

CHAPTER III

METHODOLOGY

This chapter outlines the research methodology employed in this study, including the research design, variables, population and sample, research instruments, research procedures, data collection, and data analysis techniques. The methodology is structured to ensure the validity and reliability of the research findings, which aim to determine the effectiveness of AI-Based Digital Storytelling in enhancing EFL learners' speaking proficiency. By utilizing a quasi-experimental design, the study seeks to compare the impact of AI-Based Digital Storytelling with traditional storytelling methods. The chapter also provides a comprehensive explanation of the assessment tools and statistical methods used to analyze the collected data.

A. Research Design

This study employs a quantitative methodology with a quasi-experimental design. Quantitative research is a means of testing objective theories by examining the relationship among variables that can be measured and analyzed using statistical procedures (Creswell, 2014). This approach allows the researcher to determine whether a treatment influences an outcome through statistical comparison.

Experimental research aims to examine the effect of a treatment on an outcome by comparing groups (Creswell, 2014). However, in educational settings, random assignment of participants is often not feasible because intact classes are used. Therefore, this study applies a quasi-experimental design, which involves nonrandomized groups while still allowing comparison between an experimental group and a control group. Quasi-experimental designs are commonly used in educational research when full experimental control cannot be achieved (Creswell, 2014). In this study, two groups of participants are involved: an experimental group and a control group. The experimental group engages in AI-Based Digital Storytelling activities that utilize videos to narrate their stories, while the control group employs static pictures as visual aids in storytelling. The structured implementation of this design ensures that any differences in outcomes between the two groups can be attributed to the instructional method rather than extraneous

variables. By employing controlled conditions and systematic observation, the research design ensures the validity and reliability of the findings, which are essential for drawing meaningful conclusions about the effectiveness of AI-Based Digital Storytelling in EFL learning.

Table 3.1 Group of Participants

Experimental Group	Using AI-Based Digital Storytelling as media
Control Group	Using Picture as media

B. Variables

The study examines two main variables that influence the research outcomes. The independent variable in this study is the instructional method, where the experimental group receives treatment through AI-Based Digital Storytelling, whereas the control group undergoes traditional storytelling using static images. The dependent variable is the speaking proficiency of the EFL learners, which is assessed through various criteria, including fluency, pronunciation, vocabulary usage, and grammatical accuracy.

This variable framework allows for an in-depth analysis of how AI-Based Digital Storytelling impacts different aspects of speaking performance. Mayer (2005) emphasizes that multimedia learning, which integrates both visual and auditory elements, enhances comprehension and retention in learners. By incorporating AI-based storytelling, the study investigates whether an interactive and visually engaging approach yields a more significant improvement in speaking skills compared to conventional storytelling methods.

C. Population and Sample

To ensure the validity and generalizability of the findings, the selection of participants in this study is conducted with careful consideration of their relevance to the research objectives. The chosen population should reflect the characteristics of learners who are in the process of developing their speaking skills within an educational setting. Therefore, the study specifically focuses on learners at a level where the enhancement of English oral communication skills is essential.

This study was conducted at SMK Al-Mahrusiyah Kediri, a vocational high school that integrates Islamic boarding school (pesantren) life with formal

education. The target population consists of 10th grade students enrolled in the Technology and Computer Networking (TJKT) department. A total of two classes were selected to participate in this study, namely X TJKT A and X TJKT B, with each class consisting of nearly 40 students. Thus, the overall number of participants reached approximately 80 students.

To ensure fairness and comparability, students are selected based on similar levels of English proficiency, which is assessed through a preliminary evaluation. This approach minimizes bias and ensures that differences in learning outcomes stem primarily from the instructional method rather than prior disparities in language ability. Arroba and Acosta (2021) highlight that quasi-experimental research in language education benefits from a well-matched sample, allowing for more precise measurement of instructional effects.

D. Instrument of Research

To support the objectives of the study and to measure the impact of the treatment accurately, this research employed speaking tests as the primary instruments. The instruments consisted of a pre-test and a post-test, both designed to evaluate students' speaking proficiency through topics that are familiar, meaningful, and aligned with the curriculum. The tests were assessed using a standardized speaking rubric adapted from Brown (2004), which covers four key aspects: fluency, pronunciation, vocabulary usage, and grammatical accuracy.

1. Pretest

The pre-test was administered before the treatment phase to determine the students' initial speaking proficiency. In this test, students were asked to deliver a short oral presentation of 2–3 minutes after five minutes of preparation without using a written script. The topics were chosen from daily life experiences that students could easily relate to, such as "*A memorable day in your life,*" "*A place you would love to visit and why,*" and "*Your favorite childhood memory.*" These prompts were selected to encourage natural speech and reduce speaking anxiety, while also providing a baseline measurement of students' fluency, pronunciation, vocabulary, and grammar.

2. post test

The post-test was conducted after the treatment phase to evaluate the improvement of students' speaking proficiency. The format was the same as the pre-test: students prepared for five minutes and then delivered a 2–3 minute oral presentation without relying on written notes. However, the topics were slightly varied to maintain students' interest while remaining within their scope of knowledge and experience. The topics included “*A holiday trip you will never forget,*” “*An interesting experience with your friends or family,*” and “*Your first experience trying something new.*” The results of the post-test were compared with the pre-test scores to measure students' progress and to determine the effectiveness of the treatment.

Both pre-test and post-test results were scored using the speaking rubric to ensure fairness, objectivity, and consistency in assessment. This systematic approach allowed the researcher to capture not only the students' starting point but also the extent of their improvement after being exposed to the different teaching media.

