CHAPTER III
RESEARCH METHOD

This chapter outlines the research methods used in this study and the basic principles for selecting the research design and participants. This chapter includes six main parts: (1) Research design; (2) Research site and participants; (3) Data collection techniques; (4) Data collection procedure; (5) Data analysis; (6) Concluding remarks. It attempts to answer the research questions: (a) What do ELT graduate students understand about language assessment? (b) How do the students practice their understanding of language assessment in the classroom? (c) To what extent does ELT graduate students' understanding affect their practice in assessing language assessment?

3.1 Research Design

This study used an explanatory sequential mixed-methods design. The reason was to provide the best understanding of the research issues by providing quantitative and qualitative approaches (Creswell & Creswell, 2018, p. 52). In this design, to combine the benefits of quantitative and qualitative data, the findings were described by qualitative methods (semi-structured interviews and document analysis) after conducting a quantitative method (questionnaire) (Hamied, 2017, p. 270; Leavy, 2017, p. 9). Specifically, this approach utilized the strengths of both data to comprehend research problems (Creswell, 2015).

There are several reasons why this current study meets the characteristics of an explanatory sequential mixed method design. First, the questionnaire facilitates the generalization of ELT graduate students' grasp of language assessment. Second, the interview allows a more in-depth exploration of how the students practice language assessment. Besides validating the questionnaire, the interview examines individuals' training and teaching background, understanding, and practices related to the concepts, purposes, principles, and language assessment types. Third, the analyzed documents are the supporting data to see how graduate students practice language assessment in the classroom. Figure 3.1 displays the implication of an explanatory sequential mixed-methods design in this study.

Figure 3.1 *An Explanatory Sequential Mixed-Methods Design*

Phase	Procedure	Outcome
Quantitative Data Collection	• Questionnaire (n=129) • Likert-format questionnaire (online)	• Numeric database
Quantitative Data Analysis	• Data screening • SPSS software 25.0	• Descriptive and Inferential statistics • Significance results
Results and Interpretation	• Summarize the survey results • Compare the survey results with the literature	• Presentation and interpretation of results
Qualitative Data Collection	• Semi-structured interviews (Purposefully selecting participants, n=6) • Documents	• Interview transcripts • Document descriptions
Qualitative Data Analysis	• Transcribing data • Content analysis • Coding and thematic analysis	• Quotes • Codes • Themes
Integration of Quantitative and Qualitative Findings	• Integration and explanation of Quantitative and Qualitative results	• Discussion • Implications • Future research

(Inspired by Creswell, 2015)

As shown in Figure 3.1, the first phase was collecting quantitative data using the questionnaire. The main benefit of this instrument is that it enables the researcher to get a wide range of information from many students (Leavy, 2017, p.

101). Moreover, descriptive statistics were used to address research question one, “What do ELT graduate students understand about language assessment?” and research question two, “How do the students practice their understanding of language assessment in the classroom?”. On the other hand, inferential statistics were employed to investigate research question three, “To what extend does ELT graduate students’ understanding affect their practice in assessing language assessment?” Their significant results were also interpreted from the quantitative data.

The second phase of the research was collecting qualitative data (Figure 3.1). Semi-structured interviews were conducted with six participants who were selected purposefully. Additionally, a collection of documents was analyzed. Thematic analysis was utilized to explore ELT graduate students’ understanding and practice of language assessment. Thus, these data complemented the descriptive analysis from the questionnaire. The interpretation of quantitative and qualitative data helped answer the previously mentioned research questions. Specifically, a statistical Pearson Correlation test using SPSS version 25.0 was used to investigate the relationship between ELT graduate students' understanding and practice of language assessment.

3.2 Research Site and Participants

The present study setting was a state university in West Java in the English language education program field. The rationale for choosing this site was that the university is well-known for its teachers’ education majors, especially English Language Teaching (ELT) program. The courses offered for the undergraduate and graduate programs include language assessment which is significant in this study. Additionally, this site was chosen for its accessibility and the graduate students’ willingness to participate in this study.

