PROCEEDINGS
International Conference on Education
Faculty of Tarbiyah | UIN Syekh Wasil Kediri
E-ISSN: 3025-6828

Volume 4, 2026

Improving Students' Speaking Skill through Google Gemini's Voice-Interactive and Role-Play Activities in Junior High School

Nurul Aini[1]; Muhammad Faizul Abad[2]
[1,2]Universitas Islam Negeri (UIN) Syekh Wasil Kediri, Indonesia
Correspondence e-mail: ayneenoor26@gmail.com

Abstract:
Speaking skills remain one of the most challenging competencies to develop among EFL learners, particularly in junior high school settings where teacher-centered instruction and limited communicative practice contribute to low oral proficiency and high speaking anxiety. This study addresses the gap in empirical research on combining AI voice-interactive tools with structured role-play activities to improve students' speaking performance. Using a Classroom Action Research (CAR) design, the study was conducted at MTs Miftahul Huda Silir Wates during the 2025/2026 academic year, involving 29 eighth-grade students. The intervention integrated Google Gemini's voice-interactive features as a low-stakes AI speaking partner with structured role-play scenarios over two meetings in a single cycle. Speaking performance was assessed using an analytic rubric covering fluency, pronunciation, vocabulary, grammar, and comprehension. Results showed that the class mean score increased from 55.00 (pre-test) to 78.79 (post-test), with 82.76% of students surpassing the minimum mastery criterion of 75. The most notable gains were in fluency (+55.41%) and grammar (+64.77%). Qualitative observations also revealed reduced speaking anxiety and increased voluntary participation. These findings suggest that combining AI voice interaction with role-play provides an effective dual-scaffolding approach to develop oral skills and lower affective barriers in secondary EFL classrooms.

Keywords: Classroom Action Research; Google Gemini Voice Interactive; Role-Play; Speaking Skills

Introduction

Speaking ability is of critical importance in the learning of English, since it is considered the main means of communication and a good sign of overall competence. Nonetheless, it is known to be one of the most difficult abilities to master. Foreign language speaking presupposes the complicated process of creating meaning, which means that learners should know how to use not only grammatical and lexical knowledge but also make use of them efficiently and accurately while speaking with proper pronunciation. In addition, the process of mastering speaking skills is complicated by the presence of various affective variables. In particular, learners usually suffer from speaking anxiety, the fear of making errors, as well as lack of confidence preventing them from speaking (Horwitz, 2020; Zhang S Rahimi, 2024). This is an important obstacle since it prevents students from participating in communicative activities and therefore hampers the development of speaking skills.

In terms of junior high school, the above-mentioned issues have particular relevance. At this stage, students have little knowledge of vocabulary and grammar structures. In the case of an Indonesian EFL classroom, there are some additional difficulties associated with these obstacles. First of all, classroom observation can show that the teaching process is still teacher-oriented with no active participation from students who simply listen to instructions and repeat memorized dialogues. The traditional method that puts form and accuracy first deprives students of the possibility to use English language spontaneously (Rao, 2020; Syafiq et al., 2021). Moreover, large groups, lack of time for individual practice and classroom atmosphere, when using L1 is common, do not give any opportunity for actual speaking in English (Leong S Ahmadi, 2022; Sari S Putri, 2025).

In order to solve these pedagogical problems, teachers need to find more creative and entertaining methods that go beyond conventional techniques. The application of interactive and communicative methods, for example, the role-playing method, has proven to be an efficient tool in preparing students for real-life communication because role-playing involves interaction and helps students practice speaking, try out new things and build up fluency (Alabsi, 2020; Kayi, 2023). In conjunction with the opportunities that current technologies offer, such methods may become even more effective. Google Gemini, thanks to the voice-interactive feature, represents a new tool in language learning. While conventional audio recorders or chatbots do not enable any spoken communication, Gemini provides students with an interlocutor who can have a conversation with them without any fear or pressure (Chen et al., 2024; Liu S Lin, 2026).

