CHAPTER III
RESEARCH METHODOLOGY

3.1 Introduction
This chapter discusses overview of the methodological aspects of the study. The overview comprises of seven parts; the research design, population and sampling procedures, research instruments (including a battery of self report measures and a semi-structured interview), validity and reliability of the instrument, data analysis, and concluding remarks.

3.2 Research Design
Related to the research problems mentioned in the first chapter, this study employed descriptive analysis method by using an ex post facto design. The study employed a descriptive method as it “involves collecting data in order to test hypothesis or to answer questions concerning the current status of the subject” (Gay, 1987). In this study, the researcher collected data in order to test hypothesis that there is no significant correlation between the students’ test anxiety and their academic performance. Meanwhile, an ex post facto design was used because – as stated by Ary et al. (1999, p. 298) - it is suitable to investigate attribute variables – a kind of characteristics that should be possessed by the participants of study before a researcher begins his or her study. In this study, students’ test anxiety was an attribute variable, so that the variable could not be manipulated since it has been occurred in nature.

More importantly, this research design was considered appropriate for this study because it was in line with the research objectives of this study, which were to find out the test anxiety levels experienced by EFL students, factors that have caused EFL students’ test anxiety, and correlation between students’ test anxiety (as independent variable) and their academic performance (as dependent variable). In this study, the researcher could not manipulate the independent variable by giving a treatment – as stated by Hatch and Lazaraton (1991, p. 99). Therefore, an
ex post facto design was considered appropriate to be research design of this study.

3.3 Population and Sampling Procedures

The populations of this study were the undergraduate students at a private higher education institution in Bandung. The population was taken because of two reasons. The first reason is its accessibility – as proposed by Kvale (1996). Since the researcher is one of the teachers there, she gets a permission to do the research there. The second reason is because the population can portray the varieties of abilities of the undergraduate students. It is related to their heterogeneous background and intelligence.

To determine the amount of samples involved in this study, a simple random sampling was used. By using this sampling technique, each student had an equal opportunity to be selected randomly and be representative of the population (Fraenkel and Wallen, 2007). Accordingly, the researcher randomly took 20% of the total number of students (541 students). So, about 108 undergraduate students at this higher education institution became the sample of this study. However, because of several things, such as rejection from several students to be the sample of this study, students’ leave-taking from this higher education institution, and their transfer from regular class to staff class, the sample of this study came to be about 93 students.

3.4 Research Instruments

In order to obtain adequate data for the study, the instruments used in this study were a battery of self report measures and a semi-structured interview. The self report measures included Spielberger’s Test Anxiety Inventory (TAI) developed in 1980 and a Test Anxiety Inventory retrieved and adapted from Academic Centers for Excellence. The first self report measure was employed to find out the test anxiety levels experienced by EFL students, whereas the second one was used to find out the factors that have caused EFL students’ test anxiety.
3.4.1 A Battery of Self Report Measures

The first self-report measure was Spielberger’s Test Anxiety Inventory (TAI) that was developed in 1980. It is already elaborated in the second chapter. The form of Spielberger’s TAI can be seen in Appendix 1.

Meanwhile, the second self-report measure, which was a Test Anxiety Inventory retrieved and adapted from Academic Centers for Excellence, was in response to the situation and condition of the participants. The adaptation was done based on the judgment from experts; considering the clarity, efficiency, and practicality of the self-report measure and also considering the characteristics of the participants and in line with the problems investigated. The self-report measure comprises of 23 statements that reflect the participants’ experience in test taking. If the statement reflected their experience in test taking, they must place a check mark on the line next to the number of the statement. The individual scores of this self-report measure were determined by circling the numbers to the statements that they had checked in the inventory. The areas they had answered “yes” to the most would help them identify the causes of their test anxiety in terms of test taker characteristics. The adapted self-report measure is enclosed in Appendix 2.

Regarding a battery of self-report measures employed in this study, before asking the participants to fill in the self-report measures, several clear instructions were given in both oral and written way. To avoid the participants’ misunderstanding, the self-report measures were also available in Bahasa Indonesia. The translation of self-report measures from English to Bahasa Indonesia was done under the suggestion of the experts. The translated self-report measures are presented in Appendix 3.
The battery of self report measures itself was employed in this study in the classroom; there were eight classes observed for this study and it was done in July 2012, specifically from 2nd to 5th of July, 2012.