The study participants were graduate students registered as active students in the 2022 academic year studying in the English language education program. Previous literature studies mainly investigated pre-service or in-service teachers’ perceptions of language assessment. Thus, this study aims to fill the gap in the previous studies from the perspectives of ELT graduate students.

There were 197 registered ELT graduate students as the population of the study. A sampling technique known as non-probability sampling was employed in this study as it aids researchers in choosing a sample from which to conduct the research (Dhivyadeepa, 2015, p. 88). In addition, where the goal is to quickly and cheaply determine whether a problem or issue even exists, non-probability sampling can be beneficial (Frey, 2012). Complimentary to this, non-probability sampling may be preferable in some instances during the research (Brick, 2014). Further, convenience sampling was used to produce a sample that accurately reflects the essential characteristics of the research participants and addresses the limitations of the research (Fink, 2003, p. 70; Taherdoost, 2016). Participants' availability, practicality, and accessibility to the research site supported the reasons for the selection of the sampling (Dattalo, 2008, p .6). Thus, the convenience sampling method for the quantitative phase and purposive sampling for the qualitative phase was employed to meet the study's objective.

There are several reasons to use criteria when recruiting participants, including the nature of the study, the need to focus on particular population subgroups, the significance of the study, the need to provide concrete examples, and the desire to get a general understanding of the nature of the issue being studied (Daniel, 2011). In this study, first, they were ELT graduate students whose undergraduate major was English Education/English Language Teaching or ELT graduate students who have completed *aanvullen* courses (required courses related to ELT for non-ELT majors). Second, they have attended courses related to language testing and evaluation or assessment (may have different naming) during their undergraduate or graduate study program. Third, they have had teaching experience (pre-service or in-service teaching) in formal or non-formal institutions. Thus, not all ELT graduate students fit the research criteria and objectives.

The number of participants in this research met the sample size requirement. From the literature, it was found that there are various formulas to calculate the ideal sample size (Fink, 2003, p. 76). Fink further explains that statistical calculations are the most effective technique for choosing the proper sample size (2003, p. 35). From the statistical calculation, with a confidence level of 95% and a margin of error of 5%, and assuming a population proportion of around 50%, the ideal sample size for 197 is around 129. Moreover, Isaac and Miko's table was consulted to look for specific sample size numbers (Sugiyono, 2010, p. 128). The ideal sample size for an estimated population between 190-200 is between 123-127 people (5% margin of error). Similarly, the sample size table was also found in Dhivyadeepa's work (2015, p. 20). The following table highlights the sample size calculation.

Figure 3.2 *The Sample Size Table*

Population Size	Confidence level 90%			Confidence level 95%			Confidence level 99%		
	Confidence Interval			Confidence Interval			Confidence Interval		
	5%	4%	3%	5%	4%	3%	5%	4%	3%
30	27	28	29	28	29	29	29	29	30
50	42	45	47	44	46	48	46	48	49
75	59	64	68	63	67	70	67	70	72
100	73	81	88	79	86	91	87	91	95
120	83	94	104	91	100	108	102	108	113
150	97	111	125	108	120	132	122	131	139
200	115	136	158	132	150	168	154	168	180
250	130	157	188	151	176	203	182	201	220
300	143	176	215	168	200	234	207	233	258
350	153	192	239	183	221	264	229	262	294
400	162	206	262	196	240	291	250	289	329
450	170	219	282	207	257	317	268	314	362
500	176	230	301	217	273	340	285	337	393
600	187	249	335	234	300	384	315	380	453

(source: Dhivyadeepa, 2015; Sugiyono, 2010)

However, this calculation was only valid if it was assumed that the characteristics of the chosen sample represented the overall population. In this study, specific criteria were employed to recruit the respondents. Thus, the estimated calculation might be higher than the number of qualified respondents. Out of 197 invitations sent to the population, 129 qualified respondents agreed to fill out the questionnaire. According to Hamied (2017, p. 84), having thirty out of one hundred is a reasonable sample size. Thus, a sufficient number of samples is significant in a survey questionnaire. Table 3.1 displays the respondents' distribution based on their class year.

