INVESTIGATING TEACHERS' COLLABORATION IN DEVELOPING TEST ITEMS

THESIS

Submitted by

<u>WIRDA YATUL</u>
NIM. 180203072

Student of *Fakultas Tarbiyah dan Keguruan*

Department of English Language Education

FACULTY OF TARBIYAH AND TEACHING TRAINING

AR-RANIRY STATE ISLAMIC UNIVERSITY

BANDA ACEH

2024 M / 1446 H

THESIS

Saya yang bertandatangan di bawah ini:

Name : Wirda Yatul

Reg. No :180203072

Tempat/tanggal lahir : Seubadeh/ 06 Maret 2000

Alamat : Seubadeh, Kec. Bakongan Timur, Kab. Aceh Selatan

Menyatakan dengan sesungguhnya bahwa skripsi yang berjudul:

Investigating Teachers' Collaboration in Developing Test Items

Adalah benar-benar karya saya, **kecuali semua kutipan dan referensi yang disebutkan sumbernya**. Apabila terdapat kesalahan dan kekeliruan di dalamnya,maka akan sepenuhnya menjadi tanggung jawab saya. Demikianlah surat pernyataan ini saya buat dengan sesungguhnya.

Banda Aceh, 04 April 2024

Saya yang membuat pernyataan

Wirda Yatul

It has been defended in *Sidang Munaqasyah*
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 in English Language Teaching

On:

Thursday, June 13th 2024
6 Zulhijjah 1445H

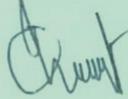In Darussalam, Banda Aceh

Board of Examiner,

Chairperson,   Secretary,

Syarifah Dahliana, M.Ag., M.Ed., Ph.D   Azizah, S.Ag., M.Pd

Member,   Member,

Chamisah, S.Ag., M.Ed   Rahmi, S.Pd.I., M.TESOL., Ph.D

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

Prof. Safrul Muluk, S.Ag., M.A., M.Ed., Ph.D
NIP. 197301021997031003

SURAT PERNYATAAN KEASLIAN

(Declaration of Originallity)

Saya yang bertandatangan di bawah ini:

Name : Wirda Yatul

Reg. No :180203072

Tempat/tanggal lahir : Seubadeh/ 06 Maret 2000

Alamat : Seubadeh, Kec. Bakongan Timur, Kab. Aceh Selatan

Menyatakan dengan sesungguhnya bahwa skripsi yang berjudul:

Investigating Teachers' Collaboration in Developing Test Items

Adalah benar-benar karya saya, **kecuali semua kutipan dan referensi yang disebutkan sumbernya**. Apabila terdapat kesalahan dan kekeliruan di dalamnya,maka akan sepenuhnya menjadi tanggung jawab saya. Demikianlah surat pernyataan ini saya buat dengan sesungguhnya.

Banda Aceh, 04 April 2024

Saya yang membuat pernyataan
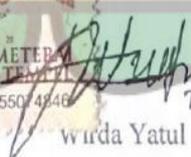
Wirda Yatul

ACKNOWLEDGMENTS

بسم الله الرحمن الرحيم

Alhamdulillahirabbil'alamin, my praise and gratitude for Allah subhanahuwa ta'ala who has given His grace, guidance, blessing, inspiration, and good health so that I can complete this thesis. Shalawat and salam may always be upon our Prophet Muhammad shalallahu 'alaihi wasallam, who has brought us from darkness to a brightness and inspired us to be good personalities in life.

First of all, my deepest gratitude and appreciation go to my thesis supervisors, Syarifah Dahliana, M.Ag., M.Ed., Ph.D. and Azizah, S.Ag., M.Pd. for their guidance, advice, support, kindness, valuable input, and useful knowledge in completing this thesis. My appreciation also goes to my academic supervisor Azizah, S.Ag., M.Pd. who has guided me since my first semester in the English Education Department. May Allah give a special place in Jannah for you. Secondly, my appreciation and gratitude for the lecturers of the English Education Department who have taught and provided a way for me to complete this thesis.

Furthermore, I owe my deepest gratitude to my beloved mother and father, Yusfarni and Sulaiman, LB, for their extraordinary kindness, infinite love, prayers, sacrifices, patience, and endless support, both moral and material, so that I can pursue my education up to this level. I would also like to thank my family members, my younger sister Imang, my sister Triska Rahmadayati and her

Finally, I realize that this thesis still needs constructive ideas to reduce its shortcomings. I hope this thesis makes a valuable contribution to the students and lecturers of the English Education Department of UIN Ar-Raniry Banda Aceh. May Allah subhanahuwa ta'ala always bless us all.

Banda Aceh, 20 April 2024
The writer,

Wirda Yatul

ABSTRACT

Name	: Wirda Yatul
NIM	: 180203072
Faculty	: Fakultas Tarbiyah dan Keguruan
Major	: Department of English Language Education
The Working Title	: Investigating Teachers' Collaboration in Developing Test Items
Main Supervisor	: Syarifah Dahliana, S.Ag., S.E., M.Ag., M.Ed., Ph.D
Co-Supervisor	: Azizah, S.Ag. M.Pd
Keyword	: Teachers' collaboration; Test items

Test item is one of the evaluation tools that provide information for teachers regarding the level of students' understanding. One of the activities to improve the quality of test items is by teachers' collaboration. Teachers' collaboration in developing test items is the process by which a number of teachers work together to develop a set of items that are relevant, meaningful, and aligned with learning objectives. The purpose of this research is to find out the teachers' collaboration activities in developing test items, teachers' awareness in constructing summative test, and challenges while collaborating to develop test items. Data were obtained from 3 English teachers in MAN Model Banda Aceh through semi-structured interview. The research design of this study was qualitative research. The result of this study showed some stages of teachers' collaboration in developing test items; these stages were conducting monthly *MGMP* meeting, conducting a deeper discussion with the same grade teacher, rechecking the test items and pretesting the students. Then, the result also showed teachers' awareness of constructing English summative test which included determining type of test, utilizing the sources, dividing the number of questions, selecting indicators, and choosing distractors. However, the teachers faced some challenges in developing test items collaboratively, they were differences in teaching approaches, challenge in time coordination, challenge in established curriculum cohesivity and difference in understanding the difficulty level of the test items.

TABLE OF CONTENT

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

This chapter presents the introduction of this research. It introduces the background of study, research question, aim of the study, significance of study and terminology.

A.   Background of Study

Learning program is influenced by various components, each of the component will require each other to achieve the learning objectives. Several components that support the learning process are objectives, materials, methods, learning tools and evaluation (Suparman, 2020). Successful of the learning process cannot be separated from the role of the teacher. The teacher's responsibility is to ensure that every component supporting the learning process must run properly. The correlation between teachers and students will determine the success of learning which affects the learning objectives. One of the ways that can be done to optimize the learning process is by improving the quality of evaluation.

Evaluation will be useful to provide information that can be used by teachers and students as internal feedback to improve themselves and teaching and learning activities (Detlev, et al. 2017). Evaluation has a very strong continuity with the learning process, because evaluation is a measuring tool used to determine the abilities that students have achieved regarding the teaching material that has been delivered and difficulties in the learning process so teachers can take advantages of it to repair materials and develop the learning programs.

Therefore, it is necessary to evaluate students in order to obtain information on the extent to which students understand the lessons taught by the teacher or not.

Evaluation is a process of collecting data to determine what and how the educational goals have been achieved (Suharsimi, 2013). The learning ability of students obtained from learning activities at school is always in line with the objectives listed on the indicators that have been planned by the teacher, where in compiling or determining the assessment and learning outcomes of indicators, the teacher refers to the taxonomy of educational objectives developed by Bloom, which consist of knowledge (cognitive domain), attitudes (affective domain), and skills (psychomotor domain), the three of which can be further detailed into various abilities that need to be developed in every learning process (Arikunto, 2005)

To achieve a good quality of learning, a good assessment system is needed as well. In order to establish the assessment work properly, in accordance with the objectives that have been stated, it is very necessary to establish an assessment standard that becomes the basis and reference for teachers and education practitioners in conducting assessment activities (Khan, 2018). To achieve a good quality of learning, it requires a good cooperation, such as teachers, students, and schools with different roles according to their respective proportions. Thus, learning evaluation plays a role in determining the efficiency of the learning process that has been implemented and the effectiveness of achieving the learning objectives.

To find out the achievement of learning objectives, it needs to be evaluated periodically and continuously. Based on teacher's role to evaluate student's competence, then the teacher must have the ability to develop a good test. Usability test is to measure the ability of students after learning process. Xu and Brown (2016) stated that successful evaluation highly depends on the evaluation tools used, one of it is a test. Test is a method to measure one's abilities and skills.

According to Sudjana (2013), the test as an evaluation tool is a question given to students to get answers from students in oral test, in written test, or in the action test. As outlined by Popham (2008), there are two primary assessment designs called summative and formative. Summative assessment appraises a student's standing in accordance with a learning target or standard (Tomlinson & Moon, 2013). On the other hand, formative assessment analyzes the current work of students and its connection to their learning objectives (Brookhart & Moss, 2019). By giving tests, it is very possible for teachers to get information related to students' abilities and the success of the techniques used by the teacher concerned in learning activities. To be able to establish test that meet the requirements is quite difficult because compiling tests requires high knowledge, skills and accuracy.

In addition, based on the test results, the teacher can find the effectiveness of the syllabus and also the method applied. Tests that aim to measure students' understanding of the material being taught must occupy the characteristics of a good test (Shohamy, 2020). In the evaluation system, learning through tests must

have a grading standards system in accordance with the regulations that have been approved at school.

The statement above explains that the test must be systematically arranged and linked to the curriculum. So that the learning process will run effectively without any overlap. Developing good test items requires careful arrangement. The test is a mandatory instrument that has good quality, because the quality of the test will affect the test output. if the test is good, then the results will be a reference for teachers in directing appropriate learning towards student achievement.

In developing test items, one of the activities that can be done by the teachers is conducting collaboration among teachers or commonly known as MGMP (*Musyawarah Guru Mata Pelajaran*). It means the preparation of test instruments is carried out simultaneously in a certain area and involves many teachers with different scopes including grid makers, question makers, question reviewers and language reviewers (Zaman, Syahiduz, et al., 2018).

The issue is teachers individually do not find it easy, fast, and precise to prepare good test and meet the characteristic of a good test. This collaboration activity is expected to help the development of test instruments in various regions goes well and appropriately according to learning evaluation standards. The practice of teacher's collaboration in developing test items has been found to produce more valid and reliable assessments (Sugiwanti, 2021). Several studies have investigated the advantages of teacher collaboration in test development. These studies suggest that involving multiple teachers in creating test items can

improve content validity, increase teacher motivation, and promote more comprehensive coverage of the curriculum (Amirullah, 2022; Fives and Buehl, 2012).

One of the previous studies was conducted by Amirullah (2022), this study identified and described the ability of teachers at SMAN 1 Riau Silip Kab. Bangka, in preparing the summative test through the MGMP work meeting. The results of this research show that the application of MGMP work meetings in preparing odd semester summative tests is very effective. The abilities of the teachers at SMAN 1 Riau Silip Kab.bangka after holding the MGMP work meeting increase rapidly, because through the MGMP work meeting, the teachers were able to be more disciplined in completing their assignments. Through MGMP work meetings, it can improve teacher competence in preparing professional tests, this can be seen in the results obtained. Through the MGMP work meeting, teachers at SMAN 1 Riau Silip, Bangka have been able to prepare summative test, based on predetermined test making criteria.

Another study by Fives and Buehl (2012) examined how collaboration can help teachers develop more authentic assessments. The authors argued that when teachers work together on test development, they are more likely to consider students' diverse experiences, strengths, and needs. It can lead to more valid and relevant assessments that reflect the complex realities of student learning. Overall, this study is investigating teacher collaboration in developing test items. It is expected to lead to more effective assessment practices. By bringing together multiple perspectives and expertise, teachers can create tests that better reflect the

needs and goals of their students and promote more comprehensive learning outcomes. Collaboration can take different forms, such as peer reviewing, item writing, or teaming up to develop a bank of test items. Teachers may face several challenges in collaborating, such as lack of time, resources, and expertise. These challenges can hinder the effectiveness of teacher collaboration, adversely impacting the quality of test items.

