

## **CHAPTER III**

### **RESEARCH METHODOLOGY**

This chapter describes the steps taken by the researcher in conducting this study. There are four sections in this chapter. The first section is research design. It presents the methods employed in this study. The second section is research site and participants. It elaborates the persons getting involved in the research and where this study is conducted. The third is data collection method. It explains the instruments and the data collection procedures of this study. The last section is data analysis. It shows how the data are analyzed in order to answer the research questions have formulated.

#### **1.1. Research design**

This study used qualitative case study design. Case study which focuses on studying social phenomena through the analysis of an individual case is relevant to this study for several reasons (Malik & Hamied, 2014).

Firstly, this study used multiple sources of data. Through case study design, this study relied on multiple data sources of evidence (Croker & Heigham, 2009). It involved both data sources which were taken from document analysis and interview (Creswell, 2012). The multiple data sources helped the researcher investigate and understand the focus of this study, which were the teachers' ability on the quality of teachers' made test and the process of constructing the reading test.

Secondly, this study concerns with a single case which is teacher competence in constructing a reading test. Through a single case, this study provides an opportunity for the intensive analysis of specific details of teachers' ability in constructing a reading test (Malik & Hamied, 2014).

Thirdly, this study used small scale in which attempts to maintain holistic approach to understand the context of teacher competence in constructing a reading test. Through the small scale, this study required the detailed and in-depth information taken from teacher's interview.

## **1.2. Research site and participants**

According to the feasibility of the study, the researcher decided to investigate teachers' competence in constructing a reading test in one of the small towns in West Java. Due to that reason, this study was conducted in Banjar.

This study involved three different teachers of English from three different vocational schools. As the name indicates, case study design deals with the total process. As it deals with the process, the site therefore was chosen with care relying on the purposive sampling techniques.

The researcher found the first participant as the key informant (Malik & Hamied, 2014). The researcher got the information herself that she is also teaching at the same vocational school with the first participant. The first participant is a fifty three year old man. He has been teaching since 1988 up to now. He has more than twenty-five years of teaching experiences.

The second participant is twenty-eight years old. She has been teaching at one of the state vocational schools in Banjar for around five years. Besides her teaching activity at vocational school, she also led and taught at the private English course. Her private English course was quite well known in Banjar. Actually, she ran her English course inherited from her parents. That English course offered services for the students from the level of kindergarten up to senior high students. The estimated number of students who were enrolling this English course was around 100 students per year.

The third participant is twenty-nine years old. She has been teaching at the private vocational school since 2014. She graduated from the same university as the second participant. All the participants to this study involved in the forum of vocational teachers of English in Banjar.

### **1.3.Data collection method**

#### **1.3.1. Interview**

This present study employed interview to the participants of this study. Interview was used to record information from the participants. The model of interview were semi structured interview. Through semi structured model interview, it allows the researcher to clarify finding which are gained from document analysis (Cohen, Manion, & Morrison, 2011).

The semi structured interview was used to see the detail exploration of teachers' understanding on the quality and the way of making a reading test. There were seven guiding questions in gaining the information of the participants about reading tests they made. Those seven questions were about background, experience, opinion, feeling, and knowledge (Patton, 1990).

The first guiding question was about the test interval. The second was about the guidance of making a reading test for teachers. The third was asking the segmentation of skill in giving the test their students. The fourth was asking about the selection of the text for the reading test, whether it is authentic text or made by the teacher. It should be asked because it dealt with the text's level of difficulty. The fifth guiding question sought the process of making reading test by the teachers. The sixth guiding question covered the scoring of the reading test. It also covered whether it is criterion or norm-reference scoring. The last guiding question was about the participants' opinion of the overall testing reading they made, whether it has measured that should be measured or not (Hughes , 2003). Last but not least, the 'guiding question' or 'interview guide' was used to get the flexibility. The flexibility is a key requirement of qualitative interviewing (King & Horrocks, 2012).

The interview ran about twenty to thirty minutes or so for each participant. The first interview was done to the first participant at the second school break around 12. a.m. The interview started with the light conversation about the topic. The second interview was completed to the third participant. Interview with this

third participant was in informal situation. The researcher decided to make this in the informal situation in order to have in depth interview without getting formal in gaining information of the selected topic. The last interview was done to the second participant after school time at the participant's school. The interview was almost the same with the first participant.

All of the interviews were done by recording them. The commonest fashion in qualitative interviewing is by recording. The recording was conducted on the purpose to avoid losing the data. Besides recording, the researcher sometimes took notes as quickly as possible. This was done in order to support specific data collected from the interview.