Table 3.1 *Participants' Class Year Distribution*

Class Year	Population		Respondents	
	Frequency (n)	Percentage (%)	Frequency (n)	Percentage (%)
2021	56	29	43	77
2020	56	29	47	84
2019	54	28	32	59
2018	31	16	7	23
Total	197	100	129	65

Data from the administration staff showed that there were 197 students registered as graduate students in the 2022 academic year from 2018 to 2021 class year. One hundred twenty-nine respondents (65%) participated in this study by filling out the questionnaire. The majority of the respondents were from the 2020 class year. Table 3.1 informs the respondents' distribution in this study. However, the significant correlation of class year distribution to the present study is not investigated.

From the demographic information asked in the questionnaire, it was found that more than half of the respondents had 1-5 teaching experiences (n=62). The second highest frequency was less than a year of teaching experience (n=31). It could be inferred that ELT graduate students who participated in this study were mostly novice teachers. Kim and Roth (2011) explain that a novice teacher is any teacher with less than five years of experience. Participants' experience in teaching is significant in this study to investigate their practice of language assessment. Only participants with more than five years of teaching experience were selected for the interviews.

Additionally, the respondents were dominated by females (n=93). However, the significant contribution of gender to this study is not explored. Table 3.2 highlights respondents' teaching experience and gender information.

Table 3.2 *Participants' Teaching Experience and Gender Information*

Teaching Experience		Gender	
Year	Frequency	Male	Female

<1 year	31	36	93
1-5 years	62		
6-10 years	19		
>10 years	17		
Total	129	129	

The questionnaire distributed to the respondents also explored their experience attending training and courses related to language testing and assessment. The result reports that 101 respondents had taken courses related to assessment during their undergraduate study. The duration of the study was mainly one semester. Those who responded with "no experience" came from the non-English Language Teaching (ELT) program. Moreover, all respondents had attended Language Testing and Evaluation (LTE) class during their graduate study.

Additionally, only seven respondents had the experience of attending training related to assessment. The duration varied from 2 weeks to 4 months and was hosted by the government or their working institution. However, this study does not investigate the relationship between assessment training and participants' understanding and practice of language assessment. Table 3.3 shows the summary of participants' backgrounds related to assessment training.

Table 3.3 *Participants' Experience in Courses and Training in Assessment*

Courses and Training	Attended (n)	Not Attended (n)
Language Testing and Assessment (Undergraduate study)	101	28
Language Testing and Assessment (Graduate Study)	129	0
Assessment Training	7	122

For the qualitative data collection, purposeful sampling was utilized. There were 54 respondents from the survey questionnaire agreed to volunteer. Nevertheless, only six participants were selected for the interviews. Specifically, two participants from each category of their questionnaire score results were recruited (Low, medium, and high achiever). There were some criteria in choosing the participants for the semi-structured interviews. First, they had more than five years of teaching experience. Second, they have attended language testing and assessment courses during their undergraduate and graduate studies. Last, they had some document artifacts of their teaching practice. The retrieved documents from these six participants were analyzed as qualitative data. Table 3.4 highlights the participants' backgrounds for the interview.

Table 3.4 *Interviewee Information*

Name (Pseudonyms)	Gender	Teaching Experience	Language Assessment Understanding and Practice Level
Abdi	Male	11 years	High achiever
Indi	Female	13 years	High achiever
Heru	Male	11 years	Medium achiever
Dini	Female	12 years	Medium achiever
Miko	Male	15 years	Low achiever
Tika	Female	7 years	Low achiever

Table 3.4 highlights three level categories of the questionnaire score result. The categorization was based on the the percentile range obtained from the questionnaire scores result. Percentiles values are used for large scale survey that divide the data into groups so that a certain percentage is above and another percentage is below (Lamprianou & Athanasou, 2009). In this study, the rank scores from the lowest to the highest were divided into three values: low, medium, and high. These categories were used to select the representation of respondents which would have implications for the analysis and interpretation the data.

3.3 Data Collection Techniques

There were three main instruments employed in the present study. The first one was a questionnaire using an online survey. The second one was from semi-structured interviews. The last data were gathered from document analyses. All instruments were triangulated to form accurate and in-depth data findings. The following discusses the description of data collection employed in this present study. They are categorized into quantitative and qualitative data collection.