After looking further, the researcher can conclude the similarities and differences between previous studies and this study. the similarity obtained from previous studies and this study is that the same research subject, the subject are teachers who incorporated of collaborative activities in developing test items. While the difference is on the research focus; previous research focused on the effectiveness of teacher collaboration in making test items so this research is classified as school action research by conducting observations and comparisons between phase I and II. It compared the results of teachers' ability in developing test items before collaborating with other teachers (phase I) and teachers' ability in developing test items after joining collaborative activities (phase II). While in this study, researcher focused on how teachers collaborate in developing test item and what challenges was experienced by teachers when collaborating, so the research method in this study used qualitative and the data obtained in the form of interviews from teachers' personal experiences.

In reality, based on researcher's experience during teaching practice, the researcher found that teachers used to create a test individually. It sometimes causes teachers rarely create appropriate test, sometimes, a teacher only uses the

existing tests, then only adjusting to the subject matter being taught. Therefore, it will affect the quality of the test items, it will cause inaccuracy between teaching materials and test items. On the other hand, some teachers are not competence to create test, so they often look for several collections of existing test items (question banks). To avoid this situation, it is very important for educational stakeholders to assist teachers to improve their abilities in developing test items. One of it is by conducting a collaboration program between English teachers. So that teacher can gather, exchange thoughts and opinions and develop test items, with expectation they will produce better quality of test items.

Considering the issue of low-quality test items is widespread in many educational systems, leading to poor student performance and inaccurate evaluation of their knowledge and skills. In order to support the quality development of better English test item, thus it is essential to investigate how teachers collaborate to develop English summative test items in MAN Model Banda Aceh.

B.   Research Question

Based on the explanation above, it can be stated that the research questions of this study are:

1.   How do English teachers collaborate to develop test items?

2.   How are teachers' awarenesses in constructing summative test?

3.   What are English teachers' challenges while collaborating to develop test items?

C.   Research Aim

Based on the research questions above, it can be concluded that this study is aimed to:

1. Find out the activities that teachers go through in collaborating to develop test items

2. Enclose teachers' awareness in constructing summative test

3. Discover the English teachers' challenges while collaborating to develop test items

D.   Significance of Study

This study is expected to have advantages for English teachers, school principal and further researchers. Hopefully, the study will give useful information regarding teachers' collaboration in developing test items which can help teachers in constructing better test items.

1.   *English Teacher*

This study is hoped to give useful information for English teachers who responsible to evaluate students' ability on how to collaborate in conducting better English test items. Thus, it can help teachers to create test items that are more structured and in accordance with the characteristics of a good test.

2.   *School Principal*

This study is expected will be useful for school principals who has the responsibility to ensure how well the English tests that are established by the English teacher.

3.   Further Researchers

This research is expected to be used by other researchers as a reference that is relevant to further research and have a significant impact on the evaluation process, especially in teacher's collaboration to develop test items.

E.   Research Terminologies

Going deeper into this study, in this part, there are some explanations from the title mentioned in the previous item which consist of teachers' collaboration and test items.

1.   Teachers' Collaboration

According to Kasmawati (2020, p. 136) teacher collaboration is "a forum for teachers to improve their competence in the form of knowledge, skills and experience". When effective collaboration occurs, teachers' knowledge and experience will be shared so that the teaching program will be improved. The experiences gained from collaborative activities will change teachers' knowledge, attitudes and confidence. In addition, collaboration that is done professionally is the key for teacher success which is beneficial in engaging and maintaining teacher professionalism, resulting to better teaching and effective student learning.

Based on statement above, it can be defined that teachers' collaboration is the efforts made by teachers in order to share knowledge, experiences and opinions to obtain better teaching and learning outcomes, especially in this research is in the field of evaluation including test item development. Indeed, this study will discuss how English teachers collaborate especially in MGMP

(Musyawarah Guru Mata Pelajaran) program to develop test items and difficulties experienced by English teachers while collaborating.

2. *Test Items*

Test items refer to questions or items in a test that are used to measure students' knowledge or abilities. Kartowagiran (2012) stated that writing good quality test items is important in developing a reliable and valid test. The quality of each test item determines the quality of the test as a whole. Test items should be written clearly and concisely, focusing on the specific objective or knowledge that should be measured.

According to Brown and Hudson (2002, p. 113), "Item analysis is usually done for the purpose of selecting which items will remain on future revised and improved version of the test." it can help teachers to revise or eliminate irrelevant items and improve the effectiveness of teachers' teaching strategies. The item is a reflection of the indicator. Therefore, each statement or item needs to be made well so that it is clear what is being asked and the answer is also clear. In this study, researcher focused on test items made by English teachers collaboratively.
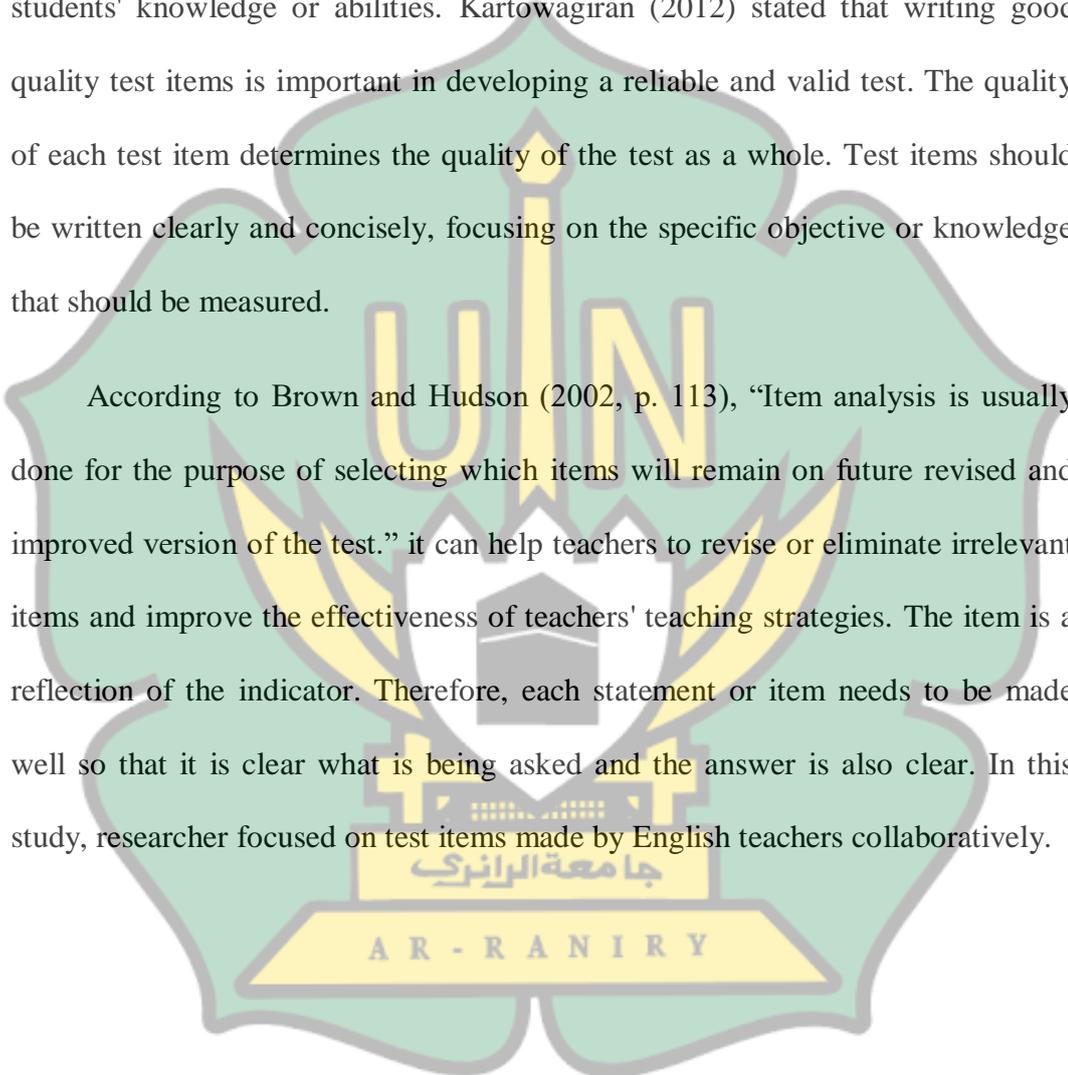
CHAPTER II
LITERATURE REVIEW

This chapter aims to explain the literature review related to topics of this study and also provide some theories to give relevant information to support this study.

A. An Overview of Test

The following below are some of overview that related to test.

1. The Definition of Test

Test is a set of questions that require answers, or a set of questions that must be answered in order to measure the ability of a person who is taken a test. According to Arikunto (2013) test is a tool or procedure used to know or measure something by rules that have been determined. While according to Arifin (2016) the test is a technique used in order to carry out measurement activities, in which there are various questions, or a series of tasks that must be done or answered by students to measure students' ability. Test is a collection of questions given to students in several forms such as oral test, written test and action test. In fact, the test is used to assess the learning progress of students related to cognitive abilities.

It can be concluded from several expert's opinions that test is a tool used to collect information for understanding, assessing and measuring student learning outcomes. It can be seen through the development of students in the understanding and mastery of materials given in the learning program.

Tests are part of the assessment, therefore, in making tests must also pay attention to the principles of assessment. Brookhart (2010) concluded that some of the principles in preparing assessments that must be considered are:

a. Determine clearly and precisely what is to be assessed.

b. design tasks or instruments that require students to demonstrate the required knowledge or skills

c. Decide what will be taken as evidence that students have used the knowledge or skills to be measured.

2. **The Function of Test**

There are two kinds of test functions according to Sudjono (2012) as follows:

a. As a student measuring tool, the tests can measure progress and improvement obtained by students after the learning process.

b. As a measuring tool for the success of teaching programs, because through tests, it will be known how far the teaching program has been achieved.

According to Purwanto (2020) in his book entitled *Prinsip-prinsip dan teknik evaluasi pengajaran*, the function of tests in learning grouped into four, those are as follows:

a. It is used to see the progress of students after learning activities

b. To observe the level of success of the teaching program, which includes the objectives, teaching materials, methods and learning evaluation tools.

c. It is used for guidance and counselling purposes

d. To improve the development of the school curriculum.

3.   **Type of Test**

Observing from its function, the test is divided into three according Arikunto's opinion (2016); they are as follows:

a.   **Diagnostic Test**

According to Widoyoko (2014) diagnostic tests are used to find the cause of learning difficulties and solutions of it. Based on the definition above, it can be concluded that diagnostic is used to find student problems in learning activities, so that solutions can be found to fix or deal with it. A diagnostic test is a test that is carried out to determine precisely the types of difficulties faced by students in a particular subject. After knowing the types of difficulties faced by students, further efforts can be made in the form of appropriate treatment.

The material asked in diagnostic test is generally emphasized on certain materials which usually according to students' experiences which are difficult to understand (Hidayati, 2013). This type of test can be carried out orally, in writing, in action or a combination of all of it. Diagnostic test can be used to identify the main problems that causing students to be unable to achieve learning outcomes. If the results of the examination show that the level of mastery of the student being

examined is low, special guidance must be given so that they can improve their level of mastery certain subjects.

b.   Formative Test

Formative test is used to monitor student progress in the learning process. Formative test is carried out at the end of each learning program which is usually in the form of daily tests (Arikunto, 2016). In addition, formative tests serve as a benchmark for students' mastery of the material on the program followed. In schools formative test is commonly known as daily test. The material of this formative test is generally emphasized on the subject matter being taught. The items in the questions consist of items, both those belonging to the easy category and those belonging to the difficult category.

According to Ismail (2022), the follow-up that needs to be done after knowing the results of the formative test are:

a. If the material being tested has been well mastered, then learning will continue with a new subject.

b. If there are parts that have not been mastered, then before proceeding with a new subject, it is first repeated or explained again the parts that have not been mastered by students.

From the description above, it becomes clear that the purpose of the formative test is to improve the level of mastery of students and at the same time to improve the learning process.

c.   Summative Test

The summative test is a learning achievement test that is carried out after a set of teaching program units has been given. Harlen (2005) stated that summative test is the process of teachers systematically collecting data on students' abilities to make conclusions about the extent of students' learning understanding. At school, this test is known as the final learning evaluation. Summative test is generally prepared on the basis of subject matter that has been given for one quarter or one semester. The summative test material is far more than the formative test material. The summative test is carried out in writing, so that all students get the same questions. The items raised in this summative test are generally also more difficult than the items in the formative test.

