CHAPTER III

Research Method

3.1 Research Design

This research employed a quantitative method. As defined by Arikunto (2006: 12), quantitative research is an approach that relies heavily on numerical data in various aspects, including the collection, interpretation, and presentation of data to derive results. In conducting the research, the researcher conducts the quantitative based on quasi experimental method. Quantitative research is based on the measurement of quantity or amount. It is applicable to phenomena that can be expressed in terms of quantity. This research adopted quasi experimental method. Experimental research is a form of comparative analysis that involves testing or attempting to prove a hypothesis by way of experimentation. This method has a control group, but cannot fully function to control external variables that affect the implementation of the experiment. The purpose of quantitative research is to obtain an explanation of a theory and laws of reality. It is developed by using mathematical models, theories or hypotheses. Thus, it means a method that emphasizes more on the aspects of objective measurement of social phenomena and further exploration in finding facts and testing theories that exist.

To collect the data, this research used a pre-test post-test design. This research conducted using two classes, one for the control class and the other for the experimental class. In the control class, students asked to do a pretest, the pretest is in the form of students learn as usual, namely the traditional class, then they did a

speaking test. In the experimental class, students treated using a flipped classroom model where material in the form of videos gave before the next class meeting. At the first meeting, students were asked to do a pretest, a pretest in the form of a responsive speaking test in English regarding the topic Report text. Students are divided into several groups and then discuss the topic. At the next meeting, students begin to be treated by being given material and group assignments before class learning begins. Students study the material that has been given together with group members. Then, students held discussions during class meetings. The treatment was carried out twice. Then, students are given a post-test at the end of the treatment by practicing speaking about the report text topic to find out whether the student's pronunciation can improve or not with examples of reading report text as in the video shared previously as a teaching medium.

3. 2 Variable of the Research

Variables are attributes as well as the objects of the research. The component is critical in generating research conclusions. It signifies that the variables are our research's limitation. It serves as a guide for researchers to focus on the objective of the research. In summary, it is anything that the researcher wishes to investigate. Variable, according to Dewberry (2017), separates into two. They are the independent and dependent variables. Independent variables are ones that the researcher manipulates or contrasts independently. The dependent variable is the variable on which the outcome of this modification or contrast is measured in the research design. It signifies that there is one variable that both influences and is impacted.

1. Independent variable (X): Flipped classroom is a class model to help students improve speaking skills.
2. Dependent variable (Y): Students' speaking skill.

3.3 Population and Sample

Population is a set with characteristics determined by the researcher in such a way that each individual/variable/data can be stated exactly whether the individual is a member or not. While the sample is representative of the population study (Rahmayani, A 2020). The population in this research are class XI students at SMA N 5 Kota Jambi. There are any 12 class and 34 in each class. The samples taken in this study were 68 students which is XI B1 is 34 students and XI B 2 is 34 students.

The researcher conducted purposive sampling technique. According to Cohen (2007) purposive sampling technique is used for a specific purpose in the research. It was in line with this research, class of XI B1 and XI B2 had the same ability in English and had the similar characteristics of scores in English.

3.4 Instrument of the Research

This is the research's input, and the outcome, the relevance and accuracy of the results, is wholly dependent on it. This research used speaking test. The speaking test is given in the pre-test and post-test in experimental class and control class. The pre-test is administered before giving the treatment to know the initial speaking of the students. Then, for the post test is managed to check what the effect of FCM in students' speaking skill.

Speaking Test

The speaking test is administered twice, namely the pre-test and the post-test. The pre-test is given before the treatment, while the post-test is conducted after the treatment has been administered. The pretest was conducted before implementing the treatment to assess the initial speaking proficiency of the students. Subsequently, the post-test was administered to determine whether the flipped classroom model had a positive impact on students' speaking abilities.

Speaking Pre-test

Instruction:

1. Make a group consist of 4-5 students
2. Join with your member group
3. Make a hortatory text
4. Compile and discuss with your group at home
5. The next meeting, your group will present the hortatory text in front of the class but speaking individually.

Task:

Free hortatory text topics.

