

CHAPTER III

RESEARCH METHOD

The current chapter provides a comprehensive overview of the methodology utilized to investigate the validity evidence in TPELT. Therefore, in providing the overview, this chapter describes the study's research design, research context and participants, data collection techniques, instrumentations, data collection procedures, and data analysis.

Additionally, it is worth noting that this research is a part of a bigger research conducted by a research team from Universitas Pendidikan Indonesia. The bigger research aims to develop a language proficiency test instrument for primary English language teachers named Test for Primary English Language Teachers (TPELT). As part of it, this study specifically aims to provide a comprehensive analysis of the evidence supporting the a priori and a posteriori validity of the TPELT reading test for primary English teachers in Indonesia.

3.1 Research Design

This study used a sequential exploratory mixed-methods design. Sequential exploratory mixed methods are a research approach that integrates both qualitative and quantitative data collection and analysis into a sequential series of phases, as discussed by Creswell and Clark (2018). The first stage of the process involves gathering qualitative data, which is then carefully analyzed. The outcome of this analysis provides valuable direction for the next stage of the research, which involves collecting and analyzing quantitative data. As highlighted by SAGE Research Methods Datasets (2019), the quantitative data may be obtained through various methods such as surveys or questionnaires. Qualitative data analysis serves a critical function in the process of formulating precise research questions for the quantitative phase, thus providing the necessary groundwork for the collection of quantitative data. In essence, this methodological design relies on the systematic integration of both qualitative and quantitative approaches in order to generate a more comprehensive understanding of the research problem.

3.2 Research Context and Participants

This study is a vital component of a broader research initiative that aims to develop a specialized proficiency test tailored for primary English teachers. The journey began in the early 2022s and was conducted for 8 (eight) months consisting of 3 (three) prominent phases according to each goal as shown in Figure 3.1. This led to the creation of a proficiency test known as the Test for Primary English Language Teachers (TPELT).

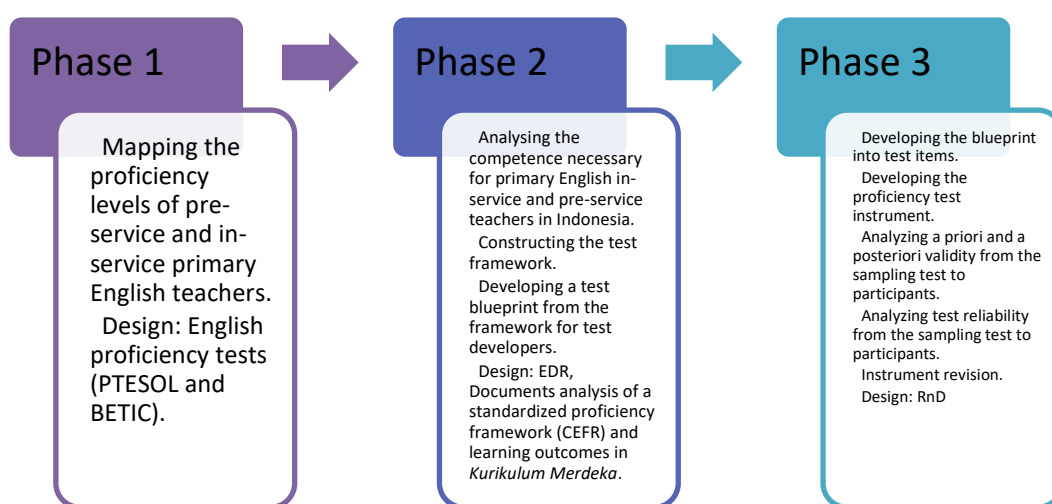


Figure 3.1 *Research Phases*

Figure 3.1 shows the overall research phases in developing a language proficiency test instrument for primary English teachers. This present study was in the Phase 3 of the bigger research project focusing on analyzing a priori and a posteriori validity from the sampling test to participants and analyzing test reliability from the sampling test to participants. This research also demonstrates a validity analysis, representing the fourth crucial phase in the comprehensive development of a test. The development of a test involves ten essential stages (Hughes, 2003). The ten stages are (1) Stating the problems, (2) Writing specifications for the test, (3) Writing and moderating items, (4) Informal trialing of items on native speakers, (5) Trialing of the test on a group of non-native speakers similar to those for whom the test is intended, (6) Analysis of result of the trial, (7) Calibration of scales, (8) Validation, (9) Writing handbooks, and (10) Training staff.

