analysis used in this research was recursive which refered to the combination of multiple regression model and mediation model through intervening variable. It is because there were three types of variable, those are independent, intervening, and dependent variable, and the independent variable expected to contribute to intervening and dependent variable directly and indirectly contribute to dependent variable through intervening variable.

B. Subject of the Study

This research was conducted in English Debate Competition of East Java English Club 4 (EJEC 4) 2020. EJEC is the name of English club association in East-Java which located at *Jalan* Bebekan No. 70 Slorok Kromengan Kabupaten Malang. EJEC 4 was held at SMK PGRI 2 Kediri, located at *Jalan* KH. Abdul Karim No. 5 Bandar Lor Mojoroto Kediri. There were 24 debate teams that join this competition and each consisted of 3 students. It means there were 72 students with 42 female and 30 male. 16 schools delegated their students to compete in EJEC 4, those are SMKN 2 Malang, SMKN 4 Kediri, SMKN 1 Kepanjen, SMKN 1 Doko, SMK Brantas Karangkates, SMAN 7 Malang, Al Fiqroh Annajiyah Malang, SMK Wahidiyah, SMKN 1 Kediri, SMKN 3 Kediri, SMK PGRI 2 Kediri, SMAN 1 Kediri, SMAN 2 Kediri, SMKN 2 Tuban, SMAN 1 Blitar, and SMKN 2 Pacitan.

The researcher got data from 67 out of 72 participants with 39 female and 28 male students. It is because 5 students did not collect the questionnaire given by researcher. This 67 students came from tenth until twelfth grades

from all schools that have mentioned before. The detail of subject can be drawn

as the table below:

Table 3.1

The Distribution of Participant

School	Female	Male
SMKN 2 Malang	2	1
SMKN 4 Malang	4	5
SMKN 1 Kepanjen	2	6
SMKN 1 Doko	1	2
SMK BrantasKarangkates	2	1
SMAN 7 Malang	2	1
Al FiqrohAnnajiyah Malang	3	-
SMK Wahidiyah	3	-
SMKN 1 Kediri	4	5
SMKN 3 Kediri	2	-
SMK PGRI 2 Kediri	3	-
SMAN 1 Kediri	1	2
SMAN 2 Kediri	4	2
SMKN 2 Tuban	3	
SMAN 1 Blitar	1	2
SMKN 2 Pacitan	2	1
Total	39	28

C. Instrument of the Data

The instrument used by researcher to collect the data for answering the research questions. There were 2 kinds of instruments used in this study. Firstly, questionnaire using "yes" or "no" answer and questionnaire using four stages of likert scale. The "Yes" or "No" questionnaire was used by researcher in Debate questionnaire. The likert scale questionnaires were used in Critical Thinking questionnaire and Self-confidence questionnaire. Secondly, speaking test using Students' Oral Language Observation Matrix (SOLOM) was also used by researcher to measure the speaking ability of the students.

In Debate questionnaire, there were 25 questionnaire items (Appendix 10). The Critical thinking questionnaire consisted of 45 question items with 36 positive questions and 9 negative questions (Appendix 10).Beside, self-confidence questionnaire consisted of 37 question items with 29 positive questions and 8 negative questions (Appendix 10). The SOLOM as the scoring indicator for speaking has 5 levels which indicated how well the speaking ability of the students (Appendix 9).

D. Validity and Reliability

Validity was used in this research to measure the instrument. Validity of research instrument assesses the extent to which the instrument measures what it is designed to measure (Robson, 2011). It is also refers to the degree to which the results are truthful. Therefore, the instrument of the research (questionnaire) must be correct to measure the concept under the study (Pallant, 2011). According to Zohrabi (2013) in (Mohajan, 2017), the

qualitative research validity is based on the fact that validity is about trustworthiness, dependability, and utility.

This research used content validity and face validity. Content validity was to measure how far the questionnaire and the score represent the possible question that should be asked based on the skill or content (Cresweel, 2005). So, it is related to the questionnaire which represents the theory or concept used. It usually depends on expert judgment in the field to measure the content because there is no statistical test to determine whether a measure adequately covers a content area (Mohajan, 2017). The content validity is used by researcher to measure the questionnaires. Questionnaire for knowing the debate background of the students was checked by English teacher and also Deputy Chairman of EJEC, Jodi CahyoWibowo, M.pd. For critical thinking and self-confidence questionnaires were checked by one of Educational Psychology Lecturer of IAIN Kediri, Irma Rosalina M.Pd.

