

**ANALYZING TEACHER-MADE TEST OF ENGLISH EXAM
QUESTIONS AT SMAN 1 BANDA ACEH**

THESIS

Submitted by

FITRIA LAILAN NUSYUR

NIM. 180203132

Student of *Fakultas Tarbiyah dan Keguruan*
Department of English Language Education



FAKULTAS TARBIYAH DAN KEGURUAN
UNIVERSITAS ISLAM NEGERI AR-RANIRY BANDA ACEH

2022M / 1443 H

It has been defended in *Sidang Monev*
in front of the board of the Examination for the working paper
and has been accepted in partial fulfillment of the requirements
for the Bachelor Degree of Education on English Language Teaching

On: 22 July 2022
Thursday, 22 Dzulkhijah 1443 H

In Darussalam, Banda Aceh

Board of Examiners,

Chairperson,

Khairiah Syahabuddin, M.Ed., M.TESOL., Ph.D

Secretary,

Asma, M.Ed

Member,

Siti Khairiah, S.Ag., M.Pd

Member,

Habiburrahim, M.Com., M.S., Ph.D

Certified by:

The Dean of *Fakultas Tarbiyah dan Keguruan*
Universitas Islam Negeri Ar-Raniry Banda Aceh

Dr. Muklis Razaq, S.H., M.A.

NIP. 195903091989031001

THESIS

Submitted to *Fakultas Tarbiyah dan Keguruan*
Universitas Islam Negeri Ar-Raniry Banda Aceh
In partial Fulfillment of the Requirements for
The Bachelor Degree of Education in English Language Teaching

by

FITRIA LAILAN NUSYLR

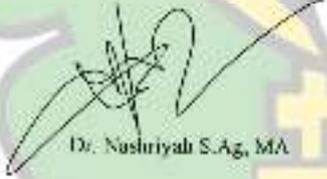
NIM 180205132

Student of *Fakultas Tarbiyah dan Keguruan*
Department of English Language Education

Approved by

Main Supervisor

Co. Supervisor


Dr. Nashriyah S.Ag, MA


Khairiah Syahabuddin, M.Hsc,ESL., M.TESOL., Ph.D

Date: / /

AR-RANIRY Date: 15 / 07 / 2022

AR-RANIRY

SURAT PERNYATAAN KEASLIAN
(Declaration of Originality)

Saya yang bertandatangan di bawah ini;

Nama : Fitria Lailan Nusyar
NIM : 180203132
Tempat/tanggal lahir : Banda Aceh, 6 Januari 2001
Alamat : Gampong Baro, Banda Aceh

Menyatakan dengan sesungguhnya bahwa skripsi yang berjudul:

**Analyzing Teacher-made Test of English Exam Questions at SMAN 1
Banda Aceh**

adalah benar-benar karya saya, kecuali semua kutipan dan referensi yang disebutkan sumbernya. Apabila terdapat kesalahan dan kekejiruan di dalamnya, maka akan sepenuhnya menjadi tanggungjawab saya. Demikianlah surat pernyataan ini saya buat dengan sesungguhnya.

Banda Aceh, 28 Juli 2022
yang membuat pernyataan.


METERAN
TEMPEL
700AKX17828579
Fitria Lailan Nusyar

AR-RANIRY

ACKNOWLEDGEMENT

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillah rabbil'amin. In the name of Allah, the Most Beneficent and The Most Merciful. All praises belong to Allah, the Lord of the world, who has bestowed us with his mercy and blessing, and who has given us strength, health, the will to learn, and intelligence. *Shalawat* and *Salam* are always granted to the best human being, the chosen one, Rasulullah Muhammad peace be upon him, along with his family and companions.

I believe that the completion of this thesis was not merely due to my efforts, but also because of the blessing of Allah and support from my parents and my beloved family and friends. First of all, I would like to express my deep gratitude to my supervisor Dr. Nashriyah, S.Ag, M.A. and Mrs. Khairiah Syahabuddin M.HSc.ESL., M.TESOL, Ph.D, who is also my academic advisor, for the guidance given to me sincerely. For all the goodness of my two supervisors, may Allah's grace always be with them.

As a form of appreciation, I would like to thank myself for doing my best, my parents and my siblings who have always prayed for and supported me, and to my beloved friends who have helped me a lot in acknowledging more knowledge for the continuity of this thesis. Thank you for all the motivation and kindness.

Banda Aceh, 28 July 2022

Fitria Lailan Nusyur

ABSTRACT

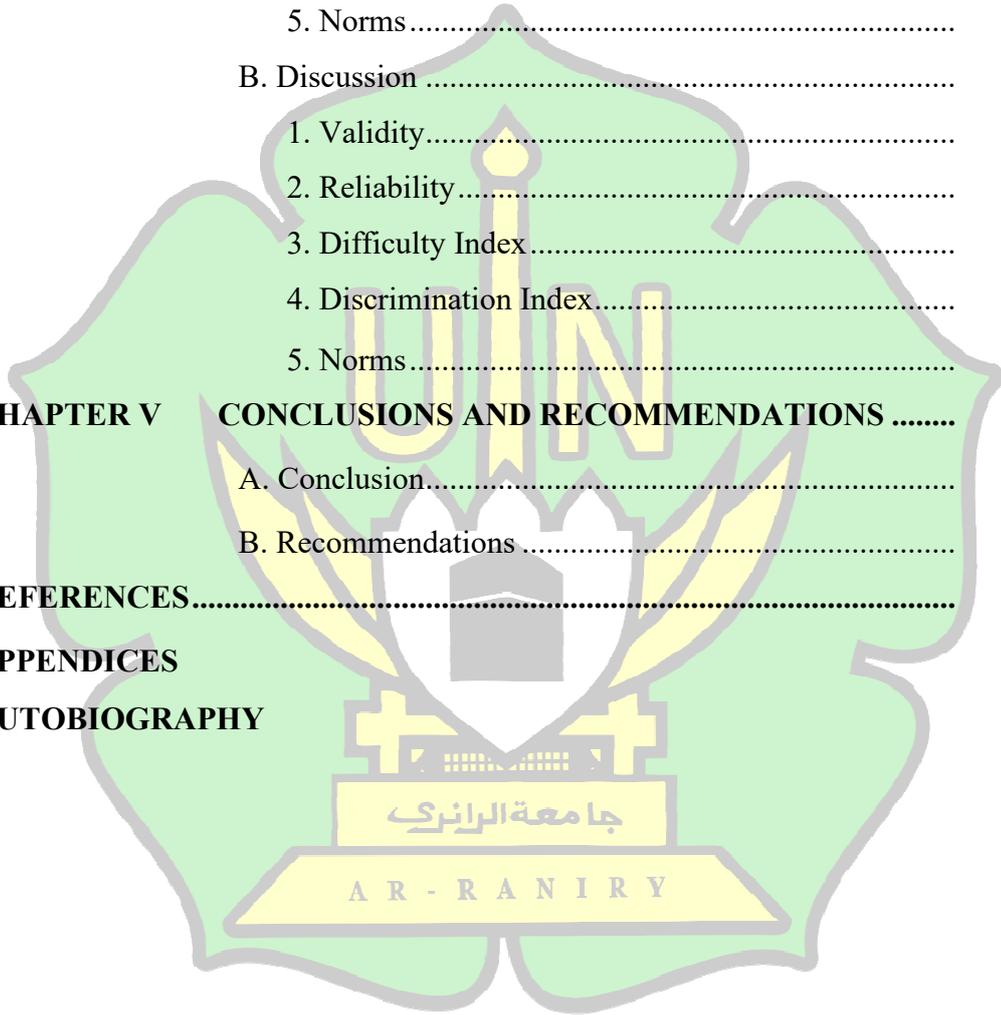
Name : Fitria Lailan Nusyur
NIM : 180203132
Faculty : *Fakultas Tarbiyah dan Keguruan*
Major : Department of English Language Education
Thesis Tittle : Analyzing Teacher-made Test of English Exam Questions at SMAN 1 Banda Aceh
Main Supervisor : Dr. Nashriyah, S.Ag, M.A
Co-Supervisor : Khairiah Syahabuddin, M.H.Sc, M. TESOL, Ph.D.
Keywords : Teacher-Made Test, Multiple-Choice Questions, Summative Examination, Difficulty Index, Discrimination Index, Norms,

This research analyzes the English summative examination for grade XI at SMAN 1 Banda Aceh. This quantitative research used a descriptive approach. The data was collected using the documentation method where the English exam questions, students' answer sheets, answer keys, and a list of students' names was collected. 250 students participated in the examination. Therefore, there were 250 answer sheets analyzed. The instrument analysis was done using IBM SPSS Statistics 25 and Microsoft Office Excel 2010. This research aims to answer the research questions about the level of difficulty and discrimination as well as to see the quality of multiple-choice questions based on the guidelines (or called norms) published by Puspendik Kemdikbud. The validity and reliability were analyzed using IBM SPSS Statistics 25 with results >0.005 , categorized as valid, and the reliability range score of 0.82, categorized as highly reliable. The difficulty index, discrimination index, and norms were analyzed using Microsoft Office Excel. The result showed that the 17 items (57%) were categorized as average, indicated that the level of difficulty was average. The result of the discrimination index showed that 8 items (27%) were categorized as bad and 22 items (73%) categorized from average to excellent, which indicated that many question items need to be revised. Moreover, for the norms, it was found that they fulfilled the norms.

TABLE OF CONTENTS

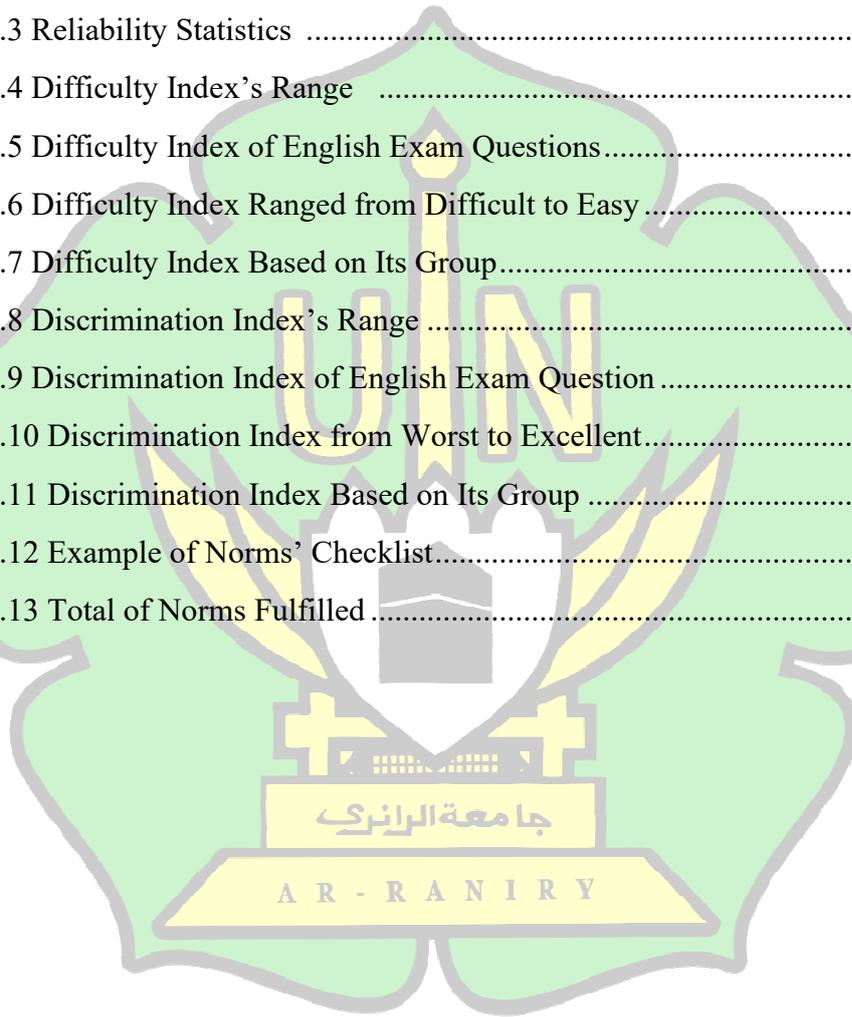
DECLARATION OF ORIGINALITY.....	i
ACKNOWLEDGEMENT.....	i
ABSTRACT.....	iii
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
LIST OF APPENDICES.....	viii
CHAPTER I INTRODUCTION.....	1
A. Background of Study.....	1
B. Research Questions.....	9
C. Research Aims.....	9
D. Significance of The Study.....	10
E. Research Terminologies.....	11
CHAPTER II LITERATURE REVIEW.....	17
A. English Examination.....	17
1. The definition of English Examination.....	17
2. The Purpose of Examination.....	20
3. The Function of Examination.....	21
4. The Technique of Evaluation.....	22
B. Multiple-choice Questions (MCQs).....	24
1. Definition of Multiple-choice Questions (MCQs).....	24
2. The Criteria of Good MCQs.....	25
3. The Level of MCQs.....	27
4. The Strengths and Limitations of MCQs.....	29
CHAPTER III RESEARCH METHODOLOGY	
A. Research Design.....	31
B. Research Location.....	32
C. Technique of Data Collection.....	33
D. Technique of Data Analysis.....	33

CHAPTER IV	FINDINGS AND DISCUSSION	38
	A. Research Findings	38
	1. Validity.....	39
	2. Reliability.....	40
	3. Difficulty Index.....	41
	4. Discrimination Index.....	47
	5. Norms.....	52
	B. Discussion	54
	1. Validity.....	56
	2. Reliability.....	57
	3. Difficulty Index.....	58
	4. Discrimination Index.....	60
	5. Norms.....	61
CHAPTER V	CONCLUSIONS AND RECOMMENDATIONS	64
	A. Conclusion.....	64
	B. Recommendations	65
REFERENCES.....		67
APPENDICES		
AUTOBIOGRAPHY		



LIST OF TABLES

Table 1.1 Norms of Good Quality Multiple-Choice Questions	16
Table 3.1 Example of Item Analysis.....	34
Table 3.2 Guidelines of A Good Multiple-Choice Question (MCQ).....	37
Table 4.1 The Validity of Multiple-Choice Questions.....	40
Table 4.2 Reliability Range.....	41
Table 4.3 Reliability Statistics	41
Table 4.4 Difficulty Index's Range	42
Table 4.5 Difficulty Index of English Exam Questions.....	44
Table 4.6 Difficulty Index Ranged from Difficult to Easy	45
Table 4.7 Difficulty Index Based on Its Group.....	46
Table 4.8 Discrimination Index's Range	47
Table 4.9 Discrimination Index of English Exam Question	49
Table 4.10 Discrimination Index from Worst to Excellent.....	50
Table 4.11 Discrimination Index Based on Its Group	51
Table 4.12 Example of Norms' Checklist.....	52
Table 4.13 Total of Norms Fulfilled	53



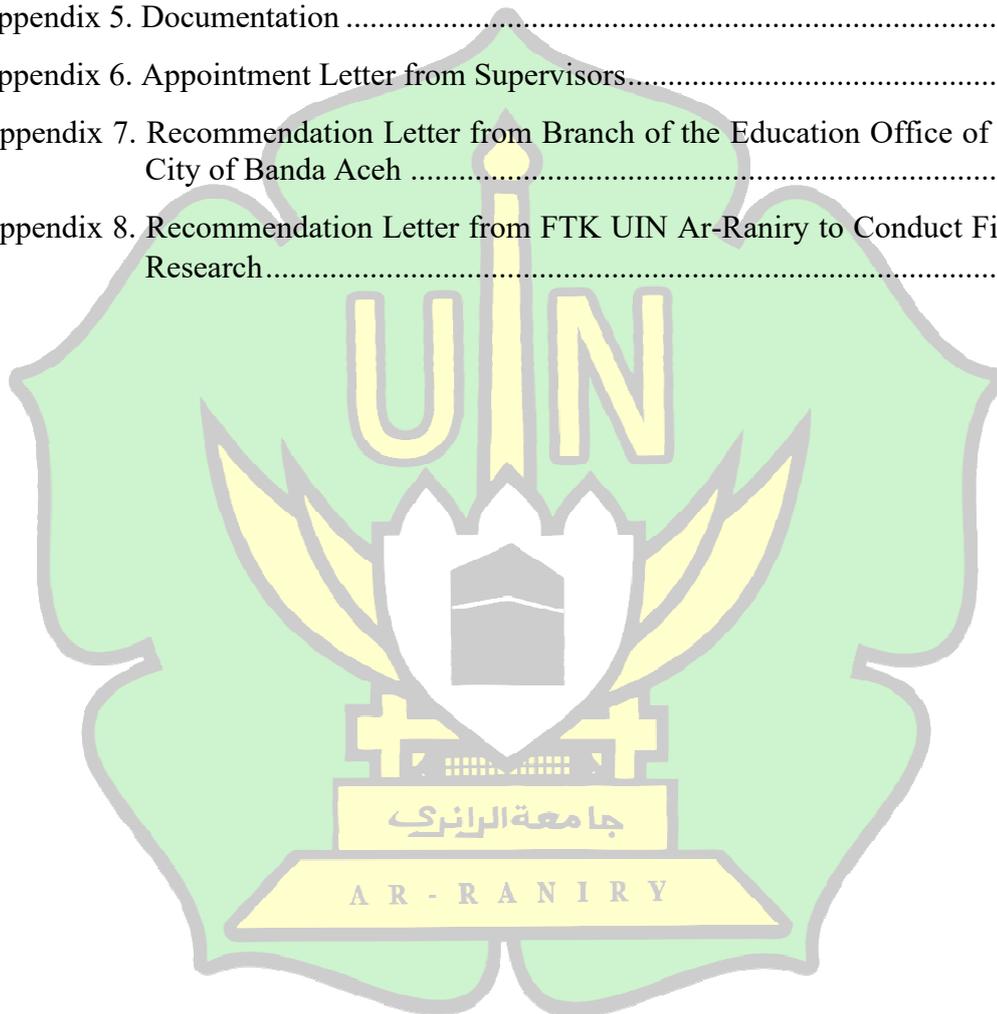
LIST OF FIGURES

Figure 2.1 An Example of Multiple-Choice Question.....	25
Figure 4.1 Students Responses on Each Item	42
Figure 4.2 Percentage of Difficulty Index.....	47
Figure 4.3 The Score Distinction of Upper and Lower Scored Students.....	48
Figure 4.4 Percentage of Discrimination Index	52



LIST OF APPENDICES

Appendix 1. English Examination Questions	73
Appendix 2. The Distribution of Difficulty and Discrimination Indices	81
Appendix 3. Validity of Question Items	89
Appendix 4. Reliability of Question Items	92
Appendix 5. Documentation	93
Appendix 6. Appointment Letter from Supervisors.....	94
Appendix 7. Recommendation Letter from Branch of the Education Office of the City of Banda Aceh	95
Appendix 8. Recommendation Letter from FTK UIN Ar-Raniry to Conduct Field Research.....	96



CHAPTER I

INTRODUCTION

This chapter provides information on the background, research questions, objective, aim, significance, and terminology of the study.

