Chapter Three

Data Collection Methodology and Research Design

3.1. Introduction
   3.1.1. Overview of the Chapter
3.2. Research Subjects
3.3. Instrumentation
   3.3.1. Designing the Decision-Making Questionnaire
   3.3.2. Designing the Authenticity Questionnaire
3.4. Reflective Tools
   3.4.1. Audio Recording
   3.4.2. Peer Observation
   3.4.3. Teacher Diary
   3.4.4. Students’ Feedback
3.5. Piloting and Trying out the Instruments
3.6. Validity of the Instruments
   3.6.1. Content Validity
   3.6.2. Criterion Validity: Predictive Validity
   3.6.3. Construct Validity
3.7. Reliability of the Instruments
3.8. Research Design
3.9. Data Collection and Data Analysis Procedures

3.9.1. Stage One
3.9.2. Stage Two
3.9.3. Stage Three
3.9.4. Stage Four
3.9.5. Stage Five
3.9.6. Statistical Methods
3.9.7. SPSS Software

3.10. Summary
3.1. Introduction

In the title of this thesis, two abbreviations—ESL and EFL have been used, which stand for English as a second language and English as a foreign language respectively. EFL has been used to refer to a context in which English language is considered as a school and an academic subject and it is not practically used in society, e.g., in a country like Iran. What remains somehow vague and needs elaboration is the controversial issue of referring to India as a context in which English is used as a second language.

According to Macmillan English Dictionary, second language is “a language that you can speak but which is not your main language.” As far as the general meaning of the term is concerned, ESL can refer to all contexts in which English is used, whether in the class or out of it, or whether accepted as the official language of the country or not. It is evident that the general meaning of the term is not the concern of this study, because it can not provide the research with any operational definition. A more concise acceptable definition mentioned in chapter one for the “ESL” abbreviation is that it refers to the teaching of English to people who live in a country in which English is the second language. According to this definition, the focus has been diverted from the English speakers to the context in which they use the language. Therefore, the researcher has to bring some evidence to show that English is the second language which is used in India.

In their study, Current Trends in Effective Communicative English Language: A Study, Lakshmi and Suneetha state that Indian people not only have accepted English language as their second language, but also they have the tendency to consider it as their first language. “From the era, where people used to hesitate to learn English, thinking it as a foreign language, India rose to that level where people prefer to speak only in English. Other side of the British advent that Indians can feel is, the rich, cherishing language they have inherited from them” (54). As Lakshmi and Suneetha explain, the Government of India has set up a formula for teaching three languages at schools including English. The fact that the authors are not satisfied with is that English is not considered as the first language to be taught and it always occupies the second or the third place in the order of priorities.
Mohanty in English in India: A Global Language with an Indian Identity states that “in terms of English speakers, the Indian subcontinent ranks third in the world after the USA and UK. Studies suggest that over 150 million people speak English in India.” According to the author, there are over 1652 mother tongues in India and three contact languages have had a major effect on the history of the country, i.e. Sanskrit, Persian and English. Considering the diversity of languages spoken across the country, “Article 343 of the Indian Constitution recognizes Hindi . . . as the official language of the country, and . . . also allows for the continuation of use of English language for all official purposes.” Mohanty asserts that “officially, English has the status of associate language in India, but arguably it is the most important language today. After Hindi, it is the most commonly spoken and understood language in India” (111).

The third scholar who supports the position of English language as the second or even first language in countries like India is Saeed. In her paper, Perspectives: Language, Peripheral Aids and the Value Question, Saeed notes that in her country, Pakistan, like many other countries of the world, English has become the “dominating language”. She believes that it is a proven fact that English is a must for developing countries “in order to keep pace with scientific and economic progress.” Therefore, in countries like Singapore, India and Pakistan “English has attained the position of first or second language” (267).

As the fourth question of the study focuses on the similarities and differences between the data obtained from the administration of reflective tools to Iranian and Indian students as learners who study English in EFL and ESL contexts, the aforementioned discussion seemed necessary to the researcher to elaborate on the issue and make it crystal clear.

3.1.1. Overview of the Chapter

The purpose of the present study is to investigate some practically untouched aspects of reflective teaching in English language teaching. The areas of interest have been summarized in the research questions which are as follows:

1. Is there any difference among the kind of data that all the tools obtain?
2. Is a single tool applicable to obtain any kind of data for a teacher to reflect on?
3. Regarding the accuracy of the obtained data, which tool has priority?
4. Regarding the first three questions, is there any difference between the obtained data from ESL and EFL contexts?

In this chapter, the researcher has attempted to elaborate the data collection methodology and research design of the study. The significant procedures followed to investigate the abovementioned questions and their related null hypotheses have been labeled as research subjects, instrumentation, reflective tools, piloting and trying out the instruments, validity of the instruments, reliability of the instruments, research design, data collection and data analysis procedures.

3.2. Research Subjects

The research sample consisted of two groups of lecturers who are referred to as class teachers in this thesis. The first group was formed of ten class teachers from the departments of English of two colleges located in Chandigarh, India: DAV College, Sector 10, and Government College for Girls (GCG), Sector 11. From each college, five teachers were randomly selected. Therefore, the sample consisted of ten teachers who had at least one class in the academic year of 2006-2007. Based on the research design, one of the classes of each teacher was selected and the total population of their classes comprised 234 students. The teachers were free to choose the class that they wanted to expose to observation. The reason behind this was to let the teacher choose a class where s/he felt more comfortable and probably there was less resistance and more cooperation from the students’ side. These ten classes formed a context in which English was taught as a second language.