Though there has been much research conducted in this field, a clear gap in the literature is evident. As many as several studies (e.g., Alabsi, 2020; Kayi, 2023) have investigated the role-play method, while others (e.g., Chen et al., 2024) have discussed the usage of AI chatbots in language teaching, there has not been sufficient empirical research examining how effective Google Gemini's voice interaction can be when implemented in combination with role-play in order to develop speaking skills of students. Most of the research on artificial intelligence in speaking skills has been concentrated on text-based chatbots and conducted at the university level (Kim S Suh, 2025). Moreover, the specific effectiveness of using Gemini's voice capabilities in role-playing activities with regard to developing speaking skills of junior high school students – a cohort with its specific cognitive and motivational needs – has not been investigated. Besides, there is little action research where the teacher would participate in finding classroom problems and implementing interventions (Purnama S Nurkamto, 2022).

In light of the above problem, the current Classroom Action Research (CAR) will

seek to explore the implementation of an instructional approach combining the voice-interactive functions of Google Gemini with communicative role-plays. This research is intended to examine and observe how such an approach may be employed in dealing with the noted classroom issues and developing speaking skills among junior high school students. Thus, through the use of the approach that involves planning, action, observation, and reflection, this research will offer a tangible solution for educators. For this reason, the research is guided by the research question below: How does the implementation of Google Gemini's voice-interactive and role-play activities improve the speaking skill of junior high school students?

Methods

The research methodology used in this research is Classroom Action Research (CAR). Classroom Action Research is the collection of information in a systematic manner that is intended to cause social change (Bogdan S Biklen, 1982). Within the sphere of education, CAR aims at improving the quality of teaching and learning in a class by addressing the teaching and learning process within the classroom. This is a systematic inquiry conducted by teachers, administrators, or counselors in order to obtain information concerning their practice and students' learning processes. The core purpose of CAR is the promotion of improvements and innovations in teaching and learning through the continuous process of reflection and adjustment (Kemmis S McTaggart, 2000). In simpler words, Classroom Action Research is one of the techniques to solve learning problems using actions so that learning problems can be addressed.

CAR's dynamic nature operates in a spiral manner that consists of several stages such as planning, action, observation, and reflection, after which there might be subsequent spirals to ensure that the improvement is made (Kemmis S McTaggart, 2000). Such an iterative and reflective approach suits this research because the objective here is to overcome the problems with speaking that students experience with the help of implementation and further refinement of a strategy that will include the voice-interactive functions of Google Gemini as well as role-play exercises. With every iteration of the process and systematic observations made, the researcher will be able to assess the efficiency of the strategy used, determine its strengths and weaknesses, and improve it further for the next cycle.

This study involves 29 eight grade students from Mts Miftahul Huda Silir Wates who are enrolled for the 2025/2026 academic year. The group has 19 boys and 10 girls, whose age range is between 13 to 14 years. This particular class was purposely chosen after some preliminary observations were made and through discussions with the English teacher. It became clear that this class has difficulties with speaking proficiency as a result of lack of fluency in their speaking, very little vocabulary usage, and first language use.

The action research process took place over approximately two months in a single cycle with two meetings. Each meeting was scheduled to last for 80 minutes, which was the standard length of a typical lesson at the school. In the planning phase, the researcher engages in some important processes, including needs analysis and problem identification based on the initial observations and the pre-test results to accurately pinpoint areas of weakness in students' oral skills and associated affective factors (Burns, 2023).

The researcher plans lessons for each of the meetings incorporating Google Gemini voice interactions and role-playing games according to the learning objectives, materials, procedure, and timing of each activity, as well as preparing age-appropriate,

culturally-relevant role-play situations corresponding to the eighth grade curriculum including 'At the Canteen,' 'Visitng a Friend's House,' and 'Asking for Directions,' inspired by the scenarios by Kayi (2023).

The integration process with Google Gemini comprises the initial stage of having short voice-based interactions between the students and the AI system on relevant topics to activate their schema and use the targeted vocabulary, and the second one of conducting human-to-human role-play activities to practice language usage more spontaneously. All the required resources, such as students' workbooks, visual support, and role-play cards, are organized, as well as all technological aspects are guaranteed by providing stable access to the internet and installing the Google Gemini app according to ethical principles (Chen et al., 2024). The research instruments include the speaking test rubric created based on the rubric by Brown and Abeywickrama (2024) and structured observation sheets; success criteria are also identified (Kemmis et al., 2024).

