# CHAPTER III

# METHODOLOGY

In recent time, the use of technology changes students' mind a lot. This circumstance advances education applications of computer that provides a rapid growing resource for language classroom. Because of that, learning language has been developed to take the advantage of technology to be applied in language classroom. Hence, it creates Computer-Assisted Language Learning or CALL. In addition, one of the features of CALL is game.

Afterward, this research investigated the use of a computer game in learning activity as a media. Thus, it was aimed to seek:

1. The effectiveness of playing a computer game to enhance students' ability to master vocabulary.
2. The responses from students toward the use of a computer game as vocabulary learning media.

This chapter discusses about methodology which was used in this research. It consists explanation of Research Methods, Hypothesis, Population and Sample, Research Instruments, Data Collection Procedures, and Data Analysis Pre and Post tests

## 3.1. Research Methods

This research employed quantitative analysis with quasi experimental design. The design was chosen to test the hypothesis. This research entangled three classes; the first class was a try-out class; it was used to seek the validity and the reliability of instruments. The second was control class and the third was experimental class (Hatsch and Farhady, 1982; Hatsch and Lazaraton, 1994; Kranszler and Moursund, 1999; Dornyei, 2007). The calculation result was analyzed and discussed to answer the research question.

Meanwhile, the questionnaire analysis was used to triangulate the data based on the statistical calculation result. The result of questionnaire and vocabulary test were compared and synthesized to get deeper interpretation of the research. In addition, that result was validated using related theoretical foundation of expert views.

## 3.2. Hypothesis

This research was started with *null*-hypothesis ($H_0$) where both classes; experimental and control were considered have no significant different in the level of mastering vocabulary.

$$H_0 : \quad \mu_{experimental} = \mu_{control}$$

In specific, the null hypothesis of this research is playing a computer game cannot develop students' ability in vocabulary mastery. It means that there is no significant difference between experimental class and control class in *mean* adjustment level (Hatsch and Farhady, 1982; Hatsch and Lazaraton, 1994; Kranszler and Moursund, 1999; Dornyei, 2007). By using null hypothesis, every possibility of the research can be shown. If the hypothesis is rejected, it can be concluded that experiment works. While, if the hypothesis is accepted, the experiment does not work.

## 3.3. Population and Sample

The main prerequisite of population and sample for this research was the sample should know how to operate computer and engage with them continually. Yet, the population and sample should not have ever played the computer game given before. Because of that, purposive sampling was used. As the result, students of one of junior high school in Sumedang were chosen.

The target population for this research is homogeny population. Therefore, the population for this research was the second grades students of junior high school. They were chosen for they become accessible population for this research. So, all second grades students in one of junior high school in Sumedang became the population of this research.

Furthermore, the population selected was narrowed to be samples. Samples for this research were taken using purposive sampling. It employed clustered intact group sampling from the population. So that, every class within the population or sample frame has the same chance to be chosen as experimental class and control class.

Since there were three classes conducted for this research, two classes were chosen first as experimental class and control class. And the other class performed as try-out class. As a result, 8a was chosen as control class, while 8b took part as experimental class. Both control class and experimental class consist of 20 students.

Since the number of samples does not reach the minimum criteria to apply several statistical calculations, two tests were employed; normal distribution test and homogeneity of variance. These tests aim to identify whether *t*-test calculation continues or not (Hatsch and Farhady, 1982; Hatsch and Lazaraton, 1994; Kranszler and Moursund, 1999; Dornyei, 2007).

### 3.3.1. Try-Out Class

The try-out class was used to analyze the validity and the reliability of the instruments, so that the instruments can be utilized as a measurement to test students' ability in mastering vocabulary. This class consists of 25 students excluded the control class and the experimental class. The same grade students

with control class and experimental class were chosen, in order to eliminate the possible issues which relate to validity and reliability of the instruments.

### 3.3.2. Experimental Class

The experimental class of this research consists of 20 students who are assumed in the same level in their vocabulary mastery. As it calls, the experiments were given to them in some segments. Generally, a computer game was given to them to be played. Thus, teacher guided them and involved with them to play the game.

In order to achieve or to finish the game, teacher did not give the participants any clues about what this game is for. In the process, teacher did not show the way how. Furthermore, teacher did not give students the definition of the unfamiliar words or direct translation. So, students could find a way to understand those words by their effort.

The experiment took eight meetings. Hence, the experiment stopped when they finished or completed the game. In addition to those eight meetings, another one meeting was held before the experiment begun, as a pre-test and another one meeting for post-test. In total, there were ten meetings held for completing this research.

