

CHAPTER III

RESEARCH METHODOLOGY

This chapter outlines the research methodology. It provides research setting that includes the site and participants, data collection procedures, and data analysis procedures.

3.1 Research Design

This study employed a qualitative case study approach to investigate the outlook of the English summative test questions following the principles and procedures that the teachers employ in constructing the tests. A qualitative case study as an approach focusing on an event, place, topic, time, or other physical limits (Creswell, 2012). A case study is a detailed description and analysis of a bounded system (Merriam, 1998). The term "bounded system" refers to a single entity or unit with defined boundaries. The most distinctive feature of case study research is the precise delimitation of the object of study or the case itself. This case could be an individual representing a specific phenomenon, a program, a group, an institution, a community, or a particular policy. The defining feature of a case study is the unit of analysis rather than the topic of investigation. Furthermore, a case study as the research process, stating that "a case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between the phenomenon and context are not evident" (Yin, 2008, p.18).

A qualitative case study approach best suited this research design as this study's concerns are how the test is structured and the procedures used by teachers align with qualitative research's emphasis on processes, meanings, and interpretations. A case study provides the opportunity to explore the "how" and "why" questions surrounding these assessment practices. Furthermore, this approach helped to collect rich, detailed data from various sources such as questionnaires, interviews, and document analyses. This gave a comprehensive view of how EFL teachers approach test construction, enhancing the reliability of the findings.

The following flow diagram shows the key components of the research design and procedures.

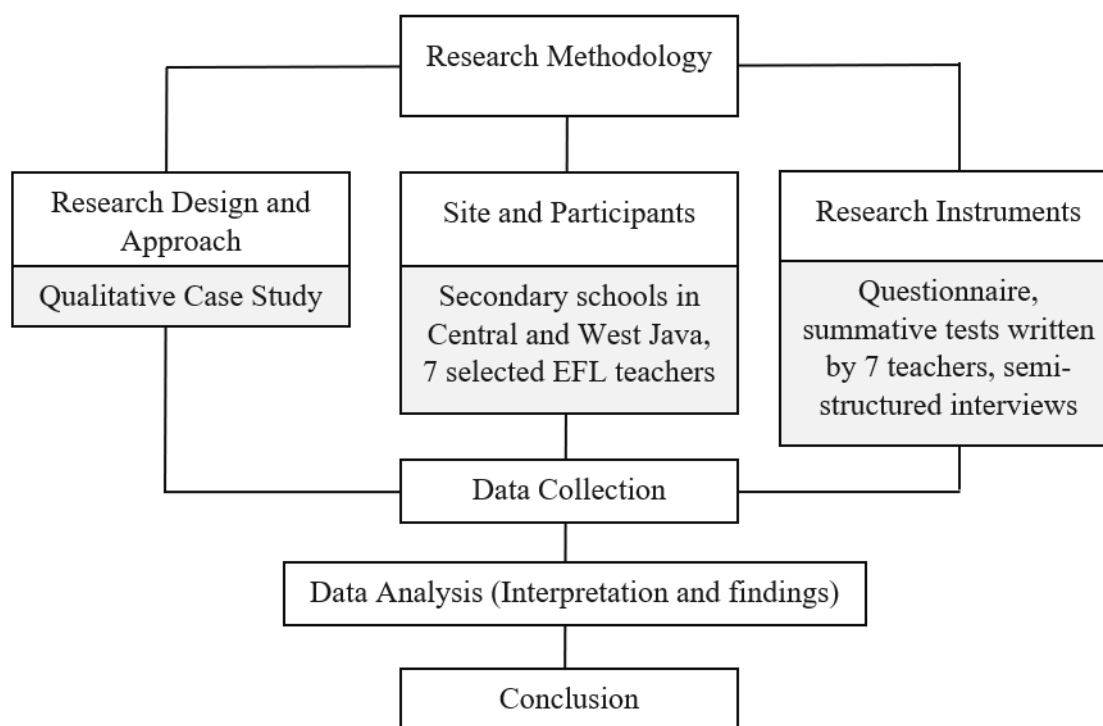


Figure 7. Key Components of the Research Methodology

3.2 Site and Participants of the Study

This study was carried out in several secondary schools located across Central Java and West Java through convenience sampling. These regions were chosen due to the researcher's ease of access, time constraints, and the feasibility of establishing communication with participating schools. As stated by Creswell (2007), this strategy allows researchers to select participants who are easy to reach in order to obtain the information needed. In Creswell's (2012) view, convenience sampling involves participants who are willing and available for research. McMillan and Schumacher (2001) also emphasise that convenience sampling is closely related to the accessibility and experience of participants. Rahi (2017) adds that this technique describes the process of collecting data from a population that can be reached easily and in practice, the selected population members usually fulfil practical criteria (Dörnyei, 2007). Additionally, the selected regions provide a range of school types and educational settings that reflect the broader diversity of Indonesia's secondary education system. This study

did not include other provinces due to practical and methodological considerations. Expanding the geographical scope beyond Central and West Java would have posed challenges in terms of travel, time management, and access to participants.