3.4.2 A Semi-structured Interview

Considering the clarity of the self report measures and the characteristics of the participants investigated, a semi-structured interview was conducted. In this study, to capture the richness and detail of test anxiety experienced by the students, a semi-structured interview was conducted on a one-to-one and face-to-face basis and recorded (as suggested by Silverman, 2005) to provide a permanent record. A semi-structured interview was considered to be the best to conduct in this study because this type of interview lets the interviewer off to look into the interviewee to be more detailed in giving responses or to follow a line of inquiry planned previously. The more responses interviewee gives, the higher possibility for the interviewer to get the data required (Heigham and Croker, 2009).

Furthermore, this type of interview offers “sufficient flexibility to probe some aspects in depth and, where necessary, to let the respondent lead in much the same way as in open interview” (Heigham and Croker, 2009, p. 186). In this regard, Merriam (1998, p. 74) says that the more flexibility the semi-structured interview offers is because of the use of less structured questions. In line with this, Gall et al. (2003, p. 240) state that “the semi-structured interview involves asking a series of structured questions and then probing more deeply using open-form questions to obtain additional information.” A list of questions that was formulated for the interview can be seen in Appendix 4.

To conduct a semi-structured interview, the interviewer focused on the analysis of self reports of the students’ test anxiety and recalled of how their test anxiety took place. Thus, based on the results of Spielberger’s TAI 1980, two EFL students who exhibited high test anxiety were interviewed according to a set of semi-structured questions to find out factors that have caused their test anxiety, how strong was each of those causal factors of test anxiety for them, and how
were their actions to each of those causal factors. In addition, two EFL students who exhibited moderate test anxiety and two EFL students who exhibited low test anxiety were interviewed to compare with the highly anxious students. The time required for the interview ranged from 11 to 20 minutes.

The detail schedules for the interview can be seen as follows.

<table>
<thead>
<tr>
<th>No.</th>
<th>Student/Subject</th>
<th>Date</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Student 1 (a student with high anxiety)</td>
<td>1st of Oct, 2012</td>
<td>12:19 minutes</td>
</tr>
<tr>
<td>2.</td>
<td>Student 2 (a student with moderate anxiety)</td>
<td>2nd of Oct, 2012</td>
<td>20:24 minutes</td>
</tr>
<tr>
<td>3.</td>
<td>Student 3 (a student with low anxiety)</td>
<td>2nd of Oct, 2012</td>
<td>11:31 minutes</td>
</tr>
<tr>
<td>4.</td>
<td>Student 4 (a student with moderate anxiety)</td>
<td>2nd of Oct, 2012</td>
<td>15:51 minutes</td>
</tr>
<tr>
<td>5.</td>
<td>Student 5 (a student with low anxiety)</td>
<td>2nd of Oct, 2012</td>
<td>12:56 minutes</td>
</tr>
<tr>
<td>6.</td>
<td>Student 6 (a student with high anxiety)</td>
<td>4th of Oct, 2012</td>
<td>10:30 minutes</td>
</tr>
</tbody>
</table>

Henceforth, the interview data were interpreted to answer research problems and compared to the theory underpinning the study (Emilia, 2008, p. 197).

Aside from the battery of self-report measures and semi-structured interview as instruments of the study, a correlation analyses was used to determine the correlation between two variables. To show the correlation between two variables more easily, as suggested by Hatch and Lazaraton (1991, p. 427), the value of one variable should be plotted against the other variable’s value. In this regard, the students’ test anxiety levels would be the variable X, whereas their academic performance would be the variable Y. These two variables, then, were calculated by using Spearman correlation. Spearman correlation was used in this study because the data were not normally distributed. It was mentioned in Hatch and Lazaraton (1991, p. 451) that most statisticians argue for Spearman correlation.
correlation when the distribution is not normal. The following table illustrates the distribution of data.

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
</tr>
<tr>
<td>Students' academic performance</td>
<td>.266</td>
<td>93</td>
</tr>
<tr>
<td>Students' test anxiety</td>
<td>.087</td>
<td>93</td>
</tr>
</tbody>
</table>

From Table 3.4, specifically Kolmogorov-Smirnov column, it could be seen that the significance value of variable X (students’ test anxiety) was 0.080, while the significance value of variable Y (students’ academic performance) was 0.000. Stated by Pallant (2011, p. 63), a result with Sig. value which is more than 0.05 indicates normality. Therefore, the variable X of this study was normally distributed. However, because the significance value of variable Y was smaller than 0.05, it could be concluded that the data was not normally distributed.

