

## **CHAPTER III**

### **RESEARCH METHODOLOGY**

This chapter presents methodology of this study. At the beginning, research design of this study is pictured to comprehend how this study was conducted. Afterwards, research site and participants are mentioned. The data collections and analyses are outlined in orders based on the research design. Then, the last section concludes this chapter.

#### **3.1. Research Design**

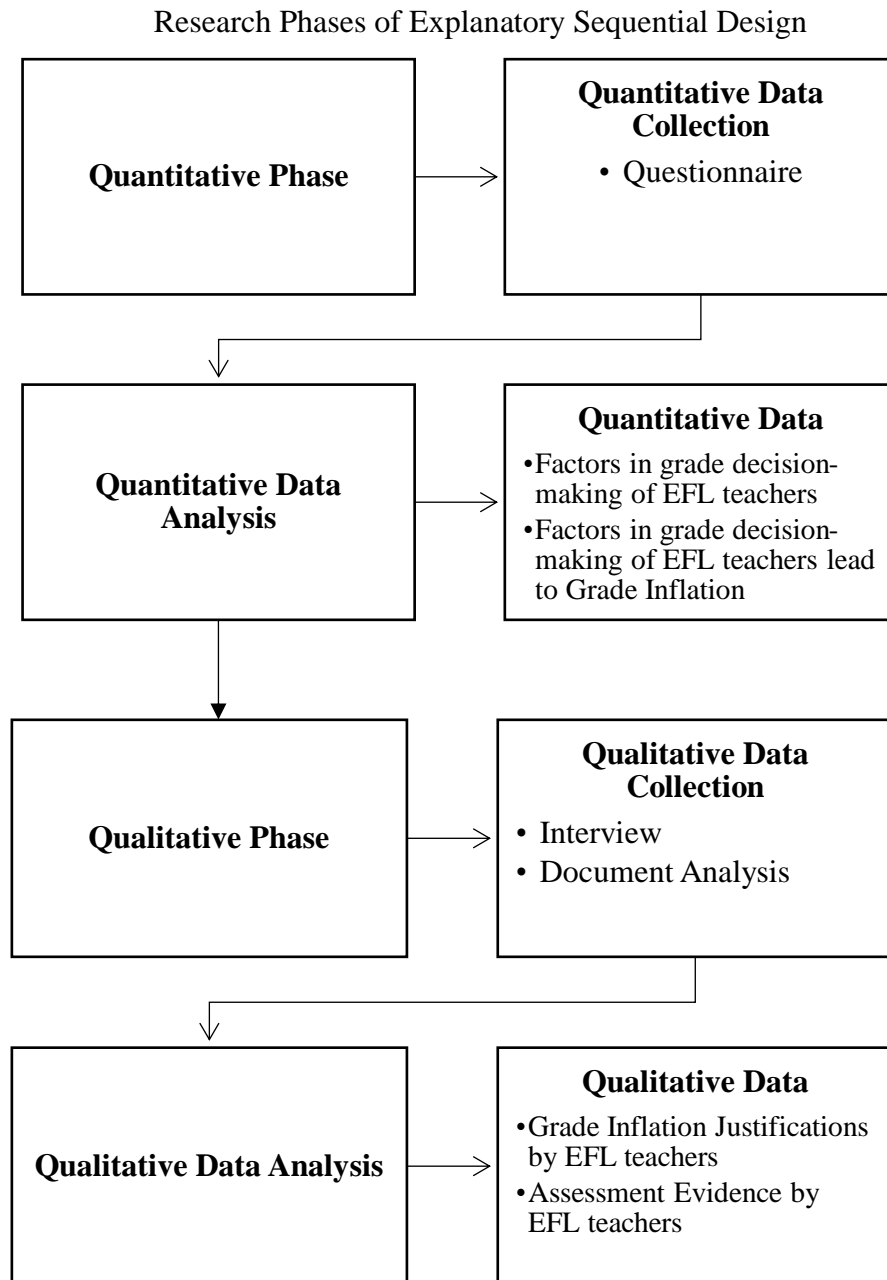
This study was conducted using an explanatory sequential mixed-methods design after contemplating three main characteristics: timing, emphasis, and mixing (Ivankova and Creswell, 2009; Jafari et al., 2015). Timing is the sequence of quantitative and qualitative data collections and analyses. In this study, quantitative data collection and analysis were initially enacted. Afterward, the data were followed up by using the qualitative phase. Emphasis refers to a method, quantitative or qualitative, that is more prioritized. This study utilized quantitative and qualitative methods since quantitative and qualitative data were needed to answer the research questions. Mixing is the phase where two quantitative and qualitative methods are integrated. As aforementioned, quantitative to qualitative phases were connected sequentially since quantitative data must be gathered at the beginning to select participants for the interview.

