CHAPTER III
METHODOLOGY

This part discusses the methodology of the research. It consists of research design, research setting, population, sample, technique of sampling, research instrument, research procedure, techniques of data collection, techniques of data analysis and data validation.

3.1 Research Design

This study adopts an explanatory sequential mixed-methods design to examine the impact of AI-integrated Mobile-Assisted Language Learning (MALL) on students' speaking mastery and to explore teachers' and students' perceptions regarding its influence on classroom dynamics. The quantitative phase applies two groups pre-test post-test design which are control group and experimental group. An experimental approach is used because the study involves the systematic implementation of a treatment and the measurement of its impact on students' speaking performance. The purpose of the experimental approach is to measure the effectiveness of the treatment in improving students' speaking performance.

Furthermore, the three classes receive speaking instruction enhanced through AI-based MALL tools such as conversational chatbots, automated pronunciation evaluators, and AI-generated speaking tasks. Prior to the intervention, students complete a pre-test measuring their speaking proficiency using an analytic rubric adapted from Luoma (2004) and validated in previous MALL studies (Lin & Lan, 2015). After several weeks of AI-supported speaking practice, a parallel post-test is administered to determine changes in learners' speaking performance. Quantitative data are analyzed using paired-samples t-tests to measure the significance of improvement and effect size. This phase aligns with recent empirical studies demonstrating the effectiveness of AI and mobile learning for oral language development (Ma & Luo, 2022; Li, 2023).

The results of this phase provide a numerical basis to answer the first research question concerning the impact of AI-enhanced MALL on speaking mastery. In the

subsequent qualitative phase, the study seeks to understand how teachers and students perceive the influence of AI-based MALL on classroom dynamics, particularly regarding interaction patterns, participation, engagement, and communicative confidence. By applying narrative approach in gathering the information. Data are collected through semi-structured interviews. The analysis employs thematic coding, drawing on frameworks such as Classroom Interactional Competence (Walsh, 2013) and technology-mediated interaction theories (e.g., Lai, 2017). By comparing emerging qualitative themes with quantitative results, the study explains why learners achieved certain levels of improvement and how AI integration influences classroom processes such as turn-taking, autonomy, and teacher–student rapport. Integration occurs at the interpretation stage through narrative explanation and thematic–statistical alignment, consistent with mixed-methods procedures used in recent technology-supported EFL research (e.g., Viberg et al., 2020). This robust design enables a deep understanding of both measurable learning outcomes and experiential classroom dynamics when AI-integrated MALL is implemented.

3.2 Research Setting

This study will be conducted at SMA NU 1 Gresik, a senior high school located in Gresik, East Java, Indonesia. The participants of this study consist of 96 students enrolled in three speaking-focused classes at SMA NU 1 Gresik. These students were selected using a purposive sampling technique, which is considered appropriate because the research specifically requires participants who have direct and continuous exposure to AI-integrated MALL activities. The criteria for selecting the 96 students are as follows: (1) they must be officially enrolled in the school's specialized English program, which integrates structured speaking activities and mobile-assisted learning tasks; (2) they must have regular access to mobile devices and be active users of the school's AI-integrated MALL application. Because all three classes fulfilled these criteria, the entire population of 96 students is included as the purposively selected sample. This approach aligns with purposive sampling logic, which prioritizes quality and relevance of information over

representativeness, enabling the researcher to collect robust data from students who directly engage with AI-mediated speaking instruction. This setting will offer a rich environment for exploring how AI-MALL will influence classroom processes and learners' speaking development.

3.3 Population, Sample and Technique of Sampling

The population in this study is determined based on the characteristics required to address the research objectives concerning the integration of AI-based Mobile-Assisted Language Learning (MALL) in speaking instruction. Since the study focuses on the impact of AI-MALL on students' speaking mastery and classroom dynamics, the population is defined as all students at SMA NU 1 Gresik who are enrolled in the speaking-focused English program (Explore Class Program) from eleventh grade. This specific group is selected as the population because they represent the learners who are systematically exposed to AI-mediated speaking tasks, making them directly relevant to the variables investigated in the study. By limiting the population to students who meet these criteria, the research ensures that all members of the population share the instructional context necessary for valid measurement of AI-MALL's impact on speaking performance.

The sample of this study consists of 96 students distributed across three speaking-focused English classes in 11 grade at SMA NU 1 Gresik . The sample was determined using a purposive sampling technique, in which participants were selected based on specific criteria relevant to the research objectives. The criteria for inclusion were: (1) students must be part of the English speaking-focused program in which AI-integrated MALL is systematically used during instruction; (2) they must have regular access to mobile devices to engage with the AI-based applications; (3) they must participate consistently in classroom activities throughout the intervention period; and (4) they must be available to complete both the pre-test and post-test of speaking mastery. Since all 96 students in the three classes met these criteria, the study employed a total sample approach, in which the entire eligible group was included as the research sample. This ensures that the

findings capture the full range of learner responses within the AI-integrated instructional context and enhances the validity of the pre–post intervention results.

