

BAB IV

FINDING AND DISCUSSION

In this research, the researcher presents the results of the study at MAN 1 Solok, focusing on the Tenth-grade students, with Class XE3 designated as the Experimental Group and Class XE5 as the Control Group. At the end of the study, both groups will be given an oral post-test to measure students' speaking achievement. After the data is complete, it will be analyzed using descriptive statistics and prerequisite analyses (Normality and Homogeneity Tests), and then further analyzed using the Independent Sample t-Test to determine the effect of the treatment.

A. FINDING

1. Description of Data

After conducting the research, two types of data were obtained: Treatment and post-test scores. Ten meetings were conducted using this technique. Students were given treatment before beginning their investigation. This technique was implemented over ten meetings. Students were given a post-test after each meeting. This study was conducted to determine the effect of using YouTube conversation videos on speaking skills at MAN 1 Solok. The goal was to determine speaking skill achievement.

The data analysis for this study is fundamentally derived from the students' scores on the post-test administered upon the conclusion of the treatment phase. A total of 55 students participated in this final oral test. The research sample was organized into two distinct groups: the Experimental Group,

comprising 28 students who received speaking instruction utilizing conversational videos, and the Control Group, consisting of 27 students who were taught using conventional methods. The scores obtained by both groups on the *post-test* constitute the primary dependent variable data used for hypothesis testing, which aims to determine the significance of the effect of the video-based intervention on students' speaking skill achievement.

The speaking test was similar, speaking results were also evaluated by considering five components based on Hughes (2003) namely Pronunciation, grammar, vocabulary, fluency, and Comprehension.

a. Treatment

The field intervention for this study was conducted in four separate face-to-face sessions, separated from the post-test period. This study adopted a no pre-test design, involving an Experimental Group (Class XE3) and a Control Group (Class XE5). The main focus of treatment in the Experimental Group was the application of Conversational YouTube Video media. Meanwhile, the Control Group (XE5) received the same material through conventional methods that did not provide intensive exposure to authentic conversational language models. This deliberate difference in treatment provides a strong basis for testing the hypothesis, namely whether Conversational YouTube Videos produce significant differences in the achievement of five aspects of speaking in the Experimental Group compared to the Control Group. The results of the speaking test were also evaluated by considering five components based on Hughes (2003), namely Pronunciation, grammar, vocabulary, fluency, and comprehension.

Table 4. 1
Distribution Data Treatment I
Descriptive Statistics

	N	Range	Min	Max	Sum	Mean	Std. Deviation	Variance
Treatment Experiment	28	29	51	80	1930	68.93	7.557	57.106
Treatment Control	27	37	38	75	1550	57.41	10.135	102.712
Valid N (listwise)	27							

The results of the descriptive statistical analysis in Treatment One showed a difference in performance between the experimental and control groups. The Treatment Experiment group (N=28) recorded a higher mean score of 68.93. Meanwhile, the Treatment Control group (N=27) only Variance 57.41. This 11.52-point difference in mean points explains the positive impact of the treatment intervention on the experimental group. In addition, the experimental group data showed a lower level of variability, with a Standard Deviation (SDD) of 7.557 and a Variance of 57.106. In contrast, the control group showed a wider spread of data, with a Standard Deviation of 10.135 and a Variance of 102.712. The results of the data explain that the treatment given in Treatment 1 not only improved the average performance but also produced data that was more concentrated around the mean, implying better consistency of results in the experimental group.