A. Background of Study

Evaluation is a vital aspect of the teaching-learning process. It becomes the parameter to test the effectiveness of all teaching-learning processes. As Liando, Setiawana, and Wuntu (2021) proposed, the evaluation helps track the progress of teaching-learning to provide a valid review regularly to create helpful feedback for future teaching-learning processes. In evaluation, the teacher who teaches and evaluates is able to see the ability and improvement of the students. Furthermore, evaluation could be an excellent process to review the teaching process done by the teachers. After the teaching-learning process has been done throughout the semester, the school authorities will start to do an evaluation. The student's ability and teachers' success in teaching will be evaluated in the evaluation. It involves creating aims, designing the process, and assessing the overall performance (Lumentut & Lengkoan, 2020). The evaluation compares what is planned for achievement (such as the learning process, attitudes, and progress in learning) and what students have gone through.

In conducting an evaluation, the teacher also conducts an assessment as a part of the evaluation to measure students' ability in learning. In its implementation, the evaluation is divided into four, namely formative, summative, diagnostic, and

placement evaluation. As a part of an evaluation, the formative, summative, diagnostic, selective, and placement evaluation has its own role in creating successful feedback for the learning process. The formative assessment is a daily assessment conducted by the teachers to monitor students' progress in daily meetings. While in the summative assessment, students are assessed at the end of the semester to evaluate the overall progress throughout the semester, which is not only evaluating students' performance and progress but also the success of the teacher in the learning process. Briefly about the diagnostic assessment, it is used to measure students' ability in a particular aspect or material and it is often done in the remedial evaluation. Furthermore, the placement evaluation is an evaluation needed to acknowledge the students' ability to fulfil the requirements of a program. In this research, the researcher will elaborate more about the summative evaluation that is regularly implemented in every school.

In the process of evaluation in the school, the most common implemented evaluations are formative evaluation and summative evaluation. Both evaluations can be formed as written or oral tests. Doing the formative evaluation, teachers assess students by their daily performances through projects, quizzes, homework, or presentation works. Summative evaluation, as the overall overview of students' progress, is conducted at the end of the semester. In summative assessment, students are given three types of tests which are multiple-choice, fill-in-the-blanks, and essays type of questions.

In its implementation in the semester exam, the three types of tests (multiple-choice, fill-in-the-blank, and essay) given by the teacher have their respective roles in measuring students' abilities and knowledge. In multiple-choice questions, students are given questions with 4 or 5 answer choices. These multiple-choice questions tend to measure students' knowledge in finding the correct answer among all the choice options. The difficulty level on multiple-choice questions is low and generally have the most items on the questions. Meanwhile, in fill-in-the-blank questions, students must fill the blank space with the correct answer. The difficulty level of this question ranges from easy to medium. Accordingly, there is an essay-typed question where students are asked to write a complete and precise description according to the question asked. Generally, the number of essay questions are the least among other questions but has the most significant weight of points. Among the three forms of questions, the researcher discusses further about the multiple-choice questions that are used to test students' knowledge and abilities in semester exams. Thus, the process of creating multiple-choice questions is an essential process in an evaluation. Therefore, the question items for summative assessment can be made by National Education Standards Agency or Badan Standar Nasional Pendidikan (BSNP) or by the teacher.

Based on its creator, there are two types of multiple-choice questions: non-teacher-made and teacher-made. According to Brown (2003), the test can be conducted in two ways: teacher-created and standard-bearer-created. Similarly, In Indonesia, the examination questions can be arranged in two ways which are arranged by the Badan Standar Nasional Pendidikan (BSNP) and arranged by

teachers. BSNP is an independent and professional institution with a mission to develop, monitor the implementation, and evaluate the implementation of national education standards. The examination questions arranged by BNSP are used for National Examination (NE).

In contrast, the multiple-choice questions that the teacher arranges are called teacher-made test. The teacher-made test is the most used type in assessing students' knowledge and ability in summative assessment. A teacher-made test is a written or oral test arranged by the teacher that is designed as a tool or media that contains indicators of learning materials to assess students' abilities and achievements against related indicators. Generally, teacher-made tests contain question items which are similar to the types of questions students practice in daily teaching and learning activities. The similarities between the examination question items and the problems students solve during daily learning makes it easier for students to answer question items because they are familiar with such form of a question. Teachers put a standard score to grade students' work by giving them scores or alphabet codes for their achievement. "To represent the students' abilities, the teachers usually use numbers such as 1 to 100, 1 to 10, or 1 to 4, or they may use category systems such as excellent, good, fair, poor, and very poor or A, B, C, D, E, or pass-fail" (Gultom, 2016, p.191). In the teacher-made test, the teacher is the one who arranges the questions and determines the range of scores for the students. Therefore, the teacher can account for both of these things.

Since the teacher-made test is used to assess students' progress throughout the semester, it must be accountable to the teacher. In the sense that teachers are responsible for the validity and reliability of a test they make, they should be able to measure the accuracy between a measurement tool and the aspect to be measured, which means the question item is proven for its validity. For the question items, they should have a good level of trust, have a minimum error, and be based on the indicators the students have learned. In learning evaluation, the measurement tool is a test, and the aspect of being measured is the student's ability to learn. The accountability also depends on the reliability of the question items, which means it holds the consistency owned by a measurement device. If two tests are carried out at different times, the measurements will be the same. In evaluating using tests, if a test carried out is repeated for testing on different subjects, and the results are the same, then a test deserves to be recognized for its reliability. To increase the value of test accountability, the calculation of the difficulty index, discrimination index, and analysis of the norms are needed.

Increasing the value of the accountability of a test, the calculation of difficulty index, discrimination index, and the analysis based on the norms become essential processes. In educational practice, it is necessary to measure the difficulty index and discrimination index (Kurniawan, Syifa, Huda, & Kesuma, 2021). The difficulty index calculates the difficulty level of a question item, whether it is easy, medium, or difficult. The higher the difficulty index, the more complex the question item is, and vice versa. By calculating the difficulty index, the teachers can distinguish between capable students and less capable ones. In comparison, the

discrimination index is an index that is used to analyze the ability of a question item to discriminate students' ability. Also, the analysis of the discrimination index is necessary to see the quality of a test, either the overall test or only specific question items that need to be analyzed (Iskandar & Rizal, 2018). Teachers can analyze the students' ability and the quality of the question items through the discrimination index. Therefore, the analysis of the difficulty index and discrimination index has become an essential issue in the field of education, where many researchers are conducting research in many different areas. Also, Susanto, Rinaldi, and Novalia (2015) believe that analyzing the question items is needed to create a well-qualified question. Well-qualified question items have the balance between the difficulty index and the discrimination index and be able to maintain their quality.

The accountability and quality of the multiple-choice questions are essential since both of them are directly related to the evaluation of the education process. The quality of the multiple-choice questions is strengthened through the norms (guidelines) used in constructing them (Burton, 1991). Norms are the criteria to assess the quality of questions based on three basic things which are, material, construction, and language. (Yanti, Paramartha, & Wahyuni, 2020). Norms are essential because in the implementation of the examination, a question item that fulfilled the norms helps the test-takers to understand the purpose of the questions being tested. Haladyna (2004) focuses on the norms in writing an excellent multiple-choice question. These norms have been issued by Puspendik Kemendikbud (2019) as an official forum for the assessment of national education

in Indonesia. In these norms, the aspects considered are writing, grammar, and clarity of the content in the questions.

Several studies were conducted to analyze the quality of test items and their validity, reliability, difficulty index, and discrimination index. Susanto et al. (2015) and Kurniawan et al. (2022) found that the test items lack validity. Effendi and Mayuni (2022) and Lestari, Ikhsanuddi, and Bunau (2014) most of the test items have a good quality of validity but lack the test items' reliability. Rahmi and Suhartono (2018) and Liando et al. (2021) showed that the overall result of the test items is good, but some revision is still needed. Setiyana (2016), Kurniawan et al. (2022), and Lestari et al. (2014) found several problems in the effectiveness of the distractors. While Prihartini, Buska, and Hasnah (2019) analyzed the test items at the level of the University and found that there are some lacks in the item difficulty and item discrimination of the teacher-made test. Furthermore, at the school level, Toksöz and Ertunç (2017) found that the test items were unable to discriminate between high-achieving students and low-achieving students. Ugwu and Mkpuma (2019) and Yanti et al. (2020) analyzed the quality of the teacher-made test based on the criteria of a good test. The two studies showed that the test items have good quality.

Along with the previous studies' results, Yanti et al. (2020), Kurniawan et al. (2022), Toksöz and Ertunç (2017), Ugwu and Mkpuma (2019), Prihartini et al. (2019), Setiyana (2016), Effendi and Mayuni (2022) provided the suggestions proposed by the teacher as the test makers regarding the improvement of test items.

The teachers suggest that the authorities provide training and seminars on how to create good test items to improve the quality of the teacher-made test. In addition, Sugianto (2011), Ugwu and Mkpuma (2019), Setiyana (2016), and Rahmi and Suhartono (2018) believe that the teacher expresses their concern about the quality of the teacher-made test by suggesting that the items should be in coherence with curriculum and indicators

Based on the previous studies, the analysis of the difficulty index and the discrimination index is the obligation for every subject teacher to do as well as the analysis of the norms. In fact, the reality is that the teachers do not run the analysis. The underlying reasons why the teachers do not operate the analysis are

1. lack of knowledge on how to do the item analysis,
2. the analysis itself burdens the teacher, and
3. the awareness of the importance of doing tests item analysis is deficient.

After learning about the reasons mentioned, the researcher believes that the analysis helps in reviewing and evaluating the quality of the question items, so in the future examination, the quality of the questions will increase. By doing the analysis, the teacher can revise the question items for upcoming examinations (Quaigrain & Arhin, 2017). This idea is similar to what Shete, Lakhar, and Khan (2015) imply, in which by doing the analysis, the quality of the question items can be evaluated, and tricky questions can be revised for future examination. Therefore, the researcher perceives the need to analyze the difficulty index, discrimination index, and norms based on a quantitative research method on English exam

questions for grade XI due to the importance of doing the analysis. Thus, this study aims to: (1) determine the difficulty index and discrimination index of the English Teacher-made test at SMAN 1 Banda Aceh and (2) determine the quality of the questions based on the norms.

Therefore, why this research is worth to be conducted is also because the researcher found that many researchers have analysed the difficulty index and discrimination index, and some have done an analysis on norms. However, researchers have not found research that combines the two, so the researchers decided to conduct this study entitled "**Analyzing Teacher-made Test of English Exam Questions at SMAN 1 Banda Aceh.**"

B. Research Questions

In the light of the research background elaboration previously, this research aims to seek the answer to the following questions as follows:

1. What is the level of difficulty and discrimination indices of the teacher-made test at SMAN 1 Banda Aceh?
2. Do the multiple-choice questions meet the norms published by *Puspendik Kemdikbud*?

C. Research Aims

1. To determine the difficulty and discrimination indices of the teacher-made test at SMAN 1 Banda Aceh.
2. To determine the quality of the multiple-choice questions based on the norms published by the *Puspendik Kemdikbud*.

D. Significance of The Study

The significant contribution of this research was at giving both theoretically and practically.

1. Theoretical Significance

Many studies have been conducted on item analysis at various levels of education, ranging from elementary school, junior high school, high school, and university. Given many studies that have been done, it proves that item analysis is a crucial process to conduct. The exam questions are directly related to the evaluation process in the education unit because it acts as a medium to measure the overall learning that has occurred while simultaneously measuring students' knowledge abilities. Therefore, to add insight into item analysis, researchers conducted this research contributing in the progress in the field of education.

2. Practical Significance

Practically, this research also has several direct impacts on students, teachers, and society. This research was conducted at SMAN 1 Banda Aceh in the 2021/2022 academic year on the odd semester exam for class XI. The students involved in this semester's exam came from 8 classes with 250 students. For all these students, this research can be a reference for them to see their development in this exam and see their classmates' overall development, which will help them to develop better in the upcoming exams.

Then for the school authorities and teachers, this research can contribute new information related to the semester exams conducted. So, if the school authorities and teachers want changes in the future, they can refer to this research. Also, to the researcher and the public, this research adds new insights related to item analysis so that this can be a reference material for future authors to conduct further research.

E. Research Terminologies

Before we examine further about this research, several terms must be explained so that there is no misleading information in understanding the content of the study.

1. Analysis

The analysis is how we deal with the issues or problems in a sequential way. According to Cambridge Dictionary, analysis is “the act of studying or examining something in detail, in order to discover or understand more about it or your opinion and judgement after doing this”. A problem will be solved from the most significant units to the minor units in conducting analysis. The analysis will examine a problem by paying attention to the details that exist in each element so that nothing is missed. The researcher must conduct the analysis sequentially, precisely, carefully, in detail, and based on the selected theory. In this case, the analysis focuses on the question items used in the examination's questions.

2. *Examination*

The examination is the process in which one is tested for what one has learned. According to Cambridge Dictionary, the examination is "the act of looking at or considering something carefully to discover something." Exams are the determinant of student success for one semester. What have been learned by students will be included in the exam questions, so students must prepare themselves and to be able to answer all the questions. It is also the determinant of the achievement of learning objectives. This can be seen from the accumulation of the number of correct questions answered by students. If most students cannot answer a particular question, then the learning objectives associated with the question will not be achieved.

3. *Standardized Test*

Standardized test is a test that is fairer in measurement than non-standardized test. Standardized tests are usually applied to large groups with the same topics given, the same processing time, and the same scoring system. It doesn't have to be an oral or a written question because it can be both, it doesn't have to be a complicated question, but it can be an easy one. But in general, the types of questions that are widely used in standardized tests are multiple-choice questions because the assessment process is easy to do. The assessment process can be done manually or using a computer so that the level of accuracy can be guaranteed. Although standardized tests can also be in the form of quizzes or daily assessments given by teachers, their application is often carried out on large-scale tests such as the National

Examination (NE). In Indonesia, the government believes that standardized test is a fair and objective assessment the ability of a large number of students (Setiawan, Garnier, & Isnaeni, 2019). “The government believes that a standardized national examination can improve the quality of education since the students and the schools will try hard to meet the demand of the examination” (Ilma, 2017, p.42).

4. *Teacher-made Test*

It is a type whose design is carried out by the subject teacher following the planned indicators and indicators that have been implemented. Teacher should learn about pedagogical competences such as teaching methodology, teaching evaluation, and curriculum analysis (Habiburrahim, 2019). Also, classroom management is a skill that need to be mastered by the teacher to create a positive and creative learning environment (Muluk Habiburrahim, Zulfikar, Akmal, Nasriyanti, & Safrul., 2021). Accordingly, the teacher also acts as an evaluator to evaluate students' achievement and understanding of specific topics being tested. A good teacher is not only able to teach well to prepare students for a test but is also able to evaluate tests made by himself (Lebagi, Sumardi, and Sudjoko., 2017). By evaluating the teacher's teacher-made test, the teacher can evaluate the learning process, the student's understanding, the implementation of the teaching methods, and the material used.

5. *Assessment*

Assessment is a process of gathering information to measure student's achievement towards learning objectives. Assessment is divided into two categories, namely:

4.1. Formative assessment

Formative assessment is a type of assessment where the data collection process is carried out while learning occurs.

4.2. Summative assessment

It is a type of assessment in which students' overall ability is measured at the end of the lesson by combining one or several essential competencies.

6. *Difficulty Index*

Difficulty index is a calculation to determine the items' level of difficulty. Finding out the level of difficulty of a question can be calculated by dividing the number of correct answers (on a question) by the total number of students. For example, in question number 1, 9 out of 10 students answered correctly. Accordingly, the formula is nine divided by ten. If a question has a difficulty index value above 0.70, then the question is categorized as an easy question. If a question has a difficulty index value below 0.30, then the question is categorized as a difficult question.

7. *Discrimination Index*

A discrimination index is a process to determine the quality of a question. To find out the discrimination index of a question, the formula that can be used is to divide the number of students into two parts, namely higher scores students and lower scores students. If there are 20 students in a class, the ten students with the top scores are categorized as higher scores students, and the ten students with the lowest scores are categorized as lower scores. Then the number of answers for the higher scores students who have been reduced by the number of answers for the lower scores students is divided by the number of students who have been divided by two. For example, 9 of the higher-scoring students answered correctly for answer number 1, and 8 of the lower scores answered correctly for answer number 1. Then nine is subtracted by eight and then divided by 10 (the number of 20 students is divided by 2). The discrimination index has a range of values between 0-1, then the value is positive. Whereas those with the value range of -1-0 have negative value. Items that have a negative value are possible to be problematic questions and must be reviewed by the subject teacher.

8. *Norms*

Norms are rules for constructing examination questions, especially multiple-choice questions. Officially in Indonesia, norms are published by Puspendik Kemdikbud (2019) as a reference for teachers as multiple-choice question makers in designing questions that have homogeneity in the writing structure, have accuracy in question delivery, and have a clear

grammatical structure. The official norms published by the Puspendik Kemdikbud (2019) show the urgency of writing excellent and correct questions.

Table 1.1

Norms of Good Quality Multiple-Choice Questions

No	Norms Description
1	Reflecting basic competencies
2	Not depending on the previous options
3	Giving content clear focus
4	Avoiding opinion-based items
5	Being grammatically correct
6	Having correct spelling
7	Not containing clues
8	Formatting options vertically
9	Having correct use of punctuation and capitalization
10	Not containing double negatives
11	Having homogenous options in content and grammatical
12	Having one correct answer
13	Having options with relatively the same length
14	Placing options in logical and numerical order
15	Not repeating options with the same words or phrase
16	Not having overlapping options
17	Having plausible distractors
18	Not using “none of the above” or “all of the above”

Source: Puspendik Kemdikbud (2019)

جامعة الرانيري

A R - R A N I R Y

CHAPTER II

LITERATURE REVIEW

This chapter presents several theories that are relevant with the research. It includes elaboration of the theory that is related to examination and multiple-choice questions.

A. English Examination

1. The definition of English Examination

English is one of the most widely spoken languages in the world for many purposes and transactions in human life (Algeo & Butcher, 2013). “English becomes the most popular language used around the world to communicate with people from different origins and linguistic backgrounds” (Munzaki et al., 106, p.12). English had been through a long process to become an international language when previously the British had used Latin and French as their official languages for several centuries before English was revived in the era of Old English, Middle English, and Modern English, which development began in 1500 regarding its grammatical, transactional, and functional (Subiyati, 1995). In some countries, English is the second language after the mother tongue.

English as a Second Language (ESL) refers to the use of English in countries where the primary language is not English, and the position of English is as a second language after the primary language or mother tongue (Carrasquillo, 2013). In Indonesia, English is also used alongside Indonesian and regional languages. The position of English Indonesia is as

the first foreign language used in Indonesia, which is included in the curriculum as a subject from elementary school to high school so that the function and use of English in Indonesia do not shift the position of Indonesian as an official language and regional languages as a language custom (Agustin, 2015). “In some EFL (English as Foreign Language) countries, learning English becomes the primary focus on learning activity, however students get difficulty and lack of interest in learning English” (Munzaki et al., 2016, p.12).