### 3.3.3. Control Class

Similar with the experimental class, the control class consists of 20 students. This class performed as a static class that controlled the vocabulary improvement of experimental class. This result of the research was analyzed by comparing the test result with the experimental class.

No experiment held in control class. In other words, no changes of way they learned vocabulary. Roughly, they learned vocabulary along with reading comprehension in the class. They read the text and try to guess the meaning of the new words they found in the passage. Otherwise, they wrote down the unknown words, gave the words to their teacher and teacher explained the meaning of the words (or direct translation).

Then, commonly, teacher asked them to produce sentences or paragraphs consist of those new words. And n the end of week or unit, students were given a quiz or test related to the vocabulary given. Again, it was along with reading the passage or reading the text.

As in experimental class this class also held 10 meeting in total. 8 meeting for delivering material to students and the other two were taken for pre-test and post-test.

## 3.4. Research Instruments

This research employed vocabulary test as the main instruments. These instruments were supported by the finding from the questionnaire. Below are the explanations of instruments for this research.

### 3.4.1. Vocabulary Test

This vocabulary test is a measurement of students' ability in mastering vocabulary. The test comprises 30 multiple choice question items which were tested to both experimental class and control class. However, to build the test as a good measurement of vocabulary mastering ability, its validity and reliability of the vocabulary tests appeared.

In the very beginning, 50 items of multiple choices question were analyzed to seek the highest validity and reliability for every question items. Before they were analyzed, those 50 items of multiple choices question are validated by experts. In consequence, several chances were done to the items. Some of them considered the words which should be familiar and localized based on the syllabus and curriculum of junior high school students in Sumedang.

In addition, the site and the condition of students were considered to make the items as valid and reliable as possible.

#### 3.4.1.1. Vocabulary Test Item

The test was given to try-out class to seek their validity and reliability of instruments before it was tested to both control class and experimental class.

The analysis began with 50 question items of vocabulary test in order to get in order to choose the better validity and reliability scores. All questions items were developed from the school syllabus, relevant theories and part of the conversations taken from the game played. They cover four basic competences and several indicators (see Appendix B).

Yet, before applying the test to try-out class, the items were analyzed by experts and document analysis by comparing the test item to the syllabus and the specific theories. Therefore, from 11 vocabulary skills which are proposed by Gairns and Redman (1986) cited from Lewis (2001), this test items are covered 10 of them. There are boundaries between conceptual meanings, polysemy, homophony, homonymy, synonymy, affective meaning, style, register and dialect, translation, chunks of language, and grammar of vocabulary. While pronunciation was not included since this skill are related to recognize and to reproduce words in speech. Meanwhile, questions items can be seen in the appendix A.

To make clear understanding about the items and prove the validity and reliability of the instruments, the specification used for constructing test can be seen on appendix B.

### 3.4.1.2. Validity test

Arikunto (1993; 63-69) explained that validity is a measurement of instrument. If the validity value of the instrument is low/poor, the instrument cannot be trusted to measure something. It means the invalid instrument cannot be applied to the respondents. Arikunto also proposed the use of *Pearson Product Moment Correlation Coefficient Values* to seek the value of instruments' validity.

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

*(Hatsch and Farhady, 1982; Hatsch and Lazaraton, 1994; Kranszler and Moursund, 1999; Riduwan, 2003; Dornyei, 2007).*

Where *r* is Pearson Product Moment Correlation Coefficient Values, *N* is number of students who are analyzed, *x* is students' vocabulary score (first variable) and *y* is students' summative score (second variable).

Every item on the instrument was calculated by using Pearson Product Moment Correlation Coefficient Values to seek correlation index for every items through correlating every single item of instrument (*x*) with total score of instrument (*y*). Next, the *r* is interpreted as follow :

**Table 3.1. Index of Validity Level**

| Coefficient Interval (*r*) | Validity Level |
|---|---|
| 0.800 – 1.000 | Excellent (Ex) |
| 0.600 – 0.799 | Good (Gd) |
| 0.400 – 0.599 | Satisfying (St) |
| 0.200 – 0.399 | Poor (Pr) |
| 0.000 – 0.199 | Very Poor (Vp) |

*(Riduwan 2004:110)*

### 3.4.1.3. Reliability test

Not only validity analysis but also reliability analyses were employed for this research, so that the effective measurement of vocabulary test can be given to gather data as exact as possible to students' ability in mastering vocabulary. This research used *KUDER RICHARDSON*-20 (KR-20) method to analyze reliability of instrument. It is because in this research instrument, every single right-answer is valued 1 and every wrong-answer is 0.