This present study is currently positioned at the eighth stage in Hughes's (2003) ten stages of test development (see **Figure 3.2**). This research critically examines the validity of TPELT, taking into account its presentation, testing processes, and evaluation techniques.

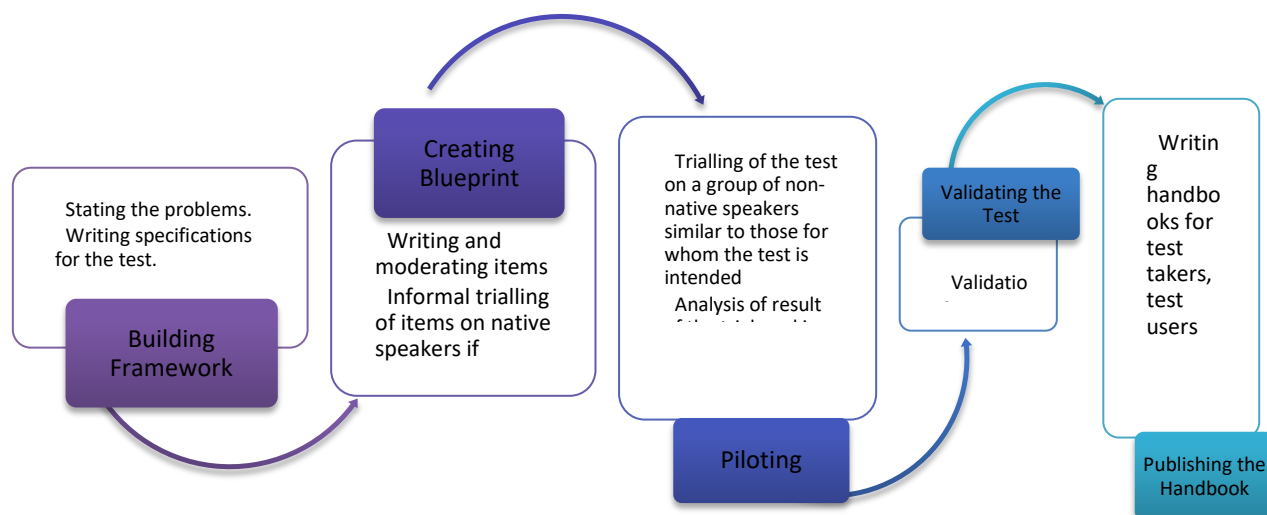


Figure 3.2 *Ten Stages of Test Development by Hughes*

It can be seen that validation takes place in a very important stage before the test is officially published. In order to confirm how valid the test is, it became my role to start conducting this research after completing Phase 1 and 2 (see **Figure 3.1**) along with the blueprint and item development with all members of the bigger research. Here are my stages and the actions conducted in this research:

Table 3.1 *Table of Researcher's Stages on the Bigger Research*

Stages	Action
Analyzing a priori and a posteriori validity from the sampling test to participants.	This validation process examines the TPELT blueprint regarding its presentation and evaluation. It includes assessing how the TPELT reading test is administered, how participants are tested, and how the test scores are evaluated.