Since English is included in the school curriculum, the teaching-learning process is constantly evaluated. Evaluation is a mechanism that will increase the quality and function of the teaching and learning process (Rea-Dickins, Germaine, & Germaine, 1992). In addition, it is a continuous process related to collecting information to assess the learning design that has been carried out (Febriana, 2021). It becomes the medium for teachers to analyze students' success at the end of the learning process (Lalogiroth, & Tatipang, 2020). A good evaluation covers many aspects, such as learning design and learning strategies, implementing planned learning programs, and evaluating the results of learning programs (Widoyoko, 2009). "The evaluation includes measurement, assessment, and testing" (Haryanto, 2020, p. 16). It is the process in which we examine some situation that will lead us to a conclusion (Kizlik, 2012). Similarly, the teacher tries to interpret the students' evaluation results into a valid score (Van, 2008).

The process of evaluating English teaching-learning could be done through the assessment process. Therefore, the following explanation elaborates more on the assessment process.

a) Types of Assessment

(1) Formative assessment

Brown (2004) mentioned that formative assessment is the assessment that is implemented during the teaching-learning process. Formative assessment involves teachers and students actively, which aims to provide feedback on learning that has already taken place (Graham, Hebert, and Haris., 2015). Formative assessment is also known as daily assessment. In Indonesia, daily assessment is called Penilaian Harian (PH). Daily assessments are carried out after the teacher has taught several essential competencies to determine the level of understanding and ability of students to the material that has been taught, as well as train students to explain, elaborate, and express opinions (Nuryati & Astriana, 2020).

(2) Summative assessment

Summative assessment is carried out at the end of the learning process to evaluate students' abilities and the programs administered (Graham et al., 2015). In Indonesia, summative assessments are called final semester assessments or Penilaian

Akhir Semester (PAS) and year-end assessments or Penilaian Akhir Tahun (PAT).

The final semester assessment (PAS) is carried out at the end of the first semester after students carry out the entire learning process to determine the students' understanding of the material in semester 1 (Rusmayani, 2020). Final semester assessment is one of the evaluation processes that aim to improve the quality of education in schools by presenting an overview of students' understanding of the material for one semester (Wahyuni & Putri, 2020). Whereas the year-end assessment or Penilaian Akhir Tahun (PAT) is part of a summative assessment accomplished routinely by the school at the end of the school year to evaluate the ongoing learning process (Latip & Sutantri, 2021).

2. The Purpose of Examination

In general, all types of examination try to evaluate specific processes that have been implemented. In education, examination tries to evaluate the abilities of students and teachers and the learning process. As proposed by Athanasou (2020), evaluations try to improve, make decisions, respond to the controversy, compare, judge, manage requirements, determine effectiveness, assess quality, and determine performance. However, the evaluation also evaluates students' understanding and attitudes in the classroom. As stated by Rahmi, Gatot, and Suhartono (2018), evaluation helps to increase the quality of

teaching-learning performance for the future teaching-learning process.

"Specifically, the purpose of the evaluation in education is to determine the level of ownership and understanding of students towards the subject matter, both in cognitive, psychomotor, and affective aspects" (Elis & Rusdiana, 2015, p.9).

3. The Function of Examination

Examination tends to figure out how well the students have understood the learning material to reach the goal of the learning process. Referring to Rukajat (2018), evaluation has several functions as follows:

1. As a tool to find out whether students have mastered the knowledge, values, and skills that a teacher has given.
2. To find out aspects of students' weaknesses in carrying out learning activities.
3. As a tool to distinguish the student's achievement in learning activities.
4. As a means of feedback for a teacher, which comes from students.
5. As a tool to determine student learning progress.
6. As the primary material of learning outcomes reported to students' parents.

Evaluation manages to raise an educational program, gauge its effectiveness, and improve its quality (Liando et al., 2021). An evaluation has covered all the functions that have been mentioned, so an evaluation must be implemented by focusing on the aspects of validity and reliability so that all these functions can be well-administered.

4. *The Technique of Evaluation*

(1) Non-test technique

Non-test technique is an assessment technique that does not involve a test. Assessment is done through utilizing observation, interview, questionnaire, case study, and collecting documents (Shobariyah, 2018). Observation is a procedure to gather information about a condition through a systematic inspection (Sudijono, 2001). Another non-test technique is the interview. The interview is a way to collect information by communicating directly or indirectly with sources to obtain information. Interview is direct communication between the information seeker and the source of information. “It is imperative to note that this framework served as a “compass” and not a rule” (Owen, 2014, p.9). Whereas the questionnaire distribution is the process in which information is collected through the answer choices given to the object. The case study is an observation of individuals to see their development. In educational practice, case studies can be applied to see students’ daily progress. Simultaneously, document collection is the process of collecting data through analysis of the available documents for evaluation purposes. The evaluation provides the students’ progress through curriculum vitae such as educational background, learning style, and achievement (Shobariyah, 2018).

(2) Test technique

1. Definition of Test

A test is a tool or procedure to measure the value of students' performance (Rukajat, 2018). What is measured in the test is the cognitive aspect of learners. Furthermore, the test is one of the most common instruments to evaluate English language teaching (Liando et al., 2021). The importance of the test, as explored by Ugwu and Mpkuma (2019), are

1. to provide an understanding of how far the objectives have been reached,
2. to ensure how adequate the methods work, and
3. to explore how advantageous the program is.

Moreover, a test should be credible. A credible test will lead the teacher to validly interpret students' abilities (Effendi & Mayuni, 2022). Further, a test should also complete the criteria of a good test.

2. Criteria of A Good Test

According to Bachman and Palmer (1996), the criteria of a good test include: (a) validity, (b) reliability, (c) practicality, (d) discrimination, and (e) authenticity.

Validity means the test successfully measured what is intended to be measured using a dependable test (Rukajat, 2018). There should be a coherency between the tool and its function to create plausible data. It should be a helpful tool for measuring students' cognitive development (Rahmi et al., 2018). Besides validity, reliability also

plays a vital role in creating a good test. Reliability is the ability of the test score to be dependent (Rogler, 2014). It holds the same meaning as the test result should be consistent even though a different person takes it at a different time (Liando et al., 2021). Therefore, if most of the items are matched, then it can be said that the test is highly reliable (Rajhy, 2014).

B. Multiple-choice Questions (MCQs)

1. Definition of Multiple-choice Questions (MCQs)

Multiple-choice Questions (MCQs) are a form of exam question that provides several answer options to choose from (Alsubait, Parsia, & Sattler., 2014). It is also known as "items," which have been widely implemented (Gajjar, Sharma, Kumar, & Rana., 2014). It has been widely used in educational assessments and has become an objective instrument for many assessments (Dehnad, Nasser, & Hosseini., 2014). In an exam that requires both measuring student understanding and knowledge and application analysis, multiple-choice questions (MCQs) are a good test to apply that can be compiled and assessed. (Abdelhameed et al., 2005). The questions used in the MCQs can be made directly by the teacher or guided by the question bank, which is the reference for preparing the questions (Abdulghani et al., 2015). MCQs can provide feedback for student development and the learning process and provide a good analysis of the items (Zaidi et al., 2018). MCQs have a pattern that begins with a question called a "stem," then the correct

answer must be selected based on the answer choices between one correct answer and several distracting options (Towns, 2014). An example of a multiple-choice question is as follows:

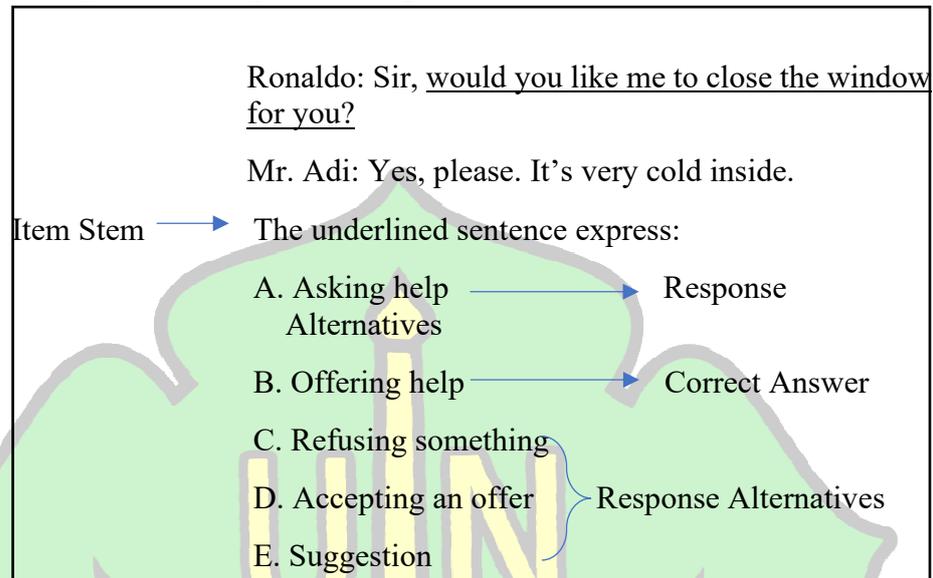


Figure 2.1. An example of multiple-choice question

2. The Criteria of Good MCQs

As proposed by Coughlin & Featherstone (2017), a good MCQ should avoid the option items that include grammatical cues, logical cues, the use of absolute terms, longer correct answers, word repeats, and the concept of convergence strategy. The more detailed elaboration of the criteria is as follows:

a) Grammatical cues

Question makers are usually more focused on the correct answer option. Therefore, to avoid the question maker entering detailed and conspicuous information on the correct answer option, the correct answer and other answers (including distractors) must have the same length (Coughlin & Featherstone, 2017).

b) Logical cues

Logical cues can be formed if several options have the same theme so that the examinees will find clues that the correct answer is in the options with the same theme (Coughlin & Featherstone, 2017).

c) The use of absolute terms

Absolute terms can be in the form of words such as "always" or "never" in the answer options so that the examinees can eliminate these options to choose the option that is more likely to be the correct answer without the presence of absolute terms (Coughlin & Featherstone, 2017).

d) Longer correct answers

Longer correct answers are related to grammatical clues where the question maker adds detailed information to the correct answer option so that it becomes longer than the other options (Coughlin & Featherstone, 2017).

e) Word repeats

Repeated words should be avoided because this can indicate the correct answer. Consequently, the possibility of participants correctly answering the question based on their knowledge is doubted (Coughlin & Featherstone, 2017).

f) Convergence strategy

A convergence strategy is a situation where a correct answer appears mostly appears in the answer options in that the question maker must provide variations on the other answer options (Coughlin & Featherstone, 2017).

3. The Level of MCQs

There are three level of multiple-choice questions based its difficulty, which are: Higher-Order Thinking Skill (HOTS), Middle-Order Thinking Skills (MOTS), and Lower-Order Thinking Skill (LOTS)

a) High-Order Thinking Skill (HOTS)

Higher-Order Thinking Skills (HOTS) questions are measurement instruments used to measure students' higher order thinking ability that measured namely the ability to think that does not just remember (recall), restate (restate), or refer without doing processing (recite) (Kemdikbud, 2017). The students are considered to have Higher-Order Thinking Skills (HOTS) when they can draw conclusion based on fact collection, generalization, and hypothesis (Fanani, 2018). Also, the students who have a Higher-Order Thinking Skills (HOTS) are able to combine their prior knowledge into the new knowledge to draw a conclusion, solve the problem, or achieve the goals (Rosnawati, 2018). The category of HOTS as proposed by King et al. (2013) are: (1) critical and logical thinking, (2) reflective thinking, (3) metacognitive thinking, and (4) creative

thinking. “The metacognitive dimension describes the ability to connect several different concepts, interpret, solve problems (problem solving), choose problem solving strategies, find (discovery) new methods, argue (reasoning), and make the right decisions” (Finani, 2018, p.61).

b) Middle-Order Thinking Skills (MOTS)

Middle-Order Thinking Skills (MOTS) are a type of question which requires an intermediate thinking ability. Its level is between Higher-Order Thinking Skills (HOTS) and Lower-Order Thinking Skills (LOTS). “A question is categorized as MOTS if the revised Bloom's taxonomy is at the level of C2 (understanding) and C3 (applying) (Himmah, 2019, p.58). When MOTS questions are found in the examination, the test takers only need the intermediate level of thinking to find the answer of the question. MOTS questions are different from Lower-Order Thinking Skills (LOTS) because in the MOTS questions, the test makers use the more advanced words to present the questions. R A N I R Y

c) Lower-Order Thinking Skills (LOTS)

Lower-Order Thinking Skills (LOTS) are a type of questions which requires low level thinking ability. Tikhonova and Kudinova (2015) stated that “Lower order thinking skills, or LOTS, involve such skills as memorizing, recalling the knowledge one already has,

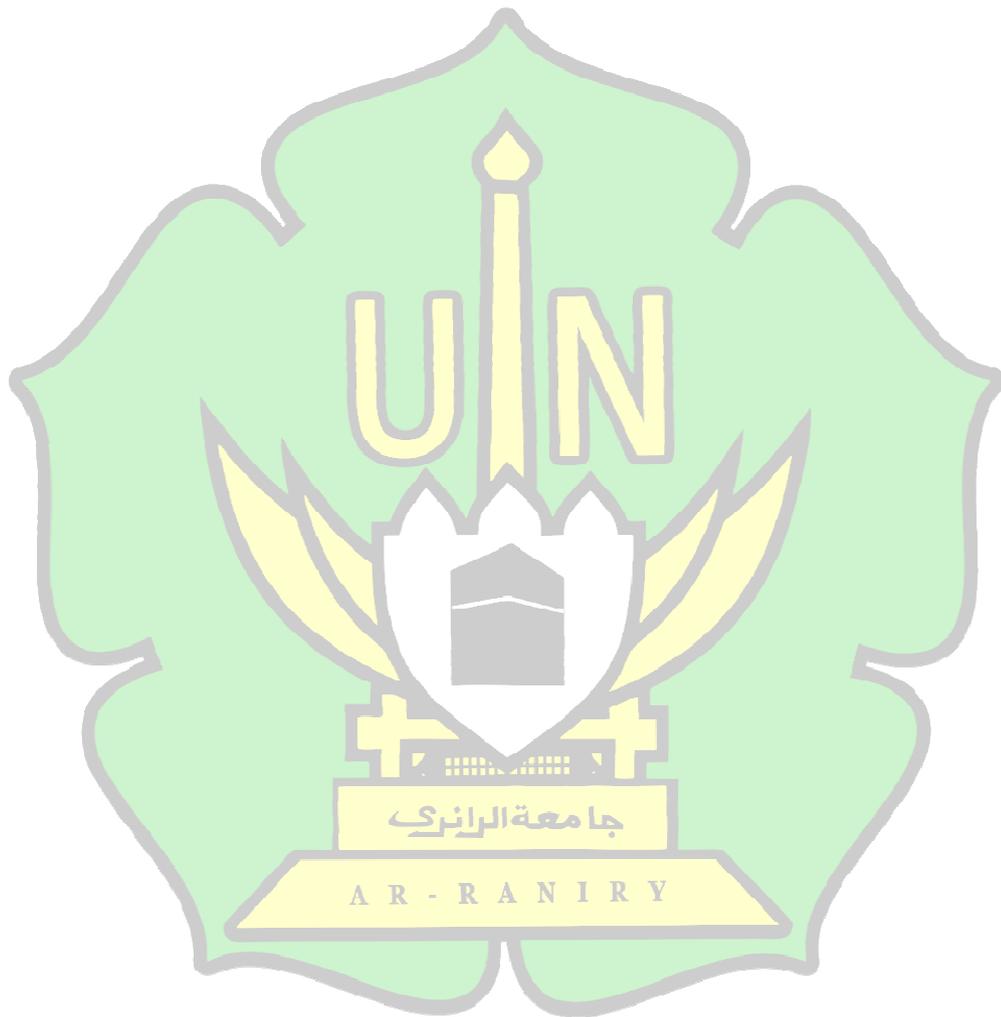
understanding what one knows, etc.” Since the LOTS is the lowest level of thinking ability, it measured the basic students’ knowledge. LOTS questions are questions that become the basis for testing students’ abilities where the abilities assessed are students’ basic abilities such as the ability to remember and recognize. The LOTS questions are also the basis for whether students can be given questions at the Middle-Order Thinking Skills (MOTS) or Higher-Order Thinking Skills (HOTS) level.

4. The Strengths and Limitations of MCQ

One of Indonesia's most common test instruments is multiple-choice questions (Yanti et al., 2020). As proposed by Coughlin and Featherstone (2017) that MCQ provides several advantages, such as making it easier for teachers to prepare blueprints, can be corrected automatically using computers, can provide feedback on examinees' performance on several different aspects, and can be a medium to measure the ability and understanding of examinees. Similarly, Towns (2014) agree that MCQ is easy to mark, especially if it is automated.

Zimmaro (2004) argues that the multiple-choice question needs longer time to construct, is hard to find plausible distractors, ineffective for problem-solving typed of questions, the test takers' reading ability may increase the score, is lack feedback, may contain a hint in the questions, inadequate to the higher levels of cognitive thinking, may have more than one possible correct answer, depends on the instructor's writing ability, does

not measure the writing ability, and may be influenced by students' guessing habit.



CHAPTER III

RESEARCH METHODOLOGY

This chapter discusses research design, research location, the technique of data collection, and the technique of data analysis.

A. Research Design

Research design tries to accommodate the research plan to achieve the goal of answering the research questions. It acts as a bridge that connects the research question and the application of the results of the research question. In this case, the researcher used a quantitative research design to determine the quality of the English teacher-made test at SMAN 1 Banda Aceh. “Descriptive research is research that seeks to describe phenomena that occur in a realistic, actual, real and current way” (Rukajat, 2018, p.1). Cresswell (1994, as cited in Senchantichai & Sukamolson, 2013) stated that quantitative research refers to the condition where the issue studied is interpreted using the numerical data calculated in statistics. This research applied quantitative methods because the data generated was formed as numbers or statistics. In quantitative research design, the researcher used the descriptive documentation study analysis method to find out each item's level of difficulty and quality.

The researcher implemented a quantitative descriptive method where the researcher obtained the data instrument for the research from the document that had been already available. In this case, the document collected from the examination were a question sheet, answer sheets, and a list of students' names. Nasution (2017) stated that descriptive statistics is a form of data that is easier to understand about

presentation, statistical, values, diagram, or chart. This analysis aimed to describe the data that has been obtained to answer the research questions which were: 1) What is the level of difficulty and discrimination indices of the teacher-made test at SMAN 1 Banda Aceh? and 2) Do the multiple-choice questions meet the norms published by *Puspendik Kemdikbud*?

B. Research Location

The research was taken place in SMAN 1 Banda Aceh in Jl. Prof. A. Majid Ibrahim II No.1, Punge Jurong, Meuraxa, Banda Aceh. The data collected were taken from the English Exam's answer sheets of grade XI in both science and social studies majors. The examination was held for two weeks. The examination of English subjects for grade XI was held on the same day as MIPA and IPS classes.