3.3.1 Quantitative Data Collection

In quantitative data collection, a survey questionnaire was employed. The result was collected and analyzed using SPSS version 25.0. Besides, the validity and reliability of the instrument were checked and validated.

3.3.1.1 Questionnaire

The questionnaire was employed as the quantitative data instrument in this study. A link to the Google forms was distributed to collect data about ELT graduate students' demographics and some close-ended responses related to students' understanding of language assessment. In a questionnaire, many people are asked about their behaviors, attitudes, and opinions (Marczyk, DeMatteo, & Festinger, 2005). Thus, in this study context, the questionnaire had a significant role in effectively collecting the data from 129 participants. A cover letter was attached to the questionnaire explaining the study's purposes and the consent form to seek the respondents' agreement before they participated in the study.

The questionnaire was developed and constructed from the theories discussed in the literature review. The concept of assessment literacy proposed by Fulcher (2012) is used as the scope and limitation of the discussion. Further, theories related to the concepts of assessment (Brown, 2003; Douglas, 2014, Fulcher & Davidson, 2007; Kahn, 2018), the purposes of assessment (Brown & Hirschfeld, 2008; Malone, 2003; William, 2013), the principles of assessment (Bachman & Palmer, 1996; Brown, 2001; Hughes, 2003) and the types of assessment (Agcam & Babanoglu, 2016; Crooks, 2002; Khan, 2018) were used as the basis in constructing the instruments related to understanding language assessment for ELT graduate students.

The questionnaire format refers to the structure used in the previous study in the related field (Hasselgreen, Carlsen, & Helness, 2004; Vogt & Tsagari, 2014). Some additional items were modified to accommodate the investigation. The statements in the questionnaire covered a variety of Language Testing and Assessment (LTA) concepts and practices. There were ten sections of the questionnaire with forty-four statements. The first part covered the introduction and general information about the study. The second section described the consent form the respondents needed to submit before filling out the questionnaire. The third section highlighted the demographic information of ELT graduate students. Those included name, email, gender, teaching experience, and courses taken related to LTA. Sections four to six covered ELT graduate students' understanding of language assessment. Sections seven to nine presented ELT graduate students' practices of language assessment. The last section was the invitation to participate in the interview.

The responses to the questionnaire in sections four to six used a four-point Likert-type scale to examine ELT graduate students' understanding level (see Table 3.5). The scores range from 1 to 4 (1= not knowledgeable at all, 2= slightly knowledgeable, 3= moderately knowledgeable, and 4= completely knowledgeable). Unipolar rating scales were chosen because the type of the survey asks the respondent to rate how much of a particular attribute or quality is present (Unipolar, 2023). In this study, the respondents self-rated their understanding of language assessment elements. The common benchmarks for a four-point Likert scale are from not at all to completely extent.

Table 3.5 *Measurement Scale for Understanding*

Scale	1	2	3	4
Description	Not knowledgeable at all	Slightly knowledgeable	Moderately knowledgeable	Completely knowledgeable

For sections seven to nine, a four-point Likert scale was used to measure the frequency of the occurrence (See Table 3.7). They were (1) never (not ever applied), (2) rarely (applied not often), (3) often (applied most of the time), (4) always (applied every time).

Table 3.6 *Measurement Scale for Practice*

Scale	1	2	3	4
Description	Never (not ever applied)	Rarely (applied not often)	Often (applied most of the time)	Always (applied every time)

A four-point Likert scale was chosen due to several reasons. First, it is easier to notify the data findings and avoid inaccurate information (Hopper, 2016). Moreover, it is helpful when precise information is required (Cornell, 2022). The data findings have to report the respondent's standpoint in this study. Therefore, the neutral answer was removed. The overall blueprint of the questionnaire is displayed in Table 3.8. The complete forms of the questionnaire are attached in Appendix 1.