	<p>Investigate the evidence using checklist forms.</p> <p>In this part, I gathered the participants' reading test scores after participating in two trial tests of TPELT as well as PTESOL.</p> <p>Analyze the scoring validity using the formula of correlation, t-test,</p>
<p>Analyzing test reliability from the sampling test to participants.</p>	<p>The scores were also analyzed to see the item reliability.</p>

However, in this study, the focus will be on the reading component. This research was conducted by performing both a priori and a posteriori validity analyses on the reading test of TPELT. A priori validity assessed whether the reading test of TPELT meets the criteria of theory-based and context validity before being administered to participants, while a posteriori validity evaluated whether the reading test of TPELT is effective in measuring the English language proficiency of teachers in Indonesia.

The participants in this study are 30 English language teachers from various regions in Indonesia. Drawing insights from Bonnet and Wright's study (2014), samples must be as small as thirty (30) to establish reliability so long as the scale items have strong inter-correlation. The participants were invited through social media platforms such as Instagram, Facebook, and WhatsApp. They were selected as the primary target users of the TPELT test. Participants were asked to take the reading test of TPELT and PTESOL as a comparison, and then provide feedback on the difficulty, readability, and appropriateness of the test items with respect to the English language context used in Indonesia.

In order to be eligible to participate in this research, prospective participants must satisfy several criteria. Firstly, they must hold a degree in English Education from any accredited (see **Appendix 1**) university. Secondly, participants must currently be teaching English at the junior high school level in Indonesia. Interested individuals must complete a registration form, providing personal

information such as their full name, date of birth, educational background, teaching experience, and the name of the school where they are employed.

3.3 Data Collection Techniques

This subsection discusses the instruments that will be used to collect the necessary data for the present study, as well as the procedures of the data collection. For a brief overview, this research employed three instruments i.e. checklist forms, questionnaires, and interviews to collect the necessary data for a qualitative method. The data retrieved from those instruments were used to investigate to what extent the evidence could be generated from the result of a priori validity analysis of the reading test of TPELT. Meanwhile, for the quantitative method, the document consisting participants' TPELT reading test scores was also collected. The data was used to fulfill the analysis process to the evidence that could be generated from the result of a posteriori validity analysis of the reading test of TPELT.

3.3.1 Instrumentation

As for the present study, to obtain the necessary data for the present study, several instruments were utilized to collect as follows:

a. Checklist Forms

In this study, certain forms by Weir (2005) were employed to examine the evidence for theory-based and contextual validity presented in the blueprint of TPELT. Specifically, these forms were used to scrutinize the contextual validity evidence embedded in TPELT and its presentation format.

There are three forms used to identify the evidence of context validity; Task Settings Form, Task Demands Form, and Settings and Test Administration Form. They consist of specific questions based on the topics.

Table 3.2 *Questions of Task Settings Form on Context Validity*

Task Setting	Questions
Rubric	<ul style="list-style-type: none"> • Does the rubric in the reading test provide clear instructions for candidates? • Is the rubric composed of concise and straightforward sentences in the reading test? • Does the rubric demonstrate grammatical accuracy? • Is the rubric free from spelling errors? • Is the rubric presented in the candidates' First Language (L1) or the Target Language (TL)? • Is the rubric recognizable and familiar to the candidates? • Does the rubric clearly state the allocated time for the reading section of the test? • If the task demands different types of responses, are separate specific instructions provided for each type of response required?
Test purpose	<ul style="list-style-type: none"> • Is the purpose of the reading test clearly and unambiguously communicated to the candidates? • Does the purpose of the reading test align appropriately with the intended goals for assessing reading proficiency?
Response format	<ul style="list-style-type: none"> • Is there any indication or evidence suggesting that the test response format may influence test takers' performances?
Known criteria	<ul style="list-style-type: none"> • Are the criteria for marking the reading test clearly stated and made explicit to both the candidates and the markers?
Weighting	<ul style="list-style-type: none"> • Is the weighting of different items adequately justified and appropriate for the reading test?
Order of item	<ul style="list-style-type: none"> • Are the items and tasks in the reading test arranged in a logical and justifiable order?
Time constraints	<ul style="list-style-type: none"> • Is the timing allocated for the reading test, including preparation and completion, appropriate and reasonable for the test takers?