The second group consisted of ten teachers chosen randomly from the department of English, Islamic Azad University, Roudehen Branch, Iran. The teachers had at least one class in the academic year of 2006-2007. Like the procedures followed for the first group, here again, the teachers had a choice to introduce the class which they preferred to be observed. The selected classes consisted of 251 students who had chosen basic or specialized English courses to
pass in the very year.

3.3. Instrumentation

As the purpose of the study was to compare the data derived from the application of different tools (i.e. Teacher Diary, Peer Observation, Students’ Feedback and Audio Recording), the researcher had to choose a uniform procedure which let him quantify the obtained data for the specified intention of the comparison and contrast of data. Inspired by Tice’s suggestions for writing a teaching diary which guides the reflective teachers to start by answering some open questions in the form of a questionnaire, the researcher designed a questionnaire (see Table 1.1 for Tice’s suggestions). In addition, as Nunan emphasizes, the advantage of the questionnaire over other data gathering instruments is that it “enables the researcher to collect data in field settings and the data themselves are more amenable to quantification than discursive data such as free-form field notes, participant observers’ journals, the transcripts of oral language” (Research Methods 143). Moreover, Best and Kahn state, “Actually the questionnaire has unique advantages, and, properly constructed and administered, it may serve as . . . [the] most appropriate and useful data gathering device in a research project” (301).

Koul in Methodology of Educational Research defines questionnaire as “a device consisting of a series of questions dealing with some psychological, social, educational, etc. topic(s) sent or given to an individual or a group of individuals, with the object of obtaining data with regard to some problems under investigation” (148). According to him, a good questionnaire should achieve two main goals. First, its items should convey the purpose of the study. Second, it should encourage the subjects to answer the questions and provide the researcher with needed data.

Therefore, the designed questionnaire was the instrument given to all research subjects (i.e. students, class teachers, colleagues and observers) to fill it out by converting their data from observation reports, recording transcripts, diary writings to one of the options linked to the questionnaire items. Although questionnaires can contain open and/or closed questions as well as statements, the researcher preferred to write the items just in the form of statements. By definition,
a closed question is an item for which researcher provides possible responses and the respondents are required to choose one amongst them. While an open question does not suggest any possible answers to the respondent and s/he decides what to write. Each type has its special advantage over the others; regarding the answers obtained from open and closed questions, Nunan states that “responses to closed questions are easier to collate and analyze, [while] one often obtains more useful information from open questions” (Research Methods 143). In addition, Hult cites Fowler who believes that if the questionnaires are supposed to be completed by the subjects, closed questions are preferable. However, Likert type statements (see subsection 3.3.1) which ask the respondents to indicate the degree of agreement to the posed idea by choosing one of the given five options, have all the advantages of closed questions. In other words, while statements are capable of expressing desired points clearly and extract the respondent’s idea, their related options can be readily converted to quantifiable data for further statistical analysis.

3.3.1. Designing the Decision-Making Questionnaire

In order to explore the teaching process through a reflective approach to teaching, the researchers should take into consideration the fact that “the questions it [a tool] explores are not linked to a particular method or view of teaching” (Richards and Lockhart 3). Therefore, the purpose is not to show an effective method of teaching English language, instead it is to reflect on the decisions that teachers make while teaching and show them the preferable tools for obtaining the required data for reflection regardless of their applied teaching methods. In addition, a tool which is supposed to gather data for reflection in the field of teaching should not be limited to a particular teaching situation, level or a special group of teachers, i.e. pre-service or in-service. Keeping in mind the abovementioned terms and the suggestions made by different experts for designing a reflective teaching tool, the researcher went through the following stages to design the desired tool.

As the focus of the study was on the area of “teacher decision making,” the researcher first reviewed the 45 questions which were suggested by Richards and Lockhart to explore the three components of decision making, i.e. planning
decisions, interactive decisions and evaluative decisions (see subsection 1.5.1). Then, the questions which were more appropriate to be investigated through using a questionnaire and/or were emphasized by other experts, i.e. Tice, Hiller, Murphy and Murdoch, were chosen and transformed into statements. The change was done based on the Likert’s suggestion related to designing the questionnaire items and responses. When a subject is asked to complete a questionnaire which is based on a five-point Likert Scale, s/he should point out his/her level of agreement to a statement by selecting one of the five options, i.e. completely disagree, agree, have no idea, disagree and completely disagree. Finally, the items were rewritten to properly address the students, teachers, observers and colleagues. For instance, the question which asked, “Did the lesson address the students’ needs?” was rewritten as follows:

- I learned something that I really needed to know. (Student’s questionnaire)
- Students learned something that they really needed to know. (Teacher, observer and colleague’s questionnaires)

3.3.2. Designing the Authenticity Questionnaire

The purpose of designing the authenticity questionnaire was to investigate the accuracy of students’ responses by asking them to express their own attitudes towards the first questionnaire. In other words, by completing the second questionnaire, the students were to state the degree of caution, seriousness, trust and efficiency that they felt filling out the first questionnaire. Hence, it was supposed to be merely administered to students. The questionnaire consisted of five tentative items which were later reviewed and modified based on the feedback obtained from the experts in the field of education and applied linguistics. The modified questions were as follows:

1. I am sure that my answers will remain confidential.
2. Because I did not write my name, I felt free to answer the questions without caution.
3. I answered the questions carefully.
4. I am sure that the result of the research will improve the quality of teaching.
5. The questions were clear to me.