This study used an explanatory sequential mixed-methods design for several reasons aligned with Creswell's explanation (2012). First, a mixed method is needed when one type of research cannot cover the research questions. A quantitative instrument, a questionnaire in this research, provides descriptive statistics to assess the frequency of EFL teachers' grade decision-making and to reveal what factors lead to grade inflation. The questionnaire findings only indicate EFL secondary school teachers performed hodgepodge grading practice, a combination of all grade factors, that makes their grade decision-making potentially lead to grade inflation (McMillan & Nash, 2000).

Second, more data are required to extend, elaborate, and explain the obtained data from the questionnaire. Although the questionnaire findings show EFL secondary teachers potentially did grade inflation, according to Brown and Abeywickrama (2010), a hodgepodge grading is acceptable if teachers transform those non-academic factors into measurable and observable evidence in assessment. Hence, it is crucial to uncover how EFL secondary teachers provided assessment evidence and their justifications by employing interviews and document analyses as qualitative data collection.

Following an explanatory sequential mixed-methods design (Creswell, 2012), quantitative data were collected and analyzed at the beginning of the study through a questionnaire to obtain descriptive statistics on the grade decision-making of EFL secondary school teachers. The twenty-five-item questionnaire was spread online using Google Forms to answer the first research question. Then, the obtained data was followed up using qualitative data collection and analysis of a semi-constructed interview and document analysis. In the end, the interpretations of data were sequentially discussed on how qualitative data explain quantitative data based on theories discussed in the literature review. The phases of this research design are illustrated in Figure 3.1.

Figure 3.1.



*(Adapted by Creswell, 2012)*

### 3.2. Research Site and Participants

In determining samples of this study, among four sampling models of mixed method: identical, parallel, nested and multilevel (Collins et al., 2006; Creswell & Guetterman, 2019; Jafari et al., 2015; Onwuegbuzie & Collins, 2007; Teddlie & yu,

Romadhon, 2023

**GRADE DECISION-MAKING AND GRADE INFLATION BY EFL SECONDARY SCHOOL TEACHERS AND THEIR MORAL JUSTIFICATIONS IN ONLINE LEARNING DURING THE PANDEMIC**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

2007), this study framed parallel samples. Since this study performed an explanatory sequential mixed-methods design, the relationship between quantitative and qualitative phases is parallel and connected. The relationship parallel means samples of quantitative and qualitative phases are different but drawn from the same population of interest. The interview participants and documents in the qualitative phase were taken from the questionnaire respondents as the population in the quantitative phase.

This study captured a non-probabilistic sampling scheme in the quantitative with several criteria to generate samples that address the research questions (Teddlie & Yu, 2007). The population refers to Indonesian EFL teachers in junior and senior high secondary schools, whether in private or public schools. This study focused on EFL teachers in Indonesian secondary schools during the pandemic since it was relatively new and more challenging for secondary schools to implement technology-based online learning than higher education.

Meanwhile, the samples of this study refer to several criteria: they had experienced teaching English online during the pandemic, with at least one semester or a half year of teaching and conducted online learning, including distance and blended learning, whether synchronously or asynchronously. Besides, the limited time and cost of collecting the questionnaire determined the number of respondents in this study. Based on the data collected from the questionnaire, as shown in Table 3.1, EFL teachers across provinces participated (N=235), with the most dominant coming from Jawa Barat (N=151). It can be seen in Table 3.1 that the respondents who participated in this study came from all over Indonesia's provinces. Also, each island in Indonesia, such as Jawa, Sumatera, Sulawesi and Kalimantan, was generally involved in representing EFL secondary school teachers.

Furthermore, in the qualitative phase, purposive sampling was administered to select five participants who: have teaching experiences of more than five years, have joined assessment training or workshop during the pandemic, represent the public or private school at any education level, and have fulfilled the interview and document analysis consent form. Five selected participants represent EFL secondary school teachers across levels and status from Islamic private junior high

school, public junior high school, Islamic private senior high school, public senior high school, and public vocational school, as seen in Table 3.10.