The researcher chose the school because the researcher found the problems related to the English exam questions. It was coherent with what the researcher wanted to obtain for the research: the problems in analyzing the difficulty index, discrimination index, and norms. The researcher took the process of finding the issues by investigating the school before choosing it by asking for some information from the English teachers. Then, the researcher found that the school had similar problems in difficulty index, discrimination index, and the quality of good MCQs. Then, the researcher perceived the need to conduct the research in SMAN 1 Banda Aceh. Therefore, the researcher chose SMAN 1 Banda Aceh as the place to conduct the study.

C. Technique of Data Collection

To collect the data for the research, the researcher used documentation as the data collection technique. “Document analysis is a systematic procedure for reviewing or evaluating documents—both printed and electronic (computer-based and Internet-transmitted) material” (Bowen, 2009, pp.27). The document collected were question sheets, answer sheets, and lists of students’ names. Then, the data from the documents were analysed and interpreted to the necessity of research.

Of all the test takers on the examination, the document collected by the researcher was only the document from grade XI of SMAN 1 Banda Aceh. In detail, there were 8 classes in XI level which were XI MIPA 1, XI MIPA 2, XI MIPA 3, XI MIPA 4, XI MIPA 5, XI MIPA 6, XI IPS 1, and XI IPS 2. The researcher believed that by including all of the XI grades’ examinations’ answer sheets in the research, the result this study would discover could be a reference for the teacher in the subsequent analysis and future research.

D. Technique of Data Analysis

The multiple-choice questions used in this research consisted of 30 questions. Then, each item of multiple-choice questions was separated from questions 1 to 30 on Microsoft Office Excel 2010. Afterward, the researcher listed the students’ names from highest-scored to lowest-scored students and their scores. Then the researcher recorded the correct answer based on each item they answered correctly. The example of data analysis was as follows.

Table 3. 1
Example of Item Analysis

	No	Name	Question 1	Question 2	Question 30
Upper	1	Student A	1	1	1
Lower	2	Student B	1	0	1

Half of the group was named upper-scored groups, and the other half was called lower-scored groups. After grouping the students into upper and lower-scored groups, the researcher processed the data using the formula of difficulty index and discrimination index. After getting the exact number of the difficulty index and the discrimination index, the researcher interpreted the score of the difficulty and discrimination indices based on their criteria. After diagnosing the data for the difficulty level, the researcher grouped it into difficult, average, and easy questions. And for the discrimination index, the researcher ranged the multiple-choice questions based on positive or negative results on the discrimination index test. Moreover, the researcher ranged the multiple-choice questions based on the percentage of the norm that question items had fulfilled. Therefore, below was the elaboration of the detail of the difficulty index, discrimination index, and quality of good multiple-choice (norms).

1. Difficulty Index (DIF) جامعة الرانري

The difficulty index is a statistical calculation that calculates the number of correct questions test-takers can answer on all multiple-choice questions (Liando *et al.*, 2021). Calculating DIF was done by counting the total number of correct answers on one number and dividing it by the total number of students. DIF represented the percentage of students who correctly answered the questions. A higher value of DIF showed that an increased number of students gave correct answers on that item. The range

of DIF ranged from 0-100%. For the formula of DIF, the researcher used the following formula to calculate the DIF.

$$DIF = \frac{\text{total of correct answer}}{\text{total students}}$$

Criteria of DIF:

- <0.30 = Difficult,
- 0.30-0.70 = Average
- >0.70 = Easy

2. Discrimination Index (DI)

Discrimination Index (DI) was a term for calculating an item's quality. By using DI, the difference between students who could answer questions and those who could not answer questions could be seen. DI shows a difference between students who are good at answering questions and students who are not good at answering a particular question (Taib et al., 2013). By applying DI analysis, teachers could determine whether students could answer questions or guess the answers and be the basis for teachers to evaluate question items and students' abilities. The contribution of this evaluation could be in the form of improvements in the quality of question items and the teaching strategy teachers apply. The range of DI ranged from 0-1. The formula that the researcher used to calculate DI was as follows:

$$DI = \frac{\text{upper} - \text{lower}}{1/2 N}$$

Criteria of DI:

>0.39	= Excellent
0.30-0.39	= Good
0.20-0.29	= Average
0.00-0.20	= Poor
<-0.01	= Worst

3. Guidelines of a Good MCQs (Norms)

Norms were a set of rules for compiling multiple-choice questions to make good ones regarding writing, grammar, and the content contained therein. Furthermore, 18 norms published by Puspendik Kemdikbud (2019) gathered the criteria for an excellent multiple-choice question. These norms were helpful so that the content presented was under what had been learned by students while making it easier for students to understand the delivery of questions created by the teacher (Yanti et al., 2020). Using a set of rules (norms) in compiling multiple-choice questions, the teacher could minimize the errors in writing exam questions to ensure all exam questions were in perfect quality to measure the overall learning process and students' abilities. The table below presented the 18 norms published by *Puspendik Kemdikbud* that the researcher used as a reference in analysing the teacher-made multiple-choice in SMAN 1 Banda Aceh.

Norms or as known as the quality of the question items is a list of criteria of a good test. All multiple-choice questions made by

teachers in Indonesia should follow the norms published by *Puspendik Kemdikbud* so that their quality is maintained (Yanti *et al.*, 2020).

Table 3.2

Guidelines of A Good Multiple-Choice Question (MCQ)

No	Norms Description
1	Reflecting basic competencies
2	Not depending on the previous options
3	Giving content clear focus
4	Avoiding opinion-based items
5	Being grammatically correct
6	Having correct spelling
7	Not containing clues
8	Formatting options vertically
9	Having correct use of punctuation and capitalization
10	Not containing double negatives
11	Having homogenous options in content and grammatical structure
12	Having one correct answer
13	Having options with relatively the same length
14	Placing options in logical and numerical order
15	Not repeating options with the same words or phrase
16	Not having overlapping options
17	Having plausible distractors
18	Not using “none of the above” or “all of the above”

Note. Source: Puspendik Kemdikbud (2019)

جامعة الرانيري

A R - R A N I R Y

CHAPTER IV

FINDINGS AND DISCUSSION

This chapter elaborates findings and discussion of this research that was found out about the range of difficulty index and discrimination index of the exam questions as well as its quality (norms).

A. Research Findings

This research was conducted to analyze the quality of question items of Teacher-made test at SMAN 1 Banda Aceh academic year 2021/2022 for grade XI in terms of validity, reliability, difficulty index, discrimination index, and the quality of question items based on guidelines (or also called norms) released by *Puspendik Kemdikbud* (2019). To analyze the question items, the researcher collected the instrument such as a question sheet, students' answer sheets, and a list of students' names. The question items that were analysed were 30 multiple-choice question items. The number of students who took the examination were 250 students from grade XI MIPA 1, XI MIPA 2, XI MIPA 3, XI MIPA 4, XI MIPA 5, XI MIPA 6, XI IPS 1, and XI IPS 2. In this research, the researcher took all of the examination sheets of students in grade XI in SMAN 1 Banda Aceh as the research instrument to research the validity, reliability, distractor effectiveness, and quality of question items based on the guidelines released by *Puspendik Kemdikbud* (2019).

Before analyzing the difficulty and discriminating indices, the researcher first analyzed the validity and reliability of the exam questions. The validity analysis was done as the parameter of the appropriateness of the question item.

Once a question item was valid, it was worth being the media to measure the students' ability for summative assessment. And reliability analysis needed to be conducted for the question items examined in this study could be proven for their level of trust. In the sense that the exam questions remained consistent in measuring what should be measured so that if the exam questions were done at different times by different test-takers, the results obtained by the two test-takers would be relatively the same. The further elaboration of validity and reliability analysis were elaborated below.

1. Validity

Analysis of validity was done using the IBM SPSS Statistics 25. Specifically, the researcher input students' scores into the application by giving code "1" for the correct answer and code "0" for the wrong answer. The value of P-value proved the validity of the question items. P-value was a value of probability that was analyzed using a statistical test. Accordingly, if the question item had a P-value below 0.005, the question item was valid. In this research, the P-value of all the 30 question items (100%) of English summative examination grade XI was below 0.005. Therefore, based on the calculation on IBM SPSS Statistics 25, the result showed that all 30 questions were valid (see table 4.1 below).

Table 4.1

The Validity of Multiple-Choice Questions

No.	Validity Index	Question Items	Total	Percentage	Status
1	< 0.005	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30	30	100%	Valid
2	> 0.005	-	0	0%	Not valid

Based on the table above, all 30 questions (100%) had a score validity index below 0.005. Therefore, it was categorized as valid. Moreover, none of the questions were classified as invalid.

2. Reliability

Based on the table above, all 30 questions (100%) had a score validity index below 0.005. Therefore, it was categorized as valid. Moreover, none of the questions were classified as invalid. The researcher analyzed the reliability of the question items through IBM SPSS Statistics 25 based on Cronbach's Alpha's value (See table 4.2). Based on the analysis, the reliability score of the English teacher-made test for grade XI in SMAN 1 Banda Aceh was scored 0.802, which can be interpreted as very high in its reliability.

Table 4.2

Reliability Range

No.	Reliability Range	Category
1	0,800 - 1,000	Very high
2	0,600 - 0,799	High
3	0,400 - 0,500	Average
4	0,200 - 0,399	Low
5	> 0,200	Very low

Table 4.3

Reliability Statistics

Cronbach's Alpha	Total Item
.802	30

Based on the reliability statistics above, the reliability value based on Cronbach's Alpha was .802 of the 30 items. So, if we look at the criteria that matched the number .802, that is, the value ranged from 0.800-1,000, where this number interpreted that the class XI English exam questions at SMAN 1 Banda Aceh had a relatively high level of reliability.

Accordingly, based on good validity and reliability results, the researchers concluded that the exam questions for the XI class semester at SMAN Banda Aceh had met the requirements to be used as a reference to become a good evaluation tool. Thus, the author's next step was to calculate the value of the difficulty index and discrimination index of the exam questions.

3. Difficulty Index

The difficulty index ranged based on three criteria, which showed the level of difficulty difference from each question item. It ranged from

below 0.30 to above 0.70 (See table 4.4). Based on the calculation using the formula of difficulty index through Microsoft Office Excel 2010, the researcher found that 4 question items (13,3%) were classified as difficult, 17 question items (56,6%) were classified as average, and 9 question items (30%) categorized as easy.

Table 4.4

Difficulty Index's Range

No	Difficulty Index	Criteria
1	<0.30	Difficult
2	0.30-0.70	Average
3	>0.70	Easy

To determine the difficulty index level of each question item, the first step that needed to be conducted was analyzing the number of students who answered correctly on each question item. The figure below presented the number of students who gave correct answers on each item:

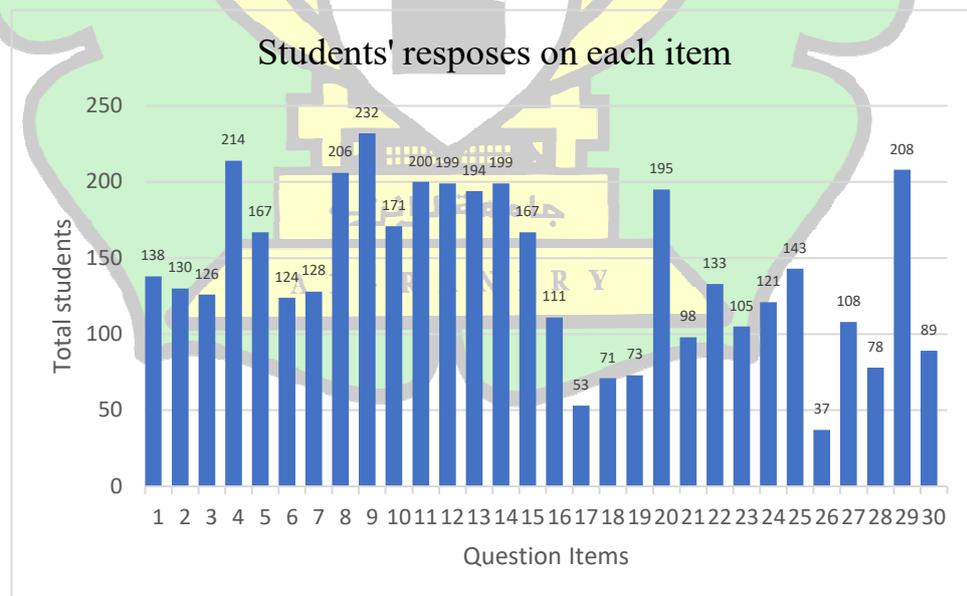


Figure 4.1 Chart of students' responses to each question item. The data resulted from calculating the Difficulty Index (DIF).

Based on the statistics chart above, we could see which question items had the highest average and lowest difficulty level. The question items that were answered by most of the students were grouped as easy question items, and the items that were answered by least of the students were grouped as difficult questions. Nevertheless, for this research that involved 250 students' responses, the simple analysis was not practical and less accurate. Therefore, the statistics calculation revealed and categorized the question items based on their criteria. Thus, the example of calculation using Brown's (2004) formula was as seen below:

1) Difficulty index question item number 1

$$\begin{aligned} \text{DIF} &= \frac{\text{total of correct answer}}{\text{total students}} \\ &= \frac{138}{250} \\ &= 0.552 \end{aligned}$$

Based on the result of the difficulty index calculation, question item number 1 had a score of 0.552, which was then categorized as "average."

2) Difficulty index question number 2

$$\begin{aligned} \text{DIF} &= \frac{\text{total of correct answer}}{\text{total students}} \\ &= \frac{130}{250} \\ &= 0.52 \end{aligned}$$

Based on the difficulty index calculation result, question item 1 had a score of 0.52, which was then categorized as "average." Thus, as seen

above, the researcher continued the calculation for all the question items to figure out the difficulty index's value. Later, the difficulty index found based on the analysis was arranged based on its numerical order and category. It was as seen below:

Table 4.5
Difficulty Index of English Exam Questions

No	Correct Responses	Total Responses	DIF	Category
1	138	250	0.55	Average
2	130	250	0.52	Average
3	126	250	0.50	Average
4	214	250	0.85	Easy
5	167	250	0.66	Average
6	124	250	0.49	Average
7	128	250	0.51	Average
8	206	250	0.82	Easy
9	232	250	0.92	Easy
10	171	250	0.68	Average
11	100	250	0.80	Easy
12	199	250	0.79	Easy
13	194	250	0.77	Easy
14	199	250	0.79	Easy
15	167	250	0.66	Average
16	111	250	0.44	Average
17	53	250	0.21	Difficult
18	71	250	0.28	Difficult
19	73	250	0.29	Difficult
20	195	250	0.79	Easy
21	98	250	0.39	Average
22	133	250	0.53	Average
23	105	250	0.42	Average
24	121	250	0.48	Average
25	143	250	0.57	Average
26	37	250	0.14	Difficult
27	108	250	0.43	Average
28	78	250	0.31	Average
29	208	250	0.83	Easy
30	89	250	0.35	Average

In reference to the table above that was arranged based on its numerical order, the researcher sorted the data according to its difficulty

level, from the highest to the lowest difficulty level. The results identified by the researcher were as follows:

Table 4.6
Difficulty Index Ranged from Difficult to Easy

No	Correct Responses	Total Responses	Difficulty index	Category
26	37	250	0.14	Difficult
17	53	250	0.21	Difficult
18	71	250	0.28	Difficult
19	73	250	0.29	Difficult
28	78	250	0.31	Average
30	89	250	0.35	Average
21	98	250	0.39	Average
23	105	250	0.42	Average
27	108	250	0.43	Average
16	111	250	0.44	Average
24	121	250	0.48	Average
6	124	250	0.49	Average
3	126	250	0.50	Average
7	128	250	0.51	Average
2	130	250	0.52	Average
22	133	250	0.53	Average
1	138	250	0.55	Average
25	143	250	0.57	Average
5	167	250	0.66	Average
15	167	250	0.66	Average
10	171	250	0.68	Average
13	194	250	0.77	Easy
12	199	250	0.79	Easy
14	199	250	0.79	Easy
20	195	250	0.79	Easy
11	100	250	0.80	Easy
8	206	250	0.82	Easy
29	208	250	0.83	Easy
4	214	250	0.85	Easy
9	232	250	0.92	Easy

Based on the arrangement above, there were 4 question items categorized as “difficult,” 17 question items categorized as “average,” and 9 question items categorized as “easy.” Thus, the researcher provided the table below to see the question items based on their group or category.

Table 4.7
Difficulty Index Based on Its Group

No	Question Items	Difficulty Index's Criteria
1	26, 17, 18, 19	Difficult
2	28, 30, 21, 23, 27, 26, 24, 6, 3, 7, 2, 22, 1, 25, 5, 15, 10	Average
3	13, 14, 12, 20, 11, 8, 29, 4, 9	Easy

Based on table 4.5, question item number 26 had the highest score of difficulty index (0.14), where 37 of the students was capable of answering it correctly, and question number 10 had the lowest difficulty index (0.92), where 232 of the students were qualified to answer it correctly. Furthermore, the question items labeled “difficult,” which were question items number 26, 17, 18, and 19, should be revised for future examination. The question labeled as “average,” which were question items number 28, 30, 31, 23, 27, 16, 24, 6, 3, 7, 2, 22, 1, 25, 5, 15, and 10 had a good difficulty index so these question items can be used as the references for future examination. Finally, the question labeled as “easy,” question items 13, 12, 13, 20, 11, 8, 29, 4, and 9 should be revised for future examination. Furthermore, below is the data of difficulty indices in the form of a pie chart to see its percentage:

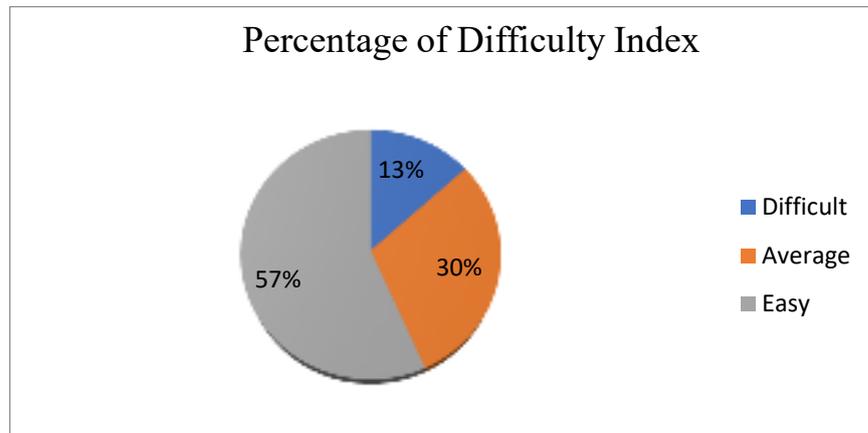


Figure 4.2 Pie chart of the percentage of criteria of Difficulty Index (DIF)

4. Discrimination Index

The discrimination index was first calculated by dividing the test-takers into two groups: upper and lower. Upper-scored students were the first half students who scored higher than the second half students. In contrast, the lower-scored students were the second half students who scored lower than the first half students. The distinction between upper and lower scored students in getting the correct answer for each question items were as seen below:

Table 4.8

Discrimination Index's Range

Discrimination Index (DI)	Criteria
>0.39	Excellent
0.30-0.39	Good
0.20-0.29	Average
0.00-0.20	Poor
< -0.01	Worst

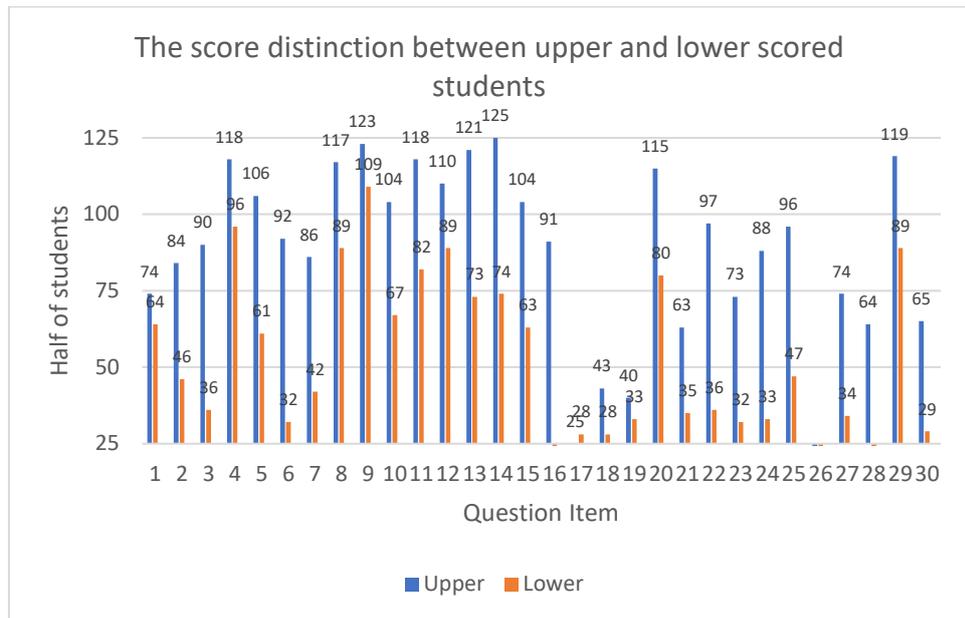


Figure 4.3 Chart of distinction between upper and lower scored students in answering each question items.