Speaking Post-test

Instruction

1. Make a group consist of 5-6 students
2. Join with your member group
3. Discuss with your group
4. Explain and share about the hortatory text spontaneously
5. Speaking individually

Task:

Free topics related to the school and the education.

The scoring rubric to assessment the students' speaking skill that adapted from David P. Harris (1977), p.84 which are the scores from pronunciation, fluency, vocabulary, accuracy and comprehension. Below are the table of test distribution and rubric the scoring.

No	Criteria	Rating Score	Description
1	Pronunciation	5 (95-100)	Has few traces of foreign language
		4 (85-94)	Always intelligible, thought one is conscious of a definite accent.
		3 (75-84)	Pronunciation problem necessities concentrated listening and occasionally lead to misunderstanding.
		2 (65-74)	Very hard to understand because of pronunciation problem, most frequently be asked to repeat.
		1 (<65)	Pronunciation problem to serve as to make speech virtually unintelligible.
2	Grammar	5 (95-100)	Make few (if any) noticeable errors of grammar and word order.
		4 (85-94)	Occasionally makes grammatical and or word orders errors that do not, however obscure meaning.
		3 (75-84)	Make frequent errors of grammar and word order, which occasionally obscure meaning.
		2 (65-74)	Grammar and word order errors make comprehension difficult, must often rephrase sentence.
		1 (<65)	Errors in grammar and word order, so severe as to make speech virtually unintelligible.
3	Vocabulary	5 (95-100)	Use of vocabulary and idioms is virtually that that of native speaker.
		4 (85-94)	Sometimes uses inappropriate terms and must rephrase ideas because of lexical and equities.

		3 (75-84)	Frequently uses the wrong words conversation somewhat limited because of inadequate vocabulary.
		2 (65-74)	Misuse of words and very limited vocabulary makes comprehension quite difficult.
		1 (<65)	Vocabulary limitation so extreme as to make conversation virtually impossible.
4	Fluency	5 (95-100)	Speech as fluent and efforts less as the of native speaker.
		4 (85-94)	Speed of speech seems to be slightly affected by language problem.
		3 (75-84)	Speed and fluency are rather strongly affected by language problem
		2 (65-74)	Usually hesitant, often force into silence by language limitation.
		1 (<65)	Speech is halting and fragmentary as to make conversation virtually imposible.
5	Comprehension	5 (95-100)	Appears to understand everything without difficulty
		4 (85-94)	Understand nearly everything at normal speed although occasionally repetition may be necessary.
		3 (75-84)	Understand most of what is said at slower than normal speed without repetiti on.
		2 (65-74)	Has great difficulty comprehend social conversation spoken slowly and with frequent repetition.
		1 (<65)	Cannot be said to understand even simple conversation.

The standard minimum proficiency level (KKM) for English mastery at SMA N 5 Kota Jambi was set at 70 for both classes. Consequently, the researcher categorized pre-test and post-test scores into three groups: low, middle, and high.

Low scores were below 70 (falling short of the minimum standard), middle scores ranged from 70 to 80, and high scores were above 80.

3.5 Data Collection Technique

Researchers obtained data by giving speaking tests to experimental classes and control classes with the aim of finding improvements in data before and after being treated with the flipped classroom method. The speaking test consists of responsive and interactive speaking between students who are observed by researchers and teachers regarding simple present tense, report and hortatory text material. When students speak, teachers and researchers give grades, and to obtain data researchers also record students' speaking using a voice recorder. In analyzing spoken English, the researcher listened carefully by playing the audio forward and backward, then completed the assessment, paying more attention to the five speaking aspects and assessed using a speaking rubric.

Therefore, this research will collect data using the following this procedure:

1. Formulate the problem

   Researcher found that current technological developments are in line with existing educational methods. The advantages of the flipped classroom method make teachers also apply it to support learning objectives. As we know, language learning, especially speaking, must be learned by increasing practice. Therefore, can this Flipped classroom mixed learning method also improve students' skills or does it not even have any effect on students' English

speaking? Researchers want to conduct this research to measure the extent of the flipped classroom effect on students' speaking skills.

2. Select the population and sample

The population of this study were class XI high school students at SMA N 5 Jambi City 370 students from 10 classes). Researcher use purposive sampling at eleventh grade. The researcher used two classes consisting of 34 experimental class and 34 control class students.