According to Brown (2004), summative test is designed to measure and summarize the learning objectives that students have achieved from the learning process. The main purpose of the summative test is to determine the value that symbolizes the success of students after they have taken the learning process for a certain period of time, so that it can be determined as follows:

a. The position of each student among of the group

b. Whether or not students can take part in the next (higher) teaching program

c. The progress of students, to be informed to parents, counselling guidance officer and other educational institutions which is stated in the form of a report card or a letter of completion.

According to Arikunto (2016), the test is divided into two forms that are as followed:

a.   Subjective Test

Teachers usually use subjective tests to test students' ability to think critically and creatively especially on the ability of analysis, problem solving, evaluation, etc., teachers can evaluate students' level of proficiency in writing, save more time in determining grades, this test also assures teachers that students not only answer the questions asked in the test by assuming, it can be a measure of student understanding (Basuki & Hariyanto, 2016).

Subjective test is generally in the form of essays or descriptions. This test is a type of a test that requires answers from students in the form of discussion and explanation in detail. In general, questions from subjective tests in the form of words such as describe, explain, why and how.

b.   Objective Test

Objective test is a test which in its examination can be carried out directly objective. Suwarto (2013) argued that subjective test has several benefits. First, this test is able to be an assessment of student understanding at a high level. second, students have a great opportunity to present answers or ideas that are considered appropriate. Third, it can support students' point of view in solving the problem in the question. Lastly, students are able to adjust complex questions of the test to the level of ability, experience, and age of the students themselves.

This objective test is used to overcome the weakness of essay test or description that can be subjective. Examination of objective test can be done with the help of a robot or computer. From the opinion of Wahyuni and Ibrahim (2012) added by Basuki and Hariyanto (2016) there are 4 types of objective test that are multiple choice test, short answer test, matching test and true-false test, the detailed explanation are as followed:

1) Multiple Choice Test

Multiple choice test is a form of test that has several choices then students are asked to choose one alternative answer (Wahyuni & Ibrahim, 2012). Multiple choice tests consist of incomplete explanation or notification of an understanding. To complete it, the test taker must choose one of the several possible answers that have been provided. This test consists of a description (stem) and a section on possible answers or alternatives (options). Possible answers consist of one correct answer, called the answer key and several distractors.

Regarding the number of alternative answers, there are actually no standard rules. The teacher can make 3, 4, or 5 alternative answers. The more the better. This is intended to reduce the guessing factor. The abilities that can be measured in the form of multiple choices questions include: knowing terms, facts, principles, methods, and procedures; identify uses of facts and principles; analysis of causal relationships and methods of assessment methods (Arifin, 2016).

The following below are some practical instructions in compiling multiple choice questions according to Arifin (2016):

a. Must refer to basic competencies and question indicators.

b. Give instructions on how to do it clearly.

c. Do not include material that is not relevant to what students have learned.

d. The statement on the problem should formulate a clear and meaningful problem.

e. Statements and choices are unbroken sentence units.

f. Alternative answers must be functional, homogeneous and logical.

g. Selection length on a featured material is shorter than the item.

h. Try not to associate reserves and options easily.

i. Alternative correct answers should not be systematic.

j. It must be believed correctly that there is only one correct answer.

The results of multiple-choice test are the students' achievement in the form of numbers or scores after being given the multiple-choice test. The scores that are obtained by students is a reference to see the understanding and mastery of learning materials.

2)   **Short Answer Test**

Weimer (2015) stated that the advantage of short answers is that the questions are easy to develop in assessing the answers. This can encourage superficial memorization of content. Short answer test requires students to provide answers thus avoid signalling. Open-ended questions that give students the

opportunity to construct answers are known as short answer questions. They are widely used in examinations to measure basic knowledge and understanding of a topic before more in-depth questions are given. Short answer tests closely resemble essay tests. They are both designed with open-ended questions. However, essay tests usually contain an extended text.

Chan (2009) explains that short answer tests have advantages and disadvantages according to their purpose as long as the questions are worded in such a way that the examiner can accept all alternative answers. this type of test is also fairly quick to score and can be scored by multiple examiners. Compared to many other evaluation approaches, it is also relatively easy to prepare. Since the structure of short answer questions is somewhat similar to exam questions, therefore it can be used in formative and summative assessments.

3) **Matching Test**

According to Sukardi (2010) matching test item is included in the objective test group. Physically, the form of matching test items consists of two parallel columns. In the first column contains a statement called a stimulus list which means a list of premises because the column contains a definition, phrase or single word and the second column contains a word phrase which is also called a list of responses or answers. Each question has an answer listed in the answer series.

Cross (as cited by Sukardi, 2010, p. 123) stated that "matching test items are appropiate for identfying the relationship things". The student's task is to find and place answers, so that they match the question. The number of answer choices

made is more than the number of questions. The difference with the multiple-choice form is the multiple choice consists of stems and options, then students only have to choose the one option that is considered the most appropriate, while the matching form consists of a collection of questions and a collection of answers, both of which are collected in two different columns.

To arrange questions of this form, Arifin (2016) provides several criteria, that are:

a. Make test instructions clear, concise, and easy to understand.

b. Customize with basic competencies and indicators.

c. The set of questions is placed on the left, while the answers are on the right.

d. The number of alternative answers should be more than the number of questions.

e. Arrange the items and alternative answers with a certain systematic. For example, before the subject matter, preceded by a stem, or it could be directly on the subject matter.

f. The entire group of questions and answers is only on one page.

g. Use sentences that are short and direct to the point.

This form of question is very good for measuring students' ability to identify information based on simple relationships to connect between two things. The

more connections between the premise and the response are made, the better the questions are presented.

4) True-False Test

True false test is test that consist of items in the form of statements followed by alternative answers or statements, this statement is either true or false (Magdalena, Mahromiyati, & Nurkamilah, 2021). The person being asked is tasked with marking each of these statements by circling (C) for the correct statement and (F) for the wrong statement. This learning approach is very useful in the learning process to bring revive the difficult and uninteresting topics. One function of the form of true-false questions is to measure students' ability to distinguish between facts and opinions. In order for the questions to function properly, the material asked should be homogeneous in terms of content. This form of questions is more widely used to measure the ability to identify information based on simple relationships.

Some practical instructions in compiling true-false questions according to Arifin (2016) are as follows:

a. In compiling the items in this true-false form, the number of items should be quite a lot, over 50 questions, so that they can be accounted for.

b. The number of correct and incorrect items should be the same.

c. Give instructions on how to do the questions that are clear and use simple sentences.

d. Avoid statements that are too general, complex, and negative.

e. Avoid using words that can give a hint about the desired answer. For example: usually, generally, always.

True-false test is quite popular as they are generally easy to write. However, this type of test should contain only one statement and be a clear true or false statement. It can also be said that these questions are not suitable for summative assessment as students will score 50% marks just by answering randomly.

4. **Criteria of Good Test**

According to Osterlind (2002) there are several criteria for a good test. First, there must be a high degree of coherence between specific items and the main objectives of the test as a whole. Second, the objectives must be clearly defined. In compiling test items, the examiner must understand what the purpose of the test is. Third, the contribution of each test item to measurement error in the test score must be minimized. fourth, the test format must be in accordance with the purpose of the test. fifth, the test is well written, the test must follow standards such as grammar, diction, spelling punctuation, usage, and syntax. Sixth, questions are legal and appropriate. This does not mean that all items must be created originally.

Miller (1991) and Gronlund & Lin (1990) stated that all good tests have to be valid, reliable and practical. Validity means that any test that have been conducted must be fit to the objectives of learning program. Reliability is in terms of the evidence it provides. While practicality means that the applicable to specific situations. All three are essential condition or something that is absolutely

necessary. Validity is considered the most important of the three elements, but reliability generally affects validity and validity cannot be fully appreciated without a basic understanding of reliability.

B.   Developing Test Items

Understanding how to develop good test items is very important for teachers because teachers will compile many test items to assess students. Here are some discussions of strategies and challenges in developing test items

1.   Strategies in Developing Test Items

Strategy refers to a planned methodological approach to a problem aimed to control and obtain information. The skill of developing test items is one that teachers should possess. Understanding the strategy of developing test items makes it easier for teachers to assess students. In developing test items, teachers can employ various steps and preparations that may differ from one another. Each teacher may have their own strategy. As stated by Brown (2000), there are several steps in developing test items, such as assessing the test situation, deciding what will be tested, balancing elements, weighting scores, and creating a functional test, the explanation are as followed:

a.   Assessing the test situation

In developing test items, teachers not only understand the test items as well as the situation in which the test takes place. McKeachie (2010) argued that it generally seems to be true that the easiest exams to create are the most difficult to grade, whereas time spent creating clear exams will save time in grading them. Understanding the situation means that teachers must grasp several aspects, for

example the situation of students' examination, how much time will be given to students when taking the test, when and where students will take the test, and how much teachers will need to grade students' test results.

b.   Deciding what to test

In deciding what to test, the teacher should have indicators of what will be included in the test. The test should be a representative sample of the overall indicators (Burker, 2009). To create indicators, teachers need an understanding of the syllabus items. By understanding the syllabus well, teachers can determine what kind of material or topics are appropriate for students. Not only determining the material but teachers also determine what skills will be tested on students. This means, the teacher decides consciously to include the skills or not such as writing or reading comprehension. By including a representative sample of the entire indicators, students' success or failure with those test items will be a good reference of how well students have understood all the material they have studied.

c.   Balancing elements

This section become a crucial aspect for teachers when developing test items. They need to consider into various factors, such as incorporating direct or indirect test items (Brown & Abeywickrama). In formulating test items, the elements must be well-balanced. Elements refer to the quantity of multiple-choice and essay test items. Moreover, the skills also should be balanced. Teachers should consider the quantity of each skill. Balancing these elements helps estimate the time that should be allocated for each section of the test.

d.   Assigning weight to scores

The weights assigned to the test items used to derive the total score have important reliability and validity implications (Kreiter et al, 2004). A teacher's perception of a student's success or failure depends on how many points will be given for each test section. For instance, the scoring for multiple-choice questions will differ from that of short essays. When assigning grades, teachers must also consider the difficulty level of each question.

e.   Conducting test trials

Teachers should test several items or the entire exam on other students who have a similar proficiency level as the prospective students before administering the actual test (Saragih, 2016). Other students may identify issues that teachers are not aware of. Additionally, they may provide answers and alternatives that were not anticipated by teachers. Afterward, teachers can revise based on the feedback from other students.

In developing test items, these strategies must be executed because they are interrelated. First, assessing the test situation means prospective teachers consider factors such as the allotted time, when students will take the test, and so on. Second, teachers must also decide what will be included in the test. Before designing the test, teachers must have a clear understanding of the syllabus. Additionally, teachers must have a list of the types of material and skills to be included in the test. Then, teachers need to balance the elements that refer to the quantity of multiple-choice and essay questions to be included in the test. After

designing the test, prospective teachers also take into account the scoring for each question. However, before administering the test to students, prospective teachers should try out several questions on other students with a similar proficiency level. Among these steps, assessing the test situation is crucial, as it marks the initial phase before test design.

2.   **Challenges in Developing Test Items**

Challenges is a difficult task or something that is hard to complete. Developing test items is a complicated responsibility (Osterlind, 2002). This happens because there are several considerations before developing test items such as understanding the level of students. This must be practiced many times in order to master items construction. Teachers may face several challenges in developing test items. In addition, it is challenging because there are several steps that must be taken when constructing test items.

It is sometimes difficult to determine that the items are good or bad. This also makes it difficult to set criteria for good test items for several reasons. One of the reasons that makes test item creation very difficult is technical skill. Each test item has different criteria for a good test. According to Brown (2004), there are several steps to develop good test items. First, obtaining test that clearly reflect the learning objectives or skills to be measured. Clear indicators help students and teachers understand what is expected to be achieved. Second, developing test specifications that describe in detail what content and skills will be measured in the test. This includes relevant learning materials and the desired level of difficulty. Third, drafting the test involves composing questions or tasks based on

the test specification. The questions should be designed in such a way that teacher can accurately measure students' understanding. Fourth, revising the test to review the test items to ensure clarity, relevance, and conformity of the test specification. Fifth, editing and typing a clear formatted test. Grammar and spelling errors should be corrected, and the test should be designed in such a way that students can answer test items clearly. Sixth, obtaining feedback from other teacher to ensure that the test has good quality. Last, utilizing feedback after giving the test to understand the extent to which the learning objectives have been achieved. This feedback can also help teachers adjust their teaching methods in the future.

