of spiders in the world

b. The analysis and prediction:

This item has ID 0.91 and DP 0.14. It discriminates students poorly and has categorized as too easy item.

The proportion of the student response who choose option A is 1, option B is 1, option C is 0, option D are 38, and option E is 0 pupil. It can be stated that the distractors are not function properly since most of the students in the lower group choose the right answer. Moreover, the test-maker use "vulnerable" which considered as a complex vocabulary which not all of the students understand the meaning.

The writer found out option A, C, D, and E are ineffective. Also, in option D, "weak" has a similar meaning with "vulnerable". Therefore option D. Spiders's silk is weak and vulnerable must be change into D. Spiders' silk is feeble and stiff. Also option C and E have no students chosen that distracter, because those distracter has clearly stated in the

stem. That is why the students easily avoid the distracter. Therefore option C and E must be changed. Option C. Some male spiders are eaten by the female after mating is changed into C. Female spiders always carry her eggs around. Option E. There are over 3000 species of spiders in the world is changed into E. Spiders are include in arachnids family.

c. The suggested:

The statements below are correct, except

- a. Female spiders produced large numbers of tiny eggs
- b. Spiders only suck the fluids of their victims
- c. Female spiders always carry her eggs around
- d. Spiders's silk is feeble and stiff *
- e. Spiders are include in arachnids family

13. Weak item number 30:

a. The copy of original item:

Spiders' food is explained in paragraph...

- a. I
- b. II
- c. III
- d. IV *
- e. V

b. The analysis and prediction:

This item has ID 0.91 and DP 0.14. It discriminates students poorly. The writer found out that the item is not clear.

The word "food" in the instruction is less appropriate. It is better to change the stem into "eating habit" since the paragraph being questioned above more tells about spiders' eating habit than kinds of food that spiders' take. It must be changed.

- c. The suggested:

Spiders' eating habit is explained in paragraph....

- | | |
|--------|---------|
| a. I | d. IV * |
| b. II | e. V |
| c. III | |

14. Weak item number 34:

- a. The copy of original item:

Student : Excuse me, may I stay here for a night, please? I have no friends to visit. Teacher: Sure.

Student: **I'm glad you don't mind. Thank you very much, Sir.**

The underline statement refers to....

- Giving advice
- Displeasure
- Pleasure *
- Pain
- Giving opinion

- b. The analysis and prediction:

This item has ID 0.88 and DP 0.19. It discriminates students poorly and it can be categorized as too easy question.

From distractor efficiency for number 34; the proportion of the student response who choose option A are 3, option B is 0, option C are 36, option D is 1, and option E is 0 pupils. The writer found out that the distracters are too obvious so the students easily get to the right answer. Also all of the distracter has grammar mistake and affect its meaning.

Moreover, option E has been chosen by no students. Therefore option A. Giving advice is changed into A. Giving

some advice; Option B. Displeasure is changed into B. Showing displeasure; Option C. Pleasure is changed into C. Showing pleasure; Option D. Pain is changed into D. Showing sense of pain; option E. Giving opinion must be changed into E. Asking for advice. In addition, the stem has grammar mistake; the stem states “The underline statement refers to...” it is supposed to be “the emotion expressed by the underlined statement above is to....”

c. The suggested:

Student : Excuse me, may I stay here for a night, please? I have no friends to visit.

Teacher: Sure.

Student: **I'm glad you don't mind. Thank you very much, Sir.**

The emotion expressed by the underlined statement above is to....

- | | |
|------------------------|--------------------------|
| a. Giving some advice | d. Showing sense of pain |
| b. Showing displeasure | e. Asking for advice |
| c. Showing pleasure * | |

15. Weak item number 35:

a. The copy of original item:

Rara : I like to be here. Bogor is great. **What do you think of it?**

Roni : I think Bogor is a traffic jam city. What is the expression of the underlined word?

- | | |
|--------------------|-------------------------|
| a. Satisfaction | d. Asking for opinion * |
| b. Dissatisfaction | e. Pleasure |
| c. Giving opinion | |

b. The analysis and prediction:

This item has ID 0.94 and DP 0.14. It discriminates students poorly and it can be categorized as too easy question. It is in line with the proportion of the student response who choose option A is 0, option B is 0, option C are 2, option D are 37, and option E is 2 pupils.