To ensure the data collected were diverse and representative, participants were selected based on a set of predetermined criteria. These included English teachers at junior and senior high school levels with teaching experiences ranging from one to ten years, aged between 26 and over 35 years, and holding educational qualifications from bachelor's to master's degrees, regardless of teaching certification status. Participation in the questionnaire was voluntary, with no fixed quota imposed. As a result, a total of 77 teachers agreed to participate and completed the questionnaire.

After the questionnaire completion, selected participants were proceeded to the in-depth interview stage and their test documents were analysed. The selection of participants for the interview and document analysis stages was carried out using a questionnaire as a recruitment tool. The selected candidate participants were those who obtained the highest total scores on the questionnaire and provided responses of “strongly agree” (score 5) or “agree” (score 4) to all items. In cases where the highest-scoring participants selected “not sure” (score 3), “disagree” (score 2), or “strongly disagree” (score 1) on any item, they were replaced by participants with lower overall scores who met the response criteria. This selection procedure was employed to ensure alignment, whether teachers who claimed to possess a good understanding and had applied assessment principles in their practice truly demonstrated these competencies. At the junior high school level, eight teachers were initially selected, comprising five from Central Java and three from West Java. However, four of them declined to proceed to the interview and document analysis stages. Consequently, only four junior secondary school teachers participated further, involving three from Central Java and one from West Java. At the senior high school level, five teachers were selected: three from West Java and two from Central Java; however, two withdrew from participation. Finally, the total number of teachers who took part in the interviews and document analysis was seven, consisting of four junior secondary school teachers and three senior secondary school teachers. The following is a breakdown of the teachers involved in the interviews and the test documents analysed.

Table 2. Details of selected participants in this study

Teacher	Name of Teachers	Gender	Age	Level of Education	Level of Teaching	Years of Teaching Experience	Teaching Certification
1	DM	Female	> 35 years old	Bachelor's Degree	Senior High School	>10 years	Certified English Teacher
2	ED	Female	> 35 years old	Master's Degree	Senior High School	>10 years	Certified English Teacher
3	AKAP	Female	26–30 years old	Bachelor's Degree	Senior High School	1–5 years	Not Certified
4	DF	Male	> 35 years old	Master's Degree	Junior High School	>10 years	Certified English Teacher
5	AF	Female	26–30 years old	Master's Degree	Junior High School	1–5 years	Not Certified
6	DD	Male	> 35 years old	Bachelor's Degree	Junior High School	>10 years	Certified English Teacher
7	FF	Female	26–30 years old	Bachelor's Degree	Junior High School	1–5 years	Not Certified

3.3 Research Instruments

The data instruments have strengths and weaknesses based on the different types of information demanded (Creswell & Creswell, 2018). Therefore, this research involved three instruments to collect qualitative data. The first is a questionnaire. As questionnaires can elicit skewed responses (Fulcher, 2012), only using them is not enough since respondents can give answers that are not by reality. Therefore, other data instruments are used to strengthen the research results. They are semi-structured interviews and some samples of summative tests made by the teachers.

The questionnaire in this study began with questions regarding the participants' self-identity and demographics. After that, the questionnaire focused on self-perceived knowledge, understanding, teachers' test-taking skills and practices in test construction, adapted from Fulcher's (2012) and Tewachew's (2024) questionnaires. In the original version, Fulcher's questionnaire asks teachers to rank the importance of topics in assessment that should be included in language testing courses. The scale used is a 1-5 Likert scale, ranging from unimportant to essential. This questionnaire has been widely adopted by previous researchers, such as Farmasari et al. (2023) as well as Herrera and Macías (2015), who utilised Fulcher's language assessment literacy questionnaire to measure satisfaction levels in Language Assessment Literacy (LAL). In this study, the questionnaire was modified by maintaining the original model but adjusting some aspects to capture different dimensions of teachers' knowledge and understanding. Out of 23 topics in language assessment, only 11 topics will be selected for use.

In terms of Tewachew's (2024) questionnaire, modifications were made by reducing the number of statements from 33 to 29 by eliminating some statements that were not relevant to the scope of the study. Some items were also adjusted in redaction to clarify their meaning. In addition, two new items were added to this section to explore the types of tests teachers commonly construct and the sources for item creation. The finished draft questionnaire was validated first to ensure it can be used in the study. The criteria for validators included English department lecturers with 10 years of teaching experience who had obtained a doctoral degree and are an expert in assessment. They are Mr. AS and Mr. AA from public universities in Bandung, West Java.

3.4 Data Collection Techniques

In collecting the data, the researcher followed several steps. The first one was distributing questionnaires to the teachers through Local Teacher Association (MGMP) WhatsApp group in Central and West Java. The two areas were chosen due to ease of access, where the researcher's domicile is Central Java and currently the researcher is continuing her studies in West Java. The questionnaire was translated into Indonesian to avoid teachers'

misunderstandings when interpreting its context. After that, the data were presented in percentage form to see the teachers' dominant patterns of understanding and practices in assessment. Teachers with 'strongly agree' and 'agree' responses to each question on the questionnaire were probed to see if the claims are true. Then, semi-structured interviews with selected participants were conducted to explore their practice on the test construction.