The Task Settings Form in TPELT underwent a comprehensive analysis, covering essential elements such as the rubric, purpose, response format, known criteria, weighting, order of items, and time constraints. Each aspect was carefully scrutinized to ensure the test's accuracy and validity, providing clear instructions and guidelines to the candidates. The analysis of the rubric aimed to make it candidate-friendly, intelligible, and explicit, avoiding any ambiguity that could lead to misinterpretation of the tasks. Moreover, the purpose and weighting of the test were established to align with the objectives of assessing the English proficiency of primary English teachers effectively.

On the other hand, the Task Demands Form in TPELT was strategically designed to assess specific aspects of candidates' language proficiency related to teaching young learners. The form included a set of thoughtfully crafted questions that covered various language skills relevant to the context of primary English teaching. By focusing on specific language aspects and teaching-related topics, the Task Demands Form aimed to create a more realistic and relevant assessment for primary English teachers. The questions were carefully chosen to ensure that candidates' familiarity with terms, topics, and situations in the texts aligns with their roles as primary English teachers, making the test content appropriate for their target situation requirements. This meticulous attention to detail in the form's construction contributes to the overall coherence and rigor of TPELT as an English proficiency test specifically tailored for primary English teachers.

Table 3.3 *Questions of Task Demands Form on Context Validity*

Task Demand	Questions
Discourse mode: Reading	• Is the discourse mode used in the reading test appropriate for the intended reading skills being assessed?
Channel of communication	• Is the channel of communication (e.g., written text) suitable for the target situation and the candidates being

	tested?
Text length	• Is the length of the reading text appropriate for the target situation and the candidates being tested?
Nature of information in the text	• Is the type of information presented in the reading test relevant to the target situation and the candidates' language abilities?
Content knowledge required	• Is the topic content of the reading test relevant and appropriate for the target situation and the candidates being tested?

The table provided below outlines the list of questions utilized in the Setting and Test Administration form. These questions are specifically designed to address critical aspects such as the physical testing environment, ensuring uniformity in test administration, and implementing security measures. By considering these factors, the TPELT test can be conducted in a standardized and controlled manner, maintaining the integrity of the assessment process for primary English teachers.

Table 3.4 *Questions of Settings and Test Administration Form on Context Validity*

Settings and Test Administration	Questions
Physical Condition	• Were the physical conditions of the reading test administration suitable and conducive to a fair testing environment?
Uniformity Administration	• Was the reading test consistently administered in the same manner across all test sites to ensure uniformity in the testing process?
Security	• Was the reading test conducted securely to prevent any unauthorized access to test materials and ensure confidentiality?

Good documentation is crucial in ensuring the validity of language tests. They said that complete and accurate documentation would help reduce the risk of errors and allow for reevaluation if necessary (Bachman and Palmer, 2010). Meanwhile, Fulcher (2015) highlighted the importance of documentation in the validity process and emphasized that complete and organized documentation would help ensure the validity of language tests. He also highlighted the importance of documentation in demonstrating transparency and accountability in the validity process. Therefore, pictures, videos, and voice recordings were also used to gather information from supporting documents of the test and how the trial tests were conducted.

Table 3.5 *Types of the Documentation*

Types	Purposes of the Documentation
Pictures, videos, and voice recordings	<ul style="list-style-type: none"> • To capture the process of testing. • To record the process of the interview.
Chat History	<ul style="list-style-type: none"> • To identify some problems experienced by the participants from the group discussion.

Certain issues might arise during the testing process; hence, this information could corroborate the outcome of the analysis.

b. Questionnaires

In this study, questionnaires were administered to investigate the background of participants prior to conducting the trial tests, as well as their perceptions of the tests after completing them. The first questionnaire aimed to verify whether prospective participants met the eligibility criteria for this research. The subsequent questionnaires were designed to obtain information from the participants' perspectives during the testing process,

thereby ensuring a balanced dataset that incorporates both the test results and participants' experiences.