Table 4. 2
Distribution Data Treatment II
Descriptive Statistics

	N	Range	Min	Max	Sum	Mean	Std. Deviation	Variance
Treatment Experiment	28	29	51	80	1918	68.50	6.963	48.481

Treatment Control	27	37	40	77	1516	56.15	9.554	91.285
Valid N (listwise)	27							

Descriptive data analysis for Treatment two shows a difference in performance between the two groups. The Treatment Experiment group (N=28) recorded an average (Mean) of 68.50, indicating a higher score than the Treatment Control group (N=27) mean of 56.15. This average difference of 12.35 points is a strong indication of the effectiveness of the implemented intervention. Then the experimental group showed a very consistent data distribution, seen in the low Standard Deviation (Std. Deviation) value (6.963) and Variance of 48.481. In contrast, the control group showed much higher variability than in treatment 1, with a Standard Deviation of 9.554 and Variance of 91.285. The high spread of data in the control group is reinforced by the low Minimum value (40) and large Range (37). Overall, the results of Treatment 2 confirm that the average values of treatments 1 and 2 have decreased. Although the mean scores decreased slightly, there was a clear improvement in terms of consistency of results (lower data variability) in both groups, especially in the experimental group, from Treatment One to Treatment Two.

Table 4. 3

Distribution Data Treatment III

Descriptive Statistics

	N	Range	Min	Max	Sum	Mean	Std. Deviation	Variance
Treatment Experiment	28	30	48	78	1790	63.93	9.063	82.143
Treatment Control	27	37	39	76	1570	58.15	10.432	108.823
Valid N (listwise)	27							

Descriptive data analysis for Treatment Three shows that the performance difference between the experimental and control groups was the narrowest among the four treatments. The Experimental Treatment group (N=28) recorded a Mean of 63.93, which, while still higher, was not significantly different from the Control Treatment group's (N=27) mean of 58.15. The mean difference was only 5.78 points, indicating that the intervention in this treatment had the weakest impact in separating the two groups. Furthermore, Treatment 3 was characterized by a high level of data variability across both groups. The control group exhibited the highest level of dispersion (Variance 108.823 and Standard Deviation 10.432) among all the control groups. The experimental group also recorded relatively high variability, with a Variance of 82.143 and Standard Deviation 9.063. This high dispersion was reinforced by the control group's very low Minimum (39) and wide range (37). Overall, the results of Treatment 3 imply that the treatment intervention still produced higher average scores, but its effectiveness in improving performance and maintaining consistency of results was much reduced compared to the previous treatment.

Table 4. 4

Distribution Data Treatment IV

Descriptive Statistics								
	N	Range	Min	Max	Sum	Mean	Std. Deviation	Variance
Treatment Experiment	28	29	57	86	1950	69.64	8.319	69.201
Treatment Control	27	32	44	76	1519	56.26	9.083	82.507
Valid N (listwise)	27							

Table 4.5

[illegible][illegible]

The descriptive statistics, calculated using the SPSS 27 software, present a preliminary overview of the speaking performance in both the experimental and control groups. In the Experimental Group (N= 28), the final scores ranged from a minimum of 44 to a maximum of 86, resulting in a range of 42. The group achieved a total score (Sum) of 1815, resulting in a mean score of 64.82. The variability within this group was measured by a Standard Deviation of 11.092 and a Variance of 123.041.

Table 4. 6
Students' Speaking Scores from Post-test of the Experimental Class

NO	NAMA SISWA	COMPONENT OF SPEAKING					
		P	G	V	F	C	
1	AFISA FAWZIYAH ADRIL	2	30	20	8	15	75
2	AGNIYATUN HANISYA	2	18	12	6	15	53
3	AIZZA AHLANA FAUZIA	3	18	12	10	23	66
4	ALLYA DILL ADHA	2	18	16	8	23	67
5	AMALATUL QUR'ANI	2	24	20	6	8	60
6	AMEGA SALSABILA	2	18	12	6	6	44
8	AYRIN SAPUTRI	2	18	16	8	19	63
9	BIMA CAHYA DININGRAT	2	24	20	6	19	69
10	BRAINLY AL FATHAN	2	18	16	6	23	61
11	CAHAYA FITRI	2	18	20	6	12	58
12	CHACA OKTAVIA	2	24	20	8	19	73
13	DINDA OKTRI RAMADHANI	2	24	20	6	19	71
14	FAIQ FADHUL AZIZ	2	12	12	6	15	47
15	GHAYS ALKHAIRI	3	30	20	10	23	86
16	HILMA AULIA	2	24	20	8	23	77
17	INTAN PERMATA SARI	2	18	16	8	23	63
18	IZZA ANANDA	3	18	16	10	23	70
19	KHANZA AZDKANISA	2	18	12	8	12	52
20	NADHIRA AMALIA	3	24	20	10	23	80