The face validity was used by researcher to evaluate the appearance of the questionnaire in the context of feasibility, readability, consistency of format and style, and the clarity of language used. Different from content validity which uses expert judgment, the face validity uses non-expert or another sample to measure the appearance of the questionnaire. In other words, it is simply refers to the subjective assessment of the researcher to validate the presentation and relevance of the questionnaire on whether or not it is appear to be relevant, clear without any ambiguity (Harmend, 2016).

The researcher conducted the tryout for the questionnaires to know the validity and reliability of the questionnaires. The tryout conducted on Saturday, 1st February 2020. The questionnaires were distributed by the researcher to students who join debate extracurricular at SMK PGRI 2 Kediri and SMKN 1 Ngasem. There were 30 students in this tryout. They filled the critical thinking and self-confidence questionnaires. For the validity of debate questionnaire, the researcher took the data from the real sample. It means the researcher did not tryout the validity, but she took the data from debate competition (EJEC 4) and analyzed the validity also from the data. Here are the explanation of validity (Appendix 5).

From the data we can see that in debate questionnaire there were 26 questionnaire items. After the validity check, the question number 7 was not valid because the r-result (0.148) is lower than r-table (0.236, N= 67). The result of critical thinking questionnaire showed that there were 54 questionnaire items before tryout. 45 questions valid and 9 questions invalid. Those questions were invalid because the r-result is lower than r-table (0.349 N= 30). For the self-confidence questionnaire, there were 45 questionnaire items before validity test. After validity test, there are 8 items wee invalid and 37 valid items.

To measure the consistency of the score, the researcher also used reliability. According to (Creswell, 2012) reliability refers to the consistency of the score obtained. It means when the students do test with the same test, the

test should yield similar result. Score should be nearly the same when the researcher administers the instrument multiple times and at the different times.

To measure the reliability of questionnaire, the researcher used coefficient alpha or Cronbach's alpha to estimate internal consistency. If the items were scored as continuous variables (e.g., strongly agree to strongly disagree), the alpha provides a coefficient to estimate consistency of score on an instrument. Therefore, in the present study, researcher used this reliability measurement.

The result of reliability test of debate questionnaire showed 0. 784 in Cronbach's alpha. This indicated that the instrument is reliable based on the fact that the alpha value is greater than 0.7.

Table 3.2

Reliability of Debate Questionnaire

Cronbach' Alpha	N of Item
.784	25

For the critical thinking questionnaire, the reliability statistic of Cronbach's alpha is 0.871. The instrument is reliable if the alpha value greater than 0.7. It means this questionnaire is reliable because the alpha value is greater than 0.7. The reliability can be seen in the table below:

Table 3.3

Reliability of Critical Thinking Questionnaire

Cronbach' Alpha	N of Item
.871	45

For the self-confidence questionnaire, the result of reliability statistic of Cronbach's alpha is 0.946. The instrument was reliable if the alpha value greater than 0.7. It means this questionnaire is reliable because the alpha value is greater than 0.7. The reliability can be seen in the table below:

Table 3.4

Reliability Self- Confidence Questionnaire

Cronbach' Alpha	N of Item
.946	37

For the reliability of speaking test, the inter-rater reliability was used by researcher in this study. It is a procedure when making observation of behavior made by two or more individuals then compare the score to know either the score is almost similar or different (Creswell, 2012). Score rater reliability refers to the situations for which reliability must be investigated, such as essay test, oral test, and observation instrument (Gay, Mills, & Peter, 2011). To measure the speaking ability of the students, researcher used SOLOM

(Students' Oral Language Observation Matrix) as the scoring indicator (Jose, 2013) cited by (Nuraeni, 2014).

The result of inter-rater correlation can be seen in the table bellow:

Table 3.5
Correlations of Inter-rater

		Rater 1	Rater 2
Rater 1	Pearson Correlation	1	.903**
	Sig. (2-tailed)		,000
	N	67	67
Rater 2	Pearson Correlation	.903**	1
	Sig. (2-tailed)	,000	
	N	67	67

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

From the data above, it can be seen that the correlation coefficient between rater 1 and 2 is 0.903 with significant level at 0.01. It means there was very strong correlation between rater 1 and 2. It also indicates that data from speaking test were valid and reliable, therefore the researcher can use the data in this research.

E. Data Collection

There are two data collection used in this study. Before collecting the data, the researcher asked the permission to the debate coach and debaters in schools who has debate extracurricular. Firstly, students responded survey questionnaires consisted of "Yes" or "No" statements and 4 points of Likert scale (strongly disagree to strongly agree). "Yes" or "No" questionnaire related

to how intensive the students in practicing debate. The likert scale questionnaire related to how debate contributes students' critical thinking and the how debate contributes to students' self-confidence. After that, the researcher tested their speaking ability in order to measure their speaking ability.

