

CHAPTER IV

FINDINGS AND DISCUSSION

This chapter consist of two parts, such as the finding of research and the discussion of the research. The finding of the research cover the description of the result of data collected through test that can be seen below.

A. Findings

1. Data Description

The findings of this research deal with the classification of the students' pre-test and post-test. To find out the question in the previous chapter, the researcher gave a test that was given twice. A pre-test was given before treatment to know the students' vocabulary mastery, while post-test was given after treatment through Velcro Dart Game and the result of the post-test of this research can answer the question of this research that aims to find out through Velcro Dart games would be able to increase the students' vocabulary mastery.

a. The Students' Score in Pre-Test and Post-Test

1. Experimental Class Result

The pre-test had done before giving concept mapping strategy. It was conducted on Wednesday, january. The students were given a pre-test. The researcher found out the result of the students pre-test based on the scoring of vocabulary before giving a treatment through Velcro Dart Game during

study which were analyzed and resulted in the information as shown in the following table:

Table. 1.1 Pre-Test and Post-Test Scores of the Experimental Group

No	Students	Pre-Test	Post-Test	Gain	X^2
		X_1	X_2	X	
1	S1	54	66	12	144
2	S2	43	61	18	324
3	S3	54	71	17	289
4	S4	53	66	13	169
5	S5	40	53	13	169
6	S6	30	55	25	625
7	S7	40	66	26	676
8	S8	46	63	17	289
9	S9	30	66	36	1296
10	S10	79	87	8	64
11	S11	43	64	21	441
12	S12	25	38	13	169
13	S13	40	74	34	1156
14	S14	64	79	15	225
15	S15	90	92	2	4
16	S16	43	71	28	784
17	S17	33	71	38	1444
18	S18	46	76	30	900
19	S19	33	40	7	49

20	S20	20	40	20	400
21	S21	40	53	13	169
22	S22	28	58	30	900
23	S23	66	71	5	25
24	S24	48	76	28	784
25	S25	51	71	20	400
26	S26	70	87	17	289
27	S27	35	40	5	25
Amount	27 Students	$\Sigma X_1 = 1244$	$\Sigma X_2 = 1755$	$\Sigma X = 511$	$\Sigma X^2 = 12209$
Mean		46.1	65	18.9	452.19

ΣX_1 = The Total Pre-test score of students in the Experimental Class

ΣX_2 = The Total Post-test score of students in the Experimental Class

ΣX = The Total gained score of students in the Experimental Class

ΣX^2 = The Square of the total gained score of students in the Experimental Class.

2. Control Class Result

The table above shows that the experimental class had distinct pre-test and post-test scores. Pre-test scores ranged from 20 to 90, with an average of 46.1. Meanwhile, the post-test average was 65 points, with the lowest score of 40 and the highest score of 92. The writer received $\Sigma X_1 =$ The overall pre-test score was 1.244, while the total post-test score was 1755 ($\Sigma X = 511$). The square of the total obtained score (ΣX_2) is equal to 1.075. This score results will then be used to determine the t-test in data analysis.

Table. 1.2 Pre-Test and Post-Test Scores of the Control Group

No	Students	Pre-Test	Post-Test	Gain	X^2
		X_1	X_2	X	
1	S1	38	59	21	441
2	S2	41	69	28	784
3	S3	48	53	5	25
4	S4	30	30	0	0
5	S5	46	46	0	0
6	S6	51	72	21	441
7	S7	38	92	54	2916
8	S8	51	82	31	961
9	S9	82	82	0	0
10	S10	46	67	21	441
11	S11	30	30	0	0
12	S12	35	35	0	0
13	S13	46	85	39	1521
14	S14	36	71	35	1225
15	S15	59	59	0	0
16	S16	87	95	8	64
17	S17	55	63	8	64
18	S18	51	82	31	961
19	S19	23	28	5	25
20	S20	36	59	23	529
21	S21	69	79	10	100

22	S22	35	35	0	0
23	S23	20	51	31	961
24	S24	31	69	38	1444
Amount	24 Students	$\Sigma X_1 = 1084$	$\Sigma X_2 = 1454$	$\Sigma X = 370$	$\Sigma X^2 = 11501$
Mean		45.1	60.5	15.4	479.2