21	NADHIRAH ZALITA	2	18	16	8	15	59
22	NAJAHUSTSANI KHAIR	2	12	16	8	19	57
21	NIA RAMADHANI	2	18	12	8	8	48
23	NURUL FAUZIAH	3	30	20	10	23	86
24	OLVIANDA MAYRETA	3	24	16	10	23	76
25	RAHMATUL HUSNA	2	18	20	6	19	65
26	RAUDHATUL MARDIAH	2	18	16	8	19	63
27	RIRIN MIFTAHUL HAYATI	2	24	16	6	23	71
28	SYAUQIA INAYAH Y.	2	18	12	8	15	55

Analysis of the post-test results in the experimental class showed that the group consisted of 28 students with a total score of 1815, resulting in an average score of 64.82. In the data obtained, the highest score achieved was 86, while the lowest score was 44. These results reflect the impact of the application of peer tutoring techniques on improving students' abilities.

The assessment components show that the Accent (Pronunciation) aspect obtained an average score of 2.21. Furthermore, the Grammar component recorded an average score of 20.57, followed by the Vocabulary component with an average of 16.57. The Fluency component showed an average score of 7.71, and finally, the Comprehension component achieved an average of 18.11.

Based on the data above, the highest component value obtained from the students' speaking scores in the experimental class is the Grammar component (20.57). Conversely, the lowest score was found in the Pronunciation component (2.21). Post-test score data of the experiment class can be seen as follows:

Figure 4. 1
Histogram of Experiment Class

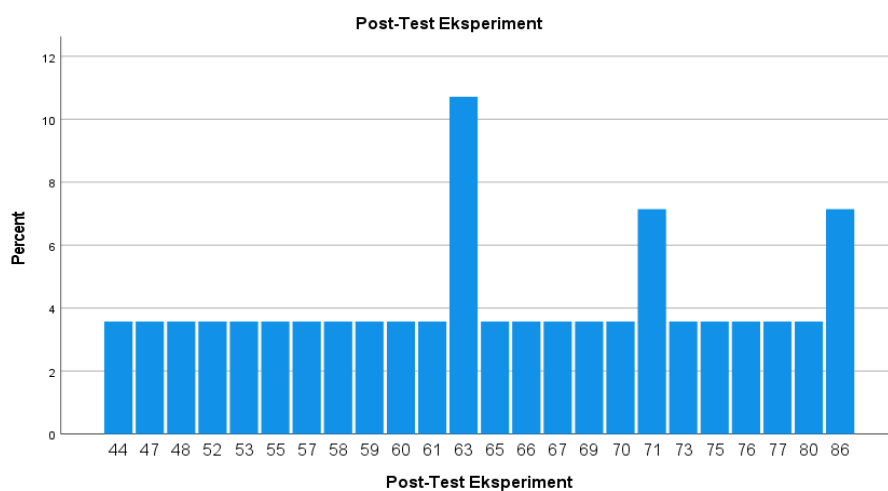


Table 4. 7
Students' Speaking Scores from the Post-test of the Control Class

NO	NAMA SISWA	COMPONENT OF SPEAKING					SCORE
		P	G	V	F	C	
1	ALMIRA NABILA	2	18	16	6	19	61
2	ARIYA ISKANDAR KELVIANSYAH	2	12	8	6	12	40
3	AULIA PRIANTI	2	24	16	4	23	69
4	AURA FATHARANI	3	26	16	8	15	68
5	CELSILLA SEBRIMA	2	12	8	4	15	41
6	DAFFA APRIANO	2	18	12	10	12	54
7	DAFFA WENFINDRA	3	24	16	12	23	78
8	DRISKA PUTRA	2	18	8	8	8	44
9	FADILAH MUHAMMAD DANI	2	12	8	6	15	43
10	FEBRIANTO HERMAN	2	18	12	6	4	44
11	GIFARI ZAKA WALI	2	18	12	6	19	48
12	HAIKAL AZMI	2	12	8	4	12	38
13	HANAFI YAHYA	2	18	12	6	15	53
14	MEY CELSA ANANDA PUTRI	3	24	16	10	19	72
15	MEY CELSI ANANDA PUTRI	2	18	16	6	8	42
16	MUHAMMAD SHENDY RIZALDI	2	18	12	6	4	42