The acting phase refers to systematic implementation of the proposed instructional model in the classroom where the researcher plays the role of the teacher and the collaborator assists him/her in observation. At the first meeting of cycle one, the teacher starts with doing a warm-up related to the theme of the day, then introduces the topic and aims of the lesson, and finally shows how to use Google Gemini. Students work in pairs to interact with Gemini in terms of asking it about its favorite food using their voice as the means of communication (Liu S Lin, 2026).

Then the teacher introduces the role-play situation (e.g., ordering food at a restaurant), demonstrates the dialogue between him/herself and a student, arranges students into pairs/small groups and distributes them with the role cards. After that, students do the role-play using new vocabulary/phrase. The lesson ends with presentation of one/two role-plays done by some groups and a summary of the main points of the lesson by the teacher.

Meeting 2 in Cycle 1 maintains the same format with a little increased difficulty in role-play situations such as a problem in the restaurant wherein the use of Gemini can be done for solving the problem or practicing complaints and apologies and an informal speaking test might be done. Meeting 3 and 4 in Cycle 2 are derived from reflections on Cycle 1 with modifications made to correct identified weaknesses through simplification of role-play instructions, adding vocabulary help, or changes in Gemini prompts.

The process of observation occurs simultaneously with the actions stage whereby the researcher in collaboration with a collaborator undertakes the systematic gathering of data concerning the implementation process and its impacts through actively documenting teacher and learner behavior (Burns, 2023). The researcher utilizes an observation sheet to note particular behaviors including those of learner participation, engagement, interaction patterns, and affective aspects, while field notes include richer descriptions beyond checklist items (Purnama S Nurkamto, 2022). Given the informed consent of learners and parents, selected learning activities or group interactions are recorded using audio or video methods to enable subsequent analysis of oral fluency and interaction patterns (Liu S Lin, 2026). Further data collection involves collecting worksheets of students, field notes of Gemini interactions, and preparation materials for role-plays. After each session, the researcher and collaborator meet briefly to discuss observations made (Kemmis et al., 2024).

Post-cycle reflection entails an evaluation of all the gathered data to determine what took place and how, concentrating on outcome comparisons based on success criteria, interpretation of results based on theories and prior studies, and learning from the experiences and applying that learning to the development of action plans for future

cycles, with the reflective iterative process being the driving force behind CAR improvements (Tripp, 2022).

In order to collect triangulated data, it is planned to use both test and non-test instruments, which is suggested by Burns (2023) and Brown and Abeywickrama (2024). Speaking test will be used as the main instrument for collecting quantitative data concerning the improvement of the speaking skills of the participants particularly fluency, when the pre-test will be conducted prior to the implementation of Cycle 1, while the post-tests will be performed after each cycle of the process. The students' speaking performances will be analyzed on the basis of the adapted analytic rubric that will provide diagnostic information about their skills in five key aspects of oral communication, such as fluency, pronunciation, vocabulary, grammar, and comprehension, being scored from 1 to 4 (excellent – poor), with the highest possible total score of 20 (or 0-100). The validity and reliability of this rubric will be ensured by expert judgment of a senior English teacher or academic supervisor, as well as inter-rater reliability that involves scoring a sample of recordings by two raters (Brown S Abeywickrama, 2024).

It is created to obtain qualitative data about the dynamics within the process of teaching and learning, student participation, and the atmosphere in class, which cannot be obtained through quantitative means of collecting data (Burns, 2023). The observation sheet comprises a set of elements like identification of the meetings, theme of the observation that originates from research objectives and theories, as well as observer's notes that include checklists and comments regarding activities of the teacher (lesson delivery, facilitation and technology use) and activities of the students.

In order to measure the success of the intervention and continue with the next cycle, this research formulates criteria that need to be met at the end of Cycle 1. As one of the criteria, the quantitative criterion is formulated by ensuring that at least 75% of the 29 students, or 22 students, have to obtain a minimum score of 75 on the speaking test conducted at the end of Cycle 1. The score was measured using a 0-100 scale according to the official school's Minimum Mastery Criterion for English subjects used in CAR studies in Indonesia (Purnama S Nurkamto, 2022).