The writer found out that the distracters are too obvious so they are not working at all and the length between each distracters are not equal. In order to distract more students in lower group, some changes in distracters are needed. Moreover, option A and B was not chosen by any of the students. It can be said that the distracters did not operate effeciently.

Therefore, option A. Satisfaction is changed into A. Showing satisfaction; option B. Dissatisfaction is changed into B. Showing dissatisfaction; option C. Giving opinion is changed into C. Giving some opinion; option D remain the same; and the last is option E. Pleasure is changed into E. Showing opinion.

c. The suggested:

Rara : I like to be here. Bogor is great. **What do you think of it?** Roni : I think Bogor is a traffic jam city.

What is the expression of the underlined word?

- | | |
|----------------------------|-------------------------|
| a. Showing satisfaction | d. Asking for opinion * |
| b. Showing dissatisfaction | e. Showing opinion |
| c. Giving some opinion | |

16. Weak item number 36:

a. The copy of original item:

Sandra: There are too many silly cartoon on TV lately. Don't you agree?

Tony: **Not at all. Actually, I like cartoons because they're very entertaining.**

From the text above, we can conclude that Tony express...

a. Disagreement	d. Warning
b. Agreement	e. relief
c. Asking opinion	

b. The analysis and prediction:

This item has ID 0.97 and DP 0.02. It discriminates students poorly and categorized as too easy item.

The writer found out that the instruction is not clear, because Tony response for Sandra's question has an ambiguous meaning. Whereas the clearness of instruction in the stem is crucial in writing items. Therefore, "From the text above, we can conclude that Tony express..." must be changed into "From the text above, we can conclude that Tony Sandra's opinion"

In addition, the proportion of students answer for option A are 37, B are 2, C is 1, D is 0, and E is 0 pupils. Based on students' response, option D and E have not been selected both by the upper and lower students, so it can be categorized as ineffective distracter. So option D. Warning must be changed into D. Satisfied with; and option E. Relief must be changed into E. Unsatisfied with.

- c. The suggested:

Sandra: There are too many silly cartoon on TV lately. Don't you agree?

Tony: **Not at all. Actually, not all of cartoons are silly.**

Some of them are very entertaining.

From the text above, we can conclude that Tony Sandra's opinion.

- | | |
|---------------------|---------------------|
| a. Dissagree with * | d. Satisfied with |
| b. Agree with | e. Unsatisfied with |
| c. Doubt with | |

17. Weak item number 37:

- a. The copy of original item:

Kania : In my opinion, the local TV program are not really good.

Jamal: **Absolutely, I personally feel the program is boring.**

From the text above, we can conclude that Jamal express...

- | | |
|-------------------|------------|
| a. Disagreement | d. Warning |
| b. Agreement * | e. relief |
| c. Asking opinion | |

- b. The analysis and prediction:

The item has DP 0.14. It discriminates students poorly, and since it has ID 0,93 the item categorized as too easy. In addition, the proportion of students answer for option A are 2, B are 37, C is 1, D is 0, and E is 0 pupils. Based on students' response, option D and E have not been selected both by the

upper and lower students, so option D and E can be categorized as ineffective distracters.

Moreover, the instruction has not clearly referred to one program. So instead of “In my opinion, the local TV program are not really good” it is better to changed to “Don’t you think the Sunday TV program, Running Man, are not really good lately?” Also, “From the text above, we can conclude that Jamal express...” it is changed into “From the text above, we can conclude that Jamal show..... expression” and the last is from “**Absolutely, I personally feel the program is boring**” is changed into “**Absolutely, I personally feel that program is not too good lately**” For option D. Warning is changed into D. Satisfied; and option E. Relief is changed into E. Unsatisfied.

c. The suggested:

Kania: Don’t you think the Sunday TV program , Running Man, are not really good lately?

Jamal: **Absolutely, I personally feel that program is not too good lately.**

From the text above, we can conclude that Jamal show..... expression.