F. Variable

Variables are the conditions or characteristics that researcher manipulates, controls, or observes (Best, 1981). Beside, variable can be defined as an attribute of a person or an object which 'varies' from the person to person or from object to object (Evelyn & Hossein, 1982). In this research, there were four variables which can be classified as:

1. Independent variable

Independent variable is the major of variable which researcher hopes to investigate. It is the variable which is selected, manipulated, and measured by the researcher. In educational research, an independent variable may be a particular teaching method, a type of teaching material, a reward, or a period of exposure to a particular condition (Best, 1981). In this study, debate was being the independent variable because it is the teaching strategy which researcher wants to investigate.

2. Intervening variable

In many types of behavioral research, the relationship between independent and dependent variables is not a simple one of stimulus to

response. Certain variables which cannot be controlled directly may have an important effect upon the outcomes. These modifying variables intervene between the cause and effect (Best, 1981). For example the study about oral fluency, some of variable have not been measured which may or may not be a part of that process such as intelligence, motivation, frustration, anxiety, and fatigue. That is intervening variable. In this research, there were two intervening variables, those were critical thinking and self-confidence because it might be a part that affected students' speaking skill.

3. Dependent variable

   Dependent variable is the variable which the researcher observes and measures to determine the effect of the independent variable. It can be the test score, number of errors, or measured speed in performing a task (Best, 1981). Speaking skill was the dependent variable in this research because the researcher aimed to know the contribution of debate as independent variable to speaking skill as dependent variable.

   Based on the explanation, it can be concluded that the independent variable (x) is debate, the intervening variables (y) are critical thinking and self-confidence, and the dependent variable (z) is speaking skill. The diagram of the research is as follows:

Diagram 3.1

Path Analysis Diagram
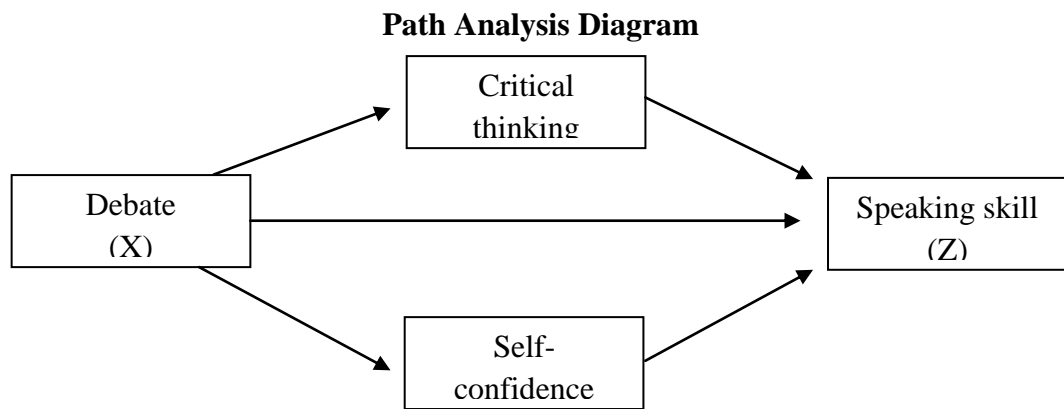
G. Data Analysis

In analyzing the data, the researcher employed SPSS 21 to gain the result of descriptive analysis and prerequisite test. Prerequisite test is divided into normality test, linearity test, heterocedasticity test, and linear regression. In linear regression, there is a part when the reseacher aimed at knowing the correlation between independent and dependent variables. Therefore, the parameter to determine whether the correlation is strong or not is needed. The table bellow shows the parameter of correlation (Machali, 2015).

Table 3.6

Coefficient Variable

No	Value (r)	Degree
1	0.00 – 0.199	Very weak
2	0.20 – 0.399	Weak
3	0.40 – 0.599	Sufficient
4	0.60 – 0.799	Strong
5	0.80 – 1.00	Very strong

To identify the direct and indirect contributions among the variables, the researcher utilized SEM (Structural Equation Model) using AMOS 24. It will assist the researcher to know the contribution through path diagram. The model of this path analysis was recursive model. It can be indicated if all arrows go toward same direction (Z). The requirements or the characteristics of this model are the arrows are going to one direction only, there is no arrows head in the reverse direction and there is only one exogenous variable and three endogenous variables. (Sarwono, 2011).