CHAPTER III
METHODOLOGY

Conducting a research in error analysis using an AWE program takes a full consideration on the research methodology selection. The selection of research methodology becomes the core of the present study since the selected methodology can influence the processes of the study and outcomes of the study; different methodologies applied will result in different outcomes. Therefore, the research methodology selected in this study was chosen based on the considerations which are in line with the objectives of the study. To give a clearer explanation, this chapter describes the research methodology applied in this study. This chapter consists of five sections explaining the process of how the study was conducted, including the research design, the participants, the procedure of data collection, the procedure of data analysis, and the research ethics.

3.1 Research Design

The design of the research becomes the first consideration taken in conducting this study. The selection of the approach needs to fit the objectives of the study which are to gain a depth understanding into the nature of errors made by the students and to explore the strengths and weaknesses of applying an AWE program for error analysis. Reflecting on the objectives proposed at the beginning of this thesis, the present study applied a qualitative descriptive research.

The qualitative descriptive research design was employed in this study to gain a depth understanding into the nature of errors made by the students and the strengths and weaknesses of AWE program as an error analyzer program. Concerning with the depth understanding into the phenomenon, the selection of this research design was derived from Creswell (2012) who explained that one of the characteristic of this research design is "exploring a problem and developing a detailed understanding of a central phenomenon" (p.16). In exploring the problem, or the phenomenon, the qualitative descriptive research design allowed the study to collect the data from various data sources, including documents, field notes, and participant interviews (Merriam, 2009) which can reveal a detailed finding for the data analysis. Moreover, this research design analyzes the finding by

identifying patterns of the data and classifying them into certain themes (Merriam, 2009) which can ease the process of error identification and AWE program's strengths and weaknesses identification. Furthermore, in reporting the data, the qualitative descriptive research design could provide a detailed description of the phenomenon investigated (Creswell, 2012). Therefore, by conducting this research design, a description of the phenomenon occurring in students' writing errors and the AWE program's strengths and weaknesses can be provided in depth.

3.2 Participants

The second consideration in conducting this study is the selection of participants. In conducting a study, the selection of the participants should fit the study contexts and objectives. Noticing that this study specifies the focus on analyzing the errors occurring in EFL undergraduate students' writings, this section explains the selection of participants, including the sampling technique, the participants' criteria, and the total number of participants.

To select the suitable participants, purposeful sampling is used in gathering the participants in this study. This technique was chosen because it is one of the most feasible ways to gather the participants who fit the context of the study. As what Merriam (2009) has mentioned in her book, this sampling strategy "involves locating a few key participants who easily meet the criteria which have established for participation in the study" (p. 79). By considering the participants based on a set of criteria, this sampling strategy can help to limit the scope of the study and eliminating unrelated data from the participants.

In this study, the second year undergraduate students of English Language Education department in a university in Bandung were considered to be the key participants (see Merriam, 2009) regarding their academic background. From their academic background, there are two criteria used to select the participants. Firstly, these students should have learned to write in a formal context. This first criterion becomes a justification for their capability in English writing and it also reflects their English writing proficiency levels. In the university, where this study took place, the teaching of formal writing is studied from the third semester which is

the *Writing in a Professional Context* course. As the study was taken in the even semester, in which the formal writing course offered is the *Writing for Academic Purposes 1* course, the participants of this study should at least have taken *Writing in a Professional Context* course and enrolled at *Writing for Academic Purposes 1* course. Secondly, these students should have learned English grammar at an advanced level. This second criterion becomes a justification on these students' understanding of the English Grammar basic. This criterion also reflects the students' background knowledge of English Grammar at an advanced level. In this university, the *Advanced Grammar* course was taught in third semester. Therefore, the participants of this study should have passed the *Advanced Grammar* course which was taken at the third semester. With these two criteria, the participants of this study can be limited to specific students who have the ability to write in a formal context and have the understanding of English grammar at an advanced level.