In addition, this method was used to gain a higher reliability value, since the result of *KR-20* tends to give a higher value than the other methods such as *KR-21*, *Anova Hoyt*, *Alpha* and so on. (Arikunto 1993: 101).

*KUDER RICHARDSON*-20 (KR-20) formula :

$$r_{11} = \left(\frac{K}{K-1}\right)\left(\frac{s^2 - \sum pq}{s^2}\right)$$

$r_{11}$ = *Internal reliability coefficient for all items*

$K$ = *sum of question item*

$p$ = *proportion of subjects who answer right*

$$p = \frac{sum\ of\ those\ who\ answer\ right\ in\ an\ item}{N(number\ of\ respondents)}$$

$q$ = proportion of subjects who answer wrong ($q = 1 - p$)

$s^2$ = variance total

$$s^2 = \frac{N\sum y^2 - (\sum y)^2}{N(N-1)}$$

Afterward, the value of $r_{11}$ is compared with index of Pearson Product Moment Correlation Coefficient Values (see Appendix E) to see whether the value is reliable or not.

### 3.4.1.4. The Result of the Try-Out test

On Monday, May 2, 2011, the instrument was applied to try out class, eight grade students of one of junior secondary school in Sumedang, to seek the validity and reliability of the instrument. The result is shown as follow.

**a. Test Items**

The calculation showed that from 50 questions items of vocabulary tested, 34 items were categorized valid and 16 items were invalid. So that 30 valid items of vocabulary test were taken as the instrument of this research and the instrument was covered as shown on the table below.

**Table 3.2. Index of Validity for Question Items (Try-Out test)**

| Coefficient Interval (*r*) | Validity Level | Question Item number |
|---|---|---|
| 0.800 – 1.000 | Excellent (Ex) | - |
| 0.600 – 0.799 | Good (Gd) | 16,20,27, |
| 0.400 – 0.599 | Satisfying (St) | 2,3,5,6,9,10,15,17,18,19,22, 25,26,31,32,33,36,37,38,39, 40,42,46,47,48. |
| 0.200 – 0.399 | Poor (Pr) | 11,21,24,28. |
| 0.000 – 0.199 | Very Poor (Vp) | 1,13. |
| <0.00 – 0.000 | Invalid | 4,7,8,12,14,23,29,30,34,35, 41,43,44,45,49,50. |

Later, the invalid items were eliminated, so that the vocabulary test item consist of three items categorized good, 25 items categorized satisfying and two items categorized poor. Furthermore, after comprising the items into 30 (see appendix I), below is the distribution table for each items of pre-test and post-test.

**Table 3.3. Index of Validity for Question Items (Pre-and Post- test)**

| Coefficient Interval (*r*) | Validity Level | Question Item number |
|---|---|---|
| 0.800 – 1.000 | Excellent (Ex) | - |
| 0.600 – 0.799 | Good (Gd) | 8,12,17 |
| 0.400 – 0.599 | Satisfying (St) | 1,2,3,4,5,6,7,9,10,11,14,15, 16,19,20,21,22,23,24,25,26, 27,28,29,30 |
| 0.200 – 0.399 | Poor (Pr) | 13,18 |
| 0.000 – 0.199 | Very Poor (Vp) | - |
| <0.00 – 0.000 | Invalid | - |

In consequence, the distribution of standard competence, basic competence, learning activities and also vocabulary skill offered by the question items were changing as in appendix I.

**b. Reliability of the Instrument**

After calculating validity of the question item using item analysis (see appendix C and D), the reliability of the instruments were analyzed to using KR-20 (Hatsch and Farhady, 1982; Hatsch and Lazaraton, 1994; Kranszler and Moursund, 1999; Dornyei, 2007).

The result of reliability question item by using *KR-20* is reliable (see appendix E). From the calculation, it is shown that the reliability value ($r_{11}$) is 0.527321 this result is reliable in level .05 which has minimum value 0.4227. It means that selected question items are valid and reliable to measure students' ability in mastering vocabulary.

### 3.4.2. Questionnaire Items

Questionnaire in this research was used to triangulate the data along with the interpretation of the test result. Generally, this questionnaire checked the result of students' vocabulary development through test from participants' point of view. This questionnaire also gave the clear reason and explanation how students could learn vocabulary.