3. Determination of materials

Researchers compiled material based on the class XI syllabus at SMA N 5 Jambi City. Researchers choose several learning videos related to the topic and students are required to study these videos before learning begins. Researchers used seven learning videos from YouTube that were appropriate to the syllabus and students' needs. Researchers took videos from the channel Travel Guide, TED Talk, TED Education, Khan Academy, Asap SCIENCE, English Today Jakarta, Tech Insider, and Bright Side.

4. Giving a pre-test

Before carrying out teaching and learning activities in this research, the researcher gave a speaking test using the responsive and interactive speaking type, namely questions and answers between students and teachers, students and students in groups. First, students are asked to form groups consisting of 5 students per group. Then, students are tasked with discussing a hortatory text together with their group members at home. Afterwards, students are assigned the task of creating a

hortatory text at home. Following this, the in-class activity involves students practicing speaking and presenting the text they have created at home with their group members in the classroom. This is done to determine their basic ability to speak English and to obtain their score data before receiving treatment. Researcher recorded the results of students' speaking tests in both the control and experimental classes. In addition to recording, both the teacher and the researcher also provide scores for students' speaking abilities using the provided rubric scoring guidelines. To ensure the accuracy of the scores given, the researcher reviews the recorded results of students' speaking in the classroom.

5. Provide treatment

At this stage, students are taught how to speak like native speakers using learning video media on YouTube. For experimental classes, students are sent learning videos long before the class day, then students are given instructions to discuss with their group friends and do the homework that has been given. For the control class, students were also given the same learning videos as the experimental classes, but the videos were given during class. then continued with group discussion. The treatment in this study was carried out three times.

6. Give a post-test

After receiving treatment, a post-test was carried out. In the last meeting, the researcher asks the students to speaking test, to get the score of post-test. The post test here still about the hortatory text and the students speak one by one, than the researcher and the teacher start to scoring. The post-test in this study aims to

evaluate students' speaking after receiving treatment. Researchers also recorded students' speaking results.

7. Assessment of students' English speaking tests

Student scores show that there are differences in students' speaking English abilities in the experimental class and the control class, before and after treatment. In assessing students' speaking skills, the researcher employs an inter-rater approach. Rater one is conducted by the researcher, while rater two is carried out by the English teacher. During the speaking test, both the teacher and the researcher approach the students and observe various aspects of their speaking. As the speaking test commences, the teacher and the researcher hold a class attendance sheet containing the names of the students and begin assigning scores as each student speaks, based on predetermined speaking aspects. If a student's pronunciation sounds clear and accurate, they receive a higher score, and the same applies to other speaking aspects. Once all students have had the opportunity to speak, the assessments by rater one and rater two are transcribed into Microsoft Excel for further data processing using SPSS, specifically employing the Paired Sample T-Test.

8. Analyze data

Researchers analyzed data from the pre-test and post-test in the experimental class and control class using the IBM SPSS Statistics 26 application to find answers to research questions. The researcher uses Paired Sample T-Test.

That is the data collection procedure for this research.

3.6 Validity

According to Timothy Teo (2013), validity is an integrated evaluative assessment of the amount to which empirical evidence and theoretical reasoning support the sufficiency and appropriateness of test results or other assessment methods. A test is said to be legitimate if it can measure exactly what should be measure in a research. The comparison of student scores demonstrates this.

1. Content Validity

   Content validity is determined by the determination or representativeness of sampling from the content to be studied (Kerlinger, 1973). Content validity is related to all the items of the instrument. In fulfilling the validity of this type, the researcher must consider at all indicators in the form of items and analyze it whether the measuring instrument as a whole can represent the material to be measured. If a measuring instrument has represented all the ideas or domains related to the material to be measured, the measuring instrument has fulfilled the aspects of content validity. In this research, the researcher used the syllabus as guidance in making the assessment of test items that are appropriate for the purpose of the test. Researchers organized the learning materials and activities based on the syllabus made by the English teacher and adjusted to the learning objectives, so that the tests given to students are based on the syllabus.