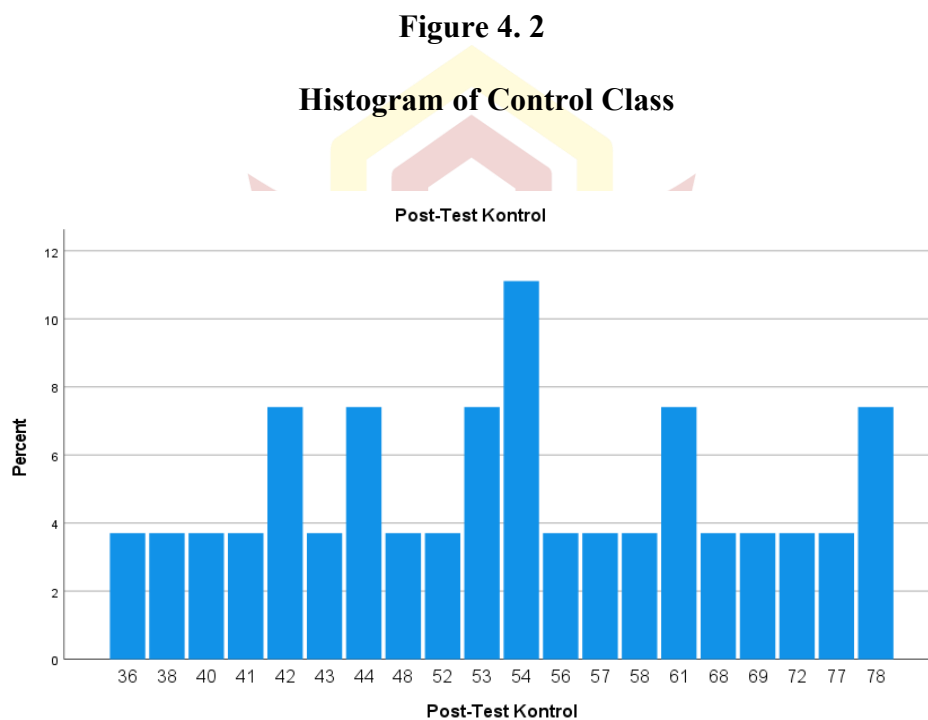
17	NADIL FAUZI	2	18	12	6	15	53
18	NAJWA INAYAH ARRASY	3	18	12	10	15	58
19	NASYAMA FALILLAH	3	12	12	10	19	56
20	NUR ASIA FADHILAH	2	18	16	6	19	61
21	RAINA ATHAYA	2	18	8	6	19	54
22	RILMA YUHENDRA	2	12	12	6	4	36
23	RISMAL GANI	2	18	12	6	19	57
24	TASRIL MULYADI	3	12	8	8	15	54
25	USMAN ALI	2	24	16	6	4	52
26	ZILDATUL HUSNA	2	24	20	8	23	77
27	ZIVANA AULIA FERROZA	3	24	16	12	23	78

The descriptive statistics for the Control Group (N=27) recorded a minimum score of 36 and a maximum score of 78, also resulting in a range of 42. The control group's total score was 1473, with a mean score of 54.56, and demonstrated higher variability, reflected by a Standard Deviation of 12.620 and a Variance of 159.256. This initial descriptive analysis indicates a noticeable difference in the final mean scores, suggesting a preliminary positive effect of the conversational video intervention on the experimental class's speaking achievement.