3.4. Reflective Tools

The reflective tools used in this research were: Audio Recording, Peer Observation, Teacher Diary and Students’ Feedback. To analyze the effectiveness of the tools in gathering appropriate data based on the reflective teachers’ needs, it was obligatory to apply all the selected tools in the same context and compare the obtained results. In this section, the procedures followed for administration of these four tools are explained.

3.4.1. Audio Recording

Audio recording procedure was done by an MP3 recorder taken to the classes. The files of recorded voices, then, were burnt onto CDs and given to colleagues who were chosen to reflect on recorded class interactions. To accomplish the purpose of the research, i.e. listening to audio files, first, the colleagues took some notes, and then filled out the related questionnaire based on their impressions.

3.4.2. Peer Observation

The researcher as a non-participant observer was present in the classes which were chosen for observation. Although, the researcher managed to jot down some of the recurring patterns of teaching and learning process in each class, based on the design of the questionnaire, the gathered data were rendered into the given options of the designed questionnaire.

3.4.3. Teacher Diary

Each teacher had a chance to look at the questionnaire which s/he was supposed to complete after the teaching session. This awareness had its own advantages and disadvantages that will be discussed in chapter five of the research.
Although, reflective teachers usually keep a diary for a period of time (i.e. one academic term for instance), the researcher asked the subjects to reflect on their own teaching and complete the designed questionnaire based on what they would have recorded if they had wanted to keep a diary of that single teaching session.

3.4.4. Students’ Feedback

Students were the only subjects who filled out two questionnaires: the decision-making questionnaire and the authenticity questionnaire. At the end of the teaching session, they were given the questionnaires in such a way that at first they filled out the decision-making questionnaire, then gave their feedback on the attitudes that they had towards it through completing the second questionnaire, i.e. the authenticity questionnaire.

3.5. Piloting and Trying out the Instruments

Under the given title, Seliger and Shohamy explain that the goal of trying out an instrument is to investigate its efficiency and do necessary modification before it is administered to the specified sample. To emphasize the importance of piloting and preliminary testing of the questionnaire, Koul states that “no matter how careful the questionnaire maker has been in phrasing his questions and designing his questionnaire, he needs to try them out with a few representative samples of the respondents before launching into the actual investigation” (153). The author explains that the first objective of pre-testing is to examine the questionnaire from the research viewpoint, while the second objective is to investigate whether the respondents can comprehend the meanings of the items and whether the questionnaire items motivate them enough to answer the questions and provide required data. Nunan considers piloting as an essential stage in the process of developing any questionnaire while Best and Kahn think, “It is also a good idea to ‘pilot test’ the instrument with a small group of persons similar to those who will be used in the study” (308). Griffee also considers piloting as a required part of constructing a questionnaire which helps the designer decide which item to keep and which item to skip.
In the present study, piloting consisted of randomly choosing two classes which comprised 98 students among the English classes held in Government College for Girls in the academic year of 2006-2007. In the next step, the two questionnaires which had been modified according to experts’ recommendations were administered to the subjects. The first questionnaire with 43 items was supposed to investigate the teaching procedure including planning, interactive and evaluative decisions made during teaching, and the second questionnaire was to investigate the extent to which the students accurately answered the questions included in the first questionnaire. Finally, the researcher utilized the extracted data to calculate the predictive validity and reliability of the designed questionnaires.

3.6. Validity of the Instruments

Validity refers to the extent to which the tool used for data collection measures what it is supposed to measure. As Seliger and Shohamy point out, “Validity cannot really be proven but it is necessary to obtain evidence of validity” (188). There are different types of validity: content validity, criterion validity and construct validity.

3.6.1. Content Validity

Koul believes that the researcher ought to get help from the experts in planning and constructing the questionnaire. Then, the items can be modified according to their criticism. Regarding the content validity of the research-instruments, a group of experts in the field of applied linguistics and education reviewed the designed questionnaires and provided the researcher with their feedback. Thereafter, the questionnaires were examined and the ambiguous items were modified or totally removed from the body of the questionnaire. The final editions of the questionnaires have been represented in Appendixes (A to D).

3.6.2. Criterion Validity: Predictive Validity

Predictive validity which is one type of criterion validity shows whether the instrument is able to predict certain behavior. In Research in Education, Best and
Kahn suggest a method for determining the predictive validity of a newly designed questionnaire. They state that, it can be done “by follow-up observations of respondent behavior at the present time or at some time in the future. In some situations overt behavior can be observed without invading the privacy of respondents” (Best and Kahn 317). To test the predictive validity of the designed questionnaires, following the Best and Kahn’s suggested procedure, the researcher chose one of the classes which had been selected for pilot study. The class consisted of 52 students. In order to make some external criterion for comparison, the researcher asked the class teacher as another source to fill out the questionnaire. In addition, the researcher as a non-participant observer attended the students’ class and gathered data to be compared with the data obtained from the students. Finally, the data gathered from the administration of the questionnaires were compared with the overall data obtained from the teacher questionnaire and observer’s to determine whether the designed instrument had the acceptable predictive validity.