- | | |
|-------------------|----------------|
| a. Disagreement | d. Satisfied |
| b. Agreement * | e. Unsatisfied |
| c. Asking opinion | |

18. Weak item number 38:

a. The copy of original item:

Laras: **What do you think about the school?**
 Kania: In my opinion, the school is the best school in this city?
 The expression above is....
 a. Asking opinion* d. Pleasure
 b. Giving advice e. Relief
 c. Pain

b. The analysis and prediction:

This item has ID 0.99 and DP 0.02. It discriminates students poorly and categorized as too easy item.

The writer found out that the item instruction is not clear yet did not use appropriate punctuation mark. Moreover, the options for this item is too obvious since the keyword also included in the distractor. In this case, the writer found out that the word “in my opinion” is clearly stated in the stem. It is better changed the instruction from into :

Laras : **What do you think about the school?**

Kania : In my opinion, the school is the best school in this city?

The expression above is.... into :

Laras : **What do you think about this school? Kania : This school is the best school in this town!**

The underlined phrases is categorized as expression

In addition, the proportion of students’ answer for option A are 39, B is 1, C is 0, D is 0, and E is 0 pupils. Based on students’ response, option C, D, and E is categorized as ineffective distracter. So, option C. Pain is changed into C. Giving opinion; option D. Pleasure is changed into D. Asking

advice; and option E. Relief is changed into E. Showing pleasure.

c. The suggested:

Laras : **What do you think about this school?**
 Kania : This school is the best school in this town!
 The underlined phrases is categorized as expression

a. Asking opinion *	d. Asking advice
b. Giving advice	e. Showing pleasure
c. Giving opinion	

19. Weak item number 39:

a. The copy of original item:

Mother: Kids, do you have any idea what we are going to do to night?

Amalia: **I suggest we go out for dinner, Mom.**

The underlined phrases express....

- | | |
|-------------------|-------------|
| a. Asking opinion | d. Pleasure |
| b. Giving advice | e. Relief |
| c. Pain | |

b. The analysis and prediction:

This item has ID 0.92 and DP 0.16. It discriminates students poorly and categorized as too easy item.

The writer found out the options for this item is too obvious since the test-maker also include some keyword in the stem “**I suggest**”, it must be changed into “**What about**”. In addition, the proportion of students answer for option A are 1, B are 37, C is 0, D is 0, and E are 2 pupils. Based on students’ response, option B and C was categorized as ineffective distracter. In order to distract students in lower group, option C. Showing

pain is changed into C. Asking advice and option B. Giving advice remains the same.

c. The suggested:

Mother: Kids, do you have any idea what we are going to do to night?

Amalia: **What about we go out for dinner, Mom?**

The underlined phrase above express....

- a. Asking opinion
- b. Giving advice *
- c. Asking advice
- d. Pleasure
- e. Relief

20. Weak item number 40:

a. The copy of original item:

Handin: How do you like your room?

Jasper: **I really like it. It's very comfortable.**

The underlined phrases express..

- a. Asking opinion
- b. Giving advice
- c. Satisfaction
- d. Dissatisfaction
- e. Asking for advice

b. The analysis and prediction:

This item has ID 0.92 and DP 0.12. It discriminates students poorly and makes the item categorized as too easy item.

The writer found out that the item instruction has an ambiguous meaning and literary translate from bahasa. Therefore the instruction of the question from "Handin: How

do you like your room?” changed into “Handin: Do you like your room?” In addition, the proportion of students answer for option A is 1, B are 2, C are 38, D is 0, and E is 0 pupils. Based on students’ response, option C, D, and E was categorized as ineffective distracter. So, writer decided to make some changes in option D and E. Option D. Dissatisfaction is changed into D. Pleasure; option E. Asking for advice is changed into E. Relief

c. The suggested:

Handin: Do you like your room?

Jasper: **I really like it. It is very comfortable.**

The underlined phrases express...

a. Asking opinion	d. Pleasure
b. Giving advice	e. Relief
c. Satisfaction *	

C. Limitations of Research

In this reserach at SMPN 12 Lebong, the researchers only limited it to certain things, namely:

1. This study only used a sample of eighth grade A students at SMPN 12 Lebong in the academic year 2020/2021.
2. Student achievement in this research was only seen from the answer sheets of summative test in the academic year 2020/2021.