Table 3.6 *Types of the Questionnaires*

Types	Purposes of the Questionnaires	Samples of Questions
Initial questionnaire	<ul style="list-style-type: none"> To learn the participants' backgrounds 	<ul style="list-style-type: none"> How long have you been teaching in Junior High School?
Test re-test questionnaire	<ul style="list-style-type: none"> To have information about the participants' experiences, feelings, and perceptions on the TPELT after doing the tests. 	<ul style="list-style-type: none"> Do the questions in the first and second tests have the same or different levels of difficulty for each item? Explain.
Parallel test questionnaire	<ul style="list-style-type: none"> To obtain information about the participants' experiences, feelings, and perceptions on PTESOL and how they compare both tests. 	<ul style="list-style-type: none"> Please provide your opinion about PTESOL and TPELT, the two types of tests you have participated in during this test trial.

These questionnaires were crucial in providing a more current and subjective analysis. As the test was a tool to measure human skills, it was imperative to consider the test takers' opinions. Their responses to the questions would ensure a well-rounded outcome. Please refer to **Appendix 2 and 3** for the complete set of questionnaires.

c. Interviews

Interviews were conducted to obtain some supporting information from experts, TPELT developers, and test participants. These have supported the analysis and provide more detailed explanation.

Table 3.7 *Types of the Interview*

Types of the Interview	Purposes of the Interview	Sample of Questions
Interviewing Test Developers	<ul style="list-style-type: none"> • To obtain some additional information about the developed TEST. • To establish the observations. 	<ul style="list-style-type: none"> • Furthermore, I would like to know how you determined the test framework for TPELT?
Interviewing a Stockholder	<ul style="list-style-type: none"> • To have an in-depth understanding of the testing development and process. • To see her opinion on the TPELT 	<ul style="list-style-type: none"> • I am interested to hear your perspective on how the reading test in TPELT compares to other proficiency tests available in the market, including your own BETIC test.

The interviews were conducted formally and informally. The data were obtained from several discussion sessions with the test developers. As the researcher was also a part of the test development team, interviews were also conducted during casual discussions that occurred during the development process. Additionally, a CEO of a language testing company was asked to test TPELT and provide her perception of the test by answering a set of questions. Her responses were recorded and included in this research. Please refer to **Appendix 4 and 5** for the script.

3.3.2 Procedures of Data Collection

A priori validity analysis applied a descriptive qualitative method which means the data gathered in this research were collected, arranged, classified, analyzed, and interpreted (Bogdan and Biklen, 1982). The data resource about TPELT which includes Test framework, TPELT blueprint, and experts' theories for this study were collected to get the information essential and prestige for educational needs. Furthermore, a set of checklist was used to observe the validity evidence of TPELT. Some interviews with the TPELT developers and the CEO of a testing company was also conducted to get detail information that was unavailable in the published

document. The initial questionnaires was also given to participants to identify their background and understanding about the test (see **Figure 3.2**).



Figure 3.2 *Timeline of Data Collection from a Priori Validity Analysis (Theory-based and Context Validity)*

Meanwhile, in a posteriori validity, the data were collected from the scores of test participants who agreed to be samples of this research (see **Figure 3.3**). Initially, the participants were given a consent form to state that they agreed if their data would be used in this research. This ethical consideration is an important step in a research publication to avoid conflicts in the future. After that, participants were invited to take the TPELT twice in two separate days in order to get data for investigating the correlation level in the validation. Another set of questionnaires was delivered after accomplishing the second test of TPELT (see **Appendix 2**).

After collecting the data from the TPELT scores and questionnaires, the participants were asked to join another test named PTESOL. The next data obtained from the participants' PTESOL scores were taken as a parallel form process in validating the test. Those scores were used for comparison. For additional final information, another set of questionnaires was also delivered to the test participants in order to get their perspective and information regarding the tests (see **Appendix 3**).