3.6.3. Construct Validity

Nunan asserts that the important point in constructing the items of a questionnaire is “to be very clear about the objectives of the study, and each item should be directly referenced against one or more of the research objectives” (Research Methods 145). Regarding the construct validity of the instrument, the researcher referred to the operational definition of teacher decision making as the mere concept that all the items of the questionnaire try to investigate. “Teachers are constantly confronted with a range of different options and are required to select from among these options the ones they think the best suited to a particular goal. The option the teacher selects is known as a decision” (Kindsvatter, Wilen and Ishler, qtd. in Richards and Lockhart 78). Further, Richards and Lockhart divide teacher decision making into three stages (i.e. planning decisions, interactive decisions and evaluative decisions). These three stages refer to the decisions that teachers make before teaching a lesson, decisions that they make when they are involved in the process of teaching and the stage of decision making which takes place after teaching each lesson to evaluate the effectiveness of teaching. As it was
mentioned in chapter one of the present study, Richards and Lockhart provide reflective teachers with three sets of questions which are linked with the three stages of teacher decision making. Tice, Hillier, Murdoch and Murphy are the other educators whose suggestions have made the basis of the designed instruments. Since, the researcher reworded and changed the structure of the suggested items and there was no report of doing any practical evaluation about the validity and the reliability of the suggested sets of items, the researcher decided to go through the process of piloting both designed questionnaires, i.e. the decision making and authenticity questionnaires.

To statistically investigate the construct validity of the questionnaires, the researcher computed the internal consistency of the items using Cronbach’s Alpha. Garson states that one of the common methods of estimating the construct validity of a questionnaire is to compute the internal consistency of the items by applying Cronbach’s Alpha. According to him, for confirmatory purposes, the consistency of .70 and .80 is considered as “acceptable” and “good” respectively.

Table 3.1
The Decision-Making Questionnaire: The Adapted Items for the Student’s Questionnaire and Their Related Sources

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The teacher began the lesson well. (Planning; Richards &amp; Lockhart)</td>
</tr>
<tr>
<td>2</td>
<td>The lesson connected to what I already knew. (Planning; Richards &amp; Lockhart, Hillier, Murdoch, Murphy)</td>
</tr>
<tr>
<td>3</td>
<td>The teacher had mastery over the content of the lesson. (Planning, interactive; Richards &amp; Lockhart, Murphy)</td>
</tr>
<tr>
<td>4</td>
<td>The teacher concluded the lesson well. (Evaluative; Richards &amp; Lockhart)</td>
</tr>
<tr>
<td>5</td>
<td>The lesson was easy for the class. (Planning, interactive and evaluative; Richards &amp; Lockhart, Tice, Murphy)</td>
</tr>
<tr>
<td>6</td>
<td>The teacher dealt with different student ability levels successfully. (Planning, interactive; Richards &amp; Lockhart)</td>
</tr>
<tr>
<td>7</td>
<td>The teacher checked on my understanding of the lesson. (Planning, interactive; Richards &amp; Lockhart)</td>
</tr>
</tbody>
</table>
8. The teacher handled the interruptions successfully. (Planning, interactive; Richards & Lockhart)
9. The teacher was not in a hurry to teach. (Planning, interactive; Richards & Lockhart, Tice, Hillier, Murphy)
10. The instructions were clear and understandable. (Interactive; Richards & Lockhart, Tice, Hillier, Murdoch, Murphy)
11. Most of the students participated in various learning activities. (Interactive and evaluative; Richards & Lockhart, Tice)
12. The teaching method was effective. The teacher does not have to try a different method. (Interactive, evaluative; Richards & Lockhart, Tice, Hillier)
13. The activities were planned. (Planning; Richards & Lockhart, Murphy)
14. The presented information was enough for us. (Interactive; Richards & Lockhart)
15. I learned something that I really needed to know. (Interactive, evaluative; Richards & Lockhart, Tice, Hillier)
16. I had the vocabulary I needed to do the task. (Interactive; Richards & Lockhart)
17. The teacher did not teach too much. I had a chance to practice. (Interactive; Richards & Lockhart, Murphy)
18. The teacher was successful in teaching the lesson. (Evaluative; Richards & Lockhart)
19. I understood and learned the lesson. I do not need to practice it any more. (Evaluative; Richards & Lockhart)
20. I was interested in the subject matter. (Evaluative; Richards & Lockhart, Tice, Murphy)
21. The teacher had enough preparation for the lesson. (Evaluative; Richards & Lockhart, Murphy)
22. I prefer the teacher to teach other lessons in the same way. (Evaluative; Richards & Lockhart, Tice)
23. The teacher does not need to re-teach any parts of the lesson. Everything was perfect. (Evaluative; Richards & Lockhart)
24. The textbook (material) was appropriate and relevant to the program objectives.
25. There were some key points to help us remember our previous knowledge. (Planning; Hillier, Murphy)