As mentioned above, teachers who are responsible for developing tests require effort and time to develop effective items that are suitable for the type of test being created, which are characterized by the main qualities of a good test and effectively realize the intended purposes of the test. Doing this job requires the test maker to be well aware of the step-by-step instructions to go deeper, further, and faster to achieve success in preparing effective test easily and conveniently.

3. **Stages of developing test items**

Preparation of tests offer a huge influence on students who will take the test. To reduce errors in measurement, tests must be planned carefully and systematically. Collaborating teachers can follow some stages in constructing good test. Sudaryono (2012) stated that for the test construction steps taken are as follows:

> 1. Set test objectives. Learning achievement can be made for

various purposes, such as tests, which aims to diagnose student learning difficulties known as a diagnostic test.

2. Curriculum analysis. Curriculum analysis aims to determine the weight of each subject that will be used as the basis for determining the number of items for each subject matter

3. Analysis of textbooks and other learning material sources.

4. Make a grid. The benefit of the grid is to guarantee a good sample of questions, in a sense of covering all subject matter proportionally. So that the test items include the whole material (topic or sub-topic), then before developing test items, it is a must to make a grid as a guide.

5. Writing test items. After the grid in the form of a specification table is available, then the next step is making test items.

6. Review the test items (validation). The questions that are made are still possible to occur deficiencies or errors which concerns the aspect of the specific ability being measured, the language used, typographical errors etc. For this reason, before being reproduced, the items must first be studied by colleagues who understand the test and the technique of writing test items to examine the surface validity of the test that have been made.

7. Limited test reproduction. The finished test is reproduced in sufficient quantity according to quantity trial sample or the number of participants who will do the test.

8. Analysis test results. Based on the test results data was analysed, especially the analysis of the items which includes item validity, level of difficulty, and deceptive functions. Based on the validity of the items The selection of questions was held using certain validity criteria.

9. Revision of questions. Valid questions based on empirical validity criteria were confirmed with a grid. If the questions have met the requirements and have represented all the material that will be tested, the questions are then assembled into a test, but if the questions are valid does not meet the requirements based on the results of confirmation with the grating, repairs can be made on the required questions.

10. Assemble questions into tests. The order of questions in a test is carried out according to the level of difficulty of the questions, from easy questions to difficult questions.

By implementing the stages that have been described, collaborating teachers can divide tasks to facilitate the process of preparing test items. These stages are expected to improve the quality of the test and ensure that the test is effective. Therefore, it can measure students' understanding in accordance with the predetermined learning objectives.

C.   Teachers' Collaboration

To ensure the relevant literature about this study, these are some discussions about teachers' collaboration.

1.   **The definition of teacher's collaboration.**

According to Gunawan (2012), collaboration means together or in a group, collaborative method is learning together or cross-training. Teachers' collaboration refers to the process of working together to achieve common educational goals. It is a coordinated effort among teachers, administrators, and other educational professionals to improve instruction, learning outcomes, and student achievement. Collaboration involves sharing knowledge, ideas, experiences, and resources to enhance the quality of teaching and learning. It can happen both within and between schools and can occur through various means, such as team meetings, professional development workshops, and shared planning time. Teachers' collaboration is essential for creating effective and efficient learning environments that benefit both teachers and students

Saud (2008) defined subject teacher collaboration is a forum for professional activities of similar subjects' teachers consisting of collaboration and subject teachers. Subject teachers are teachers who are responsible in managing subjects in accordance with the curriculum. Teachers collaborate in a multitude of ways when they interact with their colleagues to exchange ideas and resources, discuss student learning, team up for joint activities and knowledge creation. It is in these ways that teachers can co-create and enhance their learning with a shared aim to provide quality learning experiences to their students. In addition to supporting the instructional role of teachers, collaboration plays a key role in building relationships among teachers so that they feel part of a professional community and derive personal fulfilment from their work.

2.   **Stages of developing test items collaboratively**

There are some stages of teacher collaboration identified by experts:

a.   **Introduction and Relationship Building**

Bryk & Schneider (2002) stated that the importance of building strong interpersonal relationships between teachers as a foundation for effective collaboration. This stage includes getting to know the teachers involved in the collaboration. Teachers begin to build rapport, understand each other's strengths and weaknesses, and build trust with each other.

b.   **Articulation of a Shared Vision**

This stage involves discussing and defining a shared vision of what is to be achieved through collaboration. Timperley, Wilson, Barrar, & Fung (2007) discussed that the importance of developing a clear shared vision among teachers as an early stage in successful collaboration. Teachers try to find common or at least related goals, such as improving student learning outcomes, developing better teaching strategies or designing a more effective curriculum.

c.   **Planning and Formulating an Action Plan**

According to Louis, Kruse, & Marks (1996), the importance of strategic planning and the creation of concrete action plans to achieve the goals of collaboration. This stage ensures strategic planning to achieve the shared vision that has been set. Teachers formulate an action plan that includes concrete steps,

their respective responsibilities, the timing of implementation, as well as the evaluation to be conducted.

d.   Implementation

The importance of implementing the agreed action plan to achieve the set collaboration goals. This stage is the implementation of the approved action plan. Teachers work together to implement the strategies or implement the planned changes in their teaching. This involves active cooperation, monitoring of progress and adjustments if needed.

e.   Evaluation and Reflection

This stage entails evaluating the outcome of the collaboration. McLaughlin & Talbert (2006) stated that the importance of evaluation of collaboration outcomes and collective reflection to improve practice and learning outcomes. Teachers evaluate whether the set objectives have been achieved, identify what has worked and what needs to be improved. Collective reflection also occurred to understand the lessons learned from the collaborative process
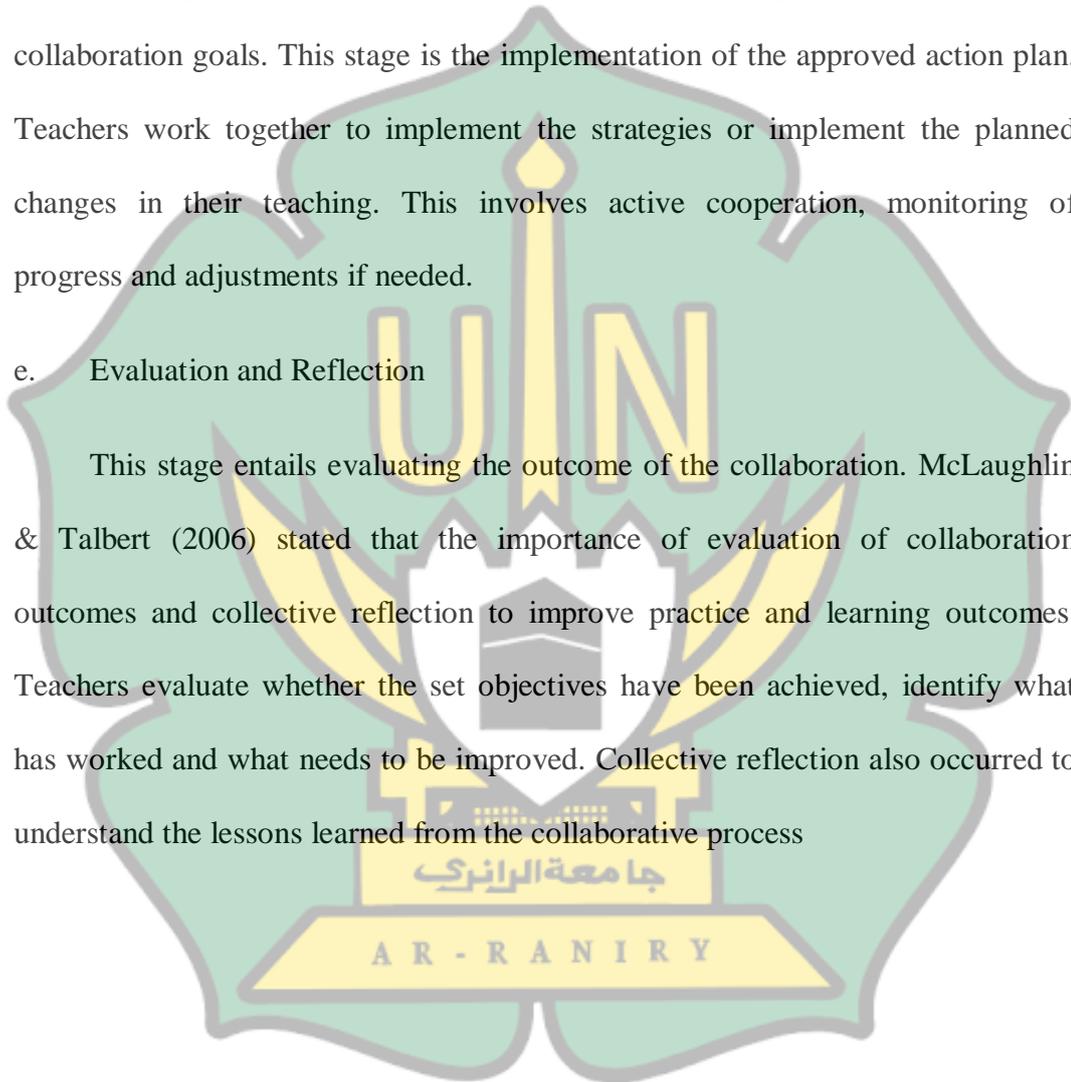
CHAPTER III
RESEARCH METHODOLOGY

This chapter discusses the methodological research procedures, including a brief description of the research design, research participants, techniques of data collection, and techniques of data analysis.

A. Research Design

This study used a qualitative descriptive method and was categorized as a case study. Researcher try to describe, explained and interpret events in accordance with the facts obtained in the field. According to Schwandt (as cited in Goodyear, 2014), case study research is a strategy for conducting social inquiry, although what constitutes of this strategy is still under discussion. The data of this research is can be a sentences, statements, documents, and other qualitative data to be analysed qualitatively.

According to Nasution (2003), qualitative research is a research procedure that produces descriptive data in the form of written or oral words, from the people and behaviours observed. In this research, it is attempted to collect as much as possible the descriptive data which will be presented in the form of reports and descriptions. Based on this understanding, it is clear that the descriptive qualitative research is a method used to collect data and information about an existing phenomenon with the formulation of written and oral words from people that will be observed, the results of this study refer to the quality of the informants' answers.

According to Creswell (2012), qualitative methods aim to investigate the detailed understanding of a central phenomenon and qualitative research subjects are individuals with small numbers. As a result, the focus of this method is to obtain deep insights through the point of views of research participants. The aim of this method for this study is to reveal the activities and challenges contained in English teachers' collaboration program in MAN Model Banda Aceh which relate to the development of test items. In qualitative research, it is more exempt, the research process provides the informants with the opportunity to answer the questions asked according to the informants' understanding.

B. Research Participants

The participant is someone or several people who are considered to have the most in-depth understanding of the object being studied (Yin, 2018). The participants of this research are teachers who are used to provide information about the situation and conditions of the research background. The participants of this study were English teachers who participate in collaboration activity in developing test item in MAN Model Banda Aceh. The researcher chose MAN Model as a research site because not all schools conduct collaborative activity in developing test items. MAN Model Banda Aceh is one of the schools that facilitated this activity.

In choosing participants of this research, purposive sampling was used by researcher. According to Sugiyono (2015), purposive sampling is a sampling technique with certain considerations and criteria. The researcher had certain criteria of the participants. First, the participants must be the teachers who teach

English and involve in English test items development at the same grade. Second, the teacher must be involved in English teachers' collaboration program. The participants who meet the characteristics can be selected to participate in this research. There are 8 English teachers in MAN Model Banda Aceh, the researcher chose 3 English teacher who collaborated in developing English summative test for eleventh grade because those 3 English teachers have fulfilled the criteria that have been determined by the researcher.

C.   Techniques of Data Collection

In this research, the main tool used to collect the data is interview. According to Uwe Flick (2009), there are three types of interviews, those are unstructured interview, a semi-structured interview, and structured interview. The interview used in this research was semi-structured interview. It means that he questions asked in semi-structured interviews have been prepared in advance, but can be changed according to what the respondent wants, based on the situation experienced by each teacher. Thus, semi-structured interviews allow researchers to collect more accurate data and also facilitate data comparison.