### **1.3.2. Reading tests**

In answering the research questions, this study involved interviews and reading test analyses. As it has been explained before, this study involved three participants. The second source data of this study is from reading tests' analyses. What being analyzed were the reading tests made by the vocational teachers of English.

At the very first time the researcher met the participant, she planned to have three reading tests containing all levels, tenth, eleventh, and twelfth. However, along with the time, the first participant gave two reading tests each for the tenth and twelfth graders. Meanwhile, the second participant also collected two reading tests at the eleventh and tenth graders. The third participant only gave one reading test; that was for eleventh graders. Before conducting interview, the researcher intended to have three different reading tests. However, after conducting the interview, the researcher knew that the participants did not teach for all grades. Accidentally, most of the participants taught at two levels, except the third participant. That was the reason why the participants collected two reading tests for two different levels and one reading test by the third participant.

Reading tests were collected to see whether the tests which were made by the teachers of English at vocational school have fulfilled the criteria of a good

reading test or not. The analyses of reading test also collected to see the how the participants perceived the reading test for their students' achievement. It was also collected to analyze things which become the gist and the process of making a reading test.

As it has been aforementioned, the first teacher collected two reading tests. The first reading test by the first teacher comprised of 20 questions. Those 20 questions belonged to two type of questions, the multiple choice questions and an alternative question, especially matching technique. The first 10 questions asked about the information lied on the text. The answers of those first ten questions covered in one text while the rest ten questions covered in 5 different dialogs. Another reading test from the first teacher consisted of 20 questions too. The different was in the sections. This reading test consisted of 3 sections; multiple choice from texts, short answer questions from advertisement, and short answer questions from announcement.

The second teacher gave two reading tests which consisted of 10 questions. The first and the second reading test by the second teacher quite having the same characteristics, like the number of the questions and the text itself. The ten questions from each test were based on one text. The type of questions from both reading tests were in short answer questions.

The third teacher only proposed one reading test as she taught at one level. This reading test comprised of 16 questions from two different type of questions. Those two type of questions consisted of two sections with the multiple choice and short answer questions. The first ten questions asked about the information from one text while the rest 6 questions asked information from a dialog.

#### **1.4.Data analyses**

The data of this study were gained from reading test constructed by the vocational teachers of English and interview. Then, there are two parts of data

analyses in this study, those are analyzing teachers' interview and analyzing the criteria for the quality of reading test made by the participants.

#### **1.4.1. Interview**

As it has been aforementioned, this interview completed in the form of semi-structured. Semi-structured interview involved guiding questions. In order to answer the first research question, concerning on the process of constructing a reading test, there were seven guiding questions.

The first step in analyzing the data from teachers' interview is transcribing. Transcribing becomes important as its roles in answering the research questions of this study. Transcription began with converting audiotape recordings into text data. Transcribing the teachers' interview takes place during the informal stage. The informal stage refers to ongoing stage (Malik & Hamied, 2014). That is why the analysis process of interview in this study is flexible.

The second step is data reduction coding. This study involved three participants. To get thing organized well, the finding of this research was categorized into eight, those are test plan, relevance, balance, efficiency, validity, reliability, adequacy of the test items, and technical sound of the test.

#### **1.4.2. Criteria checklists for reading test**

The most important data which are meant to provide answers of the research question of this study are analyzed through the criteria checklist of a good reading test. The criteria of a good reading test are combination categories from several experts, those are Alderson et al. (1995, pp. 11-38), Alderson (2000, p. 9), Hill & Parry (1994), Hughes (2003), Nation (2008), and Evaluation and Examination Service, University of Iowa in 2006. There are twenty three criteria in the checklist used. Those criteria cover the purpose, format, content, and also the scoring system of the test. The criteria checklist used in this study are formulated as follow.

**Table 3. 1**  
**Criteria of a good reading test**

No	Category	Criteria	Yes	No
1	Test Plan	Reading test is clear whether it is high stakes or not		
2		The reading test requires test takers' age, sex and grade		
3		The reading test contains time allotment		
4		The scoring is clear		
5	Relevance	The text is chosen based on the purpose of course objective		
6		The test informs the test takers type of text		
7	Balance	If there is more than one part, the test is divided into a number of sections		
8		The order of the questions is in the same order of the information occurs in the text		
9	Efficiency	The sequence of the questions does not make it more difficult to answer them		
10		The text is authentic		
11	Validity	Answering the questions requires reading of the text		
12		The questions shouldn't not use the exact words used in the text		
13		The questions should measure reasonable comprehension		
14	Reliability	The questions are easy to understand		