Table 3.1  
Province

Province	Frequency	Percent	Cumulative Percent
Aceh	1	.4	.4
Banjarmasin	1	.4	.9
Banten	6	2.5	3.0
DI Yogyakarta	3	1.3	4.7
DKI Jakarta	11	4.7	9.4
Jambi	1	.4	9.8
Jawa Barat	151	64.2	73.6
Jawa Tengah	19	8.1	81.7
Jawa Timur	4	1.7	83.4
Kalimantan Barat	3	1.3	84.7
Kalimantan Selatan	10	4.3	88.9
Kalimantan Timur	3	1.3	90.2
Kalimantan Utara	1	.4	90.6
Lampung	7	3.0	93.6
NTT	1	.4	94.0
Riau	6	2.6	96.6
Sulawesi Selatan	1	.4	97.0
Sulawesi Tengah	1	.4	97.4
Sumatera Barat	1	.4	97.9
Sumatera Selatan	1	.4	98.3
Sumatera Utara	3	1.3	99.6
Total	235	100.0	

Without relating to the data interpretation, the data recorded the gender of respondents. Female EFL teachers highly participated in the questionnaire (N=185) compared to male EFL teachers (N=50), as presented in Table 3.2.

Table 3.2

## Gender

Gender	Frequency	Percent
Male	50	21.3
Female	185	78.7
Total	235	100.0

Among all levels in the Indonesian secondary school, EFL teachers from junior high schools (SMP) were the dominant respondents (N=103). It differs highly from Islamic junior high schools (MTs), where EFL teachers participated less (N=26). Further, respondents from senior high schools were the highest (N=51), followed by vocational schools (N=43), and the lowest overall was respondents from Islamic senior high schools (N=12). The illustration can be found in Table 3.3. This study presents respondents' education levels to convey that different education levels also have other assessment policies.

Table 3.3

## Education Level

Gender	Frequency	Percent	Cumulative Percent
SMP	103	43.8	100.0
MTs	26	11.1	16.2
SMA	51	21.7	37.9
SMK	43	18.3	56.2
MA	12	5.1	5.1
Total	235	100.0	

As shown in Table 3.4, all education levels in public schools were more frequently recorded (N=120) than in private schools (N=115). The education status of respondents is also essential to convince that assessment policies and standards as external factors in each status varied.

Table 3.4

Education Status			
School Status	Frequency	Percent	Cumulative Percent
Public School	120	51.1	51.1
Private School	115	48.9	100.0
Total	235	100.0	

The most frequent respondents in the questionnaire were EFL secondary teachers with teaching experiences from one to five years (N=93). Meanwhile, the lowest percentage comes from EFL secondary schools with less than a year of teaching experience (N=14). This data is needed to correlate the grade decision-making of EFL secondary school teachers with their teaching experiences. Further detailed records can be seen in Table 3.5 below.

Table 3.5

Teaching Experiences			
Years	Frequency	Percent	Cumulative Percent
< 1 year	14	6.0	6.0
1 - 5 years	93	39.6	60.4
6 - 10 years	38	16.2	100.0
11 - 15 years	55	23.4	83.8
> 20 years	35	14.9	20.9
Total	235	100.0	

Addressing to assessment training experienced by respondents, most of them had the training (N=161) with various frequencies from once (N=87), twice (N=20), and more than twice (N=56). Although several respondents claimed they never had assessment training (N=72), they are still included in the respondent

criteria. Respondents' assessment training is essential to take into account for grade decision-making. Further details can be seen in Tables 3.6 and 3.7 below.

Table 3.6

Assessment Training			
Training	Frequency	Percent	Cumulative Percent
Never	72	31.5	31.5
Ever	163	68.5	100.0
Total	235	100.0	

Table 3.7

Assessment Training Frequency			
Experience	Frequency	Percent	Cumulative Percent
Once	87	37.0	100.0
Twice	20	8.5	39.1
More than twice	56	23.8	63.0
Never	72	30.6	30.6
Total	235	100.0	

The personal data of the respondents above are protected and guarantee research ethics; all data related to this study is under their consent, and pseudonyms are employed.