3.4 Research Instrument

Several research instruments are employed, each aligned with the objectives of examining the impact of AI-integrated Mobile-Assisted Language Learning (MALL) on students' speaking mastery and exploring teachers' and students' perceptions of classroom dynamics. The primary quantitative instrument is a speaking performance test, administered as both a pre-test and a post-test. The test consists of response-based speaking tasks like giving response to a given issue by the teacher. This type of response-based speaking task was chosen because it closely reflects authentic classroom communication, where students are required to respond spontaneously to issues or questions raised by the teacher. By asking students to give spoken responses to a given issue, the task encourages real-time language production, which is essential for measuring actual speaking ability rather than rehearsed or memorized speech. All designed to measure key components of speaking ability, including fluency, accuracy, pronunciation and vocabulary range. Students' performances are evaluated using an analytic speaking rubric adapted from Luoma (2004) and supported by criteria used in MALL-based oral assessment studies (e.g., Li, 2023; Lin & Lan, 2015). The rubric ensures detailed, reliable scoring on each sub-skill, while inter-rater reliability is maintained by involving two trained raters and calculating agreement through Cohen's Kappa or intraclass correlation.

The researcher chose Luoma's (2004) analytic speaking rubric because it is appropriate for assessing the speaking ability of senior high school students, whose English proficiency levels are generally developing and varied. Students at this level are still building control over fluency, pronunciation, grammar, and vocabulary; therefore, an analytic rubric that separates these components allows for a more accurate and fair assessment of their speaking performance. In addition, the descriptors in Luoma's rubric are clear, practical, and suitable for classroom-based

assessment, making them understandable for raters and appropriate for evaluating adolescent learners.

For the qualitative phase, several instruments are utilized to capture rich, in-depth perspectives on how AI-MALL influences classroom dynamics. Semi-structured interview guides are prepared for both teachers and students, containing open-ended questions that explore experiences with AI-based speaking tasks, perceived changes in interaction patterns, engagement levels, autonomy, and challenges faced during classroom implementation. The semi-structured format allows flexibility for participants to elaborate while ensuring consistency across interviews. Second, classroom observation sheets are used to systematically document interactional behaviors, including turn-taking, teacher–student exchanges, student collaboration, use of AI tools during activities, and overall classroom atmosphere. These observation sheets draw on the framework of Classroom Interactional Competence (Walsh, 2013).

Collectively, these instruments allow the researcher to gather comprehensive quantitative and qualitative data, ensuring that both measurable learning outcomes and experiential classroom processes are captured effectively. The combination of structured testing, rubric-based evaluation, interviews and observations enhances the validity and credibility of the study by addressing the research questions from multiple perspectives.

3.5 Research Procedure

The research procedure began with the development of research instruments, including speaking test tasks for the pre-test and post-test, an analytic speaking rubric, interview guides, and classroom observation sheets. Afterward, all AI-integrated MALL learning materials were prepared, and coordination with the English teacher was conducted to ensure smooth classroom implementation. Students' speaking performances were then scored using the analytic rubric by two trained raters, followed by the calculation of average pre-test scores and the examination of inter-rater reliability to ensure scoring consistency.

Subsequently, AI-supported speaking activities were implemented over an eight-week period as the main treatment of the study. During the intervention, the control group used AI-integrated Mobile-Assisted Language Learning (MALL) for basic speaking practice only. Students practiced pronunciation and completed short speaking tasks using AI applications; however, the activities were mainly individual, and the AI feedback provided was limited. The teacher continued to play a dominant role in guiding speaking activities, while AI functioned merely as an additional supporting tool rather than an integral part of classroom instruction.

In contrast, the experimental group used AI-MALL in a more active and integrated manner. AI was embedded directly into classroom speaking activities to provide real-time feedback, support longer and more complex speaking tasks, and facilitate interaction among students, the teacher, and the AI system. In this group, students engaged with AI not only for individual practice but also as part of classroom discussions and collaborative speaking activities, allowing AI to function as an instructional partner within the learning process.

To ensure equal treatment, both groups were taught the same speaking topics, followed the same learning objectives, spent the same amount of instructional time, and were evaluated using the same pre-test and post-test speaking tasks and analytic scoring rubric. The only difference between the two groups lay in the level of AI-MALL integration during speaking instruction, ensuring that any differences in outcomes could be attributed to the intensity of AI use rather than unequal instructional conditions.

The control group in this study was intentionally designed as an active control group rather than a no-treatment group. Both the experimental and control groups were exposed to AI-integrated MALL in order to control for general technology effects and to minimize potential novelty bias. This design ensured that any observed differences in students' speaking mastery were not simply the result of exposure to AI technology itself, but were attributable to differences in instructional implementation. Therefore, the presence of AI in both groups functioned as a methodological control rather than a source of bias.

The independent variable of the study was not the mere use of AI, but the level of AI integration within classroom-based speaking instruction. In the control group, AI was used in a limited and supplementary manner, primarily for individual practice with minimal integration into classroom interaction. In contrast, the experimental group received embedded and instructional AI integration, in which AI actively mediated speaking tasks, interaction patterns, and feedback processes alongside teacher and peer interaction. This distinction allowed the study to examine how different degrees of AI integration influenced speaking mastery, rather than comparing technology use versus non-use.