The Accent (Pronunciation) component registered a mean score of 2.26. Conversely, the Grammar component recorded the highest average score among all aspects at 18.07, followed by the Comprehension component, which achieved a mean of 14.74. The Vocabulary component obtained an average score of 12.59, while the Fluency component exhibited the lowest score at 7.11.

Based on these findings, the students' greatest relative strength in the Control Group is found in their Grammar foundation (18.07), suggesting they

possess a reasonable command of linguistic structure. In sharp contrast, the lowest scores were registered in the areas demanding mechanical mastery and spontaneous delivery: Pronunciation (2.26) and Fluency (7.11). This markedly low fluency score underscores the significant difficulty students experienced, in the absence of the intervention, with speaking smoothly and without frequent hesitation, which collectively contributed to the lower overall Post-Test score achieved by this group.



2. Data Analysis

After completing the Descriptive Statistics analysis, the next stage in the data processing process is the implementation of the Data Analysis Prerequisite Test. This step ensures that the data obtained has met the statistical assumptions underlying further inferential analysis, thus guaranteeing the validity and

reliability of the research findings. The prerequisite tests carried out include the Normality Test and the Homogeneity Test. To support this statistical computation, the researcher utilized SPSS version 27 software.

Specifically, the Normality Test aims to verify that the dependent variable data is normally distributed, while the Homogeneity Test functions to test the assumption of equality of variance between groups, which is an absolute requirement for the application of parametric *t*-Test analysis.

3. Normality Test

The Normality Test is a fundamental statistical procedure performed to verify the assumption that the dependent data score variable is normally distributed. The implementation of this prerequisite test is essential to ensure the validity of the results of subsequent inferential statistical analysis. In the context of testing, there are two main methods commonly used: Kolmogorov-Smirnov and Shapiro-Wilk. The selection method is based on the size of the research sample. The Kolmogorov-Smirnov test is more appropriately applied to large samples ($N > 50$), while the Shapiro-Wilk test is considered more effective and recommended for small samples ($N < 50$), according to references from Biu, et al. (2019) in Ahadi & Zain (2023).

Considering the total number of samples in this study was 55 students (a combination of the experimental and control groups), the researchers decided to use the Shapiro-Wilk Test to emit data normality. The criteria for determining normal distribution are if the resulting Significance (Sig.) value is greater than 0,05 ($p > 0.05$), then the data is declared normal; conversely, if the Sig. value is

greater than $\leq 0,05$, data transmission is abnormal. The test results are presented as follows:

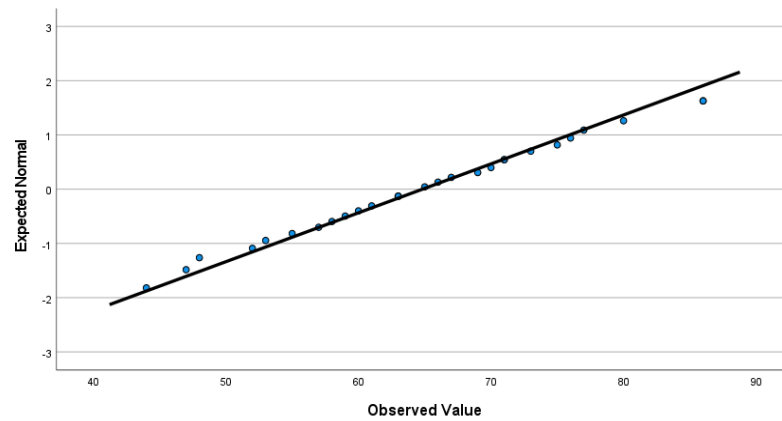
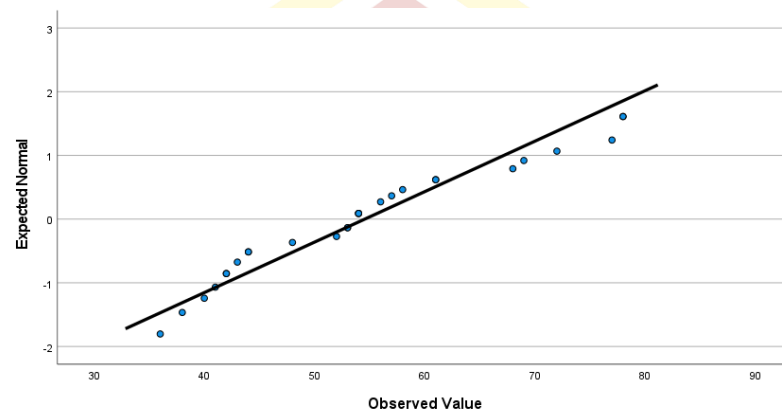
Table 4. 8
Test of Normality