CHAPTER V

CONCLUSION AND SUGGESTION

A. Conclusion

Based on the results of the data analysis, it was conclude that the item summative test English subject at SMP N 12 Lebong in academic year 2020/2021 the category are :

1. The item difficulty of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021 there was the result of the item difficulty of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021 there were 1 or (2,5%) items difficult, 6 or (15%) items moderate, and 33 or (82,5%) items easy.
2. The item discrimination power of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021, there was the result of the discrimination power item summative test English subject in Academic years 2020/2021 there were 8 or (20%) items good, 8 or (20%) items satisfactory, 22 or (55%) items poor, and 2 or (5%) items worst.
3. The item distractor efficiency of summative test items on English subject at SMP N 12 Lebong in academic 2020/2021, the researchers found there are 44 (22%) distracters with effective distractor and 156 (78%) distracters with in-effective distractor.

Thus it can be concluded that summative test items on English Subject a SMPN 12 Lebong in the academic year 2020/2021 is not good.

B. Suggestion

Based on the conclusion above, the researcher would like there are several suggestions for more item are Good quality questions can be entered into the question bank to be reused, items that are not good should be revised and corrected first to meet the quality criteria of the questions in quantitative and qualitative terms. While the questions that are not good

should be replaced by making new questions that are better. Teachers should pay more attention to the rules for making good questions such as conducting trials and analyzing questions both before and after testing, both qualitatively and quantitatively. In order to give encouragement to teachers to always learn to make questions correctly, learn to analyze questions according to the procedure of making questions so that in the future the quality of the questions is made even better.

REFERENCES

- Ahmann, J.S , & Marvin, D. G. (1967). *Evaluating Pupil Growth, Principles of Tests and Measurements*. Boston: Allyn and Bason, INC.
- Zainal, A. (2009). *Evaluasi Pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Arikunto, S. (2006). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT. Bumi Aksara.
- Bajpai, S. R., & Bajpai, R. C. (2014). Goodness of Measurement: Reliability and Validity. *International Journal of Medical Science and Public Health*, 3(2), 112-115.
https://www.researchgate.net/publication/271186978_Goodness_of_Measurement_Reliability_and_VValidity
- Brown, H. D. (2001). *Teaching by Principles*. New York: Addison Wesley Longman, Inc.
- Brown, J. D., & Thom, H. (2002). *Criterion-Referenced Language Testing*. Cambridge: Cambridge University Press.
- Evrero & Sylvanus, E. (2015). Item analysis of test of number operations. *Asian Journal of Educational Research*,3(1), 18.
<http://www.multidisciplinaryjournals.com/wpcontent/uploads/2015/01/ITEM-ANALYSIS-OF-TEST-OF-NUMBER-OPERATIONS.pdf>
- Norman, E. G., & Robert L. L. (1990). *Measurement and Evaluation in Teaching 6th Ed*. New York: Macmillan Publishing Company.
- Heaton, J.B. (1988). *Writing English Language Tests*. New York: Longman.
- Hingorjo, M.R., & Jaleel, F (2012). Analysis of One-Best MCQs: *the Difficulty Index, Discrimination Index and Distractor Efficiency*, Karachi: Journal of Pakistan Medical Association. Retrieved from
https://www.researchgate.net/publication/228111127_Analysis_of_One-Best_MCQs_the_Difficulty_Index_Discrimination_Index_and_Distractor_Efficiency
- Arthur, H. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