After splitting the data into upper and lower, the researcher operated the discrimination index formula on each item. For instance, the calculation was done on each item, as seen from the example below.

1) Discrimination index question item 1

$$\begin{aligned}
 DI &= \frac{\text{upper} - \text{lower}}{1/2 N} \\
 &= \frac{74 - 64}{125} \\
 &= \frac{10}{125} \\
 &= 0.08
 \end{aligned}$$

Based on the discrimination index calculation result, question item 1 had a score of 0.08, which was then categorized as “poor” (referred to table 4.7).

1) Discrimination index question item 2

$$\begin{aligned}
 DI &= \frac{\text{upper}-\text{lower}}{1/2 N} \\
 &= \frac{84-46}{125} \\
 &= \frac{38}{125} \\
 &= 0.304
 \end{aligned}$$

Based on the discrimination index calculation result, question item 1 had a score of 0.384, which was then categorized as “good” (referred to table 4.7).

The researcher continued calculating the data to determine each question item’s discrimination index. From question number 1-30. The researcher found out that the range discrimination index ranged from -0.02 to 0.39 meant that the English exam question grade XI in SMAN 1 Banda Aceh carried all the criteria of the discrimination index. Further data presentation was as seen below.

Table 4.9

Discrimination Index of English Exam Question

No	Discrimination Index	Criteria
1	0,08	Poor
2	0,304	Good
3	0,432	Excellent
4	0,176	Poor
5	0,36	Good
6	0,48	Excellent
7	0,352	Good
8	0,224	Average
9	0,112	Poor
10	0,296	Average
11	0,288	Average
12	0,168	Poor

13	0,384	Good
14	0,408	Excellent
15	0,328	Good
16	0,568	Excellent
17	-0,024	Worst
18	0,12	Poor
19	0,056	Poor
20	0,28	Average
21	0,224	Average
22	0,488	Excellent
23	0,328	Good
24	0,44	Excellent
25	0,392	Good
26	0,056	Poor
27	0,32	Good
28	0,4	Excellent
29	0,24	Average
30	0,288	Average

Regarding the table above that was arranged based on its numerical order, the researcher sorted the data according to its level of discrimination, from the lowest discrimination level to the highest. The results identified by the researcher were as follows:

Table 4.10

Discrimination Index from Worst to Excellent

No	Discrimination Index	Criteria
17	-0,024	Worst
19	0,056	Poor
26	0,056	Poor
1	0,08	Poor
9	0,112	Poor
18	0,12	Poor
12	0,168	Poor
4	0,176	Poor
8	0,224	Average
21	0,224	Average
29	0,24	Average
20	0,28	Average
11	0,288	Average
30	0,288	Average
10	0,296	Average

2	0,304	Good
27	0,32	Good
15	0,328	Good
23	0,328	Good
7	0,352	Good
5	0,36	Good
13	0,384	Good
25	0,392	Good
28	0,4	Excellent
14	0,408	Excellent
3	0,432	Excellent
24	0,44	Excellent
6	0,48	Excellent
22	0,488	Excellent
16	0,568	Excellent

In accordance with the data arrangement above, we found that 1 question item was categorized as “worst,” 7 question items were categorized as “poor,” 7 question items were classified as “average,” and 8 question items were categorized as “good,” and 8 question items were categorized as “excellent.” Furthermore, the table below grouped the data based on its criteria.

Table 4.11

Discrimination Index Based on Its Group

No	Question Item	Criteria
1	17	Worst
2	19, 26, 1, 9, 18, 12, 4	Poor
3	8, 21, 29, 20, 11, 30, 10	Average
4	2, 27, 15, 23, 7, 5, 13, 25	Good
5	28, 14, 3, 24, 6, 22, 16	Excellent

Referring to the data above, the researcher decided to include the percentage of the data based on each criterion. Thus, the chart below presents the percentage of the data.

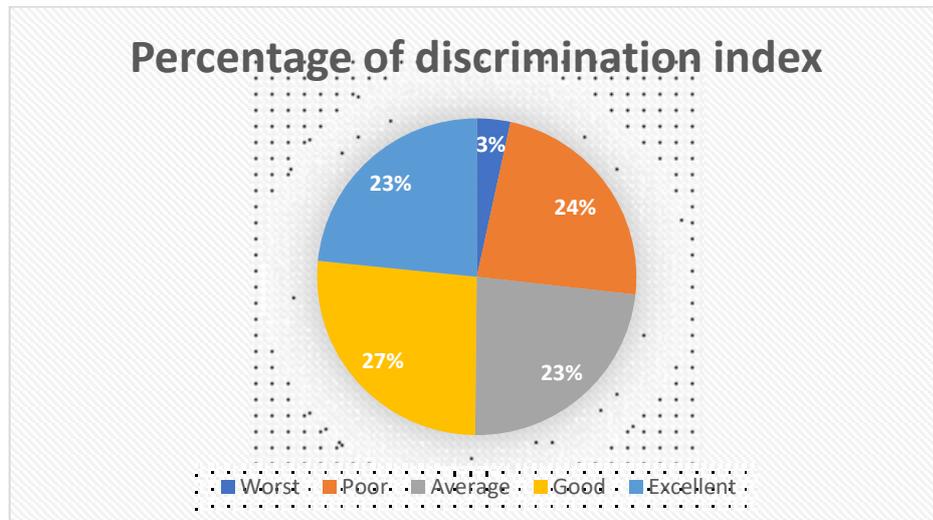


Figure 4.4 Percentage of discrimination index for English summative assessment of grade XI in SMAN 1 Banda Aceh

5. Norms

Based on the analysis on the quality of teacher-made test, the researcher found that most of the question items fulfilled the norms and none of the question items neglected more than 5 norms. To find out the norms' fulfilment from all of the question items, the researcher did a checklist to check how much norms fulfilled by each item. The table below was the example on how to do the checklist on the norms.

Table 4.12

Example of Norms' Checklist

Question Number	Norms								
	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	1	1	1	1

Question Number	Norms								
	10	11	12	13	14	15	16	17	18
1	1	1	1	1	1	1	1	1	1

Note:

- Number "1" indicated that the norm was found on the question item.
- See appendix for the description of each norm

The checklist was done for all the question items to investigate the norms' fulfilment for all question items. Then, due to the difference in the number of norms that were fulfilled differently, the researchers compiled the data based on the percentage. Where the highest percentage was worth 100% and the lowest percentage was 80%. There were 12 questions that fulfilled all norms and reached a percentage of 100%, there were 3 questions that neglected 2 norms and reached a percentage above 90%, there were 3 question items that neglected 4 to 6 norms and reached a percentage above 80%. Similarly, Yanti *et al.*, (2020) and Indrayani (2020) found that the previous research had a high percentage of multiple-choice questions quality, meant that the norms were fulfilled.

Table 4.13

Total of Norms Fulfilled

No	Norms' Description	Number of Items Fulfilling the Norms	Frequency (%)
1	Reflecting basic competencies	30	100%
2	Not depending on the previous options	30	100%
3	Giving content clear focus	28	93%
4	Avoiding opinion-based items	30	100%
5	Being grammatically correct	25	83%
6	Having correct spelling	30	100%
7	Not containing clues	30	100%
8	Formatting options vertically	30	100%
9	Having correct use of punctuation and capitalization	26	87%
10	Not containing double negatives	30	100%

11	Having homogenous options in content and grammatical structure	28	93%
12	Having one correct answer	28	93%
13	Having options with relatively the same length	24	80%
14	Placing options in logical and numerical order	30	100%
15	Not repeating options with the same words or phrase	30	100%
16	Not having overlapping options	30	100%
17	Having plausible distractors	30	100%
18	Not using “none of the above” or “all of the above”	30	100%

C. Discussion

Evaluation is a way to measure the achievement of a lesson, both teacher achievement in teaching and student achievement in understanding and mastering the material being taught. In an educational unit, these achievements can be seen through the examination process organized by the school and followed by all students. Given the importance of the role of a test or exam in the development of the quality of education, a test or exam that is routinely carried out must be tested for quality.

This quantitative research aimed to analyse English exam question at SMAN 1 Banda Aceh as well as to answer the research questions, which were: (1) “What is the level of difficulty and discrimination indices of the teacher-made test at SMAN 1 Banda Aceh” and (2) “Do the multiple-choice questions meet the norms published by Puspendik Kemdikbud?”. Accordingly, to answer the research questions the research involved all the XI grade students’ at SMAN 1 Banda Aceh.

In this research, there were 250 students of grade XI participated in the English examination. The answer sheets of 250 students were collected along with the question sheet, list of students' names, and answer key. The answer sheets were then corrected referring to the answer key to find out how many students answer each question item correctly. In this part, the researcher elaborated four points on the findings based on the research questions.

To answer the first research question “What is the level of difficulty and discrimination indices of the teacher-made test at SMAN 1 Banda Aceh?”, the researcher used two programs which were IBM SPSS Statistics 25 and Microsoft Office Excel 2010. The IBM SPSS Statistics was used to test the validity and reliability of the question items before going on to the difficulty and discrimination indices. After its validity and reliability was proven, the researcher continued the analysis using the Microsoft Office 2010 program to calculate the Difficulty and discrimination indices. Lastly, still using the Microsoft Office 2010 program, the researcher analysed how well the question items fulfilled the norms.

Consequently, based on the results of the item analysis conducted at SMAN 1 Banda Aceh on the class XI exam questions using IBM SPSS Statistics 25 and Microsoft Office Excel above, the researchers discussed in detail the validity, reliability, difficulty index, discrimination index, and norm below.

1. Validity

The validity of a test is achieved through its success in measuring learning achievement. In this research, the validity was analyzed through IBM SPSS Statistics 25, where each question item had a *P-value*. The *P-value* ranged between <0.005 and >0.005 . If the question item scored <0.005 in the *P-value* point, it is interpreted as valid. In contrast, if the question item scored >0.005 , it is interpreted as invalid.

In this research, the researcher analysed the 30 question items of summative examination for grade XI at SMAN 1 Banda Aceh. Out of the 30 questions, 1 question had a *P-value* 0.001 which interpreted as valid. Furthermore, the rest of question items had 0.00 *P-value* which also interpreted as valid. Therefore, it was concluded that in this research, all of the question items (100%) were valid. This result was similar with Liando *et al.*, (2021) and Lestari *et al.*, (2014) where the question items validity value was considered as valid. In contrast with that, Setiyana (2016) where the validity range of the second-grade students were categorized as poor.

The validity of question items was highly interconnected with the quality of the question items such as the question had a correct spelling, right use of punctuation and capitalization, the options ranged vertically, the question items were based on standard competencies, etc. Therefore, to make a test have a high validity, the question items should be good in quality.

2. Reliability

Reliability analysis is very important to run to measure the level of consistency of the overall test questions. This consistency is needed so that the exam questions can be trusted. It can be said that a test has a good reliability aspect if the test is run at different times, then the results obtained will be relatively the same (Setiyana, 2016). In this study, the researcher calculated the reliability scores for class XI exam questions at SMAN 1 Banda Aceh using IBM SPSS Statistics 25 with reference to Cronbach's Alpha scores.

The results obtained in this study were that the class XI exam questions at SMAN 1 Banda Aceh had a value of 0.802 in the Cronbach's Alpha column which based on reliability criteria was included in the range of 0.800-1.000 where the interpretation is that the test questions have a fairly high level of reliability. This was in accordance with Linn and Gronlund (2002), where the ideal range of reliability score varies between 0.60 and 0.85. Therefore, the class XI exam questions at SMAN 1 Banda Aceh had an ideal reliability value. As the research result was similar as research conducted by Setiyana (2016), Liando *et al.*, (2021), Prihartini *et al.*, (2019), where the research conducted was proven in its reliability. On the other hand, Lestari *et al.*, (2014) found that the question items were unreliable.

3. Difficulty Index

The analysis of difficulty index was done to see the level of difficulty each question items holds. In an evaluation, it is important to maintain a good difficulty index's range where the question items will be balanced. Referring to Brown (2004), a question item should not be very difficult or very easy because it cannot contrast between upper and lower groups in the test. Therefore, the exam questions should not be too difficult, nor too easy. If the questions were too difficult, students would easily give up in answering the questions because the questions were beyond their ability. On the other hand, if the questions were too easy, it would not stimulate students to think further because the questions listed were very easy.

The range for the difficulty index was varied based on different theories. But, in general, the difficulty index ranged between $<0.30 - >0.70$. Asaad and Hailaya (2005) proposed that the acceptable range of difficulty index is ranged between 0.30-0.70. Therefore, in this study, the researchers found that there were 4 items (13%) ranged below 0.30 which were classified as difficult, 17 items (57%) ranged between 0.30 until 0.70 which were classified as moderate, and there were 9 items (30%) ranged above 0.70 which were classified as easy. Hence, the question items number 1, 2, 3, 5, 6, 7, 10, 15, 16, 21, 22, 23, 24, 25, 27, and 28 that were classified as moderate can be saved as reference for future examination. Whereas, the question items number 17, 18, 19, 26, 4, 8, 9, 11, 12, 13, 14, 20, and 29 that

were classified as difficult and easy should be revised for future examination.

Drawing conclusion from the percentage, the researcher concluded that the semester exam questions for class XI at SMAN 1 Banda Aceh were not proportional. An exam question is said to be proportional if the exam questions have a ratio of 25:50:25 (Fitrianawati, 2017). It meant that 20% of the questions are difficult questions, 50% are moderate questions, and the other 25% are easy questions. While this research had a ratio of 13:57:30. Hence, it was concluded that the English exam questions for grade XI was not proportional. Similarly, Prihartini *et al.*, (2019) found that the research conducted lacked on its difficulty index. While Liando *et al.*, (2021), and Lestari *et al.*, (2014), found that the question items had a good difficulty index but some revision was still needed.

To overcome this disproportionality, several efforts were needed to achieve proportional exam questions. For exam questions that were classified as difficult, then it must be followed up on the cause behind the difficulty and changed it for future exams. Then, for questions that were classified as moderate, then the questions could be entered into the question bank. And for questions with easy categories, the weight of the questions must be increased so that the correct answers answered by students could be used as a reference for their abilities, not only because the questions were easy.

3. *Discrimination Index*

Discrimination index analysis is carried out to determine the ability of a question to distinguish between upper and lower scored students. The range of the discrimination index ranges from -1 to 1. Brown (2004) suggested that the question items that has an index near zero, the item should be discarded because it has a poor discrimination index. If an item has a negative disability index value, then the discrimination index of the item is not good, which means that the item is not able to become a medium to distinguish between capable students and those who are less able to answer questions. Furthermore, if an item has a discrimination index above 0 and is close to 1, then the item is classified as a question that has a good level of discrimination, which means that this item is capable of being a medium to distinguish between capable and disadvantaged students. Similar studies conducted on multiple choice questions that showed good discrimination index on Setiyana (2016), Liando *et al.*, (2021), Effendi and Mayuni (2022), Lestari *et al.*, (2014), Toksöz and Ertunç (2017). In contrast, Prihartini *et al.*, (2019) found that the discrimination index of the question items was bad

In this study, researchers analysed 30 semester exam questions for class XI at SMAN 1 Banda Aceh using Microsoft Office Excel program. The researcher found that 1 item (3%) was included in the worst category, 7 items (23%) were classified as poor, 7 items (23%) were classified as average, 8 items (24%) were classified as good, and 7 items were classified as good, and 7 questions (23%) are classified as excellent.

The question items that classified as “good” and “excellent” which were found in question items number 2, 27, 15, 23, 7, 5, 13, 25, 28, 14, 3, 24, 6, 22, and 16 could be saved in the question bank as reference for future examination. Whereas question items number 8, 21, 29, 20, 11, 30, and 10 classified as “average” should be revised to increase its quality in discriminating between upper and lower groups students. And lastly, question items number 17, 19, 26, 1, 9, 18, 12, and 4 that were classified as “poor” and “worst” should be discarded and not be taken as reference for future examination because it was lacking in the discriminating ability.

5. Norms

Norms or guidelines in the preparation of multiple-choice type exam questions published by Puspendik Kemdikbud (2019) in order to help teachers who also play a role as compilers of exam questions. With the norms, the subject teacher can make it a reference in preparing exam questions. In the norms there are 18 criteria which include content, context, and grammar.

In this study, the researcher examined the relationship between the test questions made by the teachers of SMAN 1 Banda Aceh for class XI with 18 good criteria. The calculation process was carried out in Microsoft Office Excel 2010.