The main instrument used in this study is a speaking assessment rubric that measures students' speaking performance based on four aspects, which include fluency, pronunciation, vocabulary usage, and grammatical accuracy. The rubric is adapted from the speaking performance scale proposed by Brown (2004), which provides clear and practical descriptors to evaluate oral language skills. Students are given speaking prompts and are required to respond within 2-3 minutes, allowing for a fair and consistent assessment of their speaking abilities. A standardized rubric ensures objectivity in scoring, reducing potential biases. The reliability of this rubric is enhanced by its alignment with established language assessment frameworks (Fu et al., 2021). This rubric is used to maintain consistency and fairness in scoring. (*See Appendix 3*)

E. Treatment Procedures

This study follows a series of structured procedures to ensure that the data collection and treatment process run smoothly and systematically. The procedures are carried out in several stages, starting from the initial assessment to the final evaluation.

At the beginning of the study, both groups—the experimental group and the control group—are given a pre-test to assess their initial speaking proficiency. After the pre-test, the treatment is implemented for each group over a period of six meetings. The experimental group receives instruction through AI-Based Digital Storytelling, while the control group learns using traditional picture-based storytelling. The learning material used throughout the treatment is Recount Text, and all teaching sessions are conducted directly by the researcher. After the treatment phase, both groups are given a post-test to measure their improvement in speaking performance.

Table 3.2. Treatment Procedures

	Experimental Group (AI-Based Digital Storytelling)	Control Group (Picture-Based Storytelling)
Pre-Teaching	Teacher introduces recount text (definition, purpose, generic structure, language features) and gives examples of simple recount texts.	Teacher introduces recount text (definition, purpose, generic structure, language features) and gives examples of simple recount texts.
Whilst Teaching	<ol style="list-style-type: none"> 1. Teacher shows an AI-based storytelling video (e.g., AI-generated animation or digital story using text-to-speech). 2. Teacher demonstrates how to retell the story using the AI-based video as a guide, focusing on pronunciation, fluency, and expression. 3. Students observe and take notes on important vocabulary and sentence structures. 	<ol style="list-style-type: none"> 1. Teacher shows a series of static pictures related to a recount text. 2. Teacher demonstrates how to retell the story based on the pictures. 3. Students observe and take notes.
Post-Teaching	Story Presentation & AI-based Pronunciation Feedback (Tools: Projector, feedback forms)	Story Presentation & Feedback (Tools: Projector, feedback forms)

F. Data Collection

Data collection focuses on pre-test and post-test assessments, which measure students' speaking proficiency before and after the treatment. The pre-test establishes a baseline performance level, while the post-test evaluates progress following the intervention. The speaking assessments are recorded and evaluated based on the scoring rubric.

This approach ensures that the data collected is both quantitative and objective, providing a reliable means of measuring the impact of AI-Based Digital Storytelling on speaking proficiency. By implementing inter-rater reliability, involving two independent scorers to evaluate the speaking performances, the study minimizes the risk of subjective evaluation errors and ensures consistency in data collection. The use of multiple scorers enhances the credibility and accuracy of the assessment outcomes.

G. Validity and Reliability

To ensure the credibility of the findings, this study addresses both the validity and reliability of the assessment instruments.

1. Validity

This study establishes content validity by ensuring that the speaking test and the scoring rubric accurately reflect the competencies intended to be measured—namely, students' speaking proficiency. The speaking test prompts are designed to elicit spontaneous and structured oral responses from students, with topics that are familiar, relevant, and appropriate for senior high school learners. This ensures that students are assessed on their actual speaking abilities rather than on their knowledge of unfamiliar content.

Moreover, the scoring rubric, adapted from Brown (2004), assesses four core aspects of speaking: fluency, pronunciation, vocabulary usage, and grammatical accuracy. These components align with the theoretical construct of speaking proficiency and with the speaking skill indicators outlined in the Indonesian senior high school English curriculum. As such, both the test and the rubric are valid in terms of content, providing a sound basis for evaluating learners' oral language performance.

2. Reliability

To enhance reliability, particularly in the scoring process, this study applies inter-rater reliability by involving two independent scorers to evaluate each student's speaking performance. The scorers are trained to use the same rubric and are instructed to score independently. Any scoring discrepancies are reviewed and resolved through discussion to reach a consensus. This method minimizes the risk of subjective judgment and enhances scoring consistency, thereby ensuring reliable and trustworthy assessment results.

H. Data Analysis

The data collected from the pre-test and post-test assessments were analyzed using both descriptive and inferential statistics. Descriptive statistics, including the calculation of mean, median, and standard deviation, were employed to summarize the overall trends of students' speaking performance in both groups. These descriptive measures provided a general overview of students' improvement before and after the treatment.

For the inferential analysis, this study applied Analysis of Covariance (ANCOVA). The use of ANCOVA allowed the researcher to compare the post-test scores of the experimental group, which was taught using AI-Based Digital Storytelling, and the control group, which was taught through picture-based storytelling, while controlling for students' pre-test scores as a covariate. By statistically adjusting for the initial differences in speaking proficiency, ANCOVA provided a more accurate estimation of the treatment effect than a simple t-test. This approach ensured that any observed differences in post-test scores were primarily the result of the instructional method rather than pre-existing disparities in students' abilities. Before hypothesis testing, normality and homogeneity tests were conducted. If the data are normally distributed ($p > 0.05$), the researcher would use ANCOVA. If the data are not normally distributed ($p < 0.05$), the researcher would use the Mann-Whitney U Test.

The analysis was conducted using IBM SPSS Statistics. The application of ANCOVA in educational research is considered appropriate for pre-test and post-test designs, as it minimizes bias and enhances the validity of findings (Dimitrov & Rumrill, 2003). While the Mann-Whitney U test is more appropriate for non-normally distributed data.