- Kashyap, S. (2015). Item analysis of multiple choice questions. *International Journal Of Current Research*,7(12), 1. Retrieved from <https://www.journalcra.com/sites/default/files/issue-pdf/12220.pdf>
- Khoshaim, H. B., & Rashid, S. (2016). Assessment of the assessment tool: analysis of items in A Non-Mcq Mathematics exam. *International Journal of Instruction*,9(1), 20. Retrieved from <https://www.semanticscholar.org/paper/Assessment-of-the-Assessment-Tool%3A-Analysis-of-in-a-Khoshaim-Rashid/cb8dc2516928f2a683352c3d4e36c3e8a51bbb27>
- Nitko, A. J. (1983). *Educational Tests and Measurement: An Introduction*. New York: Marcourt Brace Jovanovich, Inc.
- Nunnally, J. C. (1964). *Educational Measurement and Evaluation*. New York: McGrawHill, Inc.
- Putri, B. D. T. (2017). *The validity analysis of English Summative test of junior high school*. *Journal Of Languages and Language Teaching*,5(1), 6. <https://e-journal.undikma.ac.id/index.php/jollt/article/view/327>
- Rea, P., & Germaine. K. (1992). *Evaluation*. New York: Oxford University Press.
- Suruchi., S. S., & Rana. (2014). Test item analysis and relationship between difficulty level and discrimination index of test items in an achievement test in Biology. *Indian journal of research*,3(6), 56. https://www.researchgate.net/publication/337712021_Analysis_of_validity_reliability_discrimination_difficulty_and_distraction_effectiveness_in_learning_assessment
- Marrie, T. et.al. (2009). “An Assessment of Functioning and Non-functioning Distractors in Multiple-Choice Questions”. *BMC Medical Education*.
- Wilmar, T. (1988). *Evaluation of Student Achievement*. Jakarta: Departemen Pendidikan dan Kebudayaan.
- Valette, R. M. (1977). *Modern Language Testing*. New York: Harcourt Brace Jovanovich, Inc.

DINAS PENDIDIKAN DAN KEBUDAYAAN



SMP NEGERI 12 LEBONG

UJIAN AKHIR SEMESTER GANJIL

TAHUN AJARAN 2020-2021



Mata Pelajaran	: Bahasa Inggris	Kelas/Semester	: VIII/Ganjil
Hari&Tanggal :	Jum'at, 04 Desember 2020	Waktu	: 120 menit

Read the text!**The Story of the Smart Parrot**

A man in Puerto Rico had a wonderful parrot. There was no another parrot like it.

It was very, very smart. This parrot would say any word-except one. He would not say the name of the town where he was born.

The name of the town was Catano. The man tried to teach the parrot to say Catano. But the bird would not say the word. At first the man was very nice, but then he got angry. 'You are a stupid bird! Why can't you say the word? Say Catano, or I will kill you!' but the parrot would not say it. Then the man got so angry that he shouted over and over, "Say Catano, or I'll kill you!" but the bird wouldn't talk.

One day after trying for many hours to make the bird say Calano, the man got very angry. He

picked up the bird and threw him into the chicken house. "You are more stupid than the chickens. Soon I will eat them, and I will eat you, too.'

In the chicken house there are four old chickens. They were for Sunday's dinner. The man put the parrot in the chicken house and left. The next day the man came back to the chicken house. He opened the door and stopped. He was very surprised at what he saw!

He saw three dead chickens on the floor. The parrot was screaming at the fourth chicken, "Say Catano, or I'll kill you!"

Answer the questions!

- The type of the text is
 - Report
 - Narrative
 - Recount
 - Descriptive
 - News item
- The function of the text is

- a. to describe the Panot in general
 - b. to share the information about the natural
 - c. to entertain the readers
 - d. to persuade the readers
 - e. to inform the reader about the parrot
4. The generic structure of the text is ...
- f. Orientation - Events - Reorientation
 - g. Orientation - Complication – Resolution
 - h. General Classification - Description
 - i. Identification – Description
 - j. Orientation - Description – Reorientation
5. What word couldn't be said by the parrot?
- a. Catano
 - b. Parrot
 - c. Smart
 - d. Except one
 - e. town
6. Which statement is TRUE according to the text?
- f. The parrot could say Catano
 - g. catano was the name of the parrot
 - h. The man never got angry at the parrot
 - i. The man ate the parrot for dinner
 - j. At last the parrot could say catano
7. What is the main character of the story?
- d. The man
 - e. The parrot
 - f. Chicken
 - d. Catano
 - e. The dead chicken
8. "The parrot was screaming at the fourth chickens"
- What does the underlined word mean...
- a. Smiling
 - b. getting angry
 - c. laughing
 - d. shouting
 - e. crying
9. How many parrots did the man have?
- a. Only one parrot

- b. Two parrot
 - c. Three parrot
 - d. Four parrot
 - e. Five parrot
10. The statements below are correct, except..
- a. The parrots was unique
 - b. The owner was angry at last to the parrot
 - c. At the end, the panot was able to say Catano
 - d. All of the chickens was killed by the parrot
 - e. The parrot killed 3 chickens

Read the text!