ΣY_1 = The Total Pre-test score of students in the Controlled Class

ΣY_2 = The Total Post-test score of students in the Controlled Class

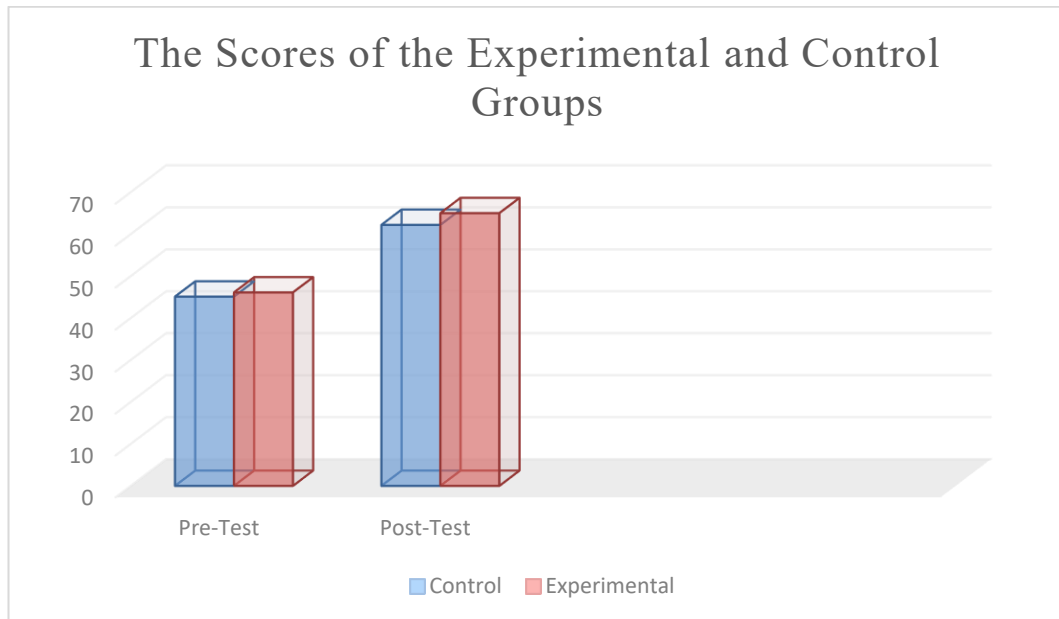
ΣY = The Total gained score of students in the Controlled Class

ΣY^2 = The Square of the total gained score of students in the Controlled Class

According to the information in the table above, it indicated that the regulated class scores from the pre-test and post-test varied. The mean for the pre-test was 45.1, with a minimum score of 20 and a maximum score of 87. Additionally, the average score of the post-test was 60.5, with the minimum score being 20 and the maximum score reaching 95. The researcher obtained $\Sigma Y_1 = 1084$ as the overall pre-test score, $\Sigma Y_2 = 1454$ as the overall post-test score, $\Sigma Y = 370$ for overall accumulated score and $\Sigma Y^2 = 1.293$ for the squared score achieved. This outcome will be utilized to determine the t-test. As described earlier, the mean score of both the experimental class and the control class saw a rise. Nonetheless, the score of the experimental group had rose a bit more than the control group. This can be seen through the range points between the two groups. The experimental group achieved

511 points while the control group achieved 370 points. The advancement of the two groups is evident as shown in diagram below.

Table. 1.3 Average Scores of the Experimental and Control Groups



2. Data Analysis

In accordance with the aforementioned collected data, the researcher examines the test scores of both the experimental and control groups through the application of the t-test formula. Prior to employing the t-test, it is imperative to ascertain the normality and homogeneity of the set data. The normality assessment is essential to determine if the data exhibit a normal distribution. Following the normality evaluation, a homogeneity test is conducted to establish whether the data are homogeneous or share a common type/format.

a. Normality Test

The normality test constitutes one of the prerequisites that must be satisfied prior to the execution of the t-test. The objective of the normality test is to ascertain whether the datasets from the two groups are normally distributed. The researcher utilized the Lilliefors table to conduct the normality assessment. The data is deemed normally distributed if it achieves a significance level of 5% (0.05), whereby L_{table} surpasses $L_{observe}$. The findings are presented on the table below.