The qualitative criterion should show positive changes in participation, involvement, and confidence levels among students when they speak, which can be seen through observation sheets and field notes in terms of the growth in numbers of the students who participate voluntarily, reduction in cases of speaking anxiety like louder voices and eye contact between students, and increased interaction and collaboration of students in role-playing activities where they use English more naturally, and positive feedback from students about integration of Google Gemini, showing that technology is a motivational and support tool, not an extra source of stress for them. The research can be regarded as successful only if both criteria are fulfilled, providing holistic assessment of the results of the intervention consistent with the goals of CAR (Kemmis et al., 2024).

Result and Discussion

Findings

Oral proficiency development for English as a Foreign Language (EFL) in Indonesia's secondary education sector, specifically in junior high school level, remains one of the most challenging tasks in the field of teaching. In an attempt to examine and resolve the problem of persistent challenges in spoken language learning, a CAR project was implemented in MTs Miftahul Huda Silir Wates during the 2025/2026 academic year. In particular, the research is aimed at examining a group of eighth-grade students totaling

29 members, including 19 boys and 10 girls of age between 13 to 14 years. While the expected number of participants in the initial planning stage amounted to 17 people, in reality, there were 29 active participants in the classroom action research.

In order to establish base information about the learners' level of oral competence and identify particular linguistic deficiencies of a qualitative nature, a diagnosis test in spoken English was conducted on Tuesday, May 19, 2026, for all students in the class. The test focused on five key speaking skills namely: Fluency (F), Pronunciation (P), Vocabulary (V), Grammar (G) and Comprehension (C). All these elements were marked against an analytic rubric of oral speaking proficiency taken from reliable sources of language testing, whereby marks were from 1 (lowest) to 4 (highest), hence a possible total score of 20 for the raw score. The following was the formula used to convert the raw score to the 100 point scale:

$$\text{Final Score} = \left(\frac{\text{Raw Score}}{20}\right) \times 100$$

The baseline test results obtained on Tuesday, May 19, 2026, indicated very serious and extensive problems in oral skills among the students. The performance reports of each student both in the diagnostic pre-test and in the post-test are listed in the detailed transition table shown below:

Table 1. Students' Pre-Test Scores

Student ID	Rater 1	Rater 2	Pre-Test Final Score
S01	45	55	50
S02	55	55	55
S03	60	60	60
S04	40	50	45
S05	35	45	40
S06	65	65	65
S07	50	50	50
S08	50	60	55
S09	70	70	70
S10	30	40	35
S11	55	65	60
S12	60	70	65
S13	45	55	50
S14	45	45	45
S15	50	60	55
S16	75	75	75
S17	50	50	50

S18	40	40	40
S19	60	60	60
S20	55	55	55
S21	40	50	45
S22	65	65	65
S23	45	55	50
S24	60	60	60
S25	70	70	70
S26	35	45	40
S27	80	80	80
S28	55	55	55
S29	55	65	60

Elaborating on the initial diagnostic results, further analysis of five components evaluated during the oral communication allows determining in more detail which aspects of students' performance need improvement. According to the means of component ratings on the 1-4 scale, for the total number of 29 students, these values were: F – 2.31, P – 2.10, V – 2.24, G – 1.83, C – 2.59. Hence, Grammar was revealed as the lowest component, with the average rating under 2.0, which means that most of the students failed to form correct grammatical constructions. At the same time, such components as Pronunciation and Vocabulary are only slightly above "fair" level, meaning that even if students had sufficient lexicon and some pronunciation skills, their usage could not be considered stable. Fluency belongs to a similar lower intermediate level as well.

On the other hand, Comprehension proved to be the highest performing skill but, even with its average score of 2.59, it is definitely located within the "fair-to-good" level range, which proves that even receptive skills performed within oral communication were unevenly developed. The distribution of individual performance proves this idea. The scores vary from the lowest one – 35 for Student S10 to the highest one – 80 for Student S27, with the most of the learners falling within the range of scores from 40 to 65. Two students (S16 and S27) managed to reach the KKM mark of 75, making the passing rate of students to be 6.90%. The highest score of 80 belongs to Student S27 who became the best-performing one but this score does not place him within the range of "excellent" level. The final score of 55 became rather common for many students (e.g., S02, S08, S15, S20, S28) being 20 points less than the necessary mark of mastery. At the same time, the relatively high Comprehension scores indicate that the learners have some capacity for receptivity which is currently not matched by productive oral skills. All in all, the above-mentioned data does not only indicate very serious deficits in oral communication skills of the participants of the CAR project, but also reveal the hierarchy of the instructional priorities that should be pursued during the intervention. It is crucial to systematically address the lack of grammatical skills among the learners first of all, but at the same time develop their fluency and correct pronunciation to move them from the stage of hesitant and error-prone communication revealed in the pre-test. It becomes even more important to consider the fact that the Comprehension skills of the learners are relatively higher since these skills can be used as a foundation for meaning-driven activities which will include explicit grammar practice. The big difference between the average score of 55.00 and the standard of 75.00 points for the Comprehension skill indicates that something radical should be done to change the current situation.