### 3.5 Validity of the Instrument

A good instrument (whether test or non-test) must be valid and reliable (Sugiyono, 2011, p. 169). This is intended to collect the valid and reliable data. Sugiyono (2011, p. 168) notes that by using the valid and reliable instruments in collecting data, it is expected that the result of the study is also valid and reliable.

In terms of validity, basically, validity is divided into three types: content validity, criterion-related validity, and construct validity (Hatch and Farhady, 1982, p. 251). Since the instrument used in this study was a non-test instrument,
which was a battery of self report measures, according to Sugiyono (2011, p. 170), to be stated as a valid instrument, it only needed to fulfill construct validity. Moreover, in this study, the two self report measures were employed to measure attitude.

Dealing with construct validity, Hadi (cited in Sugiyono, 2011, p.170) treats construct validity like logical validity or validity by definition. Thus, an instrument can be stated for having construct validity if it can be used to measure the phenomenon as its definition is. Further, to utter an appropriate definition, the researcher needs several theories to consult with. In this case, Hadi states that when the theories used to utter a definition are already appropriate, the result of measurement using the instrument which is based on those theories is already able to be stated as a valid result.

To test the construct validity, the judgment from experts was employed in this study. For expert judgment, at least three doctorate experts are needed (Sugiyono, 2011, p. 172). Hence, in this study, the judgment was done by two doctorate lecturers from Educational Psychology and Guidance program and a doctorate lecturer from Early Childhood Teacher Education program with counseling as his area of expertise. In this regard, the measured aspects of an instrument (a battery of self report measures, in this study) which were already constructed based on certain theories (i.e. Liebert and Morris, 1967; Morris et al., 1981) were consulted with the experts. Then, the experts judged whether the instrument could be used without revision, it could be used with revision, or it needed to be revised totally.

For the need of expert judgment, the rubric of instruments was developed from the operational definition of research variables. Since the self report measures would be employed to find out the test anxiety levels experienced by EFL students and to find out the factors that have caused EFL students’ test anxiety, the rubric of the instruments was developed from the operational definition of test anxiety (as independent variable) that test anxiety is comprised of two primary components; worry and emotionality (Liebert and Morris, 1967;
Morris et al., 1981). Leaving from the two-component model of test anxiety, the definition of test anxiety for this study is a feeling of apprehension and discomfort accompanied by cognitive difficulties during a test. Therefore, relevant to other theories that have been referred to in this study, for the first self report measure (i.e. Spielberger’s Test Anxiety Inventory (TAI) 1980), its rubric included two dimensions of test anxiety, namely worry and emotionality. Meanwhile, for the second one (i.e. Test Anxiety Inventory retrieved and adapted from Academic Centers for Excellence), its rubric included four sources of test anxiety; 1) concerning other people’s view, 2) concerning self-image, 3) concerning future security, 4) concerning preparation.

The rubric for those two self report measures before expert judgment is enclosed in Appendix 5. Meanwhile, the rubric for those two self report measures after expert judgment is enclosed in Appendix 6.

In short, the result of expert judgment is in the following.

<table>
<thead>
<tr>
<th>Table 3.3</th>
<th>The Result of Expert Judgment for the Second Self Report Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conclusion</td>
<td>No. of Items</td>
</tr>
<tr>
<td>Good (without revision)</td>
<td>1, 2, 3, 6, 9, 11, 15, 16, 17, 23, 24, 25, 26, 31, 32, 38, 40, 42, 46, 47, 49</td>
</tr>
<tr>
<td>Need to be revised</td>
<td>18, 33</td>
</tr>
<tr>
<td>Being dropped</td>
<td>4, 5, 7, 8, 10, 12, 13, 14, 19, 20, 21, 22, 27, 28, 29, 30, 34, 35, 36, 37, 39, 41, 43, 44, 45, 48, 50</td>
</tr>
<tr>
<td>Total of items used</td>
<td></td>
</tr>
</tbody>
</table>

Based on the table above, it could be concluded that for the second self report measure, 23 of 50 items available were stated as valid items in terms of construct validity. In other words, the 23 items have represented the indicators of the second self report measure (i.e. Test Anxiety Inventory retrieved and adapted from Academic Centers for Excellence), so the instrument was considered
appropriate to be employed in this study, which was to find out the factors that have caused EFL students’ test anxiety.