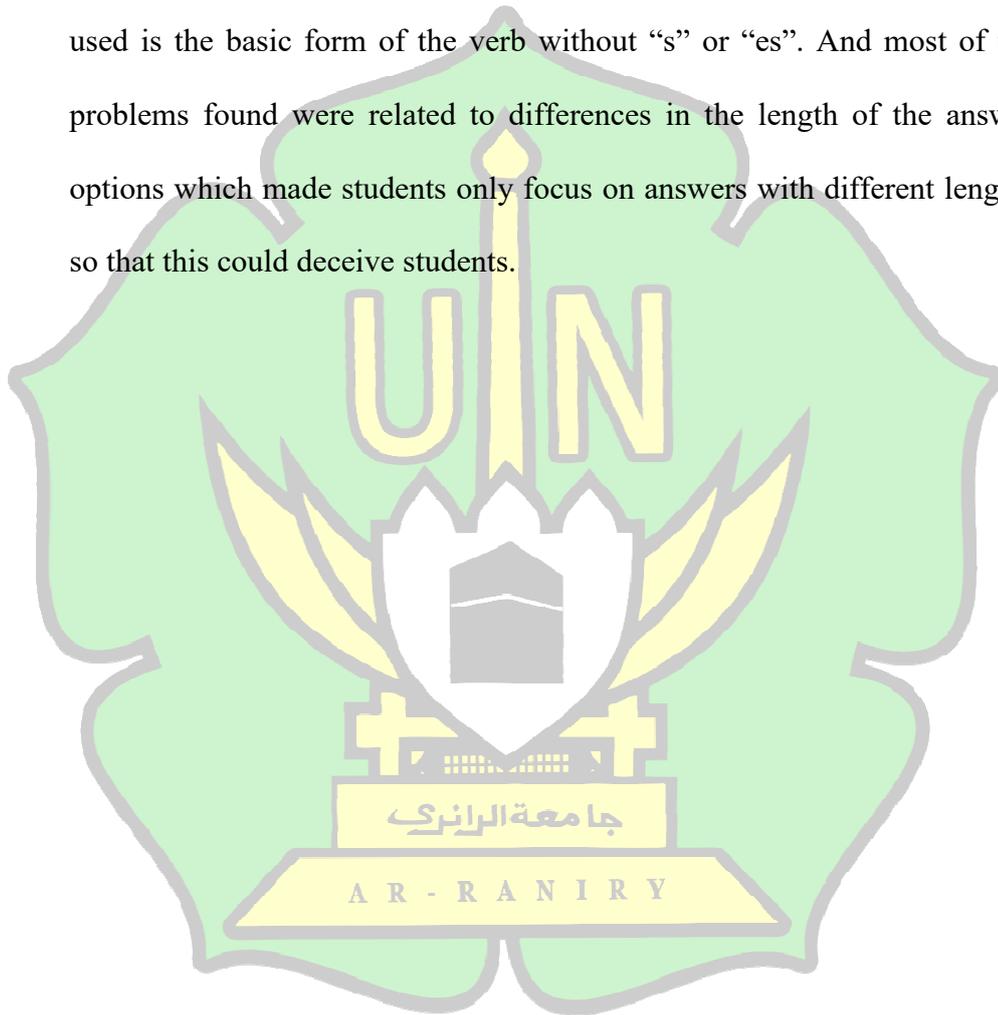
The items received 100% had fulfilled all the 18 norms where it reflected on basic competencies, not depending on the previous options, had the content clear focus, avoided opinion-based items, grammatically

correct, had correct spellings, not contained of clues, formatted options vertically, had correct use of punctuation and capitalization, not contained double negatives, had homogenous options, had one correct answer, had the same length's options, options were arranged in logical and numerical order, no repeated options, no overlapped options, had plausible distractors, no “none of the above” or “all of the above”.

The item received 93% lacked on the focus of the content, the arrangement of homogenous options, and had multiple answers. The question item was not clear about what the focus of the question due to the errors in writing. In writing answer options, the length of the answer options varied so that it allowed for a bias where students would focus their attention on one option with different lengths so that this can be a distraction on the question. And also, there were questions with which the options had two correct answers, which could trigger confusion in students.

The item received 87% faced the issues on the use of punctuation and capitalization and the arrangement of homogenous options. There were questions that misplaced capital letters, where after the answer options A, B, C, D, or E the use of letters used was lowercase where it actually should be capitalized. In writing answer options, the length of the answer options varied so that it allowed for a bias where students would focus their attention on one option with different lengths so that this can be a distraction on the question.

The item received 87% faced the issues on the grammatical correctness. There was question item that misplaced the articles "a" and "an" so that this question was not grammatically correct. And in another problem, there was an error in the verbal sentence involved the subject "she" where each verb should be followed by "s" or "es" but in this problem, the verb used is the basic form of the verb without "s" or "es". And most of the problems found were related to differences in the length of the answer options which made students only focus on answers with different lengths so that this could deceive students.



CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

This chapter presented the overall conclusions based on research findings and discussion, along with the recommendations related to the analysis of English exam questions at SMAN 1 Banda Aceh.

A. Conclusion

Based on the presentation of the data in the Findings section, the researcher concluded that the exam questions for the XI class semester at SMAN 1 Banda Aceh were valid with a validity value above 0.005 which stated that the items tested were valid. Furthermore, the analysis on reliability proved that the questions tested were reliable based on a reliability score of 0.802. Which, if translated, resulted in the conclusion that this score had a fairly high-reliability value. Then for the difficulty index value, 4 questions (13.3%) were categorized as difficult questions, 17 questions (56.6%) were categorized as moderate questions, and 9 questions (30%) were categorized as easy. This showed that most of the semester exam questions had a moderate level of difficulty. Still, some question items needed revisions to improve the quality of the questions for the upcoming exams.

Furthermore, the value of the discrimination index presented that 1 item was classified as worst, 7 questions were classified as poor, 7 questions were classified as average, 8 questions were classified as good, and 7 questions were classified as excellent. Thus, it could be concluded that the discrimination index categorized as “average”, “good”, and “excellent” could be used as a reference for the future examination. Hence, the question items categorized as “worst” and “poor” needed

several revisions to improve the discriminating ability of the items in the upcoming examination or could be discarded. Thus, for the quality of the question items (norms), we could conclude that most of the norms (12 norms) were fulfilled by the teacher-made test of exam questions of grade XI at SMAN 1 Banda Aceh. And for some items that failed in fulfilling the norms, these items should be revised by following the guidelines of creating good multiple-choice questions published by *Puspendik Kemendikbud*.

B. Recommendations

Based on the results of the research that has been done, the researcher has several recommendations that can be implemented for future research. In compiling the items, the teacher should pay more attention to the norms governing the quality of an item published by the Puspendik Kemdikbud (2019). If the preparation of the exam questions conforms to the norms contained in the guidelines, the resulting questions will have better quality in terms of content, writing, grammar, format, and so on. Then, to increase the level of difficulty index, it will be better for the teachers if they modify the syllabus and the learning materials. Hence, the students would be well-prepared for the examination. Teachers can also refer to a scale of 25:50:25 in compiling questions where 25 question items are grouped as difficult questions, 50 questions are items grouped as average questions, and 25 question items are grouped as easy questions. Also, by using this scale, the items that are made will be better able to distinguish the abilities of lower and upper students so that the level of discrimination can reach a high level.

For this reason, cooperation between the education authorities and school authorities is needed. The education authorities should be more considerable in providing training for the teachers on arranging good questions for examination. With this kind of support, the question items produce by the teacher will be more qualified and able to become proper evaluation material for both for teaching process and students' abilities.



REFERENCES

- Abdel-Hameed, A. A., Al-Faris, E. A., Alorainy, I. A., & Al-Rukban, M. O. (2005). The criteria and analysis of good multiple-choice questions in a health professional setting. *Saudi Medical Journal*, 26(10), 1505-1510.
- Abdulghani, H. M., Ahmad, F., Irshad, M., Khalil, M. S., Al-Shaikh, G. K., Syed, S., & Haque, S. (2015). Faculty development programs improve the quality of Multiple-Choice Questions items' writing. *Scientific Reports*, 5(1), 1-7.
- Agustin, Y. (2015). Kedudukan bahasa Inggris sebagai bahasa pengantar dalam dunia pendidikan. *Deiksis*, 3(04), 354-364.
- Algeo, J., & Butcher, C. A. (2013). *The origins and development of the English language*. Canada: Cengage Learning.
- Alsubait, T., Parsia, B., & Sattler, U. (2014). Generating multiple choice questions from ontologies: Lessons learnt. In *OWLED* (pp. 73-84).
- Athanasou, J. (2020). Nature and purpose of evaluation. *Understanding Adult Education and Training*, 86.
- Asaad, A., & Hailaya, W. M. (2005). *Measurement and Evaluation: Concept and Principles*. Quezon: REX Printing Company, Inc.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Badan Standar Nasional Pendidikan. (2020). *Home: Tentang BSNP*. Retrieved from: <https://bsnp-indonesia.org/tentang-bsnp-2/>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, 9(2), 27-40
- Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.
- Brown, H.D. (2003). *Language assessment principle and classroom practices*. California: Longman.
- Burton, S. J., Sudweeks, R. R., Merrill, P. F., & Wood, B. (1991). How to prepare better multiple-choice tests: Guidelines for University Faculty.
- Cambridge Dictionary. (n.d). In *cambridge.org dictionary*. Retrieved July 15, 2022, from <https://dictionary.cambridge.org/dictionary/english/analysis>
- Cambridge Dictionary. (n.d). In *cambridge.org dictionary*. Retrieved July 15, 2022, from <https://dictionary.cambridge.org/dictionary/english/examination>

- Carrasquillo, A. L. (2013). *Teaching English as a second language: A resource guide*. New York: Routledge.
- Coughlin, P. A., & Featherstone, C. R. (2017). How to write a high-quality multiple-choice question (MCQ): A guide for clinicians. *European Journal of Vascular and Endovascular Surgery*, 54(5), 654-658.
- Dehnad, A., Nasser, H., & Hosseini, A. F. (2014). A comparison between three- and four-option multiple choice questions. *Procedia-Social and Behavioral Sciences*, 98, 398-403.
- Effendi, T., & Mayuni, I. (2022). Examining teacher-made English test in a language school. *LADU: Journal of Languages and Education*, 2(2), 67-76.
- Elis R. W. E., & Rusdiana, A. (2015). *Evaluasi pembelajaran*. Bandung: Pustaka Setia.
- Fanani, M. Z. (2018). Strategi pengembangan soal hots pada kurikulum 2013. *Edudeena: Journal of Islamic Religious Education*, 2(1).
- Febriana, R. (2021). *Evaluasi pembelajaran*. Jakarta Timur: Bumi Aksara.
- Fitrianawati, M. (2015). Peran analisis butir soal guna meningkatkan kualitas butir soal, kompetensi guru dan hasil belajar peserta didik. *In Prosiding Seminar Nasional Pendidikan PGSD UMS & HDPGSDI Wilayah Jawa*, 5(3), 282-295.
- Gajjar, S., Sharma, R., Kumar, P., & Rana, M. (2014). Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. *Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive & Social Medicine*, 39(1), 17.
- Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, 115(4), 523-547.
- Gultom, E. (2016). Assessment and evaluation in EFL teaching and learning. *Proceedings of ISELT FBS Universitas Negeri Padang*, 4(1), 190-198.
- Habiburrahim, H. (2019). A review of the department of English language education curriculum development. *Englisia: journal of language, education, and humanities*, 6(1), 1-14.
- Hakim, L., & Irhamsyah, I. (2020). The analysis of the teacher-made test for senior high school at State Senior High School 1 Kutacane, Aceh Tenggara. *Jurnal Ilmiah Didaktika: Media Ilmiah Pendidikan dan Pengajaran*, 21(1), 10-20.
- Haladyna, T.M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Routledge. <https://doi.org/10.4324/9780203825945>Hall, C. & Marshall, S. (2013).

- Haryanto, M. P. (2020). *Evaluasi pembelajaran (konsep dan manajemen)*. Yogyakarta: UNY Press.
- Himmah, W. I. (2019). Analisis soal penilaian akhir semester mata pelajaran Matematika berdasarkan level berpikir. *Journal of Medives: Journal of Mathematics Education IKIP Veteran Semarang*, 3(1), 55-63.
- Ilma, R. (2017). Overview of English National Examination in Indonesian relation to standardized test. *English Empower: Journal of Linguistics and Literature*, 2(01), 42-47.
- Indrayani, Marhaeni, Paramartha, Wahyuni (2020). The analysis of the teacher-made multiple-choice tests quality for English subject. *Journal of Education Research and Evaluation*, 4(3), 272-278.
- Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 12-23.
- Kemendikbud. (2017). Modul Penyusunan Soal Higher Order Thinking Skill (HOTS). Jakarta: Direktorat Pembinaan SMA Ditjen Pendidikan Dasar dan Menengah.
- Khan, W. B., & Inamullah, H. M. (2011). A study of lower-order and higher-order questions at secondary level. *Asian Social Science*, 7(9), 149.
- King, F. J., Goodson, L., & Rohani, F. (2013). Higher order thinking skills: Definition, teaching strategies, assessment. Tallahassee, FL: Center for Advancement of Learning and Assessment.
- Kizlik, B. (2012). Measurement, assessment, and evaluation in education. Retrieved October, 10, 2015.
- Kurniawan, D. D., Syifa, A., Huda, N., & Kusuma, M. (2022). Item analysis of teacher made test in Biology subject. In *5th International Conference on Current Issues in Education (ICCIE 2021)*, (p. 312-317). Atlantis Press.
- Lalogiroth, A., & Tatipang, D. P. (2020). An analysis of English national exam and English teachers' perception using Bloom's revised taxonomy. *Journal of English Culture, Language, Literature and Education*, 8(1), 1-19.
- Latip, A., & Sutantri, N. (2021). Pelatihan pemanfaatan google form sebagai platform penilaian akhir tahun di SMA Edu Global Bandung. *Journal of Human and Education (JAHE)*, 1(1), 1-5.
- Lebagi, D., Sumardi, S., & Sudjoko, S. (2017). The quality of teacher-made test in EFL classroom at the elementary school and its washback in the learning. *Journal of English Education*, 2(2), 97-104.
- Linn, R. L., & Gronlund, N. E. (2002). *Measurement and Assessment in Teaching*. Upper Saddle River, N. J.: Prentice Hall.

- Lumentut, Y., & Lengkoan, F. (2021). The relationships of psycholinguistics in acquisition and language learning. *Journal of English Culture, Language, Literature and Education*, 9(1), 17-29.
- Lestari, W. A., Ikhsanudin, I., & Bunau, E. (2014). English listening test items evaluation: A case of a teacher-made test of SMKN 5 Pontianak 2014. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, 8(8).
- Liando, N. V., Serhalawan, E., & Wuntu, C. (2021). Analysis of teacher-made tests used in summative evaluation at SMP Negeri 1 Tompaso. *Jurnal Ilmiah Wahana Pendidikan*, 7(8), 480-493.
- Merriam-Webster Online. Reliability. In Merriam-Webster.com dictionary. Retrieved October 17, 2021, from <https://www.merriamwebster.com/dictionary/reliability>
- Muluk, S., Habiburrahim, H., Zulfikar, T., Akmal, S., Nasriyanti, N., & Safrul, M. S. (2021). Classroom management in efl speaking class: Strategies and challenges. *IJEE (Indonesian Journal of English Education)*, 1(1), 120-139.
- Munzaki, D. F., Suadah, L., & Risdaneva, R. (2016). Teaching methods used by students of department of English language education of Uin Ar-Raniry in teaching English at English course. *Englisia: Journal of Language, Education, and Humanities*, 4(1), 10-26.
- Nasution, L. M. (2020). Statistik deskriptif. *Hikmah*, 14(1), 49-55. Retrieved from <http://e-jurnal.staisumatera-medan.ac.id/index.php/hikmah/article/view/16>
- Nuryati, N., & Astriana, N. (2020). Pengaruh nilai ujian sekolah berstandar nasional terhadap prestasi belajar Pendidikan Kewarganegaraan di sekolah menengah pertama. *Academy of Education Journal*, 11(01), 52-62.
- Owen, G. T. (2014). Qualitative methods in higher education policy analysis: Using interviews and document analysis. *The qualitative report*, 19(26), 1.
- Paragae, I. G., Marhaeni, A. A. I. N., & Dantes, N. (2013). Teacher made assessment authenticity in Senior High Schools and its contribution to students' English achievement. *Jurnal Pendidikan Bahasa Inggris Indonesia*, 1.
- Prihartini, Y., Buska, W., & Hasnah, N. (2019). Analysis of test item on the final test semester exam on Arabic subjects. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 6(1), 71-92.
- Puspendik Kemendikbud. (2019). Guidelines for Implementing Test Construction Training. Jakarta, Indonesia.
- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1).

- Rahmi, A., Y. Y. Gatot & Suhartono, L. (2018). An analysis on validity and reliability of test items in final semester examination. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, 7(12).
- Rajhy, H. A. A. (2014). Five characteristics of a good language test. *National Journal of Extensive Education and Interdisciplinary*, 2(4), 61-66.
- Rea-Dickins, P., Germaine, K., & Germaine, K. P. (1992). *Evaluation*. Oxford University Press
- Rogler, D. (2014). Assessment literacy: Building a base for better teaching and learning. In *English Teaching Forum* 52(3), 2-13.
- Rosnawati, R. (2012). Enam Tahapan Aktifitas dalam Pembelajaran Matematika untuk Memberdayakan Berfikir Tingkat Tinggi Siswa (Makalah). *Diambil tanggal*, 23.
- Rukajat, A. (2018). *Pendekatan penelitian kuantitatif: quantitative research approach*. Yogyakarta: Deepublish.
- Rukajat, A. (2018). *Teknik evaluasi pembelajaran*. Yogyakarta: Deepublish.
- Rusmayani, R. (2020). Analisis butir soal penilaian akhir semester genap mata pelajaran Pendidikan Agama Islam di SMP Bintang Persada Tabanan-Bali. *Widya Balina*, 5(1), 42-49.
- Senchantichai, N., & Sukamolson, S. (2013). A study of reading strategies and effects of reading ability levels and text types on rational deletion cloze test performance of EFL University students. *rEFLECTIONS*, 16, 24-45.
- Setiawan, H., Garnier, K., & Isnaeni, W. (2019, October). Rethinking standardized test of science education in Indonesian high school. In *Journal of Physics: Conference Series* (Vol. 1321, No. 3, p. 032078). IOP Publishing.
- Setiyana, R. (2016). Analysis of summative tests for English. *English Education Journal*, 7(4), 433-447.
- Shete, A. N., Kausar, A., Lakhkar, K., & Khan, S. T. (2015). Item analysis: An evaluation of multiple-choice questions in Physiology examination. *Journal of Contemporary Medical Education*, 3(3), 106-109.
- Shobariyah, E. (2018). Teknik evaluasi non tes. *Adz-Zikr: Jurnal Pendidikan Agama Islam*, 3(2), 1-13.
- Subiyati, M. (1995). Bahasa Inggris, tumbuh mengglobal dari bahasa tertindas sampai bahasa kebutuhan intelektual. *Jurnal Cakrawala Pendidikan*, 1(1).
- Sudijono, Anas. (2009). *Pengantar evaluasi pendidikan*, Jakarta: PT. Raja Grafindo Persada.