26. I had a chance to ask questions during teaching. (Interactive; Hillier, Murdoch, Murphy)

27. The teacher asked some questions to check students’ understanding. (Interactive; Richards & Lockhart, Hillier)

28. The teacher’s tone was varied and lively. (Interactive; Hillier)

29. The teacher had eye contact with all students. (Interactive; Hillier)

30. The teacher treated us with respect. (Interactive; Hillier, Murdoch)

31. Students treated the teacher with respect. (Interactive; Hillier)

32. The teacher was successful in encouraging and facilitating students’ learning. (Evaluative; Hillier, Murdoch)

33. The time given to each activity was enough. (Planning; Richards & Lockhart)

34. The teacher provided us with different activities. (Interactive; Tice, Murdoch)

35. The teacher presented language points clearly. (Interactive, Murdoch)

36. The teacher gave us sufficient time to respond to the questions. (Interactive; Murdoch)

37. The teacher used different error correction techniques. (Interactive; Murdoch)

38. The teacher made the purpose of the activities clear to us. (Interactive; Tice, Murdoch)

39. The teacher encouraged communication among us. (Interactive; Tice, Murdoch)

40. The teacher used the board effectively. (Interactive, Hillier, Murdoch)

41. The teacher provided an exciting learning environment. (Interactive; Hillier, Murdoch, Murphy)

42. The teacher was interested in the subject. (Interactive; Hillier)

43. The teacher’s information was updated on the subject. (Evaluative; Hillier)

3.7. Reliability of the Instruments

Generally, reliability is defined as “the degree of consistency that the
instrument or procedure demonstrates: Whatever it is measuring, it does so consistently.” A reliable test is not necessarily a test which has some degree of validity. It is possible for a test to have consistency in measurement; however, it does not measure what it intended to measure. In other words, “a test must be reliable for it to be valid, but a test can be reliable and still not to be valid” (Best and Kahn 277). The higher is the coefficient of the reliability of a test; the lower will be the errors of measurement. According to Best and Kahn, there are different types of reliability which the test designers and instrument makers can use based on their special needs. They are as follows:

1. Stability over time (test-retest)
2. Stability over item samples (equivalent or parallel form)
3. Stability of items (internal consistency)
   a. Split halves
   b. Coefficient of consistency
4. Stability over scores (inter-scorer)
5. Stability over testers
6. Standard error of measurement

Each type of reliability has its own merits and demerits. For example, while the first type can measure the stability of the test over time (preferably two weeks), it requires the same subjects, i.e. test takers to get together at least after a lapse of two weeks and answer the same set of questions or fill out the same questionnaire in a situation that is exactly the same as the first time of instrument administration.

Each item in an instrument should have close relation with other items of the instrument. The reason is that when an instrument is supposed to measure a single construct, each item should measure the same thing; so, they should have high internal consistency. In this research, the internal consistency type of reliability (type 3.b above) was used to measure the coefficient of the questionnaire reliability. One of the models which are usually used to determine the extent to which the items in a designed questionnaire are related to each other is Cronbach’s Alpha. According to the SPSS help file, “this is a model of internal consistency, based on the average inter-item correlation.” By the use of Alpha model, the researcher “can
get an overall index of the repeatability or internal consistency of the scale as a whole, and . . . identify problem items that should be excluded from the scale” (SPSS for Windows).

The procedure of obtaining Alpha includes the calculation of the average correlation between pairs of items on the questionnaire. Then, this correlation which Hatch and Farhady refer to as “average item reliability” is used to compute the coefficient of Cronbach’s Alpha for the whole questionnaire (247). In the following formula which is applied to obtain the reliability of an instrument such as a test or a questionnaire, “\( r_{ii} \)” is the mean item correlation and “\( K \)” is the number of items.

\[
\text{Alpha} = \frac{K r_{ii}}{1 + (K - 1)r_{ii}}
\]

The formula which is often called Kuder-Richardson formula 20 (KR-20) enables the researchers “to estimate the entire test reliability from the mean correlation among all items of the test”. Therefore, all the items of the instrument should be the same. It means that the items ought to focus on one shared subject (Hatch and Farhady 248). Cronbach has also suggested another formula for calculating Alpha, which is slightly different from the aforementioned formula, recommended by Hatch and Farhady and used by SPSS Software (see Bachman 177 and Birjandi et al. 81).

### 3.8. Research Design

Regarding research design, this research falls into the category of descriptive (non-experimental and correlational) studies. Seliger and Shohamy define descriptive research as a kind of research referring to “investigation which utilizes already existing data or non-experimental research with a preconceived hypothesis” (117). Best and Kahn state that a descriptive study “is concerned with conditions and relationships that exist, opinions that are held, processes that are going on, effects that are evident, or trends that are developing. It is primarily concerned with the present, although it often considers past events and influences as they relate to current conditions” (114). The authors also explain that descriptive research
investigates the existing relationships among the research variables, tests the research hypotheses and generalizes the obtained results. Descriptive research manages “the development of generalizations, principles, or theories that have universal validity” (115). Descriptive research is mainly in the form of correlational research or causal-comparative research and as Best and Kahn point out there are other types of descriptive research which do not fall into these two subcategories but describe the present situation.