Table. 1.4 Normality Test Results

Statistik	Pre-Test		Post-Test	
	Experimental Group	Control Group	Experimental Group	Control Group
Sample (N)	27	24	27	24
Mean	46.1	45.1	65	62.2
L_{table}	0,168	0,176	0,168	0,176
$L_{observe}$	0,131	0,154	0,105	0,118
Conclusion	Normal		Normal	

b. Homogeneity Test

Following the normality test, a homogeneity test was conducted to determine whether the variance across both groups was equal. The criteria for

homogeneity is met if the F_{table} value exceeds the $F_{observed}$ value at a significance level of $\alpha = 0.05$. The findings are presented below:

Table. 1.5 Homogeneity Test Results

Statistik	Pre-Test		Post-Test	
	Experimental Group	Control Group	Experimental Group	Control Group
Sample (N)	27	24	27	24
Varians	273,610	276,319	210,077	411,216
L_{table}	1,988	1,988	1,988	1,988
$L_{observe}$	1,010		1,957	
Conclusion	Homogen		Homogen	

From the collected data on the table above it shows that the F_{table} was 1.988. $F_{observe}$ on pre-test is 1.010 while $F_{observe}$ after post-test is 1.957. Since the F_{table} value exceeded the $F_{observed}$ value, the pre-test and post-test data for both groups were confirmed to be **homogeneous**.

c. Hypothesis Test

Once normality and homogeneity tests are complete, a t-test is applied to determine if there is a statistically significant difference between the experimental and control groups. The analysis utilizes the t-test formula at a 5% significance level, following the steps outlined below:

1. Determine mean of Variable 1 and 2:

$$M_1 = 65$$

$$M_2 = 62.2$$

2. Determine Standard Deviation of Variable 1 and 2:

$$SD_1 = \sqrt{210.077} = 14.494$$

$$SD_2 = \sqrt{411.216} = 20.278$$

3. Determine Standard Error Mean Variable 1 and 2 with formula:

$$SEM_1 = \frac{SD_1}{\sqrt{N_1 - 1}} = \frac{14.494}{\sqrt{27 - 1}} = 2.842$$

$$SEM_2 = \frac{SD_2}{\sqrt{N_2 - 1}} = \frac{20.278}{\sqrt{24 - 1}} = 4.228$$

4. Determine standard error of difference of mean of variable 1 and 2:

$$SE_{M1-M2} = \sqrt{SEM_1^2 + SEM_2^2}$$

$$SE_{M1-M2} = \sqrt{2.842^2 + 4.228^2} = 5.094$$

5. Determining t_o :

$$t_o = \frac{(M_1 - M_2)}{SE_{M1-M2}} = \frac{(65 - 62.2)}{5.094} = 0.550$$

6. Determining t_{table} in significant level 5%, with degree of freedom:

$$df = (N_1 + N_2) - 2 = (27 + 24) - 2 = 49$$

To ensure the validity of the research findings, the researcher established the degrees of freedom (df) at 49. This calculation is derived from the total sample size adjusted for the specific statistical parameters of the test. A standard significance

level of 5% ($\alpha = 0.05$) was applied, which, according to the standard t-distribution table, results in a critical T_{table} value of 2.010.

The hypothesis testing process involved a direct comparison between the observed t-value (t_o) and the critical table value (t_{table}). This comparison determines whether the differences in student performance are statistically significant or simply the result of random chance. The testing criteria were defined as follows:

- Alternative Hypothesis (H_a): Accepted only if $t_o > t_{table}$ ($0.550 > 2.010$).

Upon analyzing the data, it was found that the t_o value of 0.550 is significantly lower than the T_{table} requirement of 2.010. This indicates that the statistical evidence is insufficient to reject the null hypothesis. Consequently, the Alternative Hypothesis (H_a) is rejected. Substantively, this result implies that there is no statistically significant difference between the Post-Test results of the experimental group and the control group.

d. Effect Size Analysis

While the t-test failed to show statistical significance, the researcher proceeded with an Effect Size analysis using Cohen's d formula. This step is crucial because, unlike p-values which are heavily influenced by sample size, the effect size provides a pure measure of the practical magnitude of the treatment's impact in a real-world classroom setting.

The calculation was performed by dividing the difference between the mean scores of both groups by the Pooled Standard Deviation:

$$d = M_1 - M_2 / SD_{\text{pooled}} = 65 - 62.2 / 17.56 = 0.156$$

The calculation resulted in a value of 0.156. Because this score falls below the 0.21–0.50 interval, it indicates that the Velcro Dart Game had a low but still meaningful effect on the students' vocabulary skills.

The final calculation yielded a value of 0.16 (rounded from 0.159). When measured against Cohen's standard criteria, this value falls below the modest interval of 0.21–0.50. Therefore, the data suggests that the use of the Velcro Dart Game has a low effect on the development of students' vocabulary skills. While the influence is statistically "small," it still represents a positive contribution to student learning, albeit one that does not drastically separate itself from traditional methods.