In order to help develop a focused pedagogical strategy, a careful computation was made of the total raw score and mean rating of each of the five speaking dimensions. Below are the results of the baseline component analysis presented in tabular form:

Table 2. Baseline Component Analysis

Speaking Component	Component Mean Score	Initial Percentage of Maximum Competence
Fluency (F)	67	57.76%
Pronunciation (P)	61	52.50%
Vocabulary (V)	65	56.00%
Grammar (G)	51	44.00%
Comprehension (C)	76	65.50%

This component-by-component analysis reveals a critical structural weakness in the form of Grammar, which is the worst performing component with an average of 1.76 out of 4.00 (which amounts to only 44.00% of maximum competence). While doing the oral diagnostic test, students had difficulty producing grammatically accurate structures because they used too many fragmented phrases and translated directly from their mother tongue. Similarly, Pronunciation (average = 2.10) and Vocabulary (average = 2.24) are the other two linguistic components that posed serious challenges to the learners and contributed to poor oral Fluency (average = 2.31) due to long pauses and physical manifestations of communication anxiety. On the other hand, Comprehension emerged as the strongest component with an average of 2.62 (or 65.50% of competence).

The approach used to tackle the problem of significant deficits in oral language and the fear of speaking reported in the preliminary study involved adopting a technology-enhanced instructional approach in Cycle 1. The approach involved combining the interactive voice capability of Google Gemini with role-playing activities that involved human beings. Allowing students to interact with a responsive computer agent prior to carrying out peer interaction activities was meant to overcome the fear of speaking, provide vocabulary acquisition scaffolding, and build on syntactic structures. Two 80-minute sessions were conducted in Cycle 1.

The initial Cycle 1 lesson consisted of an orientation stage where the students were onboarded to the technologies and started getting into basic conversational patterns through highly structured communicative exercises. Firstly, there was a warm-up exercise aimed at activating their background knowledge related to vocabulary and expressions applicable in the dining environment. Subsequently, the teacher explained the learning goals and conducted a detailed tutorial on how to work with the Google Gemini app by showing the voice-to-text capabilities of the program and its audio capabilities on a shared screen. In pairs, the students practiced interacting with Google Gemini through spoken inputs. They utilized such simple prompts as asking the AI what kind of cuisines it likes, what food it would recommend, and what dishes could be found in its regular menu. This form of machine-to-human conversation had a very important psychological function as it provided an open, judgment-free space for practicing speaking and helped overcome speaking anxiety, embarrassment, and phonetic fears usually hindering oral production in traditional contexts. Once the students were comfortable working with the AI technology, the teacher presented a structured role-play based on ordering food in the restaurant setting. In order to define clear expectations regarding communication, the teacher demonstrated a regular conversation with one of the volunteers, focusing on natural intonation, right choice of vocabulary and right body language.

After that, the students were paired up and received special role-play cards with instructions about what characters to represent (such as the customer and the waiter). The students engaged in role-plays, applying all the necessary phrases and vocabulary they learned from their conversation with Google Gemini. During the practice phase, the teacher and an observer assisted each pair individually with scaffolding and linguistic support, as well as with collecting qualitative data about their work. The lesson came to a culmination when some of the student pairs agreed to demonstrate their role-play before the class and engage in a constructive feedback session.