By taking the two criteria, the second-year undergraduate students were sorted based on their academic backgrounds. According to the students' data, there was 69 second-year undergraduate students in total. From these 69 students, there were 68 students who enrolled at *Writing for Academic Purposes 1* course which are divided into four classes; A1 consists of 15 students, A2 consists of 16 students, B1 consists of 18 students, and B2 consists of 19 students. These students were carefully checked on their Grammar advanced course, but, reflecting on students' confidential agreement, the total number of the students who passed the *Advanced Grammar* course cannot be published. After sorting these students based on their academic background, the rest of the qualified students were asked to voluntarily participate in this study. As the result, the total number of 48 students participated in this study; there are 15 students from A1, 16 students from A2, 7 students from B1, and 10 students from B2.

3.3 Data Collection

The process of data collection becomes the third consideration in this study. Data collection is a set of techniques for collecting the data which are selected and designed in order to seek for the information to the answer of the research

questions addressed in the study. In this section, the procedure of data collection is explained through subsections including the process of gaining access to participants and research site, the types of data collected in the study, and the procedures of data collection conducted.

3.3.1 Participants and Research Site Access Gaining

Before collecting the data, participants and research site selections were conducted. The selections were conducted purposively in order to achieve the goal of the study. It is in line with what has been mentioned by Creswell (2012), the success of participants and research site identifications results in an appropriate line connected to the research problem investigated in this study. As the result, the purposeful sampling was applied in selecting the participants and research site of the study to fulfill the needs of the study. In the previous section (Section 3.2), the complete justification of participants and research site selection has been explained. Meanwhile, this sub-section only focuses on explaining the administrative and procedural steps in gaining access to the participants and research site; i.e. asking permission to the head of the department, gaining access to the classroom, and asking the students to participate in the study.

The process of gaining access to the participants and research site was started by asking the permission to conduct the study from the head of the department. A letter of permission to collect the data was written and sent to the university where the study took place. Receiving a confirmation from the head of the department, a meeting was arranged with the head of the department to discuss the purpose of the study, the scope of the study, and the expected result of the study. From the result of the discussion, the head of the department gave the permission to conduct the study and asked to arrange a meeting with the lecturer who taught *Writing for Academic Purpose 1* course.

Having the permission from the head of the department, the second step, which is to gain the access to the classroom, was conducted. The lecturer who taught *Writing for Academic Purpose 1* course was contacted to discuss the purpose of the study and the process of data collection. From the result of the discussion with the lecturer, it was gained two essential information related to the

course; first, the *Writing for Academic Purpose 1* course was administered in four classes (A1, A2, B1, and B2) and second, the schedule of these classes are on Tuesday (B1 and B2) and Wednesday (A1 and A2). Thus, the lecturer gave the access to come into the classroom and select the participants of the study.

With the access given by the lecturer, finally, the last step of participants' selection was taken. The researcher came to each classroom and informed about the purpose of the study to the students. These students were asked to participate in the study and their consents were taken through a form of consent. This participants gathering process is a volunteer-based process in which the students have their free will to join the study without any pressure, as what has been mentioned in Section 3.2. As the result, the total number of 48 students participated in this study; they are 15 students from A1, 16 students from A2, 7 students from B1, and 10 students from B2.

3.3.2 The Data

Once the research questions have been proposed, the types of the data needed in this study can be identified. As one issue of data collection, the types of data collected in the study was limited to any sources of information which can answer the research questions. Noticing the questions of the study finding out the types of errors, the students' propensity of producing error, the possible causes of errors, and the strengths and weaknesses of applying Grammarly program for analyzing errors, the data collected from the participants in this study is limited to two kinds, including students' writings and students' answers on the interview.

