

## CHAPTER III

### RESEARCH METHODOLOGY

#### 3.1 Research Method

The study used mixed forms of research design (Nunan, 1993: 6); specifically the exploratory-quantitative-statistical research design. It means that the study is non-experimental in nature, while the type of data yielded by investigation is a quantitative one. In addition, the study was analyzed using a statistical analysis. In this case, the researcher will make claims about the population based on the data obtained from the sample of that population (Nunan, 1993: 28). The participants of this study involved only the subjects in Indonesia University of Education (UPI) at their sixth semester in English literature program. For a better understanding, the following section is about population and sample followed by the data collection procedures.

#### 3.2 Population and Sample

According to Kranzler and Moursund (1999: 77) population can be defined as “a large... group about which some information is desired.” Beside that, in Encyclopedia of educational evaluation, population is “a set (or collection) of all elements possessing one or more attributes of interest”. In this respect, the population is the features of sixth semester English literature students of UPI in the mastery of reading comprehension and their collocational knowledge.

In another case, sample is defined as a part or a representative of the population that is investigated (Arikunto, 2002: 109). Accordingly, the subjects taken here are 30

students of English literature students in the sixth semester. From these sample, hopefully, could give some data or information that supports the study.

Further, the sampling technique taken up uses purposive sampling. Arikunto (2002: 117) defines purposive sampling as a technique to get sample based on the purpose of the study with certain characteristics on its subjects. Thus, the sample is chosen for the following reason: firstly, the purpose of the study is to examine the correlation between knowledge of lexical collocations and their general reading comprehension. Secondly, with at least 6 semesters formal English studying in the university, UPI, the students must have learned a certain number of collocations and would have had many opportunities to practice them. Moreover, this sampling takes a random sampling by the reason that there is no differentiation of the participants engaged in this study as to gender, level of intelligent, etc. Therefore, all participants have the same privileges, although by the fact that there are the subjects who are not all smart, not all medium, and not all poor.

After highlighting the kind of method and the technique used, the following passage will then explain the steps or the procedures of collecting the data.

### **3.3 Data Collection Procedures**

This passage will concern more about the ways or steps the data are collected beginning from the preparation to analyzing the data.

#### **3.3.1 Preparing the Instrument**

In preparing the instrument, the researcher will use test as a tool of instrument, specifically the objective test that includes multiple choices. The reason is that by this type

of test, it is intended at getting scoring quickly and without any judgment and subjectivity (Maskar, 2008). The test will cover the reading comprehension test and the test that measures the knowledge of collocation. For the collocation knowledge, this study will use the Hill's (2000) collocation categories (i.e. Adj + N, N + N, V + Adj + N, V+ Prep + N, V + Adv, Adv + V, Adv + Adj, & Adj + Prep). The table below is the collocation model that will be used for this study.

Table. 4. Collocation Category

COLLOCATION TYPES	EXAMPLES
1. Adjective + Noun	<i>a huge profit</i>
2. Noun + Noun	<i>a window frame</i>
3. Verb + Adjective + Noun	<i>learn a foreign language</i>
4. Verb + Preposition + Noun	<i>speak through an interpreter</i>
5. Verb + Adverb	<i>live dangerously</i>
6. Adverb + Verb	<i>half understand</i>
7. Adverb + Adjective	<i>completely soaked</i>
8. Adjective + Preposition	<i>excited about</i>

In addition, the tests (the items) that measure subjects' mastery of reading comprehension will be taken from TOEFL (Barron's (10<sup>th</sup> edition, 2001), and Power TOEFL CDROM

(1998). The tests will include the questions such as main ideas, factual, referential, inferential, and vocabulary.

### 3.3.2 Trying Out the Instrument

After arranging the instrument, then, the tests need to be tried out in order that they are valid and reliable and suitable for the tests (see Arikunto, 2002; Harris, 1969). Besides, the tests should be practical as a prerequisite of a good test (Harris, 1969). Below will be explained about both of terms: 'validity and reliability' in turn.

#### 3.3.2.1 Validity

Validity is a measure that will judge that the tests is valid as to assessing the existing knowledge of students as they are and determining what should be measured objectively. How to determine the validity of the tests is firstly the tests are handed in to the students (students that are not the target for the study, but they are still at the same level and the same generation) and the data taken is calculated using formulation of *Pearson product-moment*. This formula will look like this.