3. Students' Difficulties During Learning Activities while using Velcro Dart Game

The second research question of this study aims to identify the difficulties that arise during students' learning activities involving the use of the Velcro Dart game. Based on the researcher's direct observation conducted across four treatment sessions in the experimental class, several recurring difficulties were identified. These difficulties are described as follows:

a. Lack of Attention During Peers' Practice

One of the most prominent difficulties observed during the treatment sessions was the tendency of students to lose focus when it was not their turn to throw the dart. While one student was actively engaging with the Velcro Dart game at the front of the class, a considerable number of their classmates did not pay sufficient attention to the activity taking place. Instead of observing and learning from their peer's performance — which is an integral part of cooperative and observational learning — many students became disengaged and diverted their attention to other things. This passive behavior reduced the overall learning opportunity that the game was intended to provide, as the vocabulary exposure during each student's turn was not fully absorbed by the rest of the class.

b. Unproductive Peer Reactions

A second difficulty that emerged was the presence of unproductive peer reactions during the activity. When a student failed to hit the target accurately — a common occurrence given the physical skill required in dart throwing — some classmates responded by jeering or mocking the student rather than offering encouragement. This kind of negative social dynamic created an uncomfortable atmosphere for students who were already less confident. According to Krashen's Affective Filter Hypothesis, when students experience anxiety, embarrassment, or a lack of self-confidence, a psychological barrier is raised that actively blocks language acquisition (Krashen, 1982). This mocking behavior therefore did not

only affect the emotional well-being of the individual student but also potentially suppressed overall vocabulary learning outcomes.

c. Insufficient Attention During Vocabulary Explanation

A third difficulty was observed in the moments following a student's successful turn at the board. When a student managed to hit a target and was asked to answer a vocabulary question — with the teacher providing additional explanation at the front of the class — the remaining students showed a lack of attentiveness toward the instructional content being delivered. Rather than listening to the vocabulary explanation, many students appeared more focused on anticipating whose turn would come next. This means that the primary educational objective of the game — vocabulary acquisition through repeated exposure and teacher explanation — was overshadowed by the students' excitement about the game mechanics themselves. As a result, the teachable moments embedded in the activity were not fully utilized.

d. Time Inefficiency Due to Poor Throwing Accuracy

The fourth difficulty identified relates to the physical dimension of the Velcro Dart game. A significant number of students had limited accuracy and throwing skill, meaning that multiple attempts were required before a student successfully hit the target on the board. This repeated trial process consumed a disproportionate amount of class time that could otherwise have been devoted to vocabulary practice and discussion. Because the treatment sessions were

constrained to 2×40 minutes per meeting, the time lost to unsuccessful throwing attempts directly reduced the frequency of vocabulary interactions each student experienced during the session. This represents a notable limitation of game-based methods that involve a physical skill component, as the physical skill itself can become a barrier to achieving the linguistic goals of the activity.

B. Discussion

1. The Effectiveness of Velcro Dart Game on Improving Students Vocabulary

The findings of this research reveal that both the experimental group taught using the Velcro Darts Game and the control group taught through conventional instruction demonstrated measurable vocabulary improvement over the study period. The experimental group's mean score rose from 46.1 on the pre-test to 65.0 on the post-test, producing a gain of 18.9 points, while the control group improved from 45.1 to 62.2, with a gain of 15.4 points. Although the experimental group showed a slightly higher rate of improvement, this margin alone is not sufficient to draw conclusions about the game's effectiveness without statistical verification.

To determine whether the difference in improvement between the two groups was statistically meaningful, an independent samples t-test was conducted at a 5% significance level with $df = 49$. The test yielded an observed value (t_o) of 0.550, which falls well below the critical t-table value of 2.010. Consequently, the Alternative Hypothesis is rejected and the Null Hypothesis is accepted, indicating that the difference in vocabulary improvement between the two groups is not

statistically significant under the conditions of this study. Despite this, the Effect Size analysis using Cohen's d formula produced a value of 0.156, which indicates a low but positive real-world contribution of the game to students' vocabulary development confirming that the game did produce some meaningful impact, even if it did not reach the threshold of statistical significance.

These results align partially with the finding of Thuraya A. Alabsi (2016), who investigated the effectiveness of role-play strategy in teaching vocabulary among EFL students using a similarly structured quasi-experimental design. Like the present study, Alabsi's pre-test revealed no significant difference between groups before the treatment began. However, unlike the present study, Alabsi's post-test results showed a statistically significant advantage for the experimental group. This contrast suggests that the depth of active language involvement in an instructional activity plays a decisive role in determining its statistical outcome. Both role-play and the Velcro Darts Game attempt to lower students' affective filters and encourage voluntary language engagement through interactive, non-traditional activities. The difference in result, however, may be explained by the nature of participation each activity demands: role-play requires every student to produce language continuously throughout the activity, whereas in the Velcro Darts Game, only one student throws at a time, which risks leaving the remaining students as passive observers if their engagement is not actively managed by the teacher.