Students' writings became the first data collected in this study. These writings were the students' assignments which were written in the previous semester, in the *Writing in a Professional Context* course. These writings were in a form of the essay with the total number of the word is around 500 to 1000 words. These writings were the final draft which had been previously revised and discussed with the lecturers for three to four times and have also been checked through peer-review. There are three genres of writings produced by the students, i.e. report, discussion, and explanation, and each student only submitted one writing in which they viewed as the most comprehensive one. These writings are

the primary data of this study since the students' writings became the main source for identifying the types and the possible causes of errors. These writings were re-submitted to Grammarly program to be analyzed and evaluated. Thus, analyzing the students' writings can reveal all the research questions' answer: the first and the second research questions which are the types of errors, the third research question which is the students' propensity of producing error, and the fourth research question which is the possible cause of errors, and fifth research question which is the strengths and weaknesses of the program.

On the other hand, the second data collected in this study was the result of the interview with 15 selected students. This interview was conducted to seek the possible causes and the factors influencing the error production by identifying the process of writing production conducted by the students and the students' error awareness. The interview was conducted after having the result of error analysis and the participants for this interview were limited to the students who have the most varieties of errors on their writings. There are 15 students who were invited to the interview to share the process of composing their writing and also to clarify their error awareness on their writings. As the result, the students' answers in the interview became the second data for justifying the possible cause of errors. Thus, the students' answers also became the answer to the fourth research question in this study.

3.3.3 Collection Procedures

After identifying the needs of the data, the process of data collection can be started. In this study, the process of data collection consisted of several techniques in taking the data from the participants. Some of these techniques required an instrument as a tool in collecting the data and thus, the instrument has been prepared and validated. The suggestions and revisions given were applied to the final version of the instrument and the final version of the instrument was applied in collecting the data. To make clearer insights on how the data was collected, the process of data collection in the present study is explained based on the kinds of data explored which are divided into two sections, i.e. students' writings and interview.

3.3.3.1 Students' Writings

As the primary data of the study, the students' writings are the first data taken in this study. After gaining the classroom access from the lecturer, the researcher came into each classroom to meet the students. The students, who participate in the study, were informed about the writing submission. Each student was asked to submit a soft file (*.doc* or *.docx* document) of an essay which they have written in the previous semester, in the *Writing in a Professional Context* course, as what has been discussed in Sub-Section 3.3.2. With the help of the classroom coordinator, the students' writings were collected and submitted through e-mail. Receiving the students' writings, each writing soft file was renamed by using *Text* as the name file and followed with a number, e.g. *Text1.doc*. Renaming files can help in sorting the students' writings based on the submission order. Besides, a title log is also created to record the title of students' writings which helps in listing the title of writing to the file name (see **Appendix 1**). Thus, these students writings become the first data collected in this study (see **Appendix 2**).

3.3.3.2 Interview

The second data taken in this study is the result of the interview with the students. The interview was conducted in this study since it is one of the useful techniques to seek information on individual's personal experiences and to gain insights into certain issues (Hamied, 2017). As what has been mentioned on the previous sub-section (see Sub-Section 3.3.2), in this study, this interview was conducted to seek the possible cause and the factors influencing the error production by identifying the process of writing production conducted by the students and the students' error awareness. To give clearer insights, this sub-section explains the process of how the set of interview guide was constructed and how the interview was conducted.

Before conducting the interview, an interview guide was constructed. This interview guide contains a set of questions which are used to gain the information from the students. The questions provided in the interview guide were the result of adaptation and modification from Brown's book (2004) on the process of writing

production. The total four key questions were created and each question has several prompts for further questions. To avoid the ambiguity in interpreting the questions (Hamied, 2017), the questions asked to the participants were constructed in participant's first language and thus the participants can respond to the interviewer more accurate. Therefore, the interview questions were constructed in the Indonesian language since the interviewees were Indonesian EFL university students (see **Appendix 9**).