Table. 5. The Formulation of Pearson Product-moment

$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$		
Where:	$r$ = correlation coefficient  $N$ = the numbers of pairs of measurements	$X$ and $Y$ = variables (variable $X$ and variable $Y$ )  $\Sigma$ = sum

### 3.3.2.2 Reliability

With reliability, it can be said that the tests are consistent in measuring the knowledge of the students. For better understanding, if the tests are given to the students more than one to the same subjects or the participants, the results will be more or less the same with the first condition. Therefore, if the tests are reliable, it is automatically considered as a good test.

To compute reliability of a test, the method to be used is *split-half method*. The method is used for the reason of simplicity. It means that the measure of reliability is gotten from one single test and then the data is divided into two parts. In that case, the first part is called x variable, and another part is named y variable. Each part is then correlated by using *Pearson product-moment* above. Then, the correlation coefficient obtained is inserted into the *Spearman-Brown* formula and then computed. The formula that is used is *Spearman-Brown* formula. The formula will look like this.

Table 6. The Formulation of Spearman-Brown

$r_{11} = \frac{2 \times r_{1/21/2}}{(1 + r_{1/21/2})}$		
Where:	$r_{11}$  $r_{1/21/2}$	= reliability  = correlation coefficient ( $r$ )

### 3.3.3 The Results of Try-Out

To try out the items of the two tests (collocation knowledge and reading comprehension), both the test were given to English students of sixth semester of UPI at the same time. The test was held on June 16, 2008.

After trying out, all items are identified in terms of validity. The test items that will be used are the items that complete the value of validity of more than 0.30 or itself (See Appendices 3 and 4). Based on the result, therefore, items of collocation tests that will be used are 28 items. Meanwhile, items of reading comprehension test that will be used are 28 items. The coverage of the tests will be shown as follow.

Table 7. Coverage of Reading Comprehension Test

No.	Coverage	Frequency	Percentage	Scores
1.	Main Ideas	2	7.14 %	2 @ 3 = 6
2.	Factual	10	35.71 %	10 @ 1 = 10
3.	Referential	2	7.14 %	2 @ 1 = 2
4.	Inferential	7	25 %	7 @ 3 = 21
5.	Vocabulary	7	25 %	7 @ 2 = 14
Total		28 items	100%	53

Table 8. Coverage of Collocation Test

No.	Coverage	Frequency	Percentage	Scores
1.	Adjective + Noun	3	10.71 %	3
2.	Noun + Noun	3	10.71 %	3
3.	Verb + Adjective + Noun	2	7.14 %	2
4.	Verb + Preposition + Noun	3	10.71 %	3
5.	Verb + Adverb	4	14.29 %	4
6.	Adverb + Verb	4	14.29 %	4
7.	Adverb + Adjective	3	10.71 %	3
8.	Adjective + Preposition	6	21.43 %	6
<b>Total</b>		28 items	100%	28

The scoring process of the two tests is different. In that case, in collocation test, the score that will be given is one (1) for each item while in the reading test, the score that will be given depends on kinds of questions. For main ideas, each score will have 3, the same criteria for the questions of inferential. While for the questions of vocabulary will get each score of 2. The rest that is the questions for factual and referential will contain the point of 1. So the total score for reading comprehension will be 53, while for collocation will be 28.

### 3.3.4 Testing the Subjects

The subjects that will be involved are 30 students of the sixth semester of English literature in UPI. The testing process covers two tests: collocation test and reading comprehension test. Both tests spend the time for 90 minutes. The testing process will be in the same day, so that it will be easier to manage the subjects.

### 3.3.5 Analyzing the Data

The data that have been collected were then processed. The data were processed in several steps. The first step, the scores and the averages were calculated. In the second step, the process revealed the normality of the distribution of both the data. And then the last step was finding out the correlation between the two variables, measuring it and then interpreting it as to significant or not.

To begin with, the first step was calculating the scores that each student had. The scorings have a range from 0 (the lowest score) to 100 (the highest score). It is done by converting each value to the multiplication of one hundred and then divided by the total numbers. So, if the score is 16 from the total number of 28, it would be 57  $((16 \cdot 100) / \text{total numbers})$ .