Do you know if you are too fat, you may have serious problem with your health? A group of doctor wrote a report about some of the effect of too much fat.

One important effect is on the heart. If you are fat, your heart has to work harder. This may lead to a heart attack; or it may lead to other heart problem.

In addition, extra fat can also change the amount of sugar in your blood. This can cause serious diseas such as diabetes.

Furthermore, high blood pressure is another possible result of being fat. More studies are needed about at all these problems. But one thing is clear, extra fat may make your life shorter.

11. The above passage is in the form of
- d. Narrative
 - d. Hortatory Exposition

- e. Descriptive Exposition e. Analytical Exposition
- f. Report
12. What is the text organization of the first paragraph?
- d. Thesis d. Orientation
- e. Argument e. Reorientation
- f. Reiteration
13. How many arguments are presented on the text?
- a. One d. Four
- b. Two e. Five
- c. Three
14. What is the function of the text above...
- a. To share the idea of being extra fat
- b. Giving information about fat in general
- c. To persuade the reader not to be overweight
- d. To entertain the reader about fat
- e. To explain how being fat
15. The generic structure of the text is ...
- f. Orientation – Events – Reorientation
- g. Orientation – Complication – Resolution
- h. General Classification – Description
- i. Thesis – Arguments – Reiteration
- j. Orientation – Description – Reorientation
16. These are the effect of too much fat, except...
- a. Cause serious disease
- b. Heart has to work harder
- c. Change the blood amount
- d. High blood pressure
- e. Causes heart problem
17. “In addition, extra fat can change the amount of sugar in your blood. This can causes serious disease such as diabetes”
The word “This” refers to ...
- a. Amount of sugar
- b. Extra fat
- c. Blood pressure
- d. Effect of too much fat
- e. Serious diseases
18. The statements below are correct, except ...
- a. Extra fat shorten our life
- b. There are several doctor did the research
- c. Extra fat causes many serious problem
- d. The research of extra fat has finally done
- e. We still need more research about the effect of too much fat
19. How many doctors wrote about the effect of being too fat?
- a. One doctors
- b. Two doctors
- c. Three doctors
- d. None
- e. Not mentioned above
20. Choose the correct statement based on the text above!
- f. The text is to persuade the readers to be fat
- g. More researches is not needed anymore
- h. The heart will work harder as the effect of being fat

- i. Being fat may not lead you to heart attack
- j. Diabetics is not serious disease

Read the text!

SPIDER

There are over 3000 species of spiders and they belong to the Arachnid family. Spiders are not insects, but are related to scorpions and other arachnid that have eight legs and no antennae. Most spiders have poor eyesight and little sense of hearing and smell. They do, however, have a well-developed sense of touch in the hairs and spines that cover their body. They may vary in size from about 1 mm to a giant 200 mm.

Spiders are known for their ability to spin silk. They are able to produce this glands in the abdomen and the silk comes out through organ called spinnerets. The silk is strong and elastic and is used to trap the spider's prey. Spiders use their webs in different ways. Some weave a net to throw over their prey, while others attach a sticky blob to the end of the silken thread and whirl around in a circle to wrap around an insect.

Spiders suck the fluids from their victims rather than eat solid food.

Then they crush the hard part, mixing them with digestive juices to make them a liquid as well. Female spiders lay large numbers of tiny egg and cover them in a silken sac. Some species carry this sac around with them. When first hatched, the young resembled adults and slowly develop to full size. Before mating, some spiders

perform courtship dances. After mating, some male spiders are eaten by the female.

Answer the questions!

21. There are over 3000 species of spiders and they belong to arachnids family. The statement tells the readers about...
 - f. Behaviour of a spider
 - g. Colony of spider
 - h. Habitat of a spider
 - i. Species of a spider
 - j. Friends of a spider
22. From the text above, we can conclude that spiders...
 - d. are not insects
 - e. eat leaves
 - f. are mammals
 - g. have least number of species
 - h. are marsupials
23. "Some **weave** a net to throw over their prey, while others attach a ticky blob to the end. The italicized word has the closest meaning to ...
 - a. Break
 - b. Knit
 - c. Make
 - d. Live
 - e. Inhabit
24. They are able to produce this from glands in the abdomen and the silk comes out through organ called spinnerets. The word abdomen has similar meaning to...
 - d. Spinnerets