2. Convergent Validity

According to Carlson (2010), convergent validity reflects the extent to which two measurements capture the same construct. Alternative measurements that provide less-than-perfect convergent validity introduce ambiguities that interfere with the development of meaningful interpretations of findings within and across studies. Convergent validity is one of the types of validity in psychological measurement that measures the extent to which a measurement tool can measure the same or similar constructs as another measurement tool that has been deemed valid (Carlson, 2010). In this context, convergent validity assesses how much the results of two different measurement tools that measure the same or similar constructs correlate or converge. Evaluate the value of Cronbach's alpha for each instrument or scale. A high Cronbach's alpha value (usually above 0.70) indicates good internal reliability and suggests that the items within the instrument or scale are well correlated. This is an indication that the instrument or scale effectively measures the same or similar constructs.

Table Validity
Pre Test

Correlations

		PrePro	PreFlu	PreVoc	PreAcc	PreCom	Jumlah
PrePro	Pearson Correlation	1	.708**	1.000**	.695**	.646**	.914**
	Sig. (2-tailed)		.000	.000	.000	.008	.000
	N	34	34	34	34	34	34
PreFlu	Pearson Correlation	.708**	1	.708**	.622**	.624	.813**
	Sig. (2-tailed)	.000		.000	.000	.042	.000
	N	34	34	34	34	34	34
PreVoc	Pearson Correlation	1.000**	.708**	1	.695**	.646**	.914**
	Sig. (2-tailed)	.000	.000		.000	.008	.000
	N	34	34	34	34	34	34
PreAcc	Pearson Correlation	.695**	.622**	.695**	1	.661**	.798**
	Sig. (2-tailed)	.000	.000	.000		.006	.000
	N	34	34	34	34	34	34
PreCom	Pearson Correlation	.646**	.624**	.646**	.661**	1	.761**

	Sig. (2-tailed)	.008	.042	.008	.006		.000
	N	34	34	34	34	34	34
Jumlah	Pearson Correlation	.914**	.813**	.914**	.798**	.661**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	34	34	34	34	34	34
**. Correlation is significant at the 0.01 level (2-tailed).							

From the table of the convergent validity test above, the value of each aspect speaks more than 0.60. This proves that the data on the pre-test of the experimental class is valid.

Validity Post Test

Correlations							
		PostPro	PostFlu	PostVoc	PostAcc	PostCom	Jumlah
PostPro	Pearson Correlation	1	.606**	.693**	.706**	.605**	.807**
	Sig. (2-tailed)		.017	.000	.000	.000	.000
	N	34	34	34	34	34	34
PostFlu	Pearson Correlation	.606**	1	.678**	.646**	.651**	.698**
	Sig. (2-tailed)	.017		.004	.001	.001	.000
	N	34	34	34	34	34	34
PostVoc	Pearson Correlation	.693**	.478**	1	.772**	.718**	.881**
	Sig. (2-tailed)	.000	.004		.000	.000	.000
	N	34	34	34	34	34	34
PostAcc	Pearson Correlation	.706**	.646**	.772**	1	.717**	.896**
	Sig. (2-tailed)	.000	.001	.000		.000	.000
	N	34	34	34	34	34	34
PostCom	Pearson Correlation	.605**	.651**	.718**	.717**	1	.883**
	Sig. (2-tailed)	.000	.001	.000	.000		.000
	N	34	34	34	34	34	34
Jumlah	Pearson Correlation	.807**	.698**	.881**	.896**	.883**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	34	34	34	34	34	34
*. Correlation is significant at the 0.05 level (2-tailed).							
**. Correlation is significant at the 0.01 level (2-tailed).							

From the table of convergent validity test on the experimental post-test above, the value of each aspect speaks more than 0.60. This proves that the data on the post-test of the experimental class is valid.