Descriptive research studies have some distinctive characteristics which differentiate them from other types of research. These distinguishing features are as follows:

1. They involve hypothesis formulation and testing.
2. They use the logical methods of inductive-deductive reasoning to arrive at the generalizations.
3. They often employ methods of randomization so that error may be estimated when population characteristics are inferred from observations of samples.
4. The variables and procedures are described as accurately and completely as possible so that the study can be replicated by other researchers. (Best and Kahn 129)

A point which is worthy of noticing is that correlational research is not considered a research methodology or design in the same way that qualitative, descriptive and experimental research are. It may be regarded as “a form of descriptive research.” Correlational research can be classified as a method of data analysis instead of being considered as a research method where it is used to analyze the data obtained from other research methods (Seliger and Shohamy 130).

As Best and Kahn sum it up, “descriptive research . . . uses quantitative methods to describe what is, describing, recording, analyzing and interpreting conditions that exist. It involves some kind of comparison or contrast and attempts to discover relationships between existing nonmanipulated variables” (22).

Following the descriptive research procedure, the researcher applied the statistical analysis to compare and contrast the data obtained from the administration
of each reflective tool with the data provided by the three other individual tools and the obtained mean of all tools.

3.9. Data Collection and Data Analysis Procedures

The procedures for carrying out a descriptive research normally include: determining the research questions and the related hypotheses, selecting the research subjects, deciding on data collection methods, gathering the data and analyzing the obtained data (Seliger and Shohamy). The present study is not an exception; so, it made use of the abovementioned procedures plus the thorough process of designing, piloting and standardizing the research instruments which were included in the procedure of “deciding on data collection method.”

3.9.1. Stage One

The first stage of doing the research, i.e. determining of the research questions and the related hypotheses was fully elaborated in chapter one and the research questions were restated in the introduction of this chapter.

3.9.2. Stage Two

The second stage consisted of selecting the research subjects. As it was explained earlier in this chapter (see 3.2), the research subjects were 20 teachers and their 485 students chosen from both ESL and EFL contexts. The points which need to be clarified are the specifications of the teachers’ classes, mainly the students and the materials. The chosen subjects were teachers of English language whose classes had the following features. Regardless of subject labels (e.g., Grammar, Phonology, Compulsory English, Functional English, Drama and Novel), and materials (e.g., textbooks, handouts and notes), the focus of the classes had to be mainly on practicing one or more of the skills or components of English language. The level of students was not controlled in this research. In other words, while the majority of students who participated in the study were undergraduates, some postgraduate students were also observed.
3.9.3. Stage Three

At the stage of data collection method, regarding the purpose of the study and the need for uniform data gathering instruments, two questionnaires were designed. The designed instruments were reviewed by experts in the fields of education and applied linguistics and the weak and ambiguous items were modified or removed. In order to investigate the validity and the reliability of the instruments, a pilot study was conducted and the obtained results underwent a series of statistical analysis. The followed procedures of the pilot study and analysis of the research validity and reliability were elaborated earlier in this chapter under the titles of “Piloting and Trying out the Instruments”, “Validity of the Instruments” and “Reliability of the Instruments.” (For the obtained indexes of different types of validity and reliability, see chapter 4, section 4.2.)

3.9.4. Stage Four

At stage four, the required data were collected. Having the permission of the principal of each college and the head of the department for conducting the research, the first step consisted of a very short orientation session for the teacher. The purpose was to make the subject familiar with the purpose and process of doing the research and agree on a teaching session for administering the research tools. Then, on the due date, in a single session, all four tools were administered for each class. It means that while the teacher started teaching, the researcher as a non-participant observer was present in the class, taking notes which he used later to complete the questionnaire. The teaching process was tape recorded in the same session. Later, the recorded tape was reviewed by a colleague and the given questionnaire was filled out accordingly. When the teacher was finished with the lesson, students were asked to give their feedback on the teaching process by completing their questionnaires. In addition to this, they were required to complete the second questionnaire which investigated the accuracy of the data provided by them when they completed the first one. The teacher was the last subject who was asked to reflect on his/her teaching by completing the first questionnaire.
3.9.5. Stage Five

Excluding the interpretation of data which is an independent part in each research, stage five is normally considered the last step in a research and it is devoted to the analysis of the collected data. In this stage, two things should be clarified: 1) What is the purpose of applying statistical analysis? Or what does the researcher want to get out of the collected data? 2) How does he want to analyze them? Or regarding the design of the study, what is the appropriate procedure to follow?

Regarding the purpose of applying statistical analysis (the first question), the researcher wanted to investigate the following:

- Taking into consideration the whole items of the decision-making questionnaire, the researcher’s purpose was to determine the difference among the means of the research samples, i.e. students, class teachers, observers and colleagues who listened to tape recordings. He also was interested in studying the coefficient of correlation between: 1) the data elicited by a single tool and the average of the whole data obtained by all four tools 2) the data obtained from ESL and EFL contexts.

- Taking into consideration the items which related to students or teachers (see Sub-scales A and B in subsection 4.4.2), the researcher wanted to explore the difference between: 1) the mean of the data collected from the students’ questionnaires (Students’ Feedback) and other tools; 2) the mean of the data collected from the teachers’ questionnaires (Teacher Diary) and the rest of the tools.

- The authenticity of the obtained data from the students’ questionnaires by investigating the students’ answers to the authenticity questionnaire.

3.9.6. Statistical Methods

Regarding the application of appropriate procedure (the second question above) the researcher used the following methods to analyze the obtained data.