This interpretation is further supported by the study of Antika Mega Mawarni, Suwandi, and Dias Andris Susanto (2021), who found that the digital

game Kahoot produced a statistically significant improvement in vocabulary mastery among senior high school students in a quasi-experimental design comparable to the present research. The key structural difference between Kahoot and the Velcro Darts Game lies in the nature of student participation: Kahoot requires every student to respond simultaneously and independently to each vocabulary question, maximizing individual engagement and ensuring that every learner practices every item. In contrast, the turn-based structure of the Velcro Darts Game concentrates active participation in one student at a time, which directly limits the total volume of vocabulary practice each individual receives per session. This structural gap in participation breadth is a plausible explanation for the difference in statistical outcome between the two studies.

The study conducted by Farideh Bavi on the effect of fun activities on vocabulary learning at the elementary level provides additional perspective. Bavi found that students who received varied, enjoyable learning activities significantly outperformed those taught through traditional methods — a finding that reinforces the general principle that engagement and positive emotional experience in the classroom are beneficial for vocabulary acquisition. What distinguishes Bavi's result from the present study is likely the variety and frequency of language engagement embedded in the activities used, which gave students more consistent and repeated exposure to target vocabulary across each session. This points to a broader conclusion: the effectiveness of any game-based vocabulary activity is not determined simply by the presence of fun or physical interaction, but by the degree

to which the activity ensures that every learner is repeatedly and actively exposed to the target language throughout the lesson.

Taken together, the results of this study and their relationship to the previous research reviewed in Chapter II including the theoretical frameworks offered by Priyasudiarja and Purwaningsih (2020), Andi (2018), Ramadhan (2021), and others suggest that the Velcro Darts Game holds genuine pedagogical value as an engaging, low-cost, and low-anxiety instructional medium for vocabulary teaching. Its limited statistical impact in this study should not be read as evidence that the game is ineffective, but rather as evidence that its effectiveness is highly conditional. The implementation challenges identified through classroom observation including student inattention during peers' turns, unproductive peer reactions, and time inefficiency caused by repeated missed throws collectively reduced the game's instructional quality in ways that directly suppressed its measurable impact on vocabulary scores. Addressing these challenges through stronger classroom management, modified game mechanics, and an increased number of treatment sessions would likely bring the game's practical potential into statistical range in future research.

2. Students' Difficulties During Velcro Dart Learning Activities

Based on direct classroom observation across four treatment sessions, several recurring difficulties were identified that collectively limited the Velcro Dart Game's pedagogical effectiveness. Students frequently lost focus when it was not their turn, missing the vocabulary exposure embedded in each peer's performance

— an issue rooted in the sequential, turn-based structure of the game that concentrates active engagement in one student at a time. Compounding this, unproductive peer reactions — particularly mocking students who missed the board or answered incorrectly — created a socially hostile atmosphere that raised students' affective filters and actively blocked language acquisition, confirming Krashen's (1982) position that emotional safety is a prerequisite for successful second language learning. The competitive nature of the game inadvertently heightened students' fear of public failure, underscoring that a supportive classroom culture must be deliberately established before any game-based activity is introduced.

Beyond attention and social dynamics, two further implementation challenges reduced the quality of vocabulary practice within each session. The excitement generated by the game mechanics consistently overshadowed the teacher's vocabulary explanations delivered mid-game, meaning that the activity's instructional moments were repeatedly lost to students' anticipation of the next turn — a problem that requires deliberate pedagogical pacing rather than relying on students to redirect their own focus voluntarily. Additionally, poor throwing accuracy caused many students to require multiple attempts before hitting the board, consuming significant portions of the 2×40 minute sessions without producing any vocabulary practice in return. Collectively, these difficulties are interconnected: reduced attention limited exposure, peer mockery suppressed participation, distraction blocked deep processing, and time inefficiency reduced practice frequency. Their combined effect directly explains why the game's positive contribution — confirmed by Cohen's $d = 0.156$ — did not reach statistical

significance, and affirms that the Velcro Dart Game's effectiveness is not inherently limited but is highly conditional on the quality of classroom management and pedagogical planning that surrounds its use.