Table 3.7 *The Questionnaire BluePrint*

Sections	Items	Total
Understanding the concepts and principles of LA	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	11
Understanding the purposes of LA	12, 13, 14, 15, 16, 17	6
Understanding the classroom-focused LA	18, 19, 20, 21, 22, 23, 24	7
Practicing the concepts and principles of LA	25, 26, 27, 28, 29, 30, 31, 32, 33	9
Practicing the purposes of LA	34, 35, 36, 37	4
Practicing the classroom-focused LA	38, 39, 40, 41, 42, 43, 44	7

Table 3.8 highlights some sample of questionnaire items exploring participants' understanding on the concepts and principles of language assessment (see Appendix 1 for the complete version). The questionnaire format refers to the structure used in the original work of Hasselgreen, Carlsen, and Helness (2004) which was updated later in the study conducted by Vogt and Tsagari (2014). Some additional items were modified to accommodate the investigation of the present study. The scales interpretation is presented in Table 3.5. The first question in the questionnaire demonstrates how the framework utilized to communicate the statement. In this statement, the frameworks proposed by Brown (2003), Douglas (2014), Fulcher and Davidson (2007), and Kahn (2018) as discussed in the previous paragraphs used as the guidelines.

Table 3.8 Sample of Questionnaire Items

No.	**Items**	**Scales**			
		1	2	3	4
1.	I can explain some concepts of language assessment.				
2.	I can demonstrate how to conduct some types of assessments. *Illustration: Formative assessment occurs when the teachers give a comment or suggestion on students' performance or call attention to an error in grammar.*				
3	I can describe how to conduct receptive skills (reading/listening) assessments.				

4	I can describe how to conduct a productive skills (speaking/writing) assessment. *Illustration: The teachers may conduct an oral interview to obtain information on students' speaking abilities and interpret students' abilities to accomplish a variety of real-world speaking tasks based on their performance in the oral interview.*

3.3.1.2 Validity and Reliability

Two experts who work closely in Language Testing and Evaluation (LTE) evaluated the questionnaire to confirm its validity and reliability. They were chosen under several criteria. First, the two experts were senior lecturers in two state universities. They have been teaching and working in ELT for many years. Second, both experts had licensed to evaluate works related to assessment. Third, they maintained professional communication and were available for consultation during the evaluation and reconstruction of the instruments. Thus, it was significant to have the instruments validated.

A pilot study was conducted with participants with similar characteristics to the present study. A survey questionnaire was distributed to the targeted participants. The pilot was conducted from July 13th to July 20th, 2022. In total, there were thirty-three valid respondents who responded to the questionnaire. The numerical database was analyzed using SPSS version 25.0. All questionnaire statements were confirmed valid. Part one of the questionnaire was exploring respondents' understanding of language assessment. The 24 items were admitted to be reliable. The Cronbach's Alpha was .959, which belonged to the excellent category (Figure 3.3).

Figure 3.3 *The Reliability of Questionnaire Items Exploring Understanding of LA*

Reliability Statistics

Cronbach's Alpha	N of Items
.959	24

The second part of the questionnaire investigated respondents' experience in practicing language assessment. There were twenty items presented using a four-point Likert scale. The Cronbach's Alpha score was .950, considered excellent (Figure 3.4). The full calculation of the pilot study questionnaire using SPSS version 25.0 is listed in Appendix 6.

Figure 3.4. *The Reliability of Questionnaire Items Exploring Practice of LA*

Reliability Statistics

Cronbach's Alpha	N of Items
.950	20

Thus, all items in the pilot questionnaire were used for the questionnaire in the current study. Table 3.8 displays the description of Cronbach's Alpha Values.

Table 3.9 *Description Value of Cronbach's Alpha*

Value of Alpha	Description
0.91 to 1.00	Excellent
0.81 to 0.90	Good
0.71 to 0.80	Acceptable
0.61 to 0.70	Questionable
0.51 to 0.60	Poor
0.01 to 0.50	Unacceptable

(source: Gliem & Gliem, 2003)

3.3.2 Qualitative Data Collection

For the qualitative data collection, semi-structured interviews and document analyses were utilized to get in-depth information.

3.3.2.1 Semi-structured Interviews

The semi-structured interviews were employed for qualitative data collection. Six participants voluntarily participated in the interview. They were selected purposively to assist the data collection process. Purposive sampling was used in qualitative data collection to identify persons with firsthand knowledge of the central phenomenon (Hamied, 2017, p. 198). The purposes of the interviews were to validate the data found in the questionnaire, explore individuals' training background in language assessment, get deeper data on graduate students' understanding of language assessment, and explore their language assessment practices.