From a pedagogical and contextual perspective, excluding AI entirely from the control group would have reduced the ecological validity of the study, as AI-integrated MALL had already been part of the instructional environment. Providing both groups with access to AI ensured instructional fairness and reflected authentic classroom practice. As a result, the research design supported a more rigorous and context-sensitive comparison, enabling the findings to highlight the pedagogical impact of AI integration intensity rather than technological exposure alone.After the intervention, a speaking post-test was administered using task types comparable to those of the pre-test to maintain measurement consistency. The pre-test and post-test results were then compared using paired-samples t-tests and effect size calculations to determine the significance and magnitude of students' improvement.

In the qualitative phase, students and English teachers who were willing to participate were selected for interviews to obtain in-depth perspectives on the use of AI-integrated MALL. Classroom dynamics were also observed periodically using structured observation sheets. Finally, all interview data were transcribed systematically for further qualitative analysis.

3.6 Technique of Data Collection

The data for this study were collected through a series of systematic procedures to ensure the credibility and depth of the findings. First, the researcher obtained formal permission from the school administration and coordinated with

the participating teachers to schedule classroom visits and research activities. Subsequently, a pre-test was administered to all 96 students from the three speaking-focused classes to measure their initial speaking performance. After the implementation of the treatment, a post-test was conducted with the same 96 students to assess changes in their speaking mastery. In addition, semi-structured interviews were carried out with selected students and three English teachers to gain deeper insights into their experiences and interpretations of the use of AI-integrated MALL in the classroom. The students selected for the interviews were those who achieved the highest post-test scores, representing approximately 20% of the total student population.

3.7 Technique of Data Analysis

The data in this study were analyzed using a combination of quantitative analysis and qualitative thematic analysis to ensure a comprehensive interpretation of the findings. The questionnaire data collected from the 96 students were first processed through descriptive analysis, including simple diagrams, to identify general patterns related to students' perceptions of AI-integrated MALL, classroom dynamics, and speaking development. This quantitative analysis provided an overview of students' responses and supported the interpretation of overall trends in the data.

Meanwhile, the interview data obtained from selected students and the four English teachers were transcribed and analyzed using thematic analysis following the coding procedures proposed by Braun and Clarke (2006), which include familiarization with the data, initial coding, categorization, theme construction, and interpretation. In addition, the interview data were classified into three types of classroom dynamics based on Walsh's Classroom Interactional Competence framework and further analyzed through interactional analysis to examine how AI-MALL was implemented and how classroom dynamics unfolded in practice. Finally, data matching and cross-checking were conducted among the pre-test and post-test results, interview responses, and classroom observation findings to identify converging and diverging patterns. This integrated analytic approach

enabled the researcher to understand not only students' and teachers' perceptions of AI-MALL but also how these perceptions aligned with actual classroom practices, thereby strengthening the validity and depth of the study's conclusions.

3.8 Data Validation

3.8.1 Quantitative validity

To ensure the robustness of the quantitative component, several procedures were implemented to establish both validity and reliability of the speaking assessment. Content validity of the pre–post speaking test tasks and the analytic scoring rubric was confirmed through expert judgment involving two specialists in English language teaching and assessment, who evaluated the clarity, relevance, and alignment of the instruments with the constructs of speaking mastery. Procedural validity was maintained by administering the pre-test and post-test under standardized conditions, using parallel task types, identical instructions, and consistent test durations. Reliability of the speaking scores was ensured through inter-rater reliability, in which two trained raters independently evaluated all student performances using the analytic rubric. The inter-rater reliability coefficient (Cohen's Kappa or ICC) was calculated to verify scoring consistency, and high agreement levels indicated that the scoring system was stable and dependable for statistical analysis. Collectively, these steps strengthened the accuracy of the quantitative findings and the validity of any observed changes in students' speaking mastery following the AI-integrated MALL intervention.

3.8.2 Qualitative validity

The qualitative phase of the study employed multiple strategies to establish trustworthiness, encompassing credibility, dependability, confirmability, and transferability. Credibility was enhanced through triangulation across interviews, classroom observations, and descriptive questionnaire data, allowing the researcher to validate emerging themes

through multiple sources. Member checking was conducted with selected students and teachers, who were invited to review and confirm the accuracy of summarized interview interpretations. Dependability was maintained through careful documentation of coding processes following Braun and Clarke's (2006) thematic analysis framework, ensuring transparency and consistency in analytic decisions. Peer debriefing with academic colleagues served to challenge and refine interpretive claims, thereby enhancing confirmability. Observation data were also cross-checked with interview findings to verify alignment between reported experiences and classroom interaction patterns, particularly those related to Walsh's (2013) Classroom Interactional Competence framework. These validation procedures collectively ensured that the qualitative findings were trustworthy.

UNIVERSITAS MUHAMMADIYAH
GRESIK