The questionnaire dealt with the students' responses toward the game as a vocabulary learning media. The questionnaire for this research consists of 15 multiple choices-questions answered by participants. It was given to the participants at the end of the program.

### 3.4.3. Learning Media (the Game)

A computer game entitled Nancy Drew: The Creature of Kapu Cave was chosen as a media for assisting students in learning vocabularies. This game was selected because it provides audio and visual aids for learner. Furthermore, this game was produced and published by English, so that, learners can study the vocabularies

given by the English native speakers. In addition, to support vocabularies learning for learners, the subtitle was provided for every character's spoken words.

Another reason this game was chosen is the game rating. Since it was directed to junior high school students, the rating of the game that can be given is labeled "E" (everyone) or "T" (teenagers) not "18+" nor "A" (adult). This Nancy Drew: The Creature of Kapu Cave has been rated by ESRB (an organization specialized to rate video games) as "T" (teenagers). It means this game is saved to be played by teens.

In a brief, this game is considered as RPG or Role Playing Game, where the player controls the main character in the game and decides what the character will do. Role-playing games requires much in the way of reading; rules are written in books after all. At it is absolute minimum, a person playing an RPG must at least read enough information to be able to create a character. Someone choosing to referee a game must do a great amount of reading; game rules, setting, history, back story, plot, character histories and descriptions, all of these must be read before a game can be properly played. It helps a lot if some of the basic facts are memorized, as well. All of this reading and rote memorization is exercise for the mind (Cameroon, 2001).

RPGs can also involve a bit of writing to go along with all of the reading. Many players write their own character history and background, to help flesh out the persona that they are portraying. A referee who decides to write his own

adventure must create a setting, plot, and characters at the very least, and usually ends up getting rather involved in writing an enjoyable story. In both cases, not only are writing skills being developed, but creative ones are as well.

Because of those reasons, a computer game entitled Nancy Drew: The Creature of Kapu Cave was chosen as a vocabulary learning media to be applied to junior high school students.

## 3.5. Data Collection Procedures

The data for this research was gathered from various resources related to it. Those data were used to create valid and reliable instruments and to provide comprehensive discussion to the research finding.

To create valid and reliable vocabulary tests, syllabus of the one of junior high schools in Sumedang were used to bond the instrument with the learning goal. So, the vocabulary tests are apart with the school learning goal. This syllabus was used as question items parameter to develop the vocabulary test.

Furthermore, the vocabulary tests were created by applying the theory of eleven vocabulary aspects by Gairns and Redman (1986) cited from Lewis (2001); Boundaries between conceptual meanings, Polysemy, Homophony, Homonymy, Synonymy, Affective meaning, Style, register, dialect, Translation, Chunks of language, Grammar of vocabulary and pronunciation. Both syllabus and vocabulary aspects became the criteria to create the vocabulary tests.

Meanwhile, the target words used in vocabulary tests were taken from several conversations that exist in the computer game (see appendix A). In sum, the vocabulary tests given to students were created from several conversations in the game with considering the learning goal in syllabus and vocabulary aspects as theory.

The other instrument of this research was questionnaire. The questionnaire for this research consisted of 15 question items. Those items were created by considering theories of vocabulary, theories of the use of game, and theories of Computer Assisted Language Learning. This questionnaire was aimed to provide students' responses toward the media.

Afterward, the computer game which was used as media entitled *"Nancy Drew; the Creature of Kapu Cave."* This game considers as RPG (Role-Playing Game) that allows players to choose the character's fate.

## 3.6. Data Analysis of Pre and Post tests

To verify the hypothesis of this research, independent *t-test* was chosen. Independent *t-test* has primary purpose to see whether the mean score of two different or independent groups differ to a statistically significant degree (Hatsch and Farhady, 1982; Hatsch and Lazaraton, 1994; Kranszler and Moursund, 1999; Dornyei, 2007). It has aim to analyze the result significance of this study. Yet, before applying the *t*-test analysis, two prerequisites need to be fulfilled; Normal distribution test and Homogeneity test.

### 3.6.1. Normal distribution (Pre-test scores)

The samples of this research were tested using pre-test question item to seek whether the distribution of the data is normal or not normal. Basically if the samples reach 30 or more, this test is not needed. Unfortunately, the samples for this research were only 20. In consequence, their score should be tested, so that, the research can be continued using parametric analysis or not-parametric analysis. The normal distribution of the sample was tested using SPSS software.