### 3.3.Data Collection Techniques

In collecting the data, following an explanatory sequential mixed-methods design, two phases were employed that consist of:

#### 3.3.1. Quantitative Instrument

In quantitative data collection, the instrument is only a questionnaire. It was conducted as the first phase of this study to get general descriptions since it provides



descriptive statistics to assess the frequent grade decision-making and to reveal major factors that lead to grade inflation of EFL secondary school teachers. Demographic categories were displayed at the beginning of the questionnaire (see Table 3.1 to Table 3.7). Participants' consent and noted descriptions were presented in the Google form to guarantee research ethics. Four categories were presented in the questionnaire. First, internal factors carry three items. Second, external factors have six items. Third, academic, cognitive, and product factors were arranged into eleven items. The last, non-academic, non-cognitive, and process factors, were converted into three items.

Interval or ratio scales (Creswell, 2012), well-known as the Likert scale, were operated to provide continuous response options. According to Creswell and Guetterman (2019), these scales may have three, four, or more response options depending on adequacy. In this study, four scales were used to avoid a neutral answer. Four scales were arranged from strongly disagree to strongly agree to measure teachers' ideology, referring to internal factors' response options. Also, four scales from always to never were set to estimate external, academic, non-academic, and other factors. The number of questionnaire items is twenty-three, with four major categories that need approximately ten to fifteen minutes to be fulfilled.

### 3.3.1.1 Validity and Reliability

The questionnaire was adopted from Cheng and Sun (2015), that studied Chinese English teachers grading decision-making: multiple influencing factors and methods in face-to-face teaching and learning. Then, it was adjusted to the Indonesian context using frameworks from several related studies (Arrafii, 2020; Isnawati & Saukah, 2017; Widiastuti, 2018; Zulaiha, 2017) and as well as in the online learning context. The questionnaire consists of two significant grading factors: external to internal, cognitive to non-cognitive, academic to non-academic, and product to process. Two experts who devote themselves to language assessment and have published journal articles related to assessment were asked for content validity (Creswell, 2012). Major revisions on questionnaire items, interview

questions, and document analysis were completed following the experts' feedback and suggestions.

Also, a validity test was performed using Statistics. As seen in Table 3.8, all questionnaire items are valid since their results of Pearson Correlation are more than .381 of 5% significance in the r table of twenty-seven respondents. This study used a two-tailed validity test because the probability came from two directions. It means that EFL secondary teachers were assumed to conduct several factors in grade decision-making without certain tendencies. Thus, questionnaire items are qualified to be used in this study.

Table 3.8  
Validity Statistics

Q1	Q2	Q3	Q4	Q5	Q6
.599**	.487**	.511**	.641**	.805**	.599**
Q7	Q8	Q9	Q10	Q11	Q12
.487**	.634**	.725**	.554**	.489**	.544**
Q13	Q14	Q15	Q16	Q17	Q18
.420*	.767**	.724**	.480*	.608**	.608**
Q19	Q20	Q21	Q22	Q23	
.787**	.707**	.798**	.743**	.562**	

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

Further, the questionnaire was piloted on twenty-seven secondary school English teachers with the same characteristics. Using SPSS, the piloted data were measured to get internal consistency reliability. It presents that all items in the questionnaire are reliable since each item reached .864, more than .60 Cronbach's Alpha, as shown in Table 3.9.

Table 3.9  
Reliability Statistics

Cronbach's Alpha	N of respondents
.864	27

### 3.3.2. Qualitative Instruments

After gaining the data from the questionnaire, the data were deepened using a semi-structured interview and document analysis.

#### 3.3.2.1. Interviews

Interviews were conducted to elaborate on the obtained data from the questionnaire and comprehend the research problem from an exhaustive to a narrow perspective. Notably, the interview is practical to explore how EFL secondary school teachers graded students, justified grade inflation, and provided measurable and observable assessments. A semi-structured was employed using open-ended questions to cover various teachers' classroom assessment practices. Merriam and Tisdell (2015) explain that this type of interview helps discover participants' perspectives about defining the world in unique ways. The interviews were employed in Indonesian as the native language to get detailed information (Alwasila, 2002) and maximize the data exploration and connectedness (Filep, 2009).

The interview questions were adopted from Yesbeck (2011) and adapted to the Indonesian context following Arrafii (2020). Unlike Arrafii (2020), which employed a group discussion besides the interview, this study focused on conducting in-depth questions. Like questionnaire items, the interview questions were developed based on two classifications of grading factors: external to internal, cognitive to non-cognitive, academic to non-academic, and product to process. Twenty-two questions took roughly twenty to forty minutes.