- Sugianto, A. (2011). Analysis of validity and reliability of English formative tests. *Journal on English as a Foreign Language*, 1(2), 87-94.
- Susanto, H., Rinaldi, A., & Novalia, N. (2015). Analisis validitas reliabilitas tingkat kesukaran dan daya beda pada butir soal ujian akhir semester ganjil mata pelajaran Matematika kelas XII IPS di SMA Negeri 12 Bandar Lampung tahun ajaran 2014/2015. *Al-Jabar: Jurnal Pendidikan Matematika*, 6(2), 203-218.
- Taib, F., & Yusoff, M. S. B. (2014). Difficulty index, discrimination index, sensitivity and specificity of long case and multiple-choice questions to predict medical students' examination performance. *Journal of Taibah University Medical Sciences*, 9(2), 110-114.
- Tikhonova, E., & Kudinova, N. (2015). Sophisticated thinking: Lower order thinking skills. In *event Conference: 2nd International Multidisciplinary Scientific Conference on Social Sciences and Arts SGEM*.
- Toksöz, S., & Ertunç, A. (2017). Item analysis of a multiple-choice exam. *Advances in Language and Literary Studies*, 8(6), 141-146.
- Towns, M. H. (2014). Guide to developing high-quality, reliable, and valid multiple-choice assessments. *Journal of Chemical Education*, 91(9), 1426-1431.
- Ugwu, N. G., & Mkpuma, S. O. (2019). Ensuring quality in education: Validity in teacher-made language tests in secondary schools in Ebonyi State. *American Journal of Educational Research*, 7(7), 518-523.
- Van Blerkom, M. (2008). *Measurement and statistics for teachers*. New York: Routledge.
- Wahyuni, S., & Putri, N. D. (2020). Karakteristik instrument penilaian akhir semester mata pelajaran Matematika di SMKN 1 Braja Selehah. *LINEAR: Journal of Mathematics Education*, 126-134.
- Widoyoko, E. P. (2009). *Evaluasi program pembelajaran*. Yogyakarta: Pustaka Pelajar, 238.
- Yanti, N. W. V. K., Paramartha, A. A. G. Y., & Wahyuni, L. G. E. (2020). The analysis of the quality of teacher-made multiple-choice tests used as summative assessment for English subject. *Indonesian Journal of Educational Research and Review*, 3(2), 35-45.
- Zaidi, N. L. B., Grob, K. L., Monrad, S. M., Kurtz, J. B., Tai, A., Ahmed, A. Z., & Santen, S. A. (2018). Pushing critical thinking skills with multiple-choice questions: Does Bloom's taxonomy work? *Academic Medicine*, 93(6), 856-859.
- Zimmaro, D. M. (2004). Writing good multiple-choice exams. *Measurement and Evaluation Centre. University of Texas at Austin. Web site*, 16.

PENILAIAN AKHIR SEMESTER GANJIL KELAS XI TA 2021/2022

FOR MULTIPLE CHOICES PLEASE CHOOSE THE CORRECT OPTION !

1. The definition of the suggestion is....
- an expression of willingness to do something.
 - an expression of an idea for someone else to do something.
 - the action of recommending something or someone.
 - a thing offered, especially as a gift or contribution
 - an expression of guidance or recommendations offered with regard to prudent future action.

2. Ronaldo: Sir, would you like me to close the window for you?

Mr. Adi: Yes, please. It's very cold inside.

The underlined sentence express:

- Asking help
- Offering help
- Refusing something
- Accepting an offer
- Suggestion

3. Romy: May I help you ?

Deny: Hmm. Don't bother yourself, thanks.

The underlined sentence express

- Asking help
- Giving help
- Offering help
- Refusing an offer
- Accepting an offer

4. Marta: It's very hot hereجامعة الرانري

Sinta: Sure. No problem

The suitable expression to complete the dialogue is .v.

- Could you close the door?
- Would you mind if I turn on the fan?
- Would you turn off the fan?
- May I leave now?
- Would you mind if I turn on the fan?

5. Ban: Nugy, where are you?

Nuy: I'm in the house. What's matter?

Ban: oh

Nuy: I'd like to help you. For what?

Complete the dialogue above with suitable expression

- Would you like some help?

- B. Sorry, I can't help you
- C. Could I possibly ask you to help me?
- D. Let me help you
- E. No, you don't need to help

6. Situation: I have toothache. What should I do?

The suggestion is proper, except...

- A. If I were you, I would go to the dentist.
- B. Why don't you go to the dentist?
- C. I would suggest to take pain killer pills. Maybe it will help?
- D. I'm sorry I never had toothache.
- E. You should visit your dentist!

7. The correct expression to give an opinion is

- A. I think that is beautify
- B. I think it's a beauty for me
- C. In my opinion it's awesomeness
- D. In my opinion it's wonderful
- E. According to me i'm wrong

8. Sista: about the math test?

Marda: It's too hard to do. I couldn't answer number 3.

- A. In your opinion
- B. Would you agree that
- C. What do you think
- D. I believe
- E. When do you attend

9. Amina: It's nearly school holiday. What are you going to do?

Marta:

Amina: Have a nice holiday!

Marta: Thank you.

- A. I think I am going to Bandung
- B. I don't know yet
- C. It's a secret
- D. What do you think about holiday?
- E. We will move

10. Ina: I like the new playground near our school. What about you?

Jun: I like it too, there is a lot of space, interesting games and a nice place to play.

We can conclude that Jun is expressing:

- A. Asking opinion
- B. expressing Doubt
- C. Disagreement
- D. Intention
- E. Giving opinion

11. Dili : What's your opinion about that novel?

Theo : I like it. It is an interesting story

From the dialogue we conclude that ...

- a. Dili is giving her opinion
- b. Theo doesn't like the novel
- c. Dili is asking Theo's opinion
- d. Dili agrees with Theo's opinion
- e. Dili is answering Theo's opinion

12. Liana: Film of Kartini it's so inspiring. And what do you think about the inspiring film of Kartini?

Bram: I think...

- A. She is a great artist
- B. She is a hero from Aceh
- C. I love Kartini's film because of her style
- D. I love Kartini's film it's so inspiring me also
- E. I love her because she is my neighbour

13. Jamilah: I think our city is saved from the virus, so we can go around.

Ellisa: But I don't think so,... Our city is not saved. We must stay at home at least in lock down time.

- A. In my opinion
- B. She forgets it
- C. See you next time
- D. You know it
- E. I am thinking of the virus

14. Marilia : Jane, I just change my hair Blunt Cut Bob style. Look at my hair...what do you think of my hair?

Jane : What a surprise! Your hair style is very good. In my mind, blunt cut bob hair style is the best trend fashion for hair right now. It's Korean girls' hair style right.

From the dialogue, we can conclude that...

- A. Jane give opinions about Marilia's hair style.
- B. Marilia give information about Jane's hair style**
- C. Jane try to response Marilia's style generally
- D. Marilia responses about the Jane's hair style.
- E. Jane give unsatisfied responses about Marilia's hair style.

READ THE TEXT BELOW TO CHOOSE THE CORRECT OPTION!

COVID-19 health protocol violators should be sanctioned. As we know, the number of people infected with the covid virus is still increasing in our country, but some people still ignore the health protocol by not wearing mask and keeping their distance.

Firstly, the violators will continue to ignore the health protocol suggested by the

government if they aren't sanctioned. By being sanctioned, they will feel deterred and ashamed so that in the future they will obey the rules given. Secondly, by sanctioning the violators, they will understand responsibility. Not only does it prevent us from being infected by the virus, but adhering to health protocols also keep others safe. We may not feel infected, but at least we prevent bad things by continuing to keep distance and wearing masks. That way, we will fulfill our responsibilities to ourselves and others. Lastly, giving sanctions to the violators will make them more aware of the danger of COVID-19. It is said that WHO declared COVID-19 a pandemic, which means that this virus is very dangerous, so it is best to continue to maintain health and avoid COVID-19.

Therefore, the health protocol violators should be sanctioned because of the reasons given above.

15. The type of the reading text above is.....

- A. Analytical exposition
- B. Hortatory exposition
- C. Procedure text
- D. Narrative text
- E. News item text

16. Situation: You want to invite your highschool teacher to join your new restaurant grand opening ceremony. The teacher's name is John Snow. How you write salutation for opening of the letter?

- A. Dear Mrs John
- B. Dear Mr Snow
- C. Dear Mrs Snow
- D. Dear John Snow
- E. Hi Mr. Jhon Snow

17. What is the complete structure of formal invitation letter?

- A. date, sender, recipient, subject, salutation, introduction, body, appreciation for the guest, salutation, signature
- B. date, sender, recipient, subject, introduction, body, appreciation for the guest, salutation, signature
- C. date, sender, recipient, subject, salutation, introduction, body, appreciation for the guest, signature
- D. date, recipient, subject, salutation, introduction, body, appreciation for the guest, salutation, signature
- E. date, recipient, subject, salutation, introduction, body, appreciation for the guest, salutation, signature, bye

18. What do we write in introduction of a formal invitation letter?

- A. introducing yourselves, from your name, your hobby, and what do you do for living.
- B. Introducing yourselves and what the invitation is about
- C. Introducing yourselves as well as the organization that you are presenting.

- D. Introducing the event of formal invitation letter
- E. Introducing the events such as birthday party event

19. The difference between formal invitation letter and the informal one, EXCEPT...

- A. Formal invitation is used to invite someone to non-formal event such as birthday party while the informal invitation is used to invite to a close friends such as wedding party
- B. Formal invitation is used to invite someone to formal event such as business gathering while informal invitation is used to invite someone to informal event.
- C. Formal invitation restricted the recipient to wear certain uniform such as suit and dress while informal one does not have such restriction.
- D. Formal invitation must be reply with RSVP respond card while an informal invitation does not require the invitee to reply the message.
- E. Formal Invitation is used the formal language while the informal viceversa

Please join us
 as we celebrate our
 40th Anniversary
 Rahmat and Wati Juminten
 Friday, May 7th
 6.30p.m -10.OOp.m
 Pondok Indah, South Jakarta
 (Dinner is available)
 RSVP 65412397
 (Nanang/ Nini/ Nita)

20. The purpose of the invitation text above is...?

- A. To tell an good relationship
- B. To describe a certain party
- C. To persuade people to have marriage
- D. To invite people to join the party
- E. To share the happiness

Dear Bianca Jodie,
 Time counted so fast
 Let's celebrate new year's eve together!
 Sunday, December 31st
 9 p.m. until drop
 at Dragonfly Bar
 Jln Gatot Subroto
 Sincerely,
 Hannah

21. From the invitation text above, we can say that?

- A. Bianca and Hannah have a good relationship

- B. Bianca Jody invites Hannah to celebrate new year's eve together
- C. The Dragonfly Bar is not at Jln. Gatot Subroto
- D. The party started from 9 o'clock in the morning
- E. The party ended at midnight

Mr. & Mrs. Kenedy
Glad to invite you to come
The marriage of their daughter
Chelsea Islan

To
Argavi Koto
at Alila Villas Uluwatu Hotel, Bali
on Friday, 1 November 2018
Started on 1 p.m

Please RSVP before 30 October to Rahmat
There will be Maroon 5 performance
We wait your presence.

22. The fact from the invitation text above is...
- A. The Chelsea island and Rahmat will get married on Bali
 - B. Mr & Mrs. Kenedy are celebrating their marriage in Alila Villas Uluwatu Hotel
 - C. You can inform Argavi Koto to attend the event
 - D. There will be Maroon 5 performance on that wedding party.
 - E. Rahmat is Kenedy's son

The Dangers of Typing SMS while Driving

The popularity of mobile devices has had some dangerous consequences. We know that mobile communications are linked to a significant increase in distracted driving which results in injury and loss of life. The National Highway Traffic Safety Administration reported that in 2010 driver distraction was the cause of 18 percent of all fatal crashes with 3,092 people killed and crashes resulting in an injury 416,000 people wounded. The Virginia Tech Transportation Institute found that text messaging creates a crash risk 23 times worse than driving while not distracted. Eleven percent of drivers aged 18 to 20 who were involved in an automobile accident and survived admitted they were sending or receiving texts when they crashed. Distracted driving endangers life and property and the current levels of injury and loss are unacceptable.

23. What is the main idea of the passage?
- A. The warning of texting and driving.
 - B. The debatable issue of texting and driving.
 - C. The involvement of mobile devices while driving.
 - D. The risks of texting while driving.
 - E. The consequences of not paying attention traffic.

24. From the text above, we know that ...
- A. The mobile phone should be banned in the street.
 - B. Distracted driving makes accident more rarely to happen.
 - C. Most of the accident caused by the condition of the road.
 - D. 18 percent of fatal crashes were caused by unnoticed street signs.
 - E. Drivers involved in car accidents admitted they were texting when they crashed.

Drug users are actually ill people who need help. Rehabilitation is one of the main things they should get. Instead of punishing drug users in prisons, the government should provide rehabilitation for them. Most users are actually victims of persuasive peddlers and they suffer from consuming the drugs. Therefore, by rehabilitating them, we are actually helping them out from traps, which they might accidentally step on. If we only arrest drug users and send them to jail, this doesn't solve the problem as drug users will still be addicted unless proper rehabilitation is imposed to them. Providing rehabilitation programs for drug users and forcing them to participate in the programs is better than only arresting them and doing nothing to their illness. Only drug dealers, traffickers, and wholesalers should be imprisoned.

25. Sending the drug users to prison is not a good solution because
- A. They can't consume harmful drugs
 - B. They are treated normally
 - C. They may still be infected with harmful drugs
 - D. They can be freed as soon as possible
 - E. They are given the whole recovery program
26. "...drug users will still be addicted unless proper rehabilitation is imposed to them." (Paragraph 3) The underlined word is closest in meaning to
- A. Influenced
 - B. Treated
 - C. Forced
 - D. Burdened
 - E. Implemented

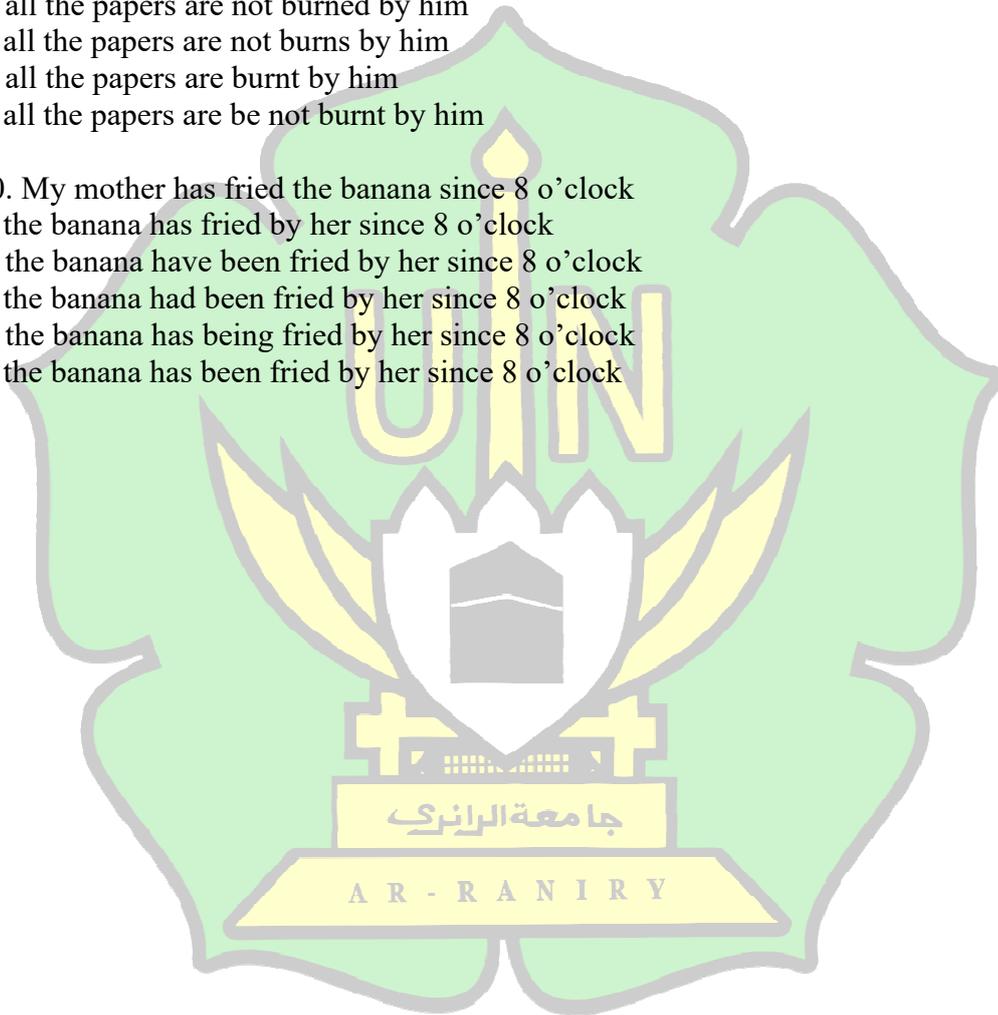
For question no 27-30 please choose the correct passive voice!

27. They are doing the homework together
- A. the homework is being did by them together
 - B. the homework is being done by them together
 - C. the homework is being do by them together
 - D. the homework is done by them together
 - E. the homework is did by them together

28. Will you leave the car here?
- a. will the car leaved by you here?
 - b. will the car left by you here?
 - c. will you be left by the car here?
 - d. will the car be left by you here?
 - e. will the car be lefting by you here?