	Kelas	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	Df	Sig.
Speaking Score	Post Test Experiment	.065	28	.200*	.985	28	.942
	Post Test Control	.132	27	.200*	.933	27	.084

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction.

Based on the Test of Normality Table presented, the significance value (Sig.) of the Post-Test for the Experimental Group using the Shapiro-Wilk test is 0,942 This value is greater than 0,05 ($0.942 > 0.05$), which means that the Experimental Group Post-Test data is normally distributed. Meanwhile, the significance value (Sig.) of the Post-Test for the Control Group also using the Shapiro-Wilk test is 0.084. This value is also greater than 0.05 ($0.084 > 0.05$), so that the Control Group Post-Test data is also declared normally distributed. Thus, because all the significance values of the Shapiro-Wilk test in both groups are above 0.05, it can be concluded that the assumption of data normality has been met.

Figure 4. 3**Normal Q-Q Plot Class Post-test Experiment Class****Figure 4. 4****Normal Q-Q Plot Class Post-test Control Class**

As a complementary and reinforcing step to the numerical tests (Shapiro-Wilk and Kolmogorov-Smirnov), the normality analysis was also confirmed through the Q-Q Plot (Quantile-Quantile Plot) visualization for both groups. Based on the graphical observation, it can be observed that the Post-Test score data points in both the Experimental and Control Groups appear to follow and are tightly distributed around the diagonal line (reference line). The distribution that approaches this line indicates that the frequency of the empirical data does not show significant deviations from the theoretical normal distribution. Thus, the Q-

Q Plot visualization results strengthen the conclusion that the Post-Test data from both groups are normally distributed, thus fulfilling the prerequisite assumptions for proceeding to the Homogeneity Test and Hypothesis Test (t-Test).

4. Homogeneity Test

After obtaining the normality test results, the next step is the homogeneity test. The homogeneity test is conducted to determine whether the data from both groups have the same level of variance. To test the homogeneity of variance between the experimental class receiving the video conversational treatment and the control class receiving the conventional treatment, Levene's statistical test was used. The results of this calculation are presented below:

Table 4. 9
Homogeneity Test

		Levene Statistic	df1	df2	Sig.
Speaking Skill	Based on Mean	.345	1	53	.560
	Based on Median	.274	1	53	.603
	Based on Median and with adjusted df	.274	1	51.614	.603
	Based on trimmed mean	.307	1	53	.582

Based on the homogeneity test table, the significant value for the homogeneity test of students' speaking scores in the post-test is 0.560. This indicates that the post-test data is homogeneous, because the value exceeds 0.05, which is the significant level determined for the homogeneity test.

5. Hypothesis Testing

After conducting the Normality Test and Homogeneity Test, the next step is to analyses the data using the T-test (Independent Sample T-test). The T- test

aims to determine whether there is a significant difference between the average (mean) of two independent data groups, namely the experimental class and the control class.