Five English teachers were selected from all respondents as representatives from public and private junior and senior high schools and vocational schools that previously fulfilled a consent form in the questionnaire. Since the relationship

between quantitative and quantitative phases is parallel and connected, among all questionnaire respondents, this study purposively filtered five English teachers who fell into several criteria: be willing to be interviewed, have more than five years of teaching experience, and have ever joined assessment training or workshops. Table 3.10 below presents the demographic backgrounds of participants whose names have changed to pseudonyms.

Table 3.10

Participants' Demographic Backgrounds

No.	Pseudonym Name	Gender	Teaching Experiences (Years)	Grades/ Semester	Education Levels	School Status	Assessment training and or workshops
1.	Mrs Ambar	Female	11 - 15	IX/ I	MTs	Private	Once
2.	Mrs Bunga	Female	11 - 15	VII/ II	SMP	Public	Once
3.	Mrs Cantika	Female	11 - 15	XII/ II	MA	Private	More than twice
4.	Mrs Dini	Female	11 - 15	X/ I	SMK	Public	More than twice
5.	Mr Ega	Male	6 - 10	X/ I	SMA	Public	Once

The interview was conducted virtually using Zoom meeting, then recorded participants' responses, and later the data were transcribed and coded using Google voice typing.

### 3.3.2.2. Document Analysis

Instructional materials such as lesson plans and syllabi were collected from selected EFL teachers. The documents related to assessment standards provide data on how the EFL teachers score and grade students in four document types: rubrics, classroom observation, tests, and students' academic reports. The data triangulated data from the interview.

### 3.3.2.3. Data Triangulation

This study implemented several strategies of data triangulation to make qualitative data credible (Merriam & Tisdell, 2015). First, this study conveyed multiple data collection methods, such as interviews and document analysis. The data gathered from documents validated data taken from the interviews. Second, the

interview also employed member checking to ensure the data interpretation met the participants' meaning and avoided misunderstanding.

### **3.4. Data Analysis**

After collecting the data, each type of data collection was analyzed, as explained in the following paragraphs.

#### **3.4.1. Quantitative Data Analysis**

The data from the questionnaire were analyzed using descriptive statistics to obtain the data frequency by calculating the mean, median, and mode of grade decision-making employed by English teachers. After that, the results were classified based on the most and the lowest frequency of grading practices (Creswell, 2012). The last, the results were described and interpreted using grade frameworks such as external to internal, cognitive to non-cognitive, academic to non-academic, and product to process (Arrafii, 2020; Brookhart et al., 2016a; Guskey & Link, 2019; Lawrence et al., 1996; McMillan et al., 2002; Yesbeck, 2011).

#### **3.4.2. Qualitative Data Analysis**

In analyzing data, two framework classifications of grading factors were used: external to internal, cognitive to non-cognitive, academic to non-academic, and product to process (Arrafii, 2020; Brookhart et al., 2016a; Guskey & Link, 2019; Lawrence et al., 1996; McMillan et al., 2002; Yesbeck, 2011). Firstly, the interview transcripts were categorized into four major categories: external factors, internal factors, cognitive to non-cognitive, academic to non-academic, and product to process factors. Then, those major themes specify into several themes taken from the interview questions as presented in Table 3.11. Second, the data were interpreted following the grade frameworks and grade guidelines to identify whether teachers provide measurable and observable evidence to prevent their grade practices from grade inflation.

Table 3.11  
Themes in the interview

Codes	Themes	Categories	
1	Teachers perform their beliefs in their assessment processes.	Internal Factors	
2	Teachers consider their values in assessing students.		
3	Teachers implement their background knowledge in assessing students.		
4	Teachers manifest their assessment training and workshops to assess students.		
5	Teachers reflect their trust in their assessment process.		
6	Teachers follow assessment standards in emergency remote teaching	External Factors	
7	Teachers maintain high-stakes tests in assigning students' grades		
8	Teachers obey regional grading policies		
9	Teachers maintain particular school policies for grading students.		
10	Teachers contemplate parents' involvement that has impacts on students' academic achievement.		
11	How the teachers assessed academic, cognitive, and product students online during the pandemic.		
12	How the teachers assessed non-academic, non-cognitive, and process students online during the pandemic.		Academic, cognitive, and product, and Non-academic, non-cognitive, and process factors
13	The most dominant domain teachers used to assess students online during the pandemic		
14	Teachers consider challenges and difficulties in assessing students in online learning during the pandemic.		

### 3.5. Concluding Remark

This chapter discusses the study's research design, site and participants, data collection, and analysis. The demographic backgrounds of the respondents are presented in the research site and participant sections. The following chapter will draw on the study's findings and discussions.