- e. Organs
 - f. Gland
 - g. Spiders
 - h. None of the options correct
25. What is the best description of the silk produce by spiders?
- a. Rigid
 - b. Not elastic
 - c. Stretchy
 - d. Made of plastics
 - e. Easily broken
26. Paragraph 4th tells us about ...
- | | |
|-----------------|-------------|
| d. Reproduction | d. behavior |
| e. Appearances | e. food |
| f. family | |
27. What paragraph should we read to find spiders' appearances?
- | | |
|--------|-------|
| a. I | d. IV |
| b. II | e. V |
| c. III | |
28. Which statement is correct based on the text above?
- a. Spiders are insects
 - b. Spiders are good in eyesight
 - c. Spiders are able to spin silk
 - d. Spider has good sense of hearing and smell
 - e. Spider eat solid food
29. Which of the following is common spiders' prey?
- a. Butterfly and snake
 - b. Flies and bird
 - c. Flies and tarantula
 - d. Flies and butterflies

- e. Ants and flies
30. The statements below are correct, except ...
- f. Female spiders lay large numbers of tiny eggs
 - g. Spiders only suck the fluids of their victims
 - h. Some male spiders are eaten by the female after mating
 - i. Spiders's silk is weak and vulnerable
 - j. There are over 3000 species of spiders in the world
31. Spiders' food is explained in paragraph...
- | | |
|--------|-------|
| d. I | d. IV |
| e. II | e. V |
| f. III | |

Answer the questions!

32. Father :, Tom! The car is running fast from the sude.
Tommy : Don't worry, dad.
- | | |
|--------------|-------------|
| a. Look up | d. Keep out |
| b. Watch out | e. Walk out |
| c. Keep up | |
33. Waty : **I feel so relieved**
Asti : Oh...I can see that.
What is the underlined words axpress?
- | | |
|-------------|-------------------|
| a. Pain | d. Warning |
| b. Pleasure | e. Giving Opinion |
| c. Relief | |
34. Daughter : **Mom, My hand...ouch, it so hurts!**
Mother : Oh...How poor you are. Just take a rest

What is the underlined phrase above?

- a. Pain
- b. Pleasure
- c. Relief
- d. Warning
- e. Giving Opinion

35. Student : Excuse me, may I stay here for a night, please? I have no friends to visit.

Teacher: Sure.

Student: **I'm glad you don't mind. Thank you very much, Sir.**

The underline statement refers to....

- a. Giving advice
- b. Displeasure
- c. Pleasure
- d. Pain
- e. Giving opinion

36. Rara : I like to be here. Bogor is great.

What do you think of it?

Roni : I think Bogor is a traffic jam city.

What is the expression of the underlined word?

- a. Satisfaction
- b. Dissatisfaction
- c. Giving opinion
- d. Asking for opinion
- e. Pleasure

37. Sandra: There are too many silly cartoon on TV lately. Don't you agree?

Tony: **Not at all. Actually, I like cartoons because they're very entertaining.**

From the text above, we can conclude that Tony express...

- a. Disagreement
- b. Agreement
- c. Asking opinion
- d. Warning
- e. relief

38. Kania : In my opinion, the local TV program are not really good.

Jamal: **Absolutely, I personally feel the program is boring.**

From the text above, we can conclude that Jamal express...

- a. Disagreement
- b. Agreement
- c. Asking opinion
- d. Warning
- e. relief

39. Laras: **What do you think about the school?**

Kania: In my opinion, the school is the best school in this city?

The expression above is....

- a. Asking opinion
- b. Giving advice
- c. Pain
- d. Pleasure
- e. Relief

40. Mother: Kids, do you have any idea what we are going to do to night?

Amalia: **I suggest we go out for dinner, Mom.**

The underlined phrases express....

- a. Asking opinion
- b. Giving advice
- c. Pain
- d. Pleasure
- e. Relief

41. Handin: How do you like your room?

Jasper: **I really like it. It's very comfortable.**

The underlined phrases express..

- a. Asking opinion
- b. Giving advice
- c. Satisfaction
- d. Dissatisfaction
- e. Asking for advice