- Analysis of Variance (ANOVA)
- Pearson’s Product-Moment Coefficient of Correlation (r)
Analysis of Variance (ANOVA)

Regarding the purpose of the study, the dominant part of the statistical analysis consists of comparing the results of administering four reflective tools and the obtained mean of all tools. Since, the t-test can merely determine the significant difference between two samples, the analysis of the variance (ANOVA) is employed to compare the means of more than two groups of subjects. Best and Kahn consider ANOVA as “an effective way to determine whether the means of more than two samples are too different to attribute to sampling error.” As the authors’ example shows, the comparison of means in a descriptive research, in which there is no treatment, can appropriately be done by utilizing the ANOVA test (409).

The following steps are followed in the analysis of variance:

1. The variance of the scores for [all] groups is combined into one composite group known as the total group variance ($V_t$).
2. The mean value of the variances of each of the [all] groups, computed separately, is known as the within-groups variance ($V_w$).
3. The difference between the total groups variance and the within-groups variance is known as the between-groups variance ($V_t - V_w = V_b$).
4. The F ratio is computed. (Best and Kahn 409)

$$F = \frac{V_b}{V_w} = \frac{(between - groups variance)}{(within - groups variance)}$$

If the obtained F-ratio exceeds the related F-critical, the difference among the groups under investigation is considered significant and the null hypothesis is
rejected at the predetermined level of significance. The F-critical values are provided in an F table in most of the research and statistics books (e.g., Best and Kahn 471).

**Pearson’s Product-Moment Coefficient of Correlation (r)**

Correlation is the relationship between two or more sets of scores or paired variables. The coefficient of correlation shows the degree of relationship between two sets of data (Baker). Among different methods of computing the correlation coefficient, a raw score method computation is more convenient and it is used when a computer software is available. The Pearson Product-Moment Coefficient of Correlation which lets the use of raw scores may be used when:

1. The data are expressed in interval or ratio score form.
2. The distribution of X and Y have a linear correlation.
3. The variances are approximately equal.
4. The distributions have no more than one mode. (Best 251)

To compute the correlation coefficient of two sets of scores by using raw scores, the following formula is used:

\[ r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}} \]

Where:
- X = scores given to the students’ responses based on the Likert scale
- Y = scores given to the teachers’ responses based on the Likert scale
- \( \Sigma \) = sum of the scores
- \( N \) = number of paired scores
- \( r \) = correlation coefficient

**Statistical Significance of a Correlation Coefficient**

As Best and Kahn put, “An observed coefficient of correlation may result from chance or sampling error and a test to determine its statistical significance is
appropriate.” The null hypothesis predicts that there is no relationship between two variables. In other words, it suggests that the coefficient of correlation between two sets of scores is zero. “Only when chance or sampling error has been discredited on a probability basis can a coefficient of correlation be accepted as statistically significant” (Best and Kahn 407).

One of the methods of determining the significance of the obtained coefficient of correlation ($r$) between two sets of scores is using the following formula:

$$t_r = \frac{r \sqrt{N - 2}}{\sqrt{1 - r^2}}$$

Where:

- $r$ = obtained coefficient of correlation
- $N$ = number of paired scores
- $t_r$ = the t-value which should be compared with the t-critical value

With N-2 degrees of freedom, a coefficient correlation is considered to be statistically significant when the observed t-value equals or exceeds the t-critical value. The t-critical values are presented in a $t$ table in most research and statistics books (e.g., Best and Kahn 469).

The statistical significance of coefficients of correlation among various sets of scores was tested to see whether on a two-tailed test at the .05 level of significance, with N-2 degrees of freedom, the related null hypothesis would be rejected (see chapter four for details).

**Divergence from Normality**

**Skewness of the Scores**

In the normal curve model, the median, the mode and the mean have the same value. In other words, all these three measures are located in the mid point of the normal curve (bell-shaped curve) and there is a complete balance between the right and the left halves of the curve. As Garrett describes, “A distribution is said to be ‘skewed’ when the mean and the median fall at different points in the distribution.
and the balance (or center of gravity) is shifted to one side or the other—to left or right.” When a distribution is normal the value of mean and median is the same and the skewness equals zero. “In fact, the greater the gap between mean and median, the greater the skewness. Moreover, when the skewness is negative, the mean lies to the left of the median; and when skewness is positive, the mean lies to the right of the median” (109). An index of skewness in a distribution is computed by the use of the following formula:

\[ SK = \frac{3(mean - median)}{\sigma} \]

Where:

\( \sigma \) = Standard Deviation

\( SK \) = Skewness

One of the ways to determine whether the obtained skewness is significantly skewed is to compare the degree of skewness with the standard error of skewness (SES). Price explains that in order for a degree of skewness not to be considered as significantly skewed, it should be within the range of (-2 SES) and (+2 SES).