15		The format of test is familiar to the test takers		
16		The question contains distractor option (Multiple choice question)		
17	Adequacy of the test items	There is question in which recalling meanings from the text		
18		There is question in which drawing inferences about the meaning of a word in a context		
19		There is question in which finding answer to question answered explicitly		
20		There is question in which finding answer to question in paraphrase		
21		There is question in which weaving together ideas in the context		
22		There is question in which drawing inference from the content		
23		The questions of the test contain level of difficulties		
24	Test is technically sound	Test was free of typing errors		
25		Instructions are clear and complete		
26		Exam copy was legible - attractive		
Percentage				

There are several steps in analyzing the checklist of the reading test made by the teachers. First, each reading test made by the participants was identified based on the criteria listed in the checklist. The criteria of the checklist are the summary



of criteria of good reading test purposed by several experts, such as Alderson, Clapham, and Wall (1995), Alderson (2000), Hill and Parry (1994), Hughes (2003), Nation (2008), and Evaluation and Examination Service, University of Iowa in 2006. There were eight categories included to examine the quality of the reading test; 1) test plan, 2) relevance, 3) balance, 4) efficiency, 5) validity, 6) reliability, 7) adequacy of the test items, and 8) technical.

Second, when the criteria found in the reading test, the raters put a tick in column “yes” and “not” if they didn’t. This step is called coding and tabulating process. In coding process, the results of reading test made by teachers were checked based on its quality. The quality of teacher made test was categorized into eight categories; 1) test plan, 2) relevance, 3) balance, 4) efficiency, 5) validity, 6) reliability, 7) adequacy of the test items, and 8) technical. Those eight categories reflected in the criteria checklists of the quality of teacher made test. Then, prior to simplifying the calculation process, the data of those eight categories were input to the table. This process is called the tabulating process.

Third, after identifying all the criteria in the checklist in the reading test, the writer drew the percentage. The percentage was drawn by counting all the “yes” and “no” answers multiplied by a hundred percent and divided by twenty six, referring to the total number of the criteria.

The last step in judging the reading tests was categorizing the percentage into five criteria; 1) Excellent, 2) Good, 3) Fairly good, 4) Poor, 5) Very poor. It is an attempt to figure out where within the scales of a good reading test this mean score places, this present study referred to a good reading test quality. This reading test criteria comprised five different levels or categories from very poor to excellent. This criteria of reading test was adapted from the common rating scales to use when writing questions in term of quality (Mahoney, 2009).

If it is found that the number of YES answers in the criteria checklist is more than 90%, the reading test is considered excellent. If it is between 80-89%, it is considered good. If it is between 70 - 79, the criteria is fairly good. If the percentage is between 65 - 69, it is categorized as poor a reading test. If the

percentage is 64 and below, it is recognized as very poor. The calculation procedure was applied to all levels and the result was summarized as follows.

**Reading test criteria (%)**

<b>1.</b>	<b>90</b>	<b>-</b>	<b>100</b>	<b>:</b>	<b>Excellent</b>
<b>2.</b>	<b>80</b>	<b>-</b>	<b>89</b>	<b>:</b>	<b>Good</b>
<b>3.</b>	<b>70</b>	<b>-</b>	<b>79</b>	<b>:</b>	<b>Fairly good</b>
<b>4.</b>	<b>65</b>	<b>-</b>	<b>69</b>	<b>:</b>	<b>Poor</b>
<b>5.</b>	<b>0</b>	<b>-</b>	<b>64</b>	<b>:</b>	<b>Very Poor</b>

Moreover, in achieving the validity of the data analyses, this study involved three raters including the researcher. Inter-rater reliability (IRR) in giving judgment on the criteria of the reading test quality was also calculated by adopted the inter-rate reliability measurement from (Gisev, Bell, & Chen, 2013).

The raters were chosen based on the researcher's assumption which stems from the fact that there are some factors affecting the Inter-Rater Reliability (IRR); rater training, rater selection, and accountability for accurate rating (Shweta, Bajpai, & Chaturvedi, 2015).

Rater training was taken place before the raters rated the reading tests. It was an attempt to put misunderstanding of the raters toward the criteria into minimum. The raters were given a sort of preliminary training in which the raters were familiarized with the concept of constructing a reading test along with the criteria listed in the checklist to be used.

Regarding rater selection, three criteria were set. The first consideration was they are English foreign language lecturers. It is assumed that they are familiar with the concept of teaching language skill including reading skill. The second consideration was that the two raters teach Language Assessment subject.

Therefore, it is assumed that that they are familiar with the concept of testing, including testing reading skill.

The three raters rated the reading test independently and separately. From the calculation, the data resulted 90% degree of agreement. This percentage indicates that the ratings given by all the raters are reliable. They share similar score to the same reading test being analyzed. Therefore, the final score coming from the calculation of the quality of the tests, which involved the computation of 'yes' rating, is trustworthy. In this study, the average score from the three raters was assigned as the final score.