Meanwhile, for the first self report measure (i.e. Spielberger’s TAI 1980), its validity has been proved. Widely employed in many studies to evaluate individual differences in test anxiety, TAI has demonstrated its validity in “before” and “after” research project. Spielberger and Vagg; Zeidner et al. (cited in Toubiana, 2005, p. 65) admit that ‘the effectiveness of different therapeutic approaches, such as systematic desensitization and cognitive behavior modification, have been evaluated.’ Therefore, this self report measure was also considered appropriate to be employed in this study, which was to find out the test anxiety levels experienced by EFL students.

3.6 Reliability of the Instrument

Reliability is “the consistency of the scores produced by a measurement tool” (Ross, 2006, p. 2). Meanwhile, Anderson et al. (cited in Arikunto, 2009, p. 87) say that ‘a reliable measure is the one that provides consistent and stable indication of the characteristic being investigated’. In other words, reliability refers to the extent to which measurements are consistent. Regarding reliability, Anderson et al. say that the validity inferred from the instruments is essential - even more crucial than the reliability. Also, it is possible for a reliable instrument to be invalid; on the contrary, a valid instrument is usually reliable. Fink (2003; 2006) also states the same thing that a valid instrument is always reliable.

Based on this assumption, a battery of self report measures employed in this study was considered to be reliable. The Spielberger’s TAI total scores have very high alpha coefficients, ranging between 0.92 and 0.96 (Spielberger, cited in Bushman et al. 2005, p. 80). It means that the TAI has high internal consistency reliability.

3.7 Data Analysis
The data taken from the battery of self report measures (consisting of Spielberger’s Test Anxiety Inventory (TAI) 1980 and a Test Anxiety Inventory retrieved and adapted from Academic Centers for Excellence) were elaborated and analyzed to answer the first and second research problem. The data analysis was organized based on each data collection technique.

In this study, the data that had been collected were analyzed based on the concept of data analysis given by Miles and Huberman (1984) in the following.

The analysis involves (1) selecting, focusing, simplifying, abstracting, and transforming the "raw" data which has been collected; (2) organizing and displaying the data, so as to allow conclusions to be drawn; and (3) drawing conclusions, by noting regularities, patterns, explanations, possible configurations, causal flows and propositions, and verifying those conclusions (pp. 21-22).

The data of Spielberger’s TAI 1980 were analyzed by finding out first the test anxiety levels of students. To find out the test anxiety levels of students, as suggested by Taty (2011), the following steps would be conducted.

1. Determining the ideal maximum score of sample.
   
   The ideal maximum score = total questions x the highest score

2. Determining the ideal minimum score of sample.

   The ideal minimum score = total questions x the lowest score

3. Determining the ideal score range of sample.

   The ideal score range = the ideal maximum score – the ideal minimum score

4. Finding the score interval.

   The score interval = \[
   \frac{\text{the ideal score range}}{3}
   \]

Then, the distribution of test anxiety level would be reflected as follows.

**Table 3.4**

The Distribution of Test Anxiety Levels
Meanwhile, the data of Test Anxiety Inventory retrieved and adapted from Academic Centers for Excellence were analyzed by finding out first the factors that have caused EFL students’ test anxiety. The areas that the participants had answered “yes” to the most would identify the causes of their test anxiety in terms of test taker characteristics.

The analysis of interview was conducted through several steps. The first step was transcribing (syntactic or orthographic transcription) the data from spoken (recorded form) to become written form. The second step was coding. The data from interview were coded according to Gail Jefferson (cited in Wray et al. 2001, pp. 202-211). The key of the convention symbols used in this study is as follows.

<table>
<thead>
<tr>
<th>No.</th>
<th>Interval</th>
<th>Test Anxiety Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20 – 40</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>40 – 60</td>
<td>Moderate</td>
</tr>
<tr>
<td>3</td>
<td>60 – 80</td>
<td>High</td>
</tr>
</tbody>
</table>

T    : Teacher  
S    : Student  
Ss   : More than one student speaks  
((pause)) : It is used if the pause occurs within a speaker’s turn.  
((gap)) : It is used if the pause occurs between different speakers’ turns.  
: A colon will be used immediately after the lengthened sound (in certain syllable), the longer the sound, the more colons will be used.  
= : This latching, equals sign, is used at the end of the first component and the beginning of the second one.  
(mekka bunit cor) : It is used when the writer can hear the sounds but cannot work out what the words are.  
// : A double oblique line put into the established speaker’s line at the point of interruption, and start the new speaker on the next line.  
[] : A double opening square bracket is used when more than one speaker speak.  
Italic : If the language used is not English, the transcription will be written in italic.  
() : The bracket is used to give more explanation on what the students do in responding the questions (transcriber’s comments), e.g. (nod).
Then, the data were categorized. The third step was counting the data, and the last one was data analysis. It was expected that interview could open up more details about some of the specific things that usually bother the participants of the study in testing situation, especially the causes of their test anxiety in terms of test characteristics.