As has been stated, the scoring of the tests was used to find out the average of the tests. The result of the average will then be used to measure the subjects' mastery of both tests: reading comprehension and the knowledge of collocation. To calculate the average will use the formula following:



Table 9. The Formulation of the Average

$Mx = \frac{\sum x}{N}; \quad My = \frac{\sum y}{N}$	
Where:	<p><math>Mx</math> = average x (grammatical knowledge)</p> <p><math>My</math> = average y (reading comprehension)</p> <p><math>\sum x</math> = the sum of x scores</p> <p><math>\sum y</math> = the sum of y scores</p> <p><math>N</math> = the number of subjects</p>

By finding out the average, it will be interpreted to what extent is the mastery of reading comprehension and the knowledge of collocation the students have. To determine the mastery and the knowledge level of the students, the result will be based on the criteria made by Harris (1969: 134) who classifies the range of scores with its probable class performance. The classification is as follow:

Table 10. Classification of the Range of Score

Test Scores	Probable Class Performance
80-100	Good to excellent
60-79	Average to good
50-59	Poor to average
0-49	Poor

The next process to be completed with after having known the level of students' mastery of reading comprehension and the level knowledge of collocation, is computing the correlation. But firstly, it should be known the normality of the distribution of the data. Computing the normality in this case will make use the software called SPSS 12 (Statistical Package for Social Sciences) with the equations of *Kolmogorov-Smirnov* and *Shapiro-Wilk*. It is done by inserting the data to the computation and automatically the results will get soon be used. The results would be in two conditions: if the data is known normal in terms of distribution of the data, the computation especially for the correlation will use parametric statistic (by using *Pearson product-moment*). However, if the condition of the data is not normally distributed, the computation of the correlation must use non-parametric statistic (by using *Spearman Correlation for Ranked Data*).

Therefore, for the data that is normally distributed, the computation employed will use *Pearson product-moment*. The formula will look like this:

$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$		
Where:	$r$ = correlation coefficient $N$ = the numbers of pairs of measurements	$X$ and $Y$ = variables (variable $X$ and variable $Y$ ) $\sum$ = sum

Meanwhile, for the data that is not normally distributed, the computation must use *Spearman Correlation for Ranked Data*. The formula will look like this.

$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$	
Where:	<p><math>D</math> = the difference between each pair of ranks</p> <p><math>D^2</math> = the square of those differences</p> <p><math>n</math> = the number of pairs of rankings</p>

Both formulations are different in nature, so are the computation. By using *Pearson product-moment*, after it has been found that the data are normally distributed, the data can be inserted to the formulation of correlation. Meanwhile, by using *Spearman* formula, the data that has been found not normally distributed must be ranked firstly before inserting to the formula. It is done by converting them to the rank data by using Microsoft Excel program 2007.

After the result has been obtained, the next step is to determine whether or not there is a correlation of the two variables. Knowing if there is a correlation or not is by consulting the result to the level of significance. Both formulations (Pearson and Spearman correlation) have each level of significance. For the formulation using Pearson will be determined by level of significance of Pearson. On the other hand, if using *Spearman* correlation, the interpretation of  $r_s$  (correlation coefficient) will be based on the table of critical values of Spearman's ranked correlation coefficient ( $r_s$ ). However, both of them will fall on the point at the level of significance .05 since this value is commonly used in social science. The difference is on the choice when entering and locating the table of significance that related directly to the numbers of the target subjects being involved ( $N$ ).

In Spearman, locating the level of significance will be the same with the total numbers of the subjects being studied. So, if the numbers of subjects ( $N$ ) is 30, the level of significance will be in that row. In contrast, with *Pearson* correlation,  $N$  must be subtracted firstly by 2 to find out  $df$  ( $df = N - 2$ ), which stands for degrees of freedom. For example, if the subjects being involved ( $N$ ) are 30 people, finding out the level of significance must be in the table of significance at item number 28 (since  $30-2$  is 28). Still at the same row, the value of significance, especially in this study, is exactly at the intersection with the level of .05 at the above column.

After locating the value, it will come into mind whether there is a correlation or not. If the value of  $r$  is equal or higher than the value of significance, it means that the two variables can be said significantly correlated and vice versa.

After finding out whether there was a significant correlation or not, the next step was interpreting and discussing the findings which are elaborated in the next chapter. The findings will be interpreted in the light of reading theories that have been previously discussed.