Interview is an activity carried out to get information directly by giving verbal questions to participants (Singarimbun & Effendi, 2009). Conversations with specific intentions, carried out by two parties between the interviewer who ask the questions and the informant as information provider. Researchers use the interview method to obtain relevant information related to English teacher activities and difficulties in collaborating to develop test items. The instrument used was interview guide which had been prepared by the

researcher as a guide for conducting interviews with English teachers who collaborated in developing test items. During the interview, a recorder was used in order to get clear data.

D.   Techniques of Data Analysis

Data analysis is the process of systematically searching and compiling data obtained from interviews, field notes and other materials, so that it can be easily understood, and the findings can be informed to others. Data analysis is done by organizing data, breaking it down into units, synthesizing, arranging into patterns, choosing which ones to study, and making conclusions (Mulyana, 2013). At this stage the data worked on and utilized in such a way to conclude truths that can be used to answer the questions or issues raised in the research.

Creswell (2012) stated that there are six steps for analyzing data which are collecting the data, organizing and preparing the data, reading all the data, coding the data, coding to build description or themes, and interpreting the findings. In analyzing the data, the researcher examines all data consisting of interviews that has been conducted. The result of the interview using coding procedures were analyzed by researchers. Coding is classifying or grouping answers from respondents. It aims to help the researcher in grouping the data from informants. The interview transcript coded by researcher will identify which information will be used or need to be reduced. By doing this, it selected specific information that focus on the study. The conclusions should be consisted of activities and difficulties of teacher's collaboration in developing test items.

CHAPTER IV

FINDING AND DISCUSSION

This chapter focuses on the data analysis to answer the research questions. It aims to elaborate findings and discussion of this research regarding English teachers' collaboration in developing test items.

a.   Research Findings

The findings of this study were in accordance with the data that has been collected. To obtain the data about teachers' collaboration in developing test items, this research interviewed 3 English teachers who taught at the same grade and involved in collaboration activity. The researcher gathered the data from 3ᵗʰ January-8ᵗʰ January 2024. These findings aimed to answer the research questions in chapter one: 1) How do English teachers collaborate to develop test items. 2) What are English teachers' challenges while collaborating to develop test items. The findings were an elaboration of these two research questions. The researcher got several findings on two main points based on the research questions. The results were shown below:

1.   Teachers' Collaboration in Developing Test Items

The first analysis was focused on how teachers collaborate in developing summative test items. There were 2 aspects that can be concluded from the results of interviews that lead to how teachers developed test items collaboratively, those were the stages or activities of teachers' collaboration in developing test items and teachers' awareness in developing test items.

According to the teachers' experiences that were retrieved from the results of interviews, there were several stages carried out by collaborative teachers in developing test items. The stages were explained below.

a.   *Conducting Monthly MGMP Meeting*

Generally, the collaboration carried out was divided into 2 different settings, those were collaboration in the implementation of MGMP (*Musywarah Guru Mata Pelajaran)* meetings and regular discussions with other teachers outside MGMP meeting.

> *"In MAN Model we coincidentally have MGMP which is quite active for English subjects, so every month we have at least one meeting as a forum for discussion and sharing". [AC]*

> *"At the beginning of semester, we already held MGMP meeting so we can discuss various problem that related to various issues" [JR]*

MGMP meetings at least was held once a month. The main purpose of MGMP was to improve the quality of learning, share experiences, and support each other among subject teachers. The content of MGMP discussions and agendas can be different depending on the policies and needs of each semester including curriculum discussion, development and preparation of learning materials, adjustment of the curriculum to the needs of students, problems and challenges experienced by teachers, learning strategies, learning evaluation, development of summative tests, etc.

MGMP meetings at the beginning of the semester for subject teachers were a good step to harmonize the understanding of the basic competencies. The teacher mentioned that it can be done by reviewing the syllabus and curriculum to identify the basic competencies that will be taught during the semester. Discussion of these basic competencies aimed to ensure that all teachers have a compatible understanding. This enabled them to determine which basic competencies that were appropriate to their needs, as well as creating an environment that supports balanced professional development among all teaching staff and certainly affects the teaching process, which is an essential requirement for a good evaluation process and good test items.

> *"The initial activities that we did as English teachers were mainly in the MGMP group, we will determine the material that will come out for the exam". [NB]*

> *"During MGMP meetings, we collaborate to discuss about material related to the questions that we will test for each semester." [JR]*

> *"For example, we conduct a meeting at the beginning of the semester to discuss together what material we will teach and the basic competence throughout the semester". [AC]*

All interviewed teachers concurred that discussing the material that will be taught throughout the semester was important which will affect the process of test items development.

In MGMP meeting, teachers also mentioned that they determined the number of test items and types of test items. It was important to create a balance

between the number of questions and the time available, as well as ensuring that the exam can accurately measure students' understanding according to the set objectives. This decision affected a number of important aspects of the measurement and evaluation process and can have a significant impact on the exam results and student experience. Determining the appropriate number of questions ensures that each section of material being tested is proportionally represented. The provisions for the number of questions and types of questions have been determined by the curriculum manager, for its application was handed over to the teacher to be adjusted to their respective needs.

> *"The stages were determining the material, overviewing the test items, and determining the number of questions and the type of question, for example, how many hots questions and how many lots questions but mostly we use hots questions.". [NB]*

> *"The number of choice and essay items had been determined by the curriculum manager, but which questions we want to make, whether hots or lots, depends on the test maker". [AC]*

After determining the number and type of tests, MGMP meting also discussed and determined the person who is fully responsible to construct test items. This is usually given to the teacher who taught the most in a particular class of a grade. There were two types of English subjects, called mandatory English and English specialization, it would be selected respectively to construct tests items of both subjects at each grade.

> *"When constructing the test, it was indeed handed over to one person of each grade, maybe not for every grade was only one person, because we also have mandatory ones as well as specializations". [AC]*

"We also determine the teachers who are taught in each grade and then one teacher will be chosen as the person who is responsible for constructing the test items". [JR]

From the results of the interview above, it can be concluded that in the MGMP meeting there are 3 discussion points related to the test items development. First, discussions related to curriculum development which includes determining the material and basic competencies in the semester. Second, discussions regarding the number of test items and types of test items. Third, determining the teacher who is responsible for constructing test items.

b.   *Conducting A Deeper Discussion with The Same Grade Teacher.*

Before constructing test items, the teachers mentioned that the test maker with other teachers who taught in the same grade have to ensure the specified material at the beginning that has been discussed and confirm that each class gets the same material.

"Before making questions, we discuss whether the basic competencies in the semester have been discussed in the class or not". [AC]

"For summative test, there are 30 choices and 5 essays questions, so we divide equally how many questions for theme 1, how many for theme 2, how many for theme 3 and so on". [NB]

Teachers indicated that discussion before the item creation process would be very helpful for the test makers to be more focused and avoid misunderstandings when typing the item script. After that, the teacher continued the process of

constructing test that refer to the material and basic competencies that have been agreed.

Then Before sending the test to the curriculum manager, the test maker will recheck the file that assisted by the teacher who taught in the same grade to check the errors and the suitability of the indicators with the questions that have been made, if there was a mismatch then the teacher who checked the test will suggest changes, if it is appropriate then no further changes were needed.

> *"First, we checked with friends, after we checked together then we should correct any mistakes, sometimes there are grammar errors or sometimes there are typos or all kinds of errors". [NB]*

> *"Sometimes according to one of us maybe the test items were still mots or medium, but according to the test maker maybe it was already hots so we discussed it again and then we told them that the right indicators". [JR]*

> *"After constructing the test items, I will send them to the group so we can discuss about the indicators. This means that the determination of the indicators was also influenced by the discussion, but if everyone feels that it was suitable then it didn't matter anymore". [AC]*

From the results above, deeper discussion with the same grade teacher was carried out before making test items which included details of the material to be tested. Then the discussion was also carried out after making test items to check and correct indicator mismatches or errors.

c.   *Rechecking The Test Items*

Teachers also stated that before being given to the curriculum manager, the test items were also distributed to the *WhatsApp* (WA) group to be rechecked by

the MGMP leader or other teachers. The checking is usually done to adjust the typing and language aspects. If it has been approved by the MGMP leader, then the questions can be submitted to the curriculum manager.

> *"After we checked the test items with a friend and consider the test was correct, we usually send the test to the MGMP WhatsApp group, there was MGMP leader, so usually she is the one who rechecks the test before it was submitted to the curriculum manager". [NB]*

> *"Before we submit the test to curriculum manager, we first send it to the MGMP group to be seen and corrected together" [AC]*

> *"After we crosscheck the test independently then we send it to the MGMP group so that all teachers can see the test items that we have made so they can also help if there are errors in writing, they can also help correct it then we will give it to the curriculum manager". [JR]*

All interviewed teachers agreed that rechecking confirmed the questions to be tested have undergone a review and verification process. It helps to ensure that the questions were in line with the curriculum and learning objectives and that mistakes were avoided.

d. *Pretesting The Students*

The teachers mentioned that the last thing must to be ensured was to provide a trial simulation to students using questions that were similar to the summative test items. Pretesting provides an opportunity for students to evaluate the extent to which they understand the material to be tested. This helps them identify areas where they may need to study further or revisit certain material before the exam.

"At the end of each chapter, we would have a daily test, it was still related to the summative test, there were a few things that we changed so there were modifications". [AC]

"Some of these questions we have already given during PTS (Penilaian Tengah Semester) and students practice in class, some of these questions we retrieved with a few changes for final exam". [JR]

"Two weeks before the exam we usually do pretest too but using different questions with the same type of questions. For example, for narrative text material, one story about Cinderella and another about Malin Kundang". [NB]

Based on teachers' responses, for the implementation of these trials was depends on each teacher, some gave trials before the implementation of the exam and some argued that the test item trials have actually been carried out by giving daily tests to students and PTS. Daily tests were also used as one of the teacher's references in constructing summative test, if the daily test were stories or texts, usually the stories or texts was replaced with other stories or changed the question instructions, these changes were made by the teacher according to their needs.

2. Teachers' awareness in constructing English summative test

Teachers' awareness in developing English summative tests is very important to ensure the effectiveness and success of the process of constructing test items. The following is the explanation of some aspects of teachers' awareness in developing test items

a. Determining Type of Test

Based on the result of the interview, the researcher discovered that all teachers have chosen multiple choice and essay as type of summative test items,

with a total of 30 multiple choices and 5 essays. The teacher mentioned the reasons for choosing the type of questions varied because different question types assess various language skills. For example, multiple-choice test may asses vocabulary and grammar knowledge, while essay questions assess writing and critical thinking skills. By incorporating a variety of question types, teachers can obtain a more comprehensive understanding of students' language proficiency. Some teachers reasoned that they chose multiple choice because it helped to give students choices and chose essays to measure students' literacy skills. Certainly, teachers said the main reason for choosing the types of question was because they were in accordance with the regulation of the curriculum management in school.

> *"For the type of test, it has been determined, the test is in the form of a multiple choices and essays". [JR]*

> *"I chose multiple choice and essay questions for the exam because I consider that multiple choice may assess vocabulary and grammar knowledge, while essay questions assess writing, literacy skills and critical thinking skills". [NB]*

b.   *Utilizing The sources*

Based on the results of the teachers' responses, all teachers used the curriculum and indicators as a guide in developing test items. Each item must be based on the material and indicators that have been approved by the MGMP group. The teachers stated that there were various sources used to support the development of test items, such as textbooks, internet, and question banks. all sources used will be adjusted to the indicators.