The main focus of this study is to answer the research question, namely: "Is there a significant influence of the use of Conversational YouTube Video media on students' speaking skills at MAN 1 Solok. In the Experiment class, researchers used conversational video media to improve students' speaking skills, while in the control class, researchers used conventional methods. The independent sample t-test was applied to determine whether there was a significant effect on students' speaking skills in the experimental class compared to the control class. The results of the t-test from the research data can be seen in the following table:

Table 4. 10
Group Statistics

	Class	N	Mean	Std. Deviation	Std. Error Mean
Speaking Score	Class Experiment	28	64.8214	11.09239	2.09626
	Class Control	27	54.5556	12.61968	2.42866

Table 4. 11
Independent Sample Test

Speaking Score	Levene's test for equality of variance	T-test for equality of means	
----------------	--	------------------------------	--

	F	Sig	T	Df	Sig(2-tailed)	Mean difference	Std. Error difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.345	.560	3.207	53	.002	10.26587	3.20061	3.84626	16.68548
Equal variances not assumed			3.200	51.595	.002	10.26587	3.20822	3.82690	16.70485

Based on the results of inferential statistical analysis, there is a significant difference in the results of speaking skills between the group of students who received the Conversational YouTube Video media treatment and the group taught using conventional methods. Descriptively statistically, the results of the Experimental Group Post-Test showed an average score of 64, 82 with a standard deviation of 11.092. This figure is superior compared to the average score of the Control Group which was recorded at 54,56 with a standard deviation of 12,620. To verify the significance of this difference, the Independent Sample t Test was used. The decision-making criteria were based on a comparison of the Significance value (Sig. 2-tailed) with a significance level α of 0, 05. From the analysis carried out, the Sig. value. The 2-tailed obtained is 0, 002. Thus, because the Significance value of 0, 002 is substantially smaller than 0.05 ($p < \alpha$), the Null Hypothesis (H_0) is rejected and the Alternative Hypothesis (H_a) is accepted. These

results definitively indicate that there is a significant influence of the use of Conversational YouTube Video media on students' speaking skills.

In conclusion, Conversational YouTube Video media is proven to have a positive and significant effect on improving students' speaking skills at MAN 1 Solok. This significant influence was observed in students' speaking skills in the Experimental Group compared to the group taught conventionally.

B. Discussion

The results of the research analysis show a significant difference in speaking skills between the group of students taught using Conversational YouTube Video media and the group taught conventionally. Descriptive statistical data supports this finding, where the average value of the Experimental Group's Post-Test reached 64, 82, far exceeding the average of the Control Group which was only 54,56. Verification through the Independent Sample *t* Test produced a Significance value (Sig. 2-tailed) of 0.002. Because the 0, 002 value is smaller than the 0, 05 significance level ($0, 002 < 0, 05$), the Alternative Hypothesis (H_a) is accepted, which clearly indicates that the use of Conversational YouTube Video media provides a positive contribution and has a significant influence on improving students' speaking skills at MAN 1 Solok.

This significant effect was manifested in an increase in average scores across all aspects of speaking ability evaluated, including grammar, vocabulary, fluency, and comprehension. Students who received video conversation instruction demonstrated superior speaking ability compared to students who

received only conventional instruction. The significant effects on each component of speaking ability are further described below.

1. Accent (Pronunciation)

Pronunciation is an essential component of speaking skills. This component refers to an individual's articulation and production of words during communication, which plays a crucial role in conveying meaning and helping listeners understand the message (Ariyanti, 2022). Based on the data obtained from this study, findings regarding pronunciation showed a different pattern from the overall results. The average post-test score for the experimental group (using conversational YouTube videos) was recorded at 2.21, while the average score for the control group (using conventional methods) was slightly higher, at 2.26.

2. Grammar

Grammar is a fundamental aspect of speaking ability because it includes the rules on how words are changed and combined to produce accurate and meaningful sentences. Good grammar mastery is very important to ensure the accuracy of the message (accuracy) delivered. Based on the analysis of the Post-Test scores, the Experimental Group showed an average score on the grammar component of 20.57, while the Control Group recorded an average score of 18.07. There is a clear positive difference of 2.50 points in favor of the Experimental Group.

3. Vocabulary

Vocabulary is a key element in speaking ability, defined as the set of words known and used by a person to communicate ideas effectively. The

richness of vocabulary directly affects the variety of expressions and the quality of communication. Based on the results of the Post-Test analysis, Vocabulary is the component that showed the most significant improvement among the five aspects assessed. The average score of the Experimental Group (which used Conversational YouTube Video media) reached 16.57, far exceeding the average score of the Control Group which was only 12.59. This average difference of 3.98 is a strong indication of the effectiveness of the media used.