Table Validity Pre Test Control Class

Correlations

		PrePro	PreFlu	PreVoc	PreAcc	PreCom	Jumlah
PrePro	Pearson Correlation	1	.246	.041	.046	.415*	.576**
	Sig. (2-tailed)		.161	.820	.794	.015	.000
	N	34	34	34	34	34	34
PreFlu	Pearson Correlation	.246	1	.641**	.258	.186	.712**
	Sig. (2-tailed)	.161		.000	.141	.291	.000
	N	34	34	34	34	34	34
PreVoc	Pearson Correlation	.041	.641**	1	.449**	.075	.649**
	Sig. (2-tailed)	.820	.000		.008	.671	.000
	N	34	34	34	34	34	34
PreAcc	Pearson Correlation	.046	.258	.449**	1	.241	.601**
	Sig. (2-tailed)	.794	.141	.008		.171	.000
	N	34	34	34	34	34	34
PreCom	Pearson Correlation	.415*	.186	.075	.241	1	.647**
	Sig. (2-tailed)	.015	.291	.671	.171		.000
	N	34	34	34	34	34	34
Jumlah	Pearson Correlation	.576**	.712**	.649**	.601**	.647**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	34	34	34	34	34	34
*. Correlation is significant at the 0.05 level (2-tailed).							
**. Correlation is significant at the 0.01 level (2-tailed).							

From the table of convergent validity test on the pre-test control class above, each aspect's value speaks more than 0.60. This proves that the data on the pre-test control class is valid.

Table validity Post-Test Control Class

Correlations							
		PostPro	PostFlu	PostVoc	PostAcc	PostCom	Jumlah
PostPro	Pearson Correlation	1	.042	.016	.003	.307	.535**
	Sig. (2-tailed)		.813	.931	.987	.078	.001
	N	34	34	34	34	34	34
PostFlu	Pearson Correlation	.042	1	.064	-.054	-.122	.265
	Sig. (2-tailed)	.813		.721	.760	.493	.129
	N	34	34	34	34	34	34
PostVoc	Pearson Correlation	.016	.064	1	.004	.178	.464**
	Sig. (2-tailed)	.931	.721		.982	.313	.006
	N	34	34	34	34	34	34
PostAcc	Pearson Correlation	.003	-.054	.004	1	.563**	.581**
	Sig. (2-tailed)	.987	.760	.982		.001	.000
	N	34	34	34	34	34	34

PostCom	Pearson Correlation	.307	-.122	.178	.563**	1	.786**
	Sig. (2-tailed)	.078	.493	.313	.001		.000
	N	34	34	34	34	34	34
Jumlah	Pearson Correlation	.535**	.265	.464**	.581**	.786**	1
	Sig. (2-tailed)	.001	.129	.006	.000	.000	
	N	34	34	34	34	34	34
**. Correlation is significant at the 0.01 level (2-tailed).							

3.8 Reliability

The instrument of this research is speaking English test. Researchers use inter-rater reliability to see test consistency. The level of agreement between several raters or judges is measured using inter-rater reliability. There will be two assessors who will assess the test, namely the researcher and the English teacher at SMA 5 Jambi City. The assessor will examine the student's speaking test based on the five aspects of speaking. Moreover, the reliability of the test in this research was measured by using SPSS version 26 Cronbach's Alpha. The data will be realible if the Alpha value is 0.70. It is shown in the following table.

Reliability of Experimental Class

Reliability pre test

Reliability Statistics	
Cronbach's Alpha	N of Items
.877	5

Reliability post test

Reliability Statistics	
Cronbach's Alpha	N of Items
.888	5

Reliability of Control Class

Reliability pre-test

Reliability Statistics	
Cronbach's Alpha	N of Items
.884	5

Reliability Post test

Reliability Statistics	
Cronbach's Alpha	N of Items
.845	5

3.9 Data Analysis Technique

Data analysis is a process of collecting data to obtain clear and understandable information. After the research data was collected, the data was then analyzed. The following are the specifications for each step in analyzing data:

1. Look at the results of the notes from researchers and English teachers in class, then play them back while sharing the students' recordings.

2. Transcribe students' speaking test scores from rater one and rater two by using in the Microsoft Excell, so that researcher can easily process data, they can search for averages and also totals in the data obtained.

3. Assessment of student pre-test and post-test using inter-rater. In both Pre-test and Post-test assessments, both Rater One and Rater Two carefully observe the students. Each rater holds the class attendance to provide speaking scores for the students when they speak. If the raters are unsure about giving a score, the teacher and the raters can also replay the student's audio recording to ensure the scores given to the students. The scoring system is based on David P Harris's rating scale for each aspect of speaking.