> *"In developing test items, usually the references were in the question bank, sometimes we searched in the textbooks too, we rarely took it from textbook,*

only 2 or 3 questions were taken, the rest were searched on the internet". [NB]

"The sources were from the textbook and some from the internet, we develop it according to the level of difficulty. First, we review the text, then we would adjust it to the indicators, not everything is directly taken from the textbook or internet". [JR]

"The references are varied because we don't have to use textbooks. We can also look for questions from the internet so we can combine the texts, we definitely take authentic ones that are clear sources". [AC]

Teachers often utilize the content covered in a standard textbook to ensure that test items match what has been taught in the classroom. This helped maintain consistency between learning and assessment materials. Teachers also utilized the internet as a source for additional ideas in constructing items, such as educational websites, articles and language learning platforms, and relevant and current content for test items. This was very important to keep the assessment up-to-date and reflective of contemporary language used. In addition, teachers also used question banks or previous exams to get inspiration and ideas for developing their own test items. However, teachers certainly modified the questions to suit the context and learning objectives in the classroom.

c.   *Dividing the number of questions*

In this section, the teachers stated that in the process of dividing the number of questions, the teachers adjusted to the school rules regarding the division of test items based on the level of difficulty. Therefore, teachers mentioned that in the process of dividing the number of questions, they referred to the material and the

level of difficulty of the questions consisting of the easiest level was called lots, the medium level was called mots and the most difficult was called hots.

> *"For example, in the test, there must be included C1 C2 C3, there must be easy one, medium one and most difficult one. how many questions are hots, mots and lots including in multiple choices and essays have been determined by the curriculum manager, but which questions we want to create depends on the question maker". [AC]*

> *"For example, how many points of questions will be made with the provisions of how many items of C1 questions, how many C2 questions, how many C3 questions, means that it is even made easier at the beginning because it has been divided". [JR]*

d.   *Selecting indicators*

In the process of selecting indicators, the first thing teachers did was discussing the designed indicators with other teachers. This can be helpful because they got input from different perspectives and helped to ensure the sustainability of the indicators. Teachers in determining indicators also considered those related to students' daily lives so as to increase the relevance of the material. Teachers also stated that each item can be directly linked to one or more indicators. Teachers agreed that the suitability of the difficulty level in the indicators reflects the level of understanding and ability that was in accordance with the level of student learning. Therefore, all indicators that have been made in the lesson plan will have the possibility to change and experience adjustments, because according to the teacher, the most important thing was to ensure that the indicators must be in accordance with the lessons taught in class.

> *"First, we discuss with MGMP friends, after make the test, we will send it to the group, we will also discuss together about the conformity indicators.*

Therefore, the determination of indicators also had influence from the discussion"

"We determine the indicators from the beginning, they are developed from the syllabus into the lesson plan, but sometimes the indicators were not suitable for students so we make test items according to their abilities so there must be revisions".

e. **Choosing distractors**

choosing distractors was an important process in designing effective multiple-choice questions. Teachers argued that some considerations in selecting distractors were that distractors should reflect common misconceptions or errors related to the objectives, ensure that the correct answer was clear and unambiguous, ensure distractors do not sound too similar to each other, distractors should not be so obvious that students can guess the correct answer without understanding the material, and requires careful consideration of the learning objectives.

Teachers also stated that the level of difficulty in choosing distractors is also influenced by the level of difficulty of the test items, distractors for hots questions were more difficult to design.

"In choosing the tractor of hots questions is a bit difficult, but if for example we stick to the text it is not too difficult because the important thing is that students should not get the answer easily". [NB]

"In choosing distractors, we make them similar so that students can be tricked and the option was not too easy to be answered. Therefore, the distractors have to be clear and unambiguous, *so the students were not confused". [JR]*

"The distractor must be appropriate, not too easy and not too difficult".
[AC]

Based on the results above, it can be concluded that some of principles that must be aware by the teachers were determining type of test, utilizing the sources, dividing the number of questions, selecting indicators, and choosing distractors.

3.   *Teachers' challenges in Developing Test Items Collaboratively*

Developing test items collaboratively can be challenging for teachers, because it involved cooperation between teachers. Some of the challenges faced in this process included coordination, communication and alignment of ideas. Based on the interviews, the following are some of the challenges that teachers experienced when collaborating to develop test items.

a.   *Differences in Teaching Approaches*

Teachers have different teaching approaches. While collaboration has the potential to produce great benefits, differences in learning approaches can create complex dynamics.

"Differences in teaching strategies also influenced the process of developing test items, because each class is not the same, teaching strategy must also be adjusted to the students' abilities and needs". [AC]

"As teachers and MGMP members, each teacher has a different teaching approach, we also have to consider how the test questions fit with the skills and abilities we emphasize in our teaching. However, if there are differences in skills emphasis between teachers, then we need to find a balance in the test questions so that they cover all important aspects of learning". [NB]

Challenges of differences in teaching approaches arose when trying to harmonize these different strategies to create consistent problems that reflect the curriculum as a whole. For example, some teachers preferred practical and project approaches, while others leaned more towards theoretical or discussion approaches. Therefore, it leaded to differences in skills emphasis between teachers.

b.   *Challenges in Time Coordination*

Based on the results of interviews, teachers stated that MGMP meeting was not always attended by all English teachers.

> *"One of the main challenges we experienced was the difficulty in finding a suitable time for all team members. Considering the busy schedules of individual teachers, it often took several tries to find a suitable time for collaborative meetings." [JR]*

> *"It is usually difficult to determine the time because we work with other people, to determine the time, sometimes we can attend the meeting but other teachers can't or vice versa". [NB]*

Some teachers have their own busy schedules, sometimes they have to make several time adjustments so that MGMP meetings can be attended by all English teachers. However, there were still those who cannot attend the meeting, it didn't matter in a way that the teacher gives an understandable reason for not being able to attend the meeting.

c.   *Curriculum Cohesivity*

Teachers have different levels of flexibility in implementing the curriculum. Some teachers more inclined to customize the curriculum, while others conformed more to the established guidelines.

"Another challenge was determining a balanced emphasis on important aspects of the curriculum. Each teacher may have different priorities and connecting these perspectives can be done through discussion." [AC]

Each teacher had different interpretations of the content and focus of the curriculum. This may lead to differences in emphasis on certain topics. As a result, questions may not reflect the curriculum as a whole. Teachers believed that developing items that were cohesive and reflective of the whole curriculum was a challenge. Differences in teachers' understanding and interpretation of the curriculum can make it difficult to achieve this cohesiveness.

d.   *Differences In Understanding the Difficulty Level of The Test Items*

Each teacher has a different perspective on the difficulty level of a test item, which can lead to inconsistencies in scoring. Students who answer an item that one teacher considers difficult may get a different assessment if tested by another teacher. Differences in perceptions of the level of difficulty can lead to an imbalance in the material tested. Some teachers preferred to create easier items, while others may be more concerned about creating difficult items. This can affect the coverage of the material measured in the exam. However, it cannot be burdened on one teacher, because the teachers who teach in their classrooms were the ones who most understand the abilities of their students. Therefore, teachers must be able to understand and accept each other's opinions.

"Some teachers may prefer to make the questions more difficult or easier, and we need to adjust so that the test reflects the different levels of student understanding. [AC]

"It was a bit hard to connect the two opinions. For example, for me this question was not HOTS yet, but for other teacher it's already HOTS. In my class maybe it was just a normal question. It turns out that in other

teacher's class the test items already difficult so we usually meet and discuss to solve the problem". [NB]

In overcoming the challenges of developing test items collaboratively requires a coordinated approach and effective cooperation between the teachers involved. Some activities that teachers can do were increasing open and regular communication between teachers to discuss differences in understanding and reach agreement on the curriculum cohesivity, level of difficulty, material focus, and learning objectives. To reduce time and space barriers, teachers can use online collaboration tools, such as online MGMP discussions via *whatsApp* group.

b.   Discussion

This section provides the discussion based on the main findings. The purpose of this qualitative research was to discover the answers of the two research questions listed in the first chapter. The method used to answer both questions was interviews.

The first research question focused on investigating how English teachers collaborate to develop test items. Based on the interview results, it was clear that collaboration activities helped to facilitate teachers in the process of developing test items. Through the collaboration activity, it was possible to learn many things from colleagues. As conveyed by Marshall (2016), in improving and maintaining a strong focus on student academic achievement, teachers need to collaborate on developing curriculum, analysing assessments, and sharing best practices and strategies. Therefore, this collaboration can encourage teachers to develop test items.

Based on research findings, some collaborative stages in developing test items consist of conducting monthly MGMP meeting that discussing curriculum development included basic competencies and materials to be taught throughout the semester, determining the number and type of questions, and determining teacher who responsible for constructing the test items. All stages carried out by teachers in attempting to develop test items were part of curriculum development. Setiawan (2020) found that teachers consider the MGMP meeting important because it provides a forum to learn from each other, share strategies, and produce exam questions that are more in line with student needs. Then, teachers also conducting deeper discussion with the same grade teacher which included details of the material to be tested and checking mismatches or errors. Next, the test items ware also rechecked by MGMP leader. Finally, before summative test was given, teacher provide pretest questions for the students.

According to Hughes (1989), the stages of developing a test are divided into three main stages. First, the writer can start by writing a test specification in accordance with several aspects including basic competencies, syllabus, learning objectives, and materials. Second, the test maker can start developing the test by paying attention to the principles of a good test and the last step is to conduct a pretest. Based on the research findings mentioned above, the stages that emerged in this study were in accordance with the stages proposed by Hughes with the implementation of collaboration that also benefits the teachers in developing test items.

When the teachers develop test items collaboratively based on those stages, they should be aware of the principles for writing good tests. According to Osterlind (2002), to construct test items, teachers should be aware of criteria of a good test. However, this study discovered that some teachers were not conscious of the criteria mentioned. One of the instances was the teachers who have no reason why they choose multiple-choice and essays questions, just because they conform to the guidelines of the school. Teachers should avoid this because being aware of the reasons for choosing the type of questions to be tested can help teachers adjust to the learning objectives.

The results of this study showed that the challenges teachers face in developing test items collaboratively included several aspects that were particularly relevant in the context of teacher collaboration. Teachers encountered different challenges in developing items collaboratively, those were differences in teaching approaches, challenge in time coordination, challenge in established curriculum cohesivity and difference in understanding the difficulty level of the test items.

The first challenge was differences in teaching approaches. Hattie (2009) highlighted that a variety of learning approaches are used by teachers in the classroom. It was found that effective approaches can differ depending on the context and learning objectives. Some teachers may prefer practical and project approaches, while others prefer theoretical or discussion approaches. Challenges arise when teachers with different approaches have to work together in developing

test items, as teachers may tend to develop items that reflect their own approaches.

The second aspect was challenge in time coordination. The difference in coordination time was one of the challenges in teacher collaboration in developing test questions collaboratively. According to Bryman (2007), time constraints are a factor that affects cooperation and collaboration. Time limitations can hinder the coordination process between teachers to develop test items collaboratively. This challenge may be more pronounced if teachers have busy schedules or if they have to overcome strict time constraints in developing test questions.

The third aspect was challenge in established curriculum cohesivity. Onti (2018) argued that some teachers more committed to following established curriculum guidelines and standards, while others more inclined to customize the curriculum according to the learning context and students' needs. This challenge points need to find a balance between flexibility in curriculum implementation and adherence to curriculum standards in test item development.

The last challenge was difference in understanding the difficulty level of the test items. Zapata-Rivera et al. (2021) stated the concept of relative difficulty in test item evaluation. This concept recognizes that the difficulty level of test items was vary depending on student characteristics and the learning context.

Based on these findings, it can be concluded that the common challenges found in developing test items collaboratively were maintaining good communication among teachers. Teachers agreed that effective communication

can facilitate harmonious collaboration in planning and drafting test items, allowing for a better understanding of student needs, learning objectives, and the focus of test items, ensuring consistency in the use of language and format of items, which is important to avoid confusion or misinterpretation, ensuring that all team members have a homogeneous understanding of learning objectives. In brief, transparent communication leads to constructive feedback among English teachers which is important in the process of test items creation.

CHAPTER V
CONCLUSIONS AND RECOMMENDATIONS

In this chapter, the researcher will summarize and reaffirm the data analysis and discussion of the research results. Some recommendations are provided for proposed research and recommend future works related to the study in the same field of authentic assessment by other researchers. The conclusion and recommendation of this research will be presented below.

A. Conclusions

Based on the research findings and data analysis in the previous chapter, several conclusions can be established. Firstly, retrieved from the findings acquired in the first research question, researcher obtained several stages or activities of teachers in developing test items collaboratively. First, teachers conducted MGMP meetings that discussing curriculum development included basic competencies and materials to be taught throughout the semester, determining the number and type of questions, determining the teacher who responsible for constructing the test items. Second, teachers at the same grade conducted deeper discussion to confirm the specified material has been taught and ensure each class at a certain grade has received the same material. Third, double-check the test items by teachers who teach in the same grade and also assisted by the MGMP leader and the last was ensuring there was a pretest that similar to summative test items to students before the implementation of the examination.