To answer the third research question, which was *is there any relationship between students’ test anxiety and their academic performance?*, a correlation analyses was used. Thus, the test anxiety levels experienced by EFL students were correlated to the students’ GPA. In this regard, the Statistical Package for Social Sciences (SPSS) version 17.0 was used to analyze and calculate the data. This tool was considered to be much more practical and efficient, and provide more accurate analysis result. Accordingly, the degree of correlation which is generally expressed in $\rho$ (Rho) was obtained. Then, this indicated both direction (whether positive or negative) and strength of the relationship between the two variables. A positive correlation indicated that as one variable increased, so did the other. Meanwhile, a negative correlation indicated that as one variable increased, the other decreased.

Henceforth, to determine the strength of the relationship between the two variables, the researcher applied Cohen’s suggestion (1988, pp. 79–81) that correlation results can be interpreted as follows.

- $r = 0.10$ to $0.29$ (small correlation)
- $r = 0.30$ to $0.49$ (medium correlation)
- $r = 0.50$ to $1.00$ (high correlation).

Other than the result of Pearson correlation, these interpretations could be applied for the result of Spearman correlation. Pallant (2011, p. 133) states that the way to interpret the result from the parametric and non-parametric approaches is not different. Furthermore, whether or not there is a negative sign out the front of $r/\rho$ value, the interpretations apply. For instance, the strength of correlation of $r/\rho = 0.5$ and $r/\rho = -0.5$ is the same, but in a different direction.
The last thing to do was assessing the significance level (listed as **Sig. 2 tailed**). The level of statistical significance indicated how much confidence the researcher should have in the results of study. More importantly, it indicated whether the null hypothesis was rejected or accepted based on the certain level of significance. O’neil (2009, p. 32) mentions that in SPSS, the *p*-value tells us whether results are statistically significant or not. Besides, for most research, the significance level of 0.05 or 0.01 is used. Thus, if the *p*-value is smaller than 0.05, the results will be statistically significant by 0.05. On the contrary, if the *p*-value is larger than 0.05, the results will not generally be accepted as “significant”. Finally, in this study, the null hypothesis was rejected when the *p*-value was smaller than 0.05. Meanwhile, when the *p*-value was larger than 0.05, the null hypothesis was accepted.

### 3.8 Concluding Remarks

This chapter explains how the data in this study were collected and analysed. The primary data were taken from a battery of self report measures (including Spielberger’s Test Anxiety Inventory (TAI) 1980 and a Test Anxiety Inventory retrieved and adapted from Academic Centers for Excellence) that was employed in the classroom - there were eight classes observed for this study - and was done in July 2012, specifically from 2nd to 5th of July, 2012.

Meanwhile, the secondary data were from the interview with six students chosen based on the results of Spielberger’s TAI 1980 (i.e. two EFL students who exhibited high test anxiety, two EFL students who exhibited moderate test anxiety, and two EFL students who exhibited low test anxiety). Then, the data gathered from interview were transcribed, coded, categorized, counted, and analyzed.

Henceforth, the data of Test Anxiety Inventory retrieved and adapted from Academic Centers for Excellence were analyzed by finding out first the factors that have caused EFL students’ test anxiety. The areas that the participants had
answered “yes” to the most would identify the causes of their test anxiety in terms of test taker characteristics.

Furthermore, a correlation analyses was used to identify the degree of correlation between students’ test anxiety and their academic performance. Regarding this, the Statistical Package for Social Sciences (SPSS) version 17.0 was used to analyze and calculate the data.

This study employed descriptive analysis method by using an ex post facto design. This study employed a descriptive method since the researcher collected data in order to test hypothesis. Meanwhile, an ex post facto design was used because it was suitable to investigate attribute variable (i.e. students’ test anxiety). Overall, this research design was considered appropriate for this study because it was in line with the research objectives of this study.