### 3.6.2. Homogeneity Variance (Pre-test scores)

To verify whether the subject of this study homogeny or not, Chi-square test and *t*-test were used. These tests are specifically used for testing the category data of hypotheses. (Hatsch and Farhady, 1982; Hatsch and Lazaraton, 1994; Kranszler and Moursund, 1999; Dornyei, 2007).

The data taken from pre-test was used to seek the homogeneity of subject by taking the pre-test variance calculation of both classes. The formula of variance for each variables and total variance are:

- Variance for variable x

$$s_x^2 = \frac{n_x \sum x^2 - (\sum x)^2}{n_x(n_x - 1)}$$

- Variance for variable y

$$s_y^2 = \frac{n_y \sum y^2 - (\sum y)^2}{n_y(n_y - 1)}$$

- Total variance

$$S_{tot}^2 = \frac{(n_x . s_x^2) + (n_y . s_y^2)}{n_x + n_y}$$

Afterward, chi-square test ($x^2$) is ready to be calculated by using the formula which is proposed by Riduwan (2004). The formula is provided below.

$$\chi^2 = (lon\ 10)(B - \sum (df . \log S_n^2)$$

Where B equals with $(\log S_{tot}^2)$ $(\sum df)$ and *df* is total subjects in one variable minus 1.

Finally, the result of $\chi^2$ is compared with table of chi-square test (see appendix g), so the criteria of subject are found. It can be described as follow:

- The result of $\chi^2 > \chi^2$on the table of chi-square test. It means that the subject is not homogeny, so the comparative test cannot be held.
- The result of $\chi^2 < \chi^2$ on the table of chi-square test. It means that the subject is homogeny, so the comparative test can be held.

While for *t-test* the formulae are;

$$t_{obs} = \frac{\bar{X}_{exp} - \bar{X}_{ctrl}}{S_{(\bar{x}_{exp} - \bar{x}_{ctrl})}}$$

$$S_{(\bar{x}_{exp} - \bar{x}_{ctrl})} = \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}$$

*Note:*

$t_{obs}$ = *the value of t-observed through the data*

$n_x$, $n_y$ = *the number of subjects in each of the two classes*

$S_x^2, S_y^2$ = *the variances of the two classes*

$\bar{X}_{exp}, \bar{X}_{ctrl}$ = *the mean of two classes*

$S_{(\bar{x}_{exp} - \bar{x}_{ctrl})}$ = *the values of standard error of differences*

And the as in chi-squared test, the result of *t*-observed is compared with critical in the *t*-table (see appendix h). So the result can be described as follow:

- The result of $t - obs > t - crit$ on the table of critical value of two-tailed test. It means that the subject is not homogeny, so the comparative test cannot be held.
- The result of $t - obs < t - crit$ on the table of critical value of two-tailed test. It means that the subject is homogeny, so the comparative test can be held.

### 3.6.3. Calculating *t*-test (post-test scores)

After the treatment was given to the experimental class, post-test score of both classes was taken. Then, those scores were calculated by using t-test for two independent samples to seek the difference between the *mean* of both classes, so that the significance of the result appeared.

To calculate the result, t-test formula which is proposed by Hatsch and Farhady, 1982; Hatsch and Lazaraton, 1994; Kranszler and Moursund, 1999; Dornyei, 2007; is used. There are four steps to calculate the result:

1. Calculate the post test variance ($S^2$) of each class by using;

$$s^2 = \frac{N\sum x^2 - (\sum x)^2}{N(N-1)}$$

Where *N* is the number of subject and *x* is the score of variable x.

2. Calculate the t obtain of the result ($t_{obt}$)

$$t_{obs} = \frac{\bar{X}_{exp} - \bar{X}_{ctrl}}{S_{(\bar{x}_{exp} - \bar{x}_{ctrl})}}$$

$$S_{(\bar{x}_{exp} - \bar{x}_{ctrl})} = \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}$$

3. After $t_{obs}$ was found, to find the meaning of the calculation, use the critical value of *t* ($t_{crit}$). To find the $t_{crit}$, see appendix H.
4. The last step was to decide whether the hypothesis is rejected or not. If $t_{obs}$ has equal value or higher than $t_{crit}$, it means that $H_0$ is rejected and vice versa. In addition, level of significance which is usually chosen is $\alpha = 0.05$

5. After testing the hypothesis, the result was interpreted, so that the difference between the experimental class and control class was shown.

The conclusion of this research was taken not only from the result of *t*-test calculation for statistical significance, but also the result of questionnaire.