4. Fluency

Fluency is defined as the speaker's ability to speak smoothly, easily, and quickly, maintaining the flow of communication without disruptive pauses, excessive repetition, or difficulty in formulating ideas. Fluency reflects cognitive and motor efficiency in language use. Based on Post-Test data, the average fluency score of the Experimental Group was 7.71, which was slightly higher than the Control Group which obtained an average of 7.11. This average increase of +0.60 points, although positive and in line with the hypothesis, was the smallest increase among the other speaking skill components that showed superiority.

5. Comprehension

Comprehension is defined as the speaker's ability to accurately process and interpret information received from the interlocutor, allowing for appropriate and relevant responses. This aspect is an important indicator that two-way communication has been successful. Based on the Post-Test score analysis, the Experimental Group showed an average comprehension score of 18.11, significantly higher than the Control Group which only obtained 14.74. This

positive difference of 3.37 makes it the second highest increase among all components, after Vocabulary.

The findings are supported and expands previous research by Lizeth Gonzales Cabanas and Patricia Nunez Mercado (2024) with research title Using YouTube Videos to Improve Students' Speaking Skills: EFL Teachers' Perceptions, it can be concluded that the use of authentic videos from YouTube has positive results in improving students' speaking skills in English. The results of the study indicate that YouTube videos not only function as a means of entertainment, but also as a source of material rich in cultural context and natural language usage, which supports the improvement of skills in the aspects of pronunciation, vocabulary, and grammatical structure. Educators see that this platform can minimize student anxiety levels and increase student motivation, thanks to the content displayed being more interesting and related to everyday life. Apart from the challenges in determining materials that are appropriate to students' ability levels and the constraints of facilities and infrastructure, this study states that YouTube plays a very effective role in creating an interesting learning environment and increasing student confidence in communication.

Rai Ayu Wilena Dewi and her colleagues (2025) in an article entitled "The Effectiveness of YouTube Video to Improve English Speaking Skills of Eighth Grade Students at SMP Negeri 19 Palu" showed effectiveness through a quasi-experimental study that the use of YouTube videos was significantly more effective than conventional approaches in optimizing students' speaking skills. This study confirmed that the coherence of YouTube video content had a

significant contribution to students' fluency and comprehension, as indicated by the average score of the experimental group which was statistically higher than the control group. This finding confirms the view that the use of YouTube-based audiovisual media not only attracts students' attention but also serves as a valid tool for achieving measurable improvements in speaking competence at the secondary school level.

Febry Siswanto and colleagues (2023) in their article entitled "Designing Transactional Conversation Videos as Supplementary Materials to Support Teaching English Speaking Skills" underscore the urgency of developing specific audiovisual materials to support oral proficiency. The study demonstrated that the implementation of transactional dialogue videos structured based on the ADDIE (Analysis, Design, Development) model was effective in accommodating students' practical communication needs. The results of this development demonstrate that the videos function not only as a means of entertainment but also as crucial supplementary materials to provide accurate language models, increase confidence in conducting structured conversations, and facilitate interactive independent learning for students.

Overall, this study confirms that the use of conversational videos is significantly effective in improving students' speaking skills, as evidenced by the positive results of the post-test hypothesis test. Although when conducting the treatment four times, the average score was recorded higher than when conducting the post-test. This can be seen in table 4.5 where the mean obtained was an average of 64.82 compared to the fourth treatment which obtained the highest

average score of 69.64. However, the post-test was designed to test speaking skills comprehensively and generally (global skills) with broader topics or topics that have never been practiced directly, which requires students to generalize their knowledge. Therefore, the decrease in scores from treatment to post-test reflects students' challenges in transferring knowledge (transfer of learning) from a specific context to a broader context, even though statistically the treatment still succeeded in significantly improving their initial abilities.