Secondly, teachers' awareness definitely takes a critical role in the effectiveness and success of constructing English summative test items, the teachers involved in collaboration activity give their concern to the requirement of making a good test item. The results of this study showed some teachers' awareness in developing test items. First, teachers determine type of test to asses students' ability based on the objectives. Second, teachers utilize the sources which included textbooks, the internet and question banks as references to assist teachers in making questions. Third, teachers divide the number of questions that are referred to the material and the level of difficulty of the questions consisting of the lots, mots, and hots levels. Forth, collaborating teachers can get suggestion from different perspectives and help to ensure the sustainability of the indicators, because indicators are one of the important things that teachers should be aware of before making questions. Last, most of the items consist of multiple-choice questions, therefore test makers should be aware of choosing distractors because it is a crucial part in designing effective multiple-choice questions.

Furthermore, there were some challenges that teachers encountered in developing test items collaboratively, they were differences in teaching approaches, challenge in time coordination, challenge in established curriculum cohesivity and difference in understanding the difficulty level of the test items. However, these challenges can certainly be overcome by communication and discussion. Effective communication can help teachers efficiently coordinate in planning and implementing the curriculum, both in the learning process and the evaluation process, including the development of good test items that will be

carried out well if teachers do it collaboratively and get support from other teachers. One of the solutions to reduce time and space barriers was using online collaboration tools, such as online MGMP discussions through *WhatsApp* groups.

B.   Recommendations

English teacher's collaboration in developing test item was expected to enhance and help teachers especially in improving the evaluation system and students' abilities. Therefore, the researcher would like to provide some recommendations. There are several possible limitations that can be seen in some parts of this study.

The first limitation may come from the research participants. In this study, I only sampled teachers from one school. Then, the data collected only came from the teachers' perspective. Therefore, for future research, I recommend researchers to add more research participants. They may be able to include students' perspectives on the appropriateness between the materials tested and the material that has been taught.

The second limitation may come from the research questions. It is hoped that future research will provide deeper research questions to collect data accurately. The last limitation comes from the data collection method, I only conducted interviews to collect data. I suggest for future research to modify the method. The research can use data analysis to provide strong evidence of the quality of test items that teachers have developed collaboratively

REFERENCES

Amirullah. (2022). Peningkatan Kompetensi Guru Dalam Menyusun Ulangan Semester Genap Melalui Rapat Kerja Mgmp Di Sman 1 Riau Silip Bangka. *Jurnal Inovasi Manajemen Dan Supervisi Pendidikan: Vol. 2 No. 2 Juni 2022, P-2797-5592*

Arifin, Z. (2016). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.

Arikunto, S. (2005). *Dasar-dasar Evaluasi Pendidikan*. Jakarta : PT. Bumi Aksara.

Applebee, A. N., & Langer, J. A. (1983). Instructional scaffolding: Reading and writing as natural language activities. *Language arts, 60*(2), 168-175

Basuki, I., & Hariyanto. (2016). *Asesmen Pembelajaran*. Bandung: PT Remaja Rosdakarya.

Bryk dan Schneider (2002) *dalam Academic Optimism of Schools: A Force for Student Achievement.*

Bryman, A. (2008) *Social research methods. 3rd Edition,* New York: Oxford University Press.

Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom.* USA: ASCD.

Brookhart, S. M., & Moss, C. M. (2019). Advancing formative assessment in every classroom: A guide for instructional leaders. Alexandria, VA: Association for Supervision and Curriculum Development

Brown, H. D. (2004). *Language assessment: Principles and classroom practices.* New York: Pearson Education, Inc.

Brown, H. D. & Abeywickrama, P. (2010). *Language assessment: principles and classroom practice* (2nd ed.). New York: Pearson Education Inc.

Brown, J. D., & Hudson, T. (2002). *Criterion-referenced Language Testing.* Cambridge: Cambridge University Press.

Burke,K. (2009) *How to Asses Authentic Learning* (5th ed). Thausand Oaks, CA; Corwin

Chan, C. (2009). Assessment: Short Answer Questions. Assessment Resources@HKU, *University of Hong Kong*. Retrieved from

https://ar.cetl.hku.hk/pdf/ShortAnsQuestions.pdf

Conti, A. (2018). The balance between curriculum compliance and flexibility: A systematic review. *Journal of Educational Research, 45(2), 123-140.*

Creswell, J. W. (2008). *Educational research: Planning, conducting,and evaluating qu antitative and qualitative research.* New Jersey: Pearson Prentice Hall.

Creswell, J. W. (2012). *Educational Research Planning.*UK Sage Publication Detlev L., Jens F., & Juliane G. E. K. (2017). Competence Assessment in

Education. https://doi.org/10.1007/978-3-319- 50030-0_16

Febrina, U. (2021). Exploring English High School Teachers' Ability in Developing English Summative Test. *Banda Aceh: Islamic University Ar-Raniry*

Fives, H., & Buehl, M. M. (2012). *Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us*? In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors. (pp. 471–499). American Psychological Association. https://doi.org/10.1037/13274-019

Goodyear, L., Barela, E., Jewiss, J., & Usinger, J. (Eds.). (2014). Qualitative inquiry in evaluation: From theory to practice (Vol. 29). John Wiley & Sons.

Gronlund, N.E. and Linn, R.L. (1990) Measurement and Evaluation in Teaching. McMillan Company, New York.

Gunawan, A. W. (2012). *Genius Learning Strategy*. Jakarta: PT Gramedia Pustaka Utama.

Harlen, W. (2005). Teacher's summative practice and assessment for learning - tensions and synergies. *The Curriculum Journal, 16*(2),207-223.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. routledge.

Hidayati, T., Nugroho, S. E., & Sudarmin, S. (2013). Pengembangan tes diagnostik untuk mengidentifikasi keterampilan proses sains dengan tema energi pada pembelajaran IPA terpadu. *Unnes Science Education Journal*, *2*(2).

Hughes, A. 1989. *Testing for Language Teacher*. New York: Cambridge University Press.

Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia*, *12*(1), 40.

Kartowagiran, B. (2012). Penulisan butir soal. *Yogyakarta: Universitas Negeri Yogyakarta*

Kasmawati, Y. (2020). Peningkatan Kompetensi Melalui Kolaborasi: Suatu Tinjauan Teoritis Terhadap Guru. *Equilibrium: Jurnal Pendidikan, 8*(2), 136-142.

Khan, R. (2018). What is assessment? Purposes of assessment and evaluation. *The TESOL Encyclopedia of English Language Teaching*, 1-7.

Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. American Educational Research Journal, 33(4), 757–798.

Kreiter, C. D, Gordon, J.A, Elliott S., & Callaway, M. (2004). Recommendations for assigning weights to component tests to derive an overall grade. *Teach Learn Med. Spring;16*(2):133-8. doi: 10.1207/s15328015tlm1602_3. PMID:15446296.

Magdalena, I., Fauzi, H. N., & Putri, R. (2020). Pentingnya evaluasi dalam pembelajaran dan akibat memanipulasinya. *Bintang*, *2*(2), 244-257.

Magdalena, I., Mahromiyati, M., & Nurkamilah, S. (2021). Analisis Instrumen Tes Sebagai Alat Evaluasi Pada Mata Pelajaran SBdP Siswa Kelas II SDN Duri Kosambi 06 Pagi. *NUSANTARA*, *3*(2), 276-287.

Marshall, J. C. (2016). *The highly effective teacher: 7 classroom-tested practices that foster student success*. ASCD.

Mckeachie, W. J (2010). *McKeachie's Teaching Tips: Strategies, Research &Theory* (13th ed). Houghton: Mifflin Co

McLaughlin, M. W., & Talbert, J. E. (2006). Building school-based teacherlearning communities: Professional strategies to improve stu-dent achievement. New York: Teacher College Press.

Miller, D. C. (1991). *Handbook of research design and social measurement* (5th ed.). Sage Publications, Inc

Nasution. (2003). *Metode Penelitian Naturalistik Kualitatif.* Bandung: Tarsito Nurgiyantoro, B. (2010). Penilaian pembelajaran sastra berbasis

kompetensi. *Yogyakarta: BPFe.*

Osterlind, S. J. (2002). *Constructing Test Items: Multiple-choice, Constructed Response, Performance, and Other Formats*. New York: Kluwer Academic Publishers.

Popham, W. J. (2008). Transformative assessment. Alexandria, VA: Association for Supervision and Curriculum Development.

Purwanto, N. (2020). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran.* Bandung:PT. Remaja Rosdakarya

Saragih, F. H. (2016). Testing and assessment in English language instruction. *Jurnal Bahas Unimed*, *27*(1), 74656.

Saud, U. S, (2013). *Pengembangan Profesi Guru* (6th ed). Bandung; Alfabeta Suherman, A., & Saondi, O. (2010). *Etika Profesi Keguruan*. Bandung: PT Refika Aditama.

Setiawan, B. (2020). The anxiety of educational reform and innovation: Bridging of top-down and bottom-up strategies within practice educational reform of curriculum in Indonesia. *Studies in Philosophy of Science and Education*, *1*(2), 87-96.

Shohamy, E. (2020). *The power of tests: A critical perspective on the uses of language tests*. Routledge.

Stufflebeam, & Coryn. (2014). Evaluation, Theory, Models, & Application *Jossey-Bass. (Second, Vol. 59)*.

Sudaryono. (2012). *Dasar-dasar Evaluasi Pembelajaran*. Yogyakarta: Graha Ilmu.

Sudjana, N. (2013). *Dasar-Dasar Proses Belajar Mengajar*. Bandung: Sinar. BaruAlgesindo.

Sugiwanti. (2021). Metode Kerja Kelompok sebagai upaya Peningkatan Kompetensi Guru dalam Menyusun Soal Tes Semester di SDN Kalibanteng Kulon 02 Semarang. *JIPS, Vol. 2 No. 2*

Suharsimi, A. (2013). *Dasar–Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara Sukardi. (2010). *Evaluasi Pendidikan Prinsip dan Operasionalnya*. Jakarta: Buma Aksara

Suparman, M. A. (2020). *Desain Instructional Modern*. Jakarta: Erlangga. Suwarto. (2013). *Pengembangan Tes Diagnostik Dalam Pembelajaran*. Yogjakarta: Pustaka Pelajar

Timperley, H., Wilson, A., Barrar, H., Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis Iteration (BES)*. Auckland: Ministry of Education and the University of Auckland.

Tomlinson, C. A., & Moon, T. R. (2013). Assessment and student success in the differentiated classroom. Alexandria, VA: Association for Supervision and Curriculum Development.

Valiathan, P. (2009). Overcoming the Limitations of MCQs. *Knowledge PlatformWhite Paper: Singapore.*

Wahyuni, S. A., & Syukur, I. (2012). *Asesmen Pembelajaran Bahasa.*

Weimer, M. (2015). Advantages and disadvantages of different types of test questions. *Faculty Focus.*

Widoyoko, E. P. (2014). *Penilaian Hasil Pembelajaran di Sekolah* (1st ed). Yogyakarta: Pustaka Pelajar.

Xu, Y., & Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. Teaching and Teacher Education. 58. 149-162

Yin, R. K. (2018). Case Study Research and Applications: Design and Methods 6th edition. Singapore: SAGE Publications, Inc

Zaini, H., Munthe, B., & Aryani, S. A. (2008). Strategi pembelajaran aktif. Yogyakarta: Pustaka Insan Madani, 89, 2008.

Zaman, S., Yaqin, A., Ghinan, A., & Surya, R. (2018). Sistem Informasi Kolaborasi Pengembangan Instrumen Tes Pada Kegiatan Musyawarah Guru Mata Pelajaran (MGMP). Konferensi Nasional Sistem Informasi (KNSI)

Zapata-Rivera, D., Andrews-Todd, J., & Oliveri, M. E. (2021). Communicating assessment information in the context of a workplace formative task. The Journal of Writing Analytics, 5, 324-341.