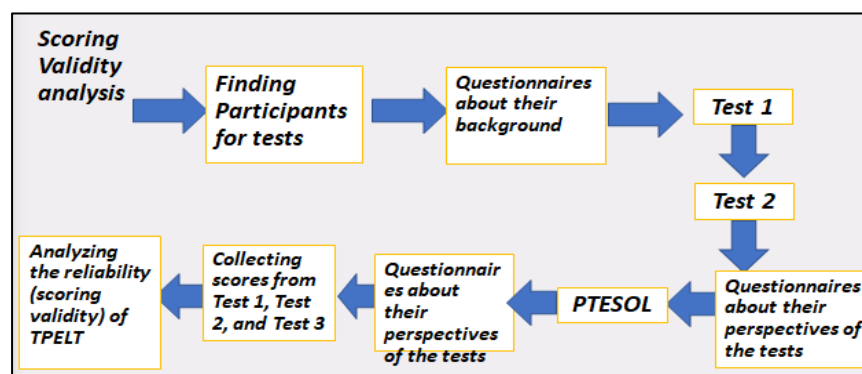


Figure 3.3 *Timeline of Data Collection from a Posteriori Validity Analysis (Scoring Validity)*

3.4 Data Analysis

In analyzing the priori validity of the test, the information obtained from the collected documents was combined and analyzed by using Wier's (2005) a priori validity and a posteriori validity framework. There were three elements of validity that became the main focus in this research i.e., theory-based validity, context validity, and scoring validity.

Initially, the blueprint of TPELT was analyzed with theory-based validity, context validity frameworks. The reading skill tested in TPELT was evaluated to confirm how the validity evidence was found and how they were presented in the test. Context validity analysis focused on how the task setting, task demands, setting and test administration of TPELT were confirmed in the test. Both priori and context validities were analyzed using checklists which refer to the Wier's (2005) theory.

Furthermore, the scores data of the participants were analyzed by SPSS for correlation, factor analysis and t-test. There were two scores of the TPELT reading test from every participant conducted at two different times, and one score of PTESOL Reading comprehension section from every participant. The result would show how reliable the reading test of TPELT is. Analyzing score validity is an important aspect of assessing the quality of a measurement instrument. The three common methods for analyzing score validity using the formulas of product moment, t-test, and Cronbach's Alpha are:

1. Product Moment Correlation

The product moment correlation coefficient (r) is used to examine the relationship between two variables. To assess score validity, we can calculate the correlation coefficient between the scores obtained from the measurement instrument and an established criterion measure that measures the same construct. The formula for the product moment correlation coefficient is:

$$r_{xy} = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{(n(\sum x_i^2) - (\sum x_i)^2)(n(\sum y_i^2) - (\sum y_i)^2)}}$$

r_{xy} = correlation coefficient *Product Moment*

n = number of participants

x_i = Item score on the first test

y_i = Item score on the second or third test

If the correlation coefficient is high (e.g., $r > 0.7$), it suggests that the measurement instrument has good score validity.

2. T-Test

The t-test is used to compare the mean scores obtained from the measurement instrument with the mean scores obtained from the criterion measure. If the mean scores from both measures are not significantly different, it suggests that the measurement instrument has good score validity. The formula for the t-test is:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

t = t score

r = correlation coefficient

n = number of participants

If the t-value is not significant (e.g., $p > 0.05$), it suggests that the measurement instrument has good score validity.

3. Cronbach's Alpha

Cronbach's alpha is a statistical measure used to assess the internal consistency or reliability of a test. The primary purpose of Cronbach's alpha is to determine the extent to which all the items within a scale or

instrument measure the same underlying construct. It is calculated using the following formula:

$$r_{11} = \left(\frac{n}{n-1} \right) \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right)$$

r_{11} = Reliability

n = number of items

$\sum \sigma_i^2$ = number of variant scores in every item

σ_t^2 = variants of total scores

If the value is high (e.g., > 0.7), it suggests that the items in the measurement instrument has good score validity.

Overall, using a combination of these methods can provide a more comprehensive assessment of score validity for a given measurement instrument.