After constructing the interview guide, the interview with the students was conducted. The students were firstly sorted based on the type of errors detected on their writings. As what has been mentioned on the previous sub-section (see Sub-Section 3.3.2), the students who became the interviewees were selected based on the varieties of the error types. With this criterion, it was selected 15 students who have the highest varieties of the error types, i.e. 4 students came from 4A1, 6 students came from 4A2, 2 students came from 4B1, and 3 students came from 4B2. These students were contacted and confirmed on their participation in the interview. With the help of the class coordinators, several meetings was arranged for the interview in which the interview was conducted in different time for each class; the interview with 4A1 at April, 19th 2018, 4B1 and 4B2 at April, 20th 2018, and 4A2 at April, 23rd 2018. Each of the students got their pseudonyms (see **Appendix 11**). During the interview, the students' voices were being recorded with a tape recorder and the transcriptions of the interview were created at the end of the interview (see **Appendix 12**). The result of transcriptions became the database of the data analysis of the interview (Merriam, 2009) and thus, the result of the interview becomes the additional data for justifying the error production and clarifying the possible causes of errors occurring in students' writings.

3.4 Data Analysis

The process of data analysis is the last part of the research methodology in this study. In general, the data analysis consists of a set of techniques in analyzing the data which are selected and designed in order to answer the research questions addressed in the study. Since there are five research questions proposed in the study, the process of data analysis is divided into five parts: the types of detected

errors, the types of undetected errors, the students' propensity of producing error, the possible causes of errors, and the strengths and weaknesses of Grammarly program as an error analyzer. Therefore, this section of data analysis explains the procedure of data analysis which is divided into five sub-sections based on the research questions addressed in the study.

3.4.1 Types of Detected Errors

Analyzing the types of detected errors is the first part of the data analysis conducted in this study. This analysis was conducted in order to answer the first research question of the study, which is to identify the types of errors detected by Grammarly program. In this part of data analysis, the students' writings become the main source of the data. The students' writings were submitted to Grammarly program and evaluated by the program. The result of Grammarly evaluation revealed the errors occurring in students' writings and thus, these errors found in students' writings were analyzed to find its types.

In this phase, each of errors found by the Grammarly program was classified based on Linguistic Category and Surface Structure Taxonomies. Firstly, the process of the error classification based on Linguistic Category Taxonomy was conducted by classifying the errors based on the explanation provided by the Grammarly program in the feedback box. Secondly, after classifying the types of error based on the Linguistic Category Taxonomy, the classification process was continued to the Surface Structure Taxonomy classification which classifies the errors into four types: omission, addition, misformation, and misordering (Dulay, Burt, & Krashen, 1982).

As the errors have been classified, the results of the types of detected errors found were recorded in a writing error log. This writing error log was adapted from Hirschel's grammar log (2011). In this study, the log was modified and it consisted of five columns: 1) the first column displays the sentence which consists of errors made by the students, 2) the second column displays the correction by the Grammarly program evaluation, 3) the third column displays the types of errors based on the linguistic category from Grammarly program's explanation, 4) the fourth column displays the types of errors based on Surface

Structure Taxonomy, and 5) the fifth column displays the source of the text, each student's writing's file name, e.g. *"Text 1"*.

After the errors being recorded into the writing error log, two inter-raters were invited to re-check and re-evaluate the analysis result of the detected errors. Further discussion on the errors found was also conducted between the researcher and the inter-raters. After receiving revision and suggestion, the result of the analysis of the detected errors was revised. As the result, a list of the types of the detected errors was created.

In displaying the result of data analysis, the list of the types of detected errors is displayed in a form of a table. This table consists of eight columns: the first column displays the number of errors type found, the second column displays the types of errors based on the Linguistic Category Taxonomy (the Grammarly program's explanation), the third to the sixth columns display the types of errors based on Surface Structure Taxonomy's classification, the seventh column displays the frequency of error production, and the eighth column displays the percentage of the frequency of error production. Furthermore, a further interpretation is also made in line with the types of errors detected by Grammarly program.

3.4.2 Types of Undetected Errors

Analyzing the types of undetected errors is the second part of the data analysis conducted in this study. This analysis was conducted in order to answer the second research question proposed in the study. This analysis also uses the students' writings as the main source of the data, but the process of the data analysis did not involve the AWE program utilization; in this part of data analysis, the students' writings were analyzed manually to identify any errors which are left undetected by the Grammarly program. Meanwhile, the process of the error type analysis was similar to previous data analysis which uses Linguistic Category and Surface Structure Taxonomies to classify the errors.