APPENDICES

Appendix A

Appointment Letter of Supervisors

KEPUTUSAN DEKAN FAKULTAS TARBIYAH DAN KEGURUAN UIN AR-RANIRY BANDA ACEH
NOMOR : B-4474/Un.08/FTK/Kp.07.6/06/2024

TENTANG:
PENGANGKATAN PEMBIMBING SKRIPSI MAHASISWA

DENGAN RAHMAT TUHAN YANG MAHA ESA

DEKAN FAKULTAS TARBIYAH DAN KEGURUAN UIN AR-RANIRY BANDA ACEH

Menimbang :
a. bahwa untuk kelancaran bimbingan skripsi mahasiswa pada Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry Banda Aceh maka dipandang perlu menunjuk Pembimbing skripsi;
b. bahwa yang namanya tersebut dalam Surat Keputusan ini dianggap cakap dan mampu untuk diangkat dalam jabatan sebagai Pembimbing skripsi Mahasiswa;
c. bahwa berdasarkan pertimbangan sebagaimana dimaksud dalam huruf a dan huruf b, perlu menetapkan Keputusan Dekan Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry Banda Aceh.

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 Presiden 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 Nomor 64 Tahun 2013, tentang perubahan Institusi Agama Islam negeri Ar-Raniry Banda Aceh Menjadi Universitas Islam Negeri Ar-Raniry Banda Aceh;
7. Peraturan Menteri Agama RI Nomor 44 Tahun 2022, tentang Organisasi & Tata Kerja UIN Ar-Raniry Banda Aceh;
8. Peraturan Menteri Agama Nomor 14 Tahun 2022, tentang Statuta UIN Ar-Raniry Banda Aceh;
9. Keputusan Menteri Agama Nomor 492 Tahun 2003, tentang Pendelegasian Wewenang Pengakatan, Pemindahan dan Pemberhentian PNS di Lingkungan Depag RI;
10. Keputusan Menteri Keuangan Nomor 293/Kmk.05/2011, tentang penetapan institusi agama Islam Negeri UIN Ar-Raniry Banda Aceh pada Kementerian Agama sebagai Instansi Pemerintah yang menerapkan Pengelolaan Badan Layanan Umum;
11. Surat Keputusan Rektor UIN Ar-Raniry Nomor 01 Tahun 2015, Tentang Pendelegasian Wewenang kepada Dekan dan Direktur Pascasarjana di Lingkungan UIN Ar-Raniry Banda Aceh.

MEMUTUSKAN

Menetapkan : Keputusan Dekan Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry Banda Aceh tentang Pembimbing Skripsi Mahasiswa

KESATU : Mencabut Keputusan Dekan FTK UIN Ar-Raniry No: B-5156/Un.08/FTK/Kp.07.6/03/2023

KEDUA : Menunjuk Saudara :

1. Syarifah Dahliana, M.Ag., M.Ed., Ph.D   Sebagai Pembimbing pertama
2. Azizah, S.Ag., M.Pd   Sebagai Pembibing kedua

Untuk membimbing Skripsi

Nama : Wirda Yatul
NIM : 180203072
Program Studi : Pendidikan Bahasa Inggris
Judul Skripsi : Investigating Teachers' Collaboration in Developing Test Items

KETIGA : Kepada pembimbing yang tercantum namanya diatas diberikan honorarium sesuai dengan peraturan perundang-undangan yang berlaku;

KEEMPAT : Pembiayaan akibat keputusan ini dibebankan pada DIPA UIN Ar-Raniry Banda Aceh Nomor SP DIPA-025.04.2.423925/2024 Tanggal 24 November 2023 Tahun Anggaran 2024;

KELIMA : Surat Keputusan ini berlaku selama enam bulan sejak ditetapkan;

KEENAM : Surat Keputusan ini berlaku sejak tanggal ditetapkan dengan ketentuan bahwa segala sesuatu akan dirubah dan diperbaiki kembali sebagaimana mestinya, apabila kemudian hari ternyata terdapat kekeliruan dalam Surat Keputusan ini.

Ditetapkan di : Banda Aceh
Pada tanggal : 04 Juni 2024
Dekan,

Safrul Muluk

Tembusan
1. Sekjen Kementerian Agama RI di Jakarta;
2. Dirjen Pendidikan Islam Kementerian Agama RI di Jakarta;
3. Direktur Perguruan Tinggi Agama Islam Kementerian Agama RI di Jakarta;
4. Kantor Pelayanan Perbendaharaan Negara (KPPN), di Banda Aceh;
5. Rektor UIN Ar-Raniry di Banda Aceh
6. Kepala Bagian Keuangan dan Akuntansi UIN Ar-Raniry Banda Aceh;
7. Yang bersangkutan;
8. Arsip.

Appendix B

Recommendation Letter from The Fakultas Tarbiyah dan Keguruan to conduct field research

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI AR-RANIRY
FAKULTAS TARBIYAH DAN KEGURUAN
Jl. Syeikh Abdur Rauf Kopelma Darussalam Banda Aceh
Telepon : 0651- 7557321, Email : uin@ar-raniy.ac.id

Nomor   : B-12434/Un.08/FTK.1/TL.00/12/2023
Lamp   : -
Hal   : *Penelitian Ilmiah Mahasiswa*

Kepada Yth,

1. Kepala Kantor Kementerian Agama Kota Banda Aceh
2. Kepala MAN Model Banda Aceh

Assalamu'alaikum Wr.Wb.
Pimpinan Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry dengan ini menerangkan bahwa:

Nama/NIM   : WIRDA YATUL / 180203072
Semester/Jurusan : XI / Pendidikan Bahasa Inggris
Alamat sekarang : Blangkrueng

Saudara yang tersebut namanya diatas benar mahasiswa Fakultas Tarbiyah dan Keguruan bermaksud melakukan penelitian ilmiah di lembaga yang Bapak/Ibu pimpin dalam rangka penulisan Skripsi dengan judul *Investigating Teachers' Collaboration in Developing Test Items*

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

Banda Aceh, 04 Desember 2023
an. Dekan
Wakil Dekan Bidang Akademik dan Kelembagaan,

Berlaku sampai : 17 Januari 2024

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

Appendix C

Recommendation Letter from Kementrian Agama Kota Banda Aceh to conduct field research

KEMENTERIAN AGAMA REPUBLIK INDONESIA
KANTOR KEMENTERIAN AGAMA KOTA BANDA ACEH
Jalan Mohd. Jam No. 29 Telp 6300597 Fax. 22907 Banda Aceh Kode Pos 23242
Website : kemenagbna.web.id

Nomor	:	B -7604 /Kk.01.07/4/TL.00/12/2023	06 Desember 2023
Sifat	:	Biasa	
Lampiran	:	Nihil	
Hal	:	Rekomendasi Melakukan Penelitian	

Yth, Kepala MAN Model
Kota Banda Aceh

Assalamu'alaikum warahmatullahi wabarakatuh

Sehubungan dengan surat dari Dekan Fakultas Tarbiyah dan Keguruan Program Studi Pendidikan Bahasa Inggris Universitas Islam Negeri Ar-Raniry, nomor : B-12434/Un.08/FTK.1/TL.00/12/2023 tanggal 04 Desember 2023, perihal sebagaimana tersebut dipokok surat, maka dengan ini kami mohon bantuan saudara untuk dapat memberikan data maupun informasi lainnya yang dibutuhkan dalam rangka memenuhi persyaratan bahan penulisan Skripsi, kepada saudara/i :

Nama	:	Wirda Yatul
NIM	:	180203072
Prodi/Jurusan	:	Pendidikan Bahasa Inggris
Semester	:	XI

Dengan ketentuan sebagai berikut :

1. Harus berkonsultasi langsung dengan Kepala Madrasah yang bersangkutan dan sepanjang tidak mengganggu proses belajar mengajar.
2. Tidak memberatkan Madrasah.
3. Tidak menimbulkan keresahan-keresahan lainnya di Madrasah.
4. Tetap mematuhi protokol kesehatan yang berlaku di Madrasah.
5. Bagi yang bersangkutan supaya menyampaikan foto copy hasil penelitian sebanyak 1 (satu) eksemplar ke Kantor Kementerian Agama Kota Banda Aceh.

Demikian rekomendasi ini kami keluarkan, atas perhatian dan kerja sama yang baik kami ucapkan terima kasih.

Wassalamu'alaikum warahmatullahi wabarakatuh

Kepala

H. Salman, S.Pd, M.Ag
NIP. 197001021997031005

Tembusan :

1. Kepala Kantor Wilayah Kementerian Agama Provinsi Aceh;
2. Dekan Fakultas Tarbiyah dan Keguruan Program Studi Pendidikan Bahasa Inggris Universitas Islam Negeri Ar-Raniry;
3. Mahasiswa Yang Bersangkutan.

Appendix D

Confirmation Letter of Conducted Research from the School of MAN Model
Banda Aceh

KEMENTERIAN AGAMA REPUBLIK INDONESIA
KANTOR KEMENTERIAN AGAMA KOTA BANDA ACEH
MADRASAH ALIYAH NEGERI 1 BANDA ACEH
Jalan Pocut Baren No. 116 Keuramat Banda Aceh
Telp. 0651-636804 Fax. 0651-29466
Website: manmodelbna.sch.id, Email: mandelbandaaceh@gmail.com

Nomor	: B-837/Ma.01.90/TL.00/4/2024		1 April 2024
Lamp	: -		
Hal	: Telah Melakukan Penelitian		

Yth. Dekan Fakultas Tarbiyah dan Keguruan UIN Ar-Raniry
di-
Tempat

Assalamu'alaikum warahmatullahiwabarakatuh.

Dengan hormat,

Memenuhi maksud surat Wakil Dekan Bidang Akademik dan Kelembagaan Fakultas
Tarbiyah dan Keguruan Universitas Islam Negeri Ar-Raniry Nomor : B-12434/Un.08/
FTK-I/TL.00/12/2023 tanggal, 4 Desember 2023, dan Surat Rekommendasi Kepala
Kantor Kementerian Agama Kota Banda Aceh Nomor : B-7604/Kk.01.07/
4/TL.00/12/2023 tanggal 6 Desember 2023 perihal Rekomendasi Melakukan Penelitian,
maka dengan ini menyatakan bahwa:

Nama	:	Wirda Yatul
N I M	:	180203072
Program Studi	:	Pendidikan Bahasa Inggris
Semester	:	XI / sembilan
Fakultas	:	Tarbiyah dan Keguruan Universitas Islam Negeri Ar-Raniry
Jenjang	:	S1

Telah melaksanakan tugas melakukan penelitian untuk mengumpulkan data Skripsi
dengan judul *"Investigating Teachers' Collaboration In Developing Test Items"*,
pada Madrasah Aliyah Negeri 1 Banda Aceh.

Demikian surat ini kami sampaikan untuk dapat dipergunakan seperlunya.

Kepala,

Nursiah

Appendix D

Interview Protocol

1. What are the initial activities that you do in English teacher collaboration?

2. What are the stages of teacher collaboration in the process of developing test items?

3. What is the difference between making items individually and collaboratively?

4. When collaborating, what types of questions do you usually use in making items and why do you choose these types of questions?

5. How do you determine the indicators that are suitable for the items that have been made?

6. Are the questions that you make the question indicators already integrated critical thinking skills?

7. In the process of making questions, what references do you usually use?

8. What are the difficulties experienced when making questions collaboratively?

9. In collaborative activities, do you have difficulty in determining the distribution of the number of questions?

10. Do you have difficulty in determining the distractors?

11. Do you have difficulty in determining the value of each item?

12. Is there an examination of the items that have been made before they are given to students? How is the process?

13. Is there a trial of questions to students first before finally the questions are tested on students?

Appendix E

Documentation

AUTOBIOGRAPHY

Personal Identity

Name   : Wirda yatul

Place / Date of Birth : Seubadeh/06 Maret 2000

Sex   : Female

Religion   : Islam

Nationality   : Indonesia

Address   : Seubadeh, Kec. Bakongan Timur, Aceh Selatan

Email   : 180203072@student.ar-raniry.ac.id

Occupation   : Student of Department of English Language Education,

UIN Ar-Raniry

Parents' Bio

Father's Name   : Sulaiman, LB

Mother's Name   : Yusfarni

Father's Occupation   : Civil Servent

Mother's Occupation   : Civil Servent

Address   : Seubadeh, Kec. Bakongan Timur, Aceh Selatan

Educational Background

Elementary School   : SDN 1 Seubadeh

Junior High School : MTsS Al-munjiya Labuhan haji

Senior High School : MAS Ash-habul Yamin Bakongan

University   : Department of English Language Education, Fakultas

Tarbiyah dan Keguruan, UIN Ar-Raniry