The next step was to collect the samples of the summative test made by the selected teachers being interviewed beforehand. The test items were analysed through guideline rubrics, which the researcher developed from the theories of some experts.

3.5 Data Analysis Techniques

To answer research question 1, the participants were asked to fill the questionnaire/survey to explore their knowledge, understanding, and practices in classroom assessments and test construction procedures. Besides, selected participants were interviewed regarding the procedures they employed in constructing summative tests, and then the recordings were transcribed. After that, the researcher analyzed and explored the teachers' literacy assessment.

On questionnaire results, the data was presented as percentages which represent teachers' confidence based on personal claims of their knowledge and skills in test design. Responses of strongly agree and agree indicate teachers' belief and active engagement in practice, and are therefore categorized as indicators of teachers who have adequate understanding in test construction. In contrast, teachers who responded disagree and strongly disagree were less likely to be confident in their knowledge and abilities. They indicate limitations in knowledge and practice. Meanwhile, some teachers who responded unsure were in the midst of uncertainty between whether or not they understood and whether or not they practised certain steps in test construction. This reflected their confusion about how well they had mastered the skills or whether they had actually applied certain steps effectively.

To answer research question 2, the teacher-made tests were categorized according to the principles as shown in the table below, which was developed by combining guidelines proposed by Brown and Abeywickrama (2019), Cunningham (1998), Hughes and Hughes

(2020), Kubiszyn and Borich (2016), Popham (2017), Pusmendik (2019), and also Tewachew et al. (2024).

Table 3. Rubric criteria for summative test writing developed by the researcher

Aspect	Criteria
General	Each item matches the blueprints.
	Test items are original, adapted with substantial modification
	The items formulate a problem with clear instructions as well.
	Each question connects to the text stimuli.
	Each question requires test takers to read the text (stimuli) to answer the questions.
	Test items are free from ambiguity.
	The design of each item ensures that no clues lead to the correct answer.
	The items are grammatically correct.
	The items have correct spelling, punctuation, capitalization, and word choice.
	Each question has an identified and valid correct answer(s).
	Each answer is formulated independently of prior questions.
	There is a time allotment in the test.
	The test uses a variety of item testing formats.
	The tests range in complexity from easy to complex and cover both LOTS and HOTS.
Specific	MCQ and CMCQ/MCMA
	Answer choices are homogeneous.
	The distractors are plausible.
	Test items avoid using of ‘all of these’ or ‘all the above’ as well as ‘none of the above’ as an alternative option.

	Matching
	Answer choices are homogeneous.
	The number of responses is more than the number of premises.
	True False
	Each item is clearly true or false.
	Each item focuses on a single clear idea within the statement.
	Short answer
	The answers required by the questions are short and definite, i.e. in the form of a word, phrase, numbers, symbols, places, or times.
	Essay
	The questions are constructed in a way that prompts students to provide elaborative responses.

If the test or a particular test item meets the aspect or criterion, it will be marked with a checklist in the table of analysis. For specific aspects, the assessment of teacher-made tests was carried out based on the type of questions used in the test. That is, if a test only contains multiple choice questions, then the assessment is only focused on the form of multiple-choice questions; the same applies to other forms of questions such as essays, matching, short answer, and so on. After completing the checklist, each test was further described qualitatively in terms of its overall adherence to the established requirements as well as any issues identified.

3.6 Concluding remarks

This study used qualitative case study design and approach which concerns on how the test is structured aligned with qualitative research's emphasis on processes, meanings, and interpretations. This research involved English teachers at junior and senior high school levels with teaching experiences ranging from one to ten years, aged between 26 and over 35 years, and holding educational qualifications from bachelor's to master's degrees, regardless of teaching certification status. The instruments used in this study were questionnaires adapted

from Fulcher (2012) and Tewachew (2024), semi-structured interviews, and teacher-made tests. The questionnaire has been validated by two experts who were English department lecturers with 10 years of teaching experience who had obtained a doctoral degree and are an expert in assessment. A total of 77 teachers agreed to complete the questionnaire and 7 were proceeded into interviews. They were teachers with ‘strongly agree’ and ‘agree’ responses to each question on the questionnaire who were probed to see if the claims are true.

In analyzing teacher-made tests, a scoring rubric was used by combining expert theories of Brown and Abeywickrama (2019), Cunningham (1998), Hughes and Hughes (2020), Kubiszyn and Borich (2016), Popham (2017), Pasmendik (2019), and also Tewachew et al. (2024) which categorized general and specific aspects of assessment.

In the next section, the findings of the study were elaborated with reference to the planning, designing, and try-out processes as well as the analysis of the teacher-made tests.