Firstly, the process of the error classification based on Linguistic Category Taxonomy was conducted by analyzing the linguistic item in the English language structure which is affected by the error. Secondly, after classifying the types based

on the linguistic category, the classification process continued to the Surface Structure Taxonomy classification which classifies the errors into four types: omission, addition, misformation, and misordering (Dulay, Burt, & Krashen, 1982). As the errors have been classified, the results of the types of undetected errors found were also recorded in a writing error log. This writing error log was a modification of previous error log; the previous error log is added with one column which displays the undetected errors.

After the undetected errors being recorded into the writing error log, two inter-raters were invited to re-check and re-evaluate the analysis result of the undetected errors. Further discussion on the undetected errors found was also conducted between the researcher and the inter-raters. After receiving revision and suggestion, the result of the analysis of the undetected errors was revised. As the result, a list of the types of the undetected errors was created.

Similar to previous data analysis, the list of the types of detected errors is displayed in a form of a table. In this context, the table also consists of eight columns: the first column displays the number of error type found, the second column displays the types of errors based on the Linguistic Category Taxonomy, the third to the sixth column display the types of errors based on its surface structure (Surface Structure Taxonomy's classification), the seventh column displays the frequency of error production, and the eighth column displays the percentage of the frequency of error production. Furthermore, a further interpretation is also made in line with the types of errors which are left undetected by Grammarly program.

3.4.3 Students' Propensity of Producing Errors

Analyzing the students' propensity of producing errors is the third part of the data analysis conducted in this study. This analysis was conducted in order to answer the third research question of the study and also to identify the tendency of errors produced by the students in their writings. In this study, the process of the analysis was conducted by listing the types of errors found in students' writing and summing up the total of errors found.

Firstly, the list of the detected and the undetected errors types were combined to create a new list of errors produced by the students. Any similar types of errors which were occurred in both detected and undetected errors were reduced to avoid double listing. After listing the types of error, the total of each error type was calculated by summing up the number of detected errors and undetected errors under each category. As the result, a list of errors produced by the students in writing was created.

In displaying the data analysis result, the list of errors produced by the students in writing is displayed in a form of a table. In this context, the table consists of four columns: the first column displays the number of error type found, the second column displays the types of errors based on the Linguistic Category Taxonomy, the third column displays the total of errors produced by the students, and the fourth column displays the percentage of the frequency of error production. Furthermore, a further interpretation is also made in line with the students' propensity of producing errors.

3.4.4 Possible Causes of Errors

Analyzing the possible causes of errors is the fourth part of the data analysis conducted in this study. This analysis was conducted to answer the fourth research question of the study which is to identify the possible causes of errors occurring in students' writing. In this part of data analysis, the students' writings and the students' answers on interview became the source of data for the possible error causes data analysis. In data analysis, the identification of the possible cause of the errors was conducted through three processes.

Firstly, the process of identifying the possible causes of errors was conducted using Richards' classification (1971). As has been stated in Chapter II (see Section 2.2), this study employed two of three classifications proposed by Richards (1971) which is the interlingual error and intralingual error; the third classification, the developmental error, was omitted since these students have learned and acquired English grammar knowledge at an advanced level. By using these classifications, the errors found were classified into some possible causes, whether intralingual or interlingual error.

Secondly, after sorting the errors based on its causes, the result of possible error causes analysis was validated with the result of interview regarding the factor influencing the error production. As the result, a list of possible error causes was created and it was recorded into the writing error log; this writing error log is a modification of the previous writing error log which has been added with one column; this column displays the possible cause of errors (see **Appendix 3**).

Lastly, two inter-raters were invited to re-check and re-evaluate the analysis result. Further discussion on the possible causes of errors found was also conducted between the researcher and the inter-raters. After the discussion, the inter-raters gave revision and suggestion toward the result of the analysis of the possible causes of errors. After being revised, a list of possible causes of errors was created.