Therefore, to determine the significance of the skewness of a set of scores, one has to compute the standard error of skewness. As Brown in Skewness and kurtosis mentions, the approximate standard error of skewness can be estimated by the following simple formula:

\[ SES = \sqrt{\frac{6}{N}} \]

Where:

\( N \) = Number of Scores

\( SES \) = Standard Error of Skewness

In the present study, the researcher computed the skewness of the obtained scores and its related SES from the research sample to determine the amount of divergence of the selected distribution from normality. The analysis of the degree of skewness was presented in chapter four.
Kurtosis of the Scores

Another descriptive value that can be obtained to describe a distribution is called kurtosis. Kurtosis of a frequency distribution is its peakedness or flatness compared with the normal curve. A frequency distribution which is more picked than the normal is labeled as “leptokurtic,” and a distribution which is flatter than the normal is said to be “platykurtic” (Garrett 101; Singh Sidhu 87). A formula for measuring the kurtosis of a frequency distribution based on its percentiles is as follows:

\[
Ku = \frac{Q}{(P_{90} - P_{10})}
\]

Where:
- \(Q\) = Quartile (0.6745 X Standard Deviation)
- \(P\) = Percentiles
- \(Ku\) = Kurtosis

The same numerical process which is used to determine whether the derived skewness is significant can be applicable to check if the kurtosis is not significantly normal. “A normal distribution will have Kurtosis value of zero. So again we construct a range of ‘normality’ by multiplying the Std. Error of Kurtosis by 2 and going from minus that value to plus that value” (Price, School of Psychology, UNE).

Here again, computing the standard error of Kurtosis (SEK) is considered as a prerequisite for determining the significance of the given value. To find the SEK, a similar formula is pointed out by Brown which is as follows:

\[
SEK = \sqrt{\frac{24}{N}}
\]

Where:
- \(N\) = Number of Scores
- \(SEK\) = Standard Error of Kurtosis

In this study, the researcher computed the kurtosis and its related SEK of the scores obtained from the research sample to determine the amount of divergence of
the frequency distribution from normality. The analysis of the degree of kurtosis was presented in chapter four.

3.9.7. SPSS Software

Since one of the purposes of conducting the present research is to show the reflective teachers how to apply reflective tools in their classes and statistics is eventually an essential part of studies of this sort, the process of statistical computation which has been done by the use of SPSS software is elaborated in the final section of the present chapter.

To handle the data generated by administering the questionnaires, first the researcher should decide how to give numerical values to the answers. The process includes coding the questionnaire responses in such a way that the numerical codes substitute the original responses. Then, statistical programs like SPSS can access the created numerical codes without any problems. In this thesis, the researcher represented the responses as follows: completely agree by “5”, agree by “4”, have no idea by “3”, disagree by “2” and completely disagree by “1”.

The main window of the software provides a matrix in which each row represents an individual who completes the questionnaire and each column represents one of the items as well as other variables like the individual’s sex, job, nationality, age, etc. Entering the data into the matrix, the researcher can analyze them, using different statistical procedures, and represent them through various graphs. The procedures which the researcher followed in order to analyze and represent the research obtained data are illustrated here:

**Frequencies:** Choose “Analyze” option from the Menu Bar > Select “Descriptive Statistics” > Pick “Frequencies” > Choose “Variables” > Click “Ok.”

**Mean, Median and Standard deviation:** Choose “Analyze” option from the Menu Bar > Select “Descriptive Statistics” > Pick “Descriptives” > Choose “Variables” > Click “Ok.”

**The Comparison of More than Two Means (ANOVA):** Choose “Analyze” option from the Menu Bar > Select “Compare Means” > Pick “One-Way ANOVA” > Choose “Variables” > Decide on the method of Post Hoc Multiple Comparisons >
Tick “Means Plot” among the options > Click “Ok.”

**Coefficient of Correlation and Its Significance:** Choose “Analyze” option from the Menu Bar > Select “Correlate” > Pick “Bivariate” > Tick off “Pearson” > Tick off “One-tailed” or “Two-tailed” > Choose “Variables” > Click “Ok.”

**Cronbach’s Alpha:** Choose “Analyze” option from the Menu Bar > Select “Scale” > Choose “Reliability Analysis” > Pick “Alpha” and “Variables” > Click “Ok.”

**Merging Related Variables:** Choose “Transform” option from the Menu Bar > Select “Compute” > Pick “Variables” > Define “Variables” > Click “Ok.”

**Skewness and Kurtosis:** Choose “Analyze” option from the Menu Bar > Select “Descriptive Statistics” > Pick “Descriptives” > Tick off “Skewness” and “Kurtosis” > Click “Continue” > Choose “Variables” > Click “Ok.”

**Histograms:** Choose “Graphs” option from the Menu Bar > Select “Histogram” > Pick “Variable” > Click “Ok.”

**Pie Charts:** Choose “Graphs” option from the Menu Bar > Select “Pie” > Pick the variable to be defined by slices > Click “Ok.”

**Scattered Diagrams:** Choose “Graphs” from the Menu Bar > Select “Scatter” > Pick the desired diagram and Click “Define” > Assign the variables to “X” and “Y” Axes > Click “Ok.”

### 3.10. Summary

Chapter three represented the data collection methodology and research design of the study. In different sections and subsections of the chapter, the researcher attempted to give a thorough account of the procedures followed for conducting the study. The issues which were included in this chapter were of two kinds: those which were discussed and concluded in the very chapter, e.g., topics like “research subjects” and; those topics which were just introduced, and they had to be practically investigated, e.g., “piloting of the study” and statistical analyses of the data collected from the subjects. In the next chapter, the researcher continues to work on those posed topics (i.e. the second type) which need to be statistically investigated.