The interview questions were structured based on general responses from the questionnaire exploring the ELT graduate students' understanding and practices of language assessment. Additionally, two language testing and assessment experts were consulted to check the validity and clarity of the interview questions. Moreover, the questions were categorized into four sections: demographic information, the concepts and principles of LA, the purposes of LA, and the classroom-focused LA. The interview concerned eighteen critical questions with various prompts (Appendix 2). Table 3.6 shows the coverage of interview questions sections.

Table 3.10*The Interview Questions BluePrint*

Sections	Items	Total
Demographic information	1, 2, 3	3
The concepts and principles of LA	5, 9, 11, 12, 13, 15	6
The purposes of LA	7, 8, 14	3
The classroom-focused LA	4, 6, 10, 16, 17, 18	6

The interviews were conducted and recorded through zoom or WhatsApp application. The length of the interviews varied from 40 to 55 minutes per session. The participants' consent was requested before the interview began. Besides, the participants were also assigned pseudonyms names for confidentiality. They were also informed about the purposes of the study. Additionally, the participants were offered to have the interview using Indonesian or English.

3.3.2.2 Document

Some documents were analyzed in this present study. The reasons for analyzing the documents were to see the practice of language assessment and reflect on the data found in the questionnaire and interviews. Documents such as lesson plans, tests, quizzes, and scoring rubrics were collected from graduate students with teaching experience in ELT. Thus, the data collected from the questionnaire, interviews, and document analyses were identified to reveal graduate students' understanding and practices of language assessment. Document analysis checklist rubric was used to analyze the practice of language assessment. Each item from the rubric score was examined using a 4-Likert type scale: (1) Poor, (2) average, (3) good, and (4) excellent (see table 3.10).

Table 3.11 *Measurement Scale for Document Analysis*

Sections	Total
The concepts and principles of LA	6
The purposes of LA	3
The classroom-focused LA	5

Table 3.11 shows the blueprint of the checklist rubric for document analysis. Similar to the themes presented for the questionnaire and interviews, the documents were analyzed from LA's concepts and principles of LA, the purposes of LA, and the classroom-focused LA. The finding result of the purposefully selected participants is enclosed in Appendix 5.

Table 3.12 *The Document Analysis Checklist Rubric BluePrint*

Scale	Level	Description
1	Poor	No evidence is available due to the participants' limited ability to conduct the given topic related to language assessment.
2	Average	Evidence of the topic appears incomplete because participants have little knowledge of conducting the given topic related to language assessment.
3	Good	Evidence of the topic appears in every part of the document collection because participants have enough knowledge to conduct a given topic related to language assessment.
4	Excellent	Evidence of the topic appears frequently and consistently since participants have complete knowledge of conducting the given topic related to language assessment.

3.3.2.3 Data Triangulation

Multiple sources of data were used to triangulate the data. The functions are strengthening validity, presenting a better overview of a research problem, and examining various approaches to answer research questions (Creswell & Creswell, 2018, p. 290). Triangulation frequently assists in validating study findings by ensuring that several methods of the same topic generate the same outcomes (Hatch, 2002, p. 121). In this study, a complementarity type of triangulation was employed. By allowing the outcomes from several methodologies to inform one another, complementarity triangulation creates a deeper picture of the research findings. Data collected from the questionnaire, interviews, and document analysis were analyzed and interpreted to conclude the findings.

3.3.3 Data Collection Procedures

The data employed in this study were collected from multiple sources. Before distributing the questionnaire, a formal letter requesting permission to conduct the study in the English education program was requested. Then, the link to the questionnaire was distributed through WhatsApp group and participants' email. A message accompanied the link to the survey, encouraging the members to fill out the survey and spread the link to other ELT graduate students studying in the English education program. In section one, the participants were required to read the consent form and submit it before they agreed to participate in the survey. They were asked to fill out some demographic information in section two. The purpose was to avoid double data and better describe respondents' demographic backgrounds. Their data were confidential and guaranteed to be used only for this study. In section three, the respondents were required to complete the questionnaire honestly. There were some close-ended questions. In the last section of the survey, the participants were asked to participate in an interview voluntarily.