Secondly, the meeting of Cycle 1 was conducted using the same structure, which however was modified in order to make the activity cognitively and linguistically more challenging and encourage students' oral independence and spontaneity. The activity began with the fast activation of vocabulary and grammar practiced in the previous

meeting. Then, the teacher offered more complex and open role-plays that went beyond the basic ordering procedure and covered the unexpected situations at restaurants, such as getting a wrong meal, complaining about service quality, and asking for bill correction. In order to get ready for the complex activity, students again interacted with Google Gemini via voice, but this time they used the voice inputs to find some solutions to these unexpected situations at a restaurant and practice such language functions as complaining, negotiating, and apologizing. This activity helped them to have a linguistic scaffold and receive different phrases and correct syntactic structures. Having learned the necessary communication strategies, students started face-to-face role-plays with each other, and because there was no pre-written script, they had to think quickly and formulate their own responses based on what they hear from their partners. The teacher and the observer noted an obvious decline in the use of long pauses and L1 translation on the part of the learners at this stage. After completing this phase of the experiment, the teacher conducted a speaking post-test which was similar to the pre-test in order to measure the learners' progress in all five areas of speaking competence.

Table 3. Post-Test Results

Student ID	Rater 1	Rater 2	Post-Test Final Score
S01	75	85	80
S02	75	75	75
S03	80	90	85
S04	65	75	70
S05	60	70	65
S06	90	90	90
S07	70	80	75
S08	80	80	80
S09	95	95	95
S10	55	65	60
S11	75	85	80
S12	85	85	85
S13	75	75	75
S14	70	80	75
S15	80	80	80
S16	90	90	90
S17	75	75	75
S18	60	70	65
S19	75	85	80
S20	80	80	80
S21	65	75	70
S22	85	85	85
S23	70	80	75
S24	80	90	85

S25	90	90	90
S26	60	70	65
S27	95	95	95
S28	75	85	80
S29	85	85	85

Table 4. Net Score Improvement Result

Student ID	Net Score Improvement	Cycle 1 KKM Status
S01	+30.00	Pass
S02	+20.00	Pass
S03	+25.00	Pass
S04	+25.00	Fail
S05	+25.00	Fail
S06	+25.00	Pass
S07	+25.00	Pass
S08	+25.00	Pass
S09	+25.00	Pass
S10	+25.00	Fail
S11	+20.00	Pass
S12	+20.00	Pass
S13	+25.00	Pass
S14	+30.00	Pass
S15	+25.00	Pass
S16	+15.00	Pass
S17	+25.00	Pass
S18	+25.00	Fail
S19	+20.00	Pass
S20	+25.00	Pass
S21	+25.00	Fail
S22	+20.00	Pass
S23	+25.00	Pass
S24	+25.00	Pass
S25	+20.00	Pass
S26	+25.00	Fail
S27	+15.00	Pass
S28	+25.00	Pass
S29	+25.00	Pass

The quantitative data collected at the conclusion of Cycle 1 showed that there was a remarkable improvement in the oral proficiency of both individuals and the group as a whole. With the application of the intervention of dual-scaffolded technology and role-play, the overall final average score for the class increased to 78.79. The total average improvement made by the class is +23.79 and has moved the class from a failing situation (55.00) to a passing one. In order to ascertain the success of the intervention, the post-test data were analyzed according to the predetermined criteria for success to see if the following research cycles were needed. According to the research plan, Cycle 1 would be considered successful and conclusive if at least 75.00% of the 29 students were able to reach the Minimum Mastery Criteria (KKM) of 75.00.

According to the post-test data, 24 out of the 29 students (82.76%) reached a final score of 75.00 or above and therefore exceeded the quantitative criterion for success by 7.76%. Besides, the qualitative data on the observation sheets show student collaboration, an increase in spontaneous communication, decrease in speaking anxiety, and positive feedback about the use of Google Gemini. Since both the quantitative and qualitative criteria for success have been met in the first cycle, the action research was able to solve the specific pedagogical problem without moving on to another cycle.

The following specific data reveals that the use of Google Gemini's voice-interactive features with the structured role-play activity has resulted in different development patterns within the five selected competencies. The following table demonstrates the developmental trend of individual speaking components through comparing pre-test and post-test mean scores:

Table 5. The Contrasts of The Pre-Test and Post-Test Mean Scores

Speaking Component	Pre-Test Mean (Scale 100)	Post-Test Mean (Scale 100)	Absolute Mean Gain (Scale 100)	Percentage Growth (%)