4. Calculate all data on the pretest and posttest in the control and experimental classes to find out whether there is an improvement in each aspect of speaking. Find the overall total score in each control and experimental class,

   average, difference and percentage difference. In calculating the students' scores in this research, the researcher uses Microsoft Excel to find the total, average, lowest score, and also the highest score. However, to test the significance, the researcher uses the Paired Sample T-Test in the SPSS application.

5. Investigate the calculation of student speaking statistics. Researchers used the Paired Sampled T-test in SPSS to obtain statistical data calculations.

6. Obtain data to find out whether there is an improvement in students' speaking after they were taught using the flipped classroom method and to

find out whether there is a significant difference in students' speaking abilities in flipped classroom learning.

7. Arrange discussions based on the results. This is obtained by comparing student scores on the pre-test post-test experimental class and pre-test post-test control class.

8. Get answers to research questions by concluding results based on analysis.

The steps above are the process of analyzing this research data.

3.9.1 Normality Test

Paired sample t-test can be used if the data come from a normal distribution. Normality test is conducted in order to know whether the data is normally distributed or not. The followings are the hypotheses of the normality test:

H0: The distribution of the data is normal.

H1: The distribution of the data is not normal

The null hypothesis (H0) is accepted if the significant level of the normality test is higher than 0.05, whereas the alternative hypothesis is accepted if the significant level of the normality test is lower than 0.05. The normality test is shown in the table below:

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Class	Statistic	df	Sig.	Statistic	df	Sig.
Students Score	Pre Test Experimental Class	.114	34	.200*	.962	34	.269
	Post Test Experimental Class	.116	34	.200*	.975	34	.625

	Pre Test Control Class	.107	34	.200*	.990	34	.988
	Post Test Control Class	.123	34	.200*	.975	34	.596
*. This is a lower bound of the true significance.							
a. Lilliefors Significance Correction							

Saphiro-Wilk formula is used as the significant values in this test because the element (df) is 34. From the table above, it indicates that the significant level of pretest and posttest in experimental class are higher than 0.05 (0.269 > 0.05 ; 0.625> 0.05). In control class also higher than 0,5 (0.988 > 0.05 ; 0.596 > 0.05). Therefore, the null hypothesis (H0) is accepted, it means that all of the data are normally distributed.

3.9.2 Homogeneity Test

Homogeneity test is used to know whether the data come from the homogeneous variance or not. To calculate the data, the researcher uses IBM SPSS version 26.

Test of Homogeneity of Variances					
		Levene Statistic	df1	df2	Sig.
Output students	Based on Mean	.382	1	66	.539
	Based on Median	.405	1	66	.527
	Based on Median and with adjusted df	.405	1	65.288	.527
	Based on trimmed mean	.408	1	66	.525

Based on the output above, the sig value is known. Based on the mean for the speaking learning outcome variable, it is 0.539, where this figure is above 0.05. so, it can be concluded that the variance in the data on students' speaking learning outcomes in the experimental class and control class is homogeneous.

3.10 Hypothesis Testing

Hypotheses were analyzed by using Paired Sample T-test of Statistical Package for Social Science (SPSS) version 26. SPSS used to know how the effect of flipped classroom model in English speaking skill. Paired sample t-test is a test of the mean difference for two paired samples. This test was used to analyze the pre-post research model in experimental and control class. Paired sample t-test was used to evaluate certain treatments on the same sample. According to Widiyanto (2013:35), paired sample t-test is one of the testing methods used to assess the effectiveness of the treatment, marked by differences in the average before and after treatment in experimental class and control class. The basis for taking the decision to accept or reject H0 in this test is as follows.

1. If T-value > T-table and probability (Asymp.Sig) < 0.05, then H0 is rejected and H1 is accepted.
2. If T-value < T-table and probability (Asymp.Sig) > 0.05, then H0 is accepted and H1 is rejected. The following is the formula for testing this research hypothesis:

$$H1 = Sig. < 0.05$$

H0: There is no statistically significant difference between the pre test and post test in the control class.

H1: There is a statistically significant difference in the pre-test and post-test in the experimental class.