Second, semi-structured interviews were conducted via WhatsApp call or zoom application. The participants were chosen by using purposeful sampling. It is believed that the better the position of the participants concerning the topic, the richer the data will be (Leavy, 2017, p. 79). The participants' volunteerism and responses to the close-ended questionnaire were the rationales for choosing them. Later, they were requested to share some documents related to their teaching materials. Their data were kept confidential and only used for this study's purpose.

3.4 Data Analysis

This study was analyzed quantitatively and qualitatively using the explanatory sequential mixed-methods design. First, the data from the questionnaire were described using descriptive statistics to count the number of occurrences of a topic being asked (Leavy, 2017, pp. 111-112). SPSS version 25.0 was used to calculate the descriptive statistics, reliability, and Pearson's correlation analysis. In the descriptive analysis, the average response (mean), the frequency and percentage, and the standard deviation (SD) were calculated (Cooper & Schindler, 2008). Moreover, the reliability analysis was measured to check the consistency of the result using Cronbach's Alpha (Gliem & Gliem, 2003; Taber, 2018). The internal consistency of variables was checked during the pilot and actual study. Table 3.12 shows Cronbach's Alpha coefficient and the interpretation. Additionally, Pearson's correlation analysis was employed to investigate the relationship between ELT graduate students' understanding of language assessment and their practice in this study (Hair et al., 2007; Zaiontz, 2016).

Table 3.13 *Pearson's Correlation Coefficient Interpretation*

Coefficient Range	Description
0.91 to 1.00	Very strong
0.71 to 0.90	High
0.41 to 0.70	Moderate
0.21 to 0.40	Small but definite relationship
0.00 to 0.20	Slight, almost negligible

(source: Hair et al., 2007)

Then, the data were analyzed under language assessment theories. After that, the audios from the semi-structured interviews were transcribed. The interview transcripts were delivered to the participants to confirm their accuracy. Next, it was analyzed qualitatively under four major themes. They were the concepts of LA, the purposes of LA, the principles of LA, and the types of LA.

The analyses followed the language assessment theories discussed in the literature review. Qualitative Data Analysis Software (QDAS), NVivo 12 Plus, was used to code the data. This software has received positive claims in interpreting data across research instruments (Creswell, 2007, p. 167; Hatch, 2002; Merriam &

Tisdell, 2016). The procedures in coding the data using NVIVO 12 plus are as following: a. importing the trancript files, b. open-coding the entire interview transcripts, c. making a node, d. create thematic nodes according to the research frameworks, e. opening previously coded nodes, f. doing systhematic code to confirm the presence of already identified codes (Allsop et al., 2022).

After that, some documents were analyzed qualitatively to confirm the findings in the questionnaires and interviews. The document analyses rubrics were used to evaluate some documents such as lesson plans, the syllabus, assessment worksheet, scoring rubrics, samples of teaching materials, and sample students' worksheets. Thus, both the quantitative and qualitative were interpreted and identified to reveal graduate students' understanding and practices of language assessment.

3.5 Ethical Considerations

Ethical considerations are one of the most crucial factors in research, especially human study. This set of principles guides the research design and practice. In this study, the ethical considerations follow the guidance proposed by Bryman and Bell (2011). First, the participants were voluntarily recruited. Second, the consent form was informed before the participants agreed to fill in the questionnaire and participate in the interview. Then, the confidentiality of participants' information was guaranteed and was only shared with the supervisor. For the interview, participants were assigned pseudonyms names to ensure anonymity. Next, the participants were not subjected to harmful acts, and their dignity was prioritized. Also, the source of funding was declared. Last, the results of interview transcriptions were sent back to the interviewees to avoid misleading information.

3.6 Concluding Remarks

Chapter three presents the research methodology employed in this study. The method, research site, and participants have been elaborated. Data collection techniques that involve quantitative and qualitative data have also been discussed.

Further, the data collected were also analyzed to address the research questions. The findings and discussion of this study are presented in chapter four.