29. He doesn't burn all the papers
- a. all the papers are not burnt by him
 - b. all the papers are not burned by him
 - c. all the papers are not burns by him
 - d. all the papers are burnt by him
 - e. all the papers are be not burnt by him

30. My mother has fried the banana since 8 o'clock
- a. the banana has fried by her since 8 o'clock
 - b. the banana have been fried by her since 8 o'clock
 - c. the banana had been fried by her since 8 o'clock
 - d. the banana has being fried by her since 8 o'clock
 - e. the banana has been fried by her since 8 o'clock



Appendix 2. The distribution of Difficulty and Discrimination Indices

NO	Question Item																														Total			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30				
1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	26		
2	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	25		
3	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	25		
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	0	1	0	1	0	1	1	25	
5	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1	0	24		
6	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1	1	1	24		
7	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	0	1	24		
8	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	0	1	24		
9	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	24		
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	0	1	24		
11	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	0	1	1	1	0	1	24		
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	0	1	0	1	0	1	24		
13	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	1	1	1	1	1	0	1	1	1	0	1	23		
14	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0	0	0	1	1	1	1	1	23	
15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	23		
16	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	1	1	1	1	0	1	0	1	1	1	23		
17	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	0	1	0	1	0	1	23	
18	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	1	1	0	0	1	1	1	1	1	1	1	23	
19	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1	1	1	23
20	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	0	1	1	1	0	1	23	
21	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	23	
22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0	1	0	0	1	23	
23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	0	0	1	0	0	1	23	
24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	1	0	0	1	1	0	23	
25	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	0	1	1	1	23	
26	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	0	1	1	0	1	1	0	1	1	1	1	23	
27	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	23	
28	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	0	0	1	1	0	1	1	1	1	1	23	
29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	1	0	1	0	1	1	23	
30	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	0	1	0	0	1	0	0	1	1	1	23	
31	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	23	
32	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	23
33	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	0	1	0	23	
34	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	23	
35	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	23	
36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	0	1	1	0	1	23	
37	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	1	0	1	1	1	1	0	1	23	
38	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	0	1	0	1	1	1	0	23	
39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0	1	1	0	1	1	0	1	1	1	0	1	23	

4	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	0	1	0	0	1	1	1	2
4	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	0	1	0	1	1	1	1	0	0	1	1	0	2	
4	2	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0	1	0	0	1	1	1	1	2		
4	3	1	0	1	1	1	0	1	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	0	1	0	0	1	1	1	2		
4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	0	1	1	0	0	0	1	1	2		
4	5	1	1	0	1	1	1	0	1	1	0	1	1	1	1	1	0	0	0	1	1	1	0	0	1	1	1	1	1	1	2		
4	6	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	0	1	0	1	0	1	1	0	1	1	1	2		
4	7	1	0	1	1	0	0	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	0	1	1	2		
4	8	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	0	0	1	1	0	0	0	1	1	1	1	1	1	1	2		
4	9	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	0	1	0	1	0	1	1	0	1	1	1	2		
5	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	0	0	1	1	1	2			
5	1	1	1	1	1	0	1	1	1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	1	1	1	1	2		
5	2	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	0	1	1	2			
5	3	0	0	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	1	1	1	1	1	0	1	0	1	1	2			
5	4	0	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1	1	0	1	0	1	1	2				
5	5	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	0	0	1	1	0	2				
5	6	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	1	0	1	1	1	2				
5	7	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	0	1	2			
5	8	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	0	1	1	0	1	1	1	0	2			
5	9	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	0	0	1	0	1	1	1	2				
6	0	0	1	0	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	1	1	1	1	0	1	1	2				
6	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	0	0	1	0	1	0	1	1	0	1	1	1	1	2			
6	2	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	1	1	0	0	0	1	0	2			
6	3	0	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	1	0	1	1	1	0	1	0	1	0	1	2			
6	4	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	0	0	1	1	0	1	2			
6	5	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	0	1	0	1	2			
6	6	1	1	0	1	1	1	0	1	1	1	1	1	1	0	0	1	0	1	1	0	1	1	1	1	0	1	0	1	2			
6	7	1	1	0	0	1	1	1	1	0	1	1	1	1	0	1	0	1	0	1	1	0	1	1	0	1	1	1	1	2			
6	8	1	1	1	1	0	1	1	1	1	0	1	1	1	0	1	0	1	0	0	0	1	1	0	1	1	1	1	1	2			
6	9	1	1	0	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	1	1	1	1	1	2			
7	0	0	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1	0	1	0	1	1	2				
7	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	1	1	0	1	1	0	1	0	1	2			
7	2	1	1	1	1	0	0	1	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0	1	0	1	0	2			
7	3	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1	1	0	0	0	1	0	1	2			
7	4	0	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	0	1	1	1	0	1	1	1	0	1	2			
7	5	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	0	1	0	0	0	1	0	1	2			
7	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	0	1	0	0	0	1	1	0	2			
7	7	0	1	1	1	1	0	1	0	1	0	1	1	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1	2			
7	8	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0	1	0	0	0	1	0	1	1	1	0	2			
7	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	2			
8	0	0	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1	0	1	0	1	0	2			
8	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	1	1	1	0	1	0	0	1	0	1	0	2			

147	1	0	0	1	0	1	0	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	0	1	0	0	1	0	1	1	1	1	6	
148	1	1	0	1	1	1	0	1	1	1	0	1	1	0	0	0	0	1	0	1	1	0	0	1	0	0	0	0	1	1	1	6		
149	1	1	0	0	0	0	1	1	1	1	1	0	1	1	1	0	1	0	0	1	0	0	1	1	0	0	0	0	1	1	1	6		
150	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	1	0	0	0	0	1	0	1	6		
151	1	1	1	1	1	0	0	0	1	0	0	0	1	1	0	0	1	0	0	1	1	1	0	1	1	0	1	0	1	0	1	6		
152	1	0	0	1	1	1	0	1	1	1	1	0	1	1	1	0	0	0	1	0	1	0	1	0	1	0	0	0	1	0	1	6		
153	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0	1	0	1	6	
154	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	1	0	0	0	0	0	1	6	
155	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0	6	
156	1	1	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	1	6	
157	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	1	0	1	6	
158	1	1	1	1	1	1	0	0	1	1	1	0	0	0	1	1	0	0	0	1	0	0	0	1	1	0	0	0	1	0	1	0	5	
159	1	0	1	1	0	0	1	1	1	0	1	1	1	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0	1	1	1	5		
160	1	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	5	
161	1	1	0	1	0	0	1	0	1	0	0	1	0	1	1	0	1	1	1	1	0	0	0	0	1	0	1	0	1	0	1	0	5	
162	0	1	1	1	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	0	1	0	0	0	0	1	0	1	0	5	
163	1	0	1	1	0	0	0	1	1	0	1	1	1	1	0	0	0	0	0	1	1	0	1	0	1	0	0	0	1	1	1	5		
164	0	0	0	1	1	0	0	1	1	0	1	1	0	1	1	1	0	0	0	1	1	1	0	1	0	0	0	1	0	1	0	1	5	
165	0	0	0	1	1	0	0	1	1	0	1	1	0	1	1	1	0	0	0	1	1	1	1	0	1	0	0	0	1	0	1	0	5	
166	0	1	0	1	1	0	0	1	1	0	1	1	0	0	1	1	0	0	0	1	1	1	0	1	0	0	0	1	0	1	0	1	5	
167	1	0	0	0	1	0	0	1	1	0	1	1	1	1	1	0	0	1	0	1	0	0	1	1	0	0	1	0	1	0	1	0	5	
168	1	1	1	0	0	0	1	1	1	0	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0	1	1	1	0	1	5	
169	0	0	0	1	1	1	0	1	1	1	1	1	1	1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	1	1	5		
170	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	1	0	5	
171	1	1	0	1	1	0	1	0	1	0	1	1	1	1	0	0	0	1	0	1	0	0	0	0	1	0	0	1	1	0	1	0	5	
172	0	0	1	1	0	0	1	1	1	1	1	1	0	1	1	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	5
173	1	0	0	1	1	0	0	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1	4		
174	0	0	0	1	0	0	0	1	1	1	1	0	1	1	1	0	0	0	0	1	0	1	0	1	1	0	0	0	1	1	1	4		
175	1	1	0	1	1	1	0	1	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	4	
176	0	0	0	1	1	0	0	1	1	0	0	0	0	1	1	0	0	0	1	0	1	1	1	1	1	1	0	0	1	0	1	0	4	

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
	Sig	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	N	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255		
29	PC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	Sig	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	N	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	
30	PC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	Sig	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	N	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	
Jumiah	PC	2.000	3.000	5.000	3.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4.000
	Sig	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	N	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255	255

*. Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Appendix 4. Reliability of Question Items

Reliability Statistics	
Cronbach's Alpha	N of Items
.802	30

جامعة الرانري

AR - RANIRY



Appendix 6. Appointment Letter from Supervisors

SURAT KEPUTUSAN DEKAN FAKULTAS TARBİYAH DAN KEGURUAN UIN AR-RANIRY
 Nomor : B-2228/UIN.08/PTK/KP.07.692/2022

TENTANG
PENGANGKATAN PEMBIMBING SKRIPSI MAHASISWA FAKULTAS TARBİYAH DAN KEGURUAN
UIN AR-RANIRY

DEKAN FAKULTAS TARBİYAH DAN KEGURUAN UIN AR-RANIRY

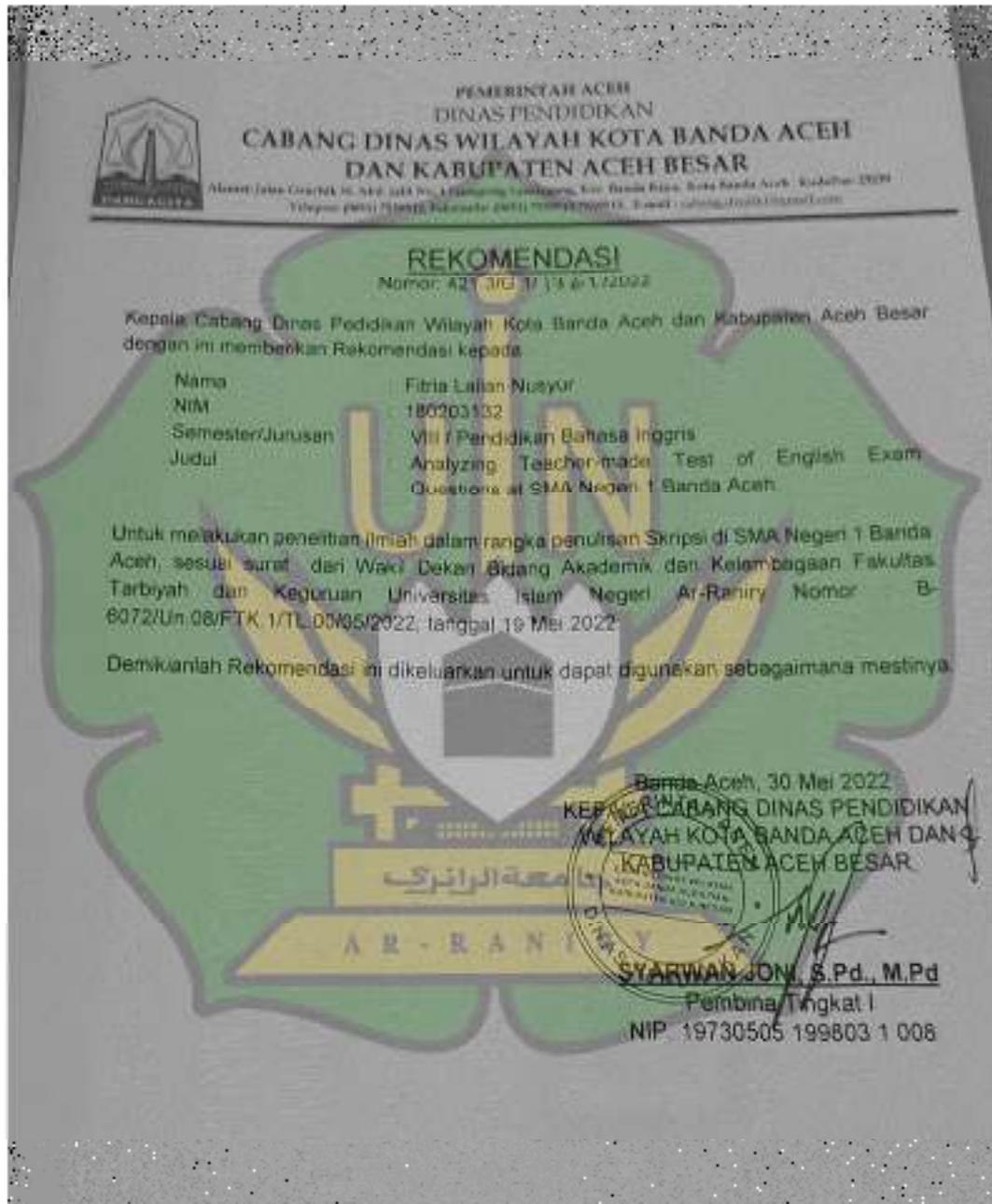
- Menimbang** :
- a. bahwa untuk kelancaran bimbingan skripsi dan ujian monevnya mahasiswa pada Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry Banda Aceh, maka dipandang perlu menunjuk pembimbing skripsi tersebut yang dituangkan dalam Surat Keputusan Dekan;
 - b. bahwa saudara yang tersebut namanya dalam surat keputusan ini dipandang cakap dan memenuhi syarat untuk diangkat sebagai pembimbing skripsi.
- Mengingat** :
- 1. Undang-undang Nomor 20 Tahun 2003, tentang Sistem Pendidikan Nasional;
 - 2. Undang-undang Nomor 14 Tahun 2005, tentang Guru dan Dosen;
 - 3. Undang-undang Nomor 12 Tahun 2012, tentang Pendidikan Tinggi;
 - 4. Peraturan Pemerintah Nomor 74 Tahun 2012 tentang Perubahan atas Peraturan Pemerintah RI Nomor 23 Tahun 2005 tentang Pengelolaan Keuangan Badan Layanan Umum;
 - 5. Peraturan Pemerintah Nomor 4 Tahun 2014, tentang Penyelenggaraan Pendidikan Tinggi dan Pengelolaan Perguruan Tinggi;
 - 6. Peraturan Presiden RI Nomor 64 Tahun 2013, tentang Perubahan IAIN Ar-Raniry Banda Aceh Menjadi UIN Ar-Raniry Banda Aceh;
 - 7. Peraturan Menteri Agama RI Nomor 12 Tahun 2014, tentang Organisasi dan Tata Kerja UIN Ar-Raniry Banda Aceh;
 - 8. Peraturan Menteri Republik Indonesia No. 31 Tahun 2015, tentang Statuta UIN Ar-Raniry;
 - 9. Keputusan Menteri Agama Nomor 492 Tahun 2003, tentang Pendelegasian Wewenang, Pengangkatan, Pemindahan dan Pemberhentian PNS di Lingkungan Departemen Agama Republik Indonesia;
 - 10. Keputusan Menteri Keuangan Nomor 193/KM.K.05/2011, tentang Penetapan Institut Agama Islam Negeri Ar-Raniry Banda Aceh pada Kementerian Agama sebagai Instansi Pemerintah yang Menerapkan Pengelolaan Badan Layanan Umum;
 - 11. Keputusan Rektor UIN Ar-Raniry Nomor 01 Tahun 2013, tentang Pendelegasian Wewenang kepada Dekan dan Direktur Pascasarjana di Lingkungan UIN Ar-Raniry Banda Aceh.
- Memperhatikan** :
- Keputusan Sidang Seminar Proposal Skripsi Prodi Pendidikan Bahasa Inggris Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry Tanggal 04 Januari 2022
- MEMUTUSKAN**
- Menetapkan**
PERTAMA :
- | | |
|--|--|
| Menunjuk Saudara:
1. Dr. Nashriyah, M.A.
2. Khairiah Syahabuddin, M.HSc, EdS., M.TESOL., Ph.D. | Sebagai Pembimbing Pertama
Sebagai Pembimbing Kedua |
|--|--|
- Untuk membimbing Skripsi :
- | | |
|---|--|
| Nama : FHRiz Ladan Nasyar
NIM : 180203132
Program Studi : Pendidikan Bahasa Inggris
Judul Skripsi : Analyzing Teacher-Made Test of English Exam Questions at SMAN 1 Banda Aceh | Sebagai Pembimbing Pertama
Sebagai Pembimbing Kedua |
|---|--|
- KEDUA** : Pembagian tugas dan pembimbing pertama dan kedua tersebut diatas ditetapkan pada DIPA UIN Ar-Raniry Banda Aceh Tahun 2020, No.025.04.3.431925/2020, tanggal 12 November 2019.
- KETIGA** : Surat keputusan ini berlaku sejak akhir semester Ganjil Tahun Akademik 2022/2023
- KEEMPAT** : Surat Keputusan ini berlaku sejak tanggal ditetapkan dengan ketentuan apabila keesokan harinya diubah dan diperbaiki kembali sebagaimana mestinya apabila kemudian hari ternyata ada terdapat kekeliruan dalam penetapan ini.

Ditetapkan di: Banda Aceh
 Pada Tanggal: 14 Februari 2022
 Dekan,


 Muslim Razali

Tembusan
 1. Dekan UIN Ar-Raniry (sebagai laporan)

Appendix 7. Recommendation Letter from Branch of the Education Office of the City of Banda Aceh



Appendix 8. Recommendation Letter from FTK
UIN AR-Raniry to Conduct Field Research

6/29/22, 2:27 PM

Document



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI AR-RANIRY
FAKULTAS TARBIIYAH DAN KEGURUAN
Jl. Syekh Abdur Rauf Kapelton Universitas Banda Aceh
Telepon : 0651-7577321, Email : ar@uinaraniry.ac.id

Nome : B-6072/Un.06/FTK.1/TL.00/05/2022
Lamp : -
Hal : *Penelitian Ilmiah Mahasiswa*

Kepada Yth,

1. Kepala Dinas Pendidikan Kota Banda Aceh
2. Kepala Sekolah SMAN 1 Banda Aceh

Assalamu'alaikum Wa. Wa.

Pimpinan Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry dengan ini menerangkan bahwa:

Nama/NIM : **FERRIA LAILAN NUSYUR / 180203132**
Semester/Jurusan : VIII / Pendidikan Bahasa Inggris
Alamat sekarang : Jl. Mantri Zaimin No. 1 Dusun Delima Gampoeng Baro Kec. Meuraxa, Banda Aceh.

Saudara yang tersebut namanya diatas benar mahasiswa Fakultas Tarbiyah dan Keguruan bernaksud melakukan penelitian ilmiah di keluarga yang Bapak/Tu sebagai dalam rangka penelitian Skripsi dengan judul *Analyzing Teacher-made Test of English Exam Questions at SMAN 1 Banda Aceh*

Demikian surat ini kami sampaikan atas perhatian dan kerjasama yang baik, kami mengucapkan terimakasih.

Banda Aceh, 19 Mei 2022
an. Dekan
Wakil Dekan Bidang Akademik dan Kelembagaan,



Berlaku sampai : 19 Juni 2022

Dr. M. Chalis, M.Ag.

جامعة الرانيري
AR - RANIRY

AUTOBIOGRAPHY

1. Name : Fitria Lailan Nusyur
2. Place/Date of Birth : Banda Aceh, 06 January 2001
3. Gender : Female
4. Religion : Islam
5. Nationality : Indonesian/Acehnese
6. Address : Gampong Baro, Banda Aceh
7. Marital Status : Single
8. Occupation/NIM : Student/180203132
9. The Parent
 - a. Father's Name : Marzuki A.B
 - b. Mother's name : Rahmiati, S.Pd.
 - c. Occupation : Retired
 - d. Address : Gampong Baro, Banda Aceh
10. Siblings
 - a. Sister : Vira Maghfirah
 - b. Brother : Ferry Firdaus, S.Tr,Pi.
11. Educational Background
 - a. Elementary School : MIN Model Banda Aceh (2006-2012)
 - b. Junior High School : MTsN Model Banda aceh (2012-2016)
 - c. Senior High School: SMAN 1 Banda Aceh (2015-2018)
 - d. University : UIN Ar-Raniry Banda Aceh (2018-2022)

Banda Aceh, 28 Juli 2022

Fitria Lailan Nusyur