In displaying the result of data analysis, the list of the possible causes of errors is displayed in a form of a table. This table consists of fifth columns; the first column displays the number of error type found, the second column displays the types of errors based on the linguistic category (Grammarly program's classification), the third column displays the number of intralingual error found, the fourth column displays the number of interlingual error found, and the fifth column displays the frequency of the possible cause of error. Furthermore, a further interpretation was also made in line with the possible causes of errors occurring in the EFL university students' writings.

3.4.5 Grammarly Program's Strengths and Weaknesses

Analyzing the strengths and weaknesses of Grammarly program is the last part of the data analysis conducted in this study. This analysis was conducted to identify the strengths and the weaknesses in utilizing the Grammarly program as an error analyzer. In this part of data analysis, the result of Grammarly evaluation became the main source of the data in which the present study re-evaluated the evaluation to specifically find its accuracy and deviancy in detecting errors. Any findings related to its strengths and weaknesses were collected and decoded into certain themes; any detailed and correct evaluations were classified as its strengths and any undetailed evaluation, misleading identification, and undetected errors were

classified as its weaknesses. Furthermore, each theme of the evaluation result was described and interpreted, completed with the case found in the study. As the result, a description of the strengths and weaknesses of Grammarly program was created.

3.5 Research Ethics

The ethical issue in conducting a study is one of the major concerns gotten into the researcher's anticipation. The consideration of ethical issues which might arise during the study needs to be taken into account since it can influence the credibility, authenticity, and trustworthiness of one study (Creswell, 2014). As what has been mentioned by Creswell (2014), a researcher needs to concern with their ethical issues in conducting a study, especially in dealing with the permission, participants, and research methodologies.

Concerning with the potential issues in conducting the study, there were 10 ethical issues has been identified during the study. These ethical issues occurred in five parts of the process of this study, including the prior to conducting the study, beginning of the study, the process of collecting the data, the process of analyzing the data, and the process of reporting, sharing, and sorting the data (**Table 3.1**).

Table 3.1 clearly displays the process of how the ethical issues found in the study were addressed. Each of ethical issue found was addressed by fulfilling the requirement to conduct the study, building positive relationships with the participants by keeping their personal information safe and not giving any pressure to them in participating in the study, analyzing the data accurately, reporting the result of the study honestly, and avoiding any plagiarism acts. Thus, each of ethical issue has been addressed properly and it indicates that the present study has fulfilled the ethics in conducting the study.

Table 3.1
Ethical Issues Addressed in This Study

Where in the Process of Research the Ethical Issue Occurs	Type of Ethical Issue	How to Address the Issue
1. Prior to conducting the study	a) Seek university approval.	1) Hold a proposal seminar 2) Submit the revised proposal to Postgraduate School's staff.
	b) Seek sponsor approval.	1) Submit a copy of the research proposal to LPDP.
	c) Gain the research site's approval.	1) Arrange a meeting with the head of the department. 2) Arrange a meeting with the lecturer who teaches academic writing.
2. Beginning of the study	a) Avoid giving any pressure to the participants to participate the study	1) Inform the participants that they are not forced to participate in the study.
3. Process of collecting the data	a) Keep the participants' personal information	1) Remove the participants' name, students' number, and class on their writing file.
	b) Concern with participants' confidential to be interviewed and avoid forcing them to be interviewed	1) Inform the participants that they are not forced to participate in the interview. 2) Ask the participants' approval on recording their voice.
	c) Avoid exploiting the participants only.	1) Giving reward for the participants.
4. Process of analyzing the data	a) Concern with the data analysis accuracy.	1) Use the framework in analyzing the data. 2) Have an expert to re-check the results of data analysis.
5. Process of reporting, sharing, and storing the data	a) Avoid hiding the unexpected result.	1) Report the results honestly.
	b) Avoid plagiarism	1) Paraphrase the sentences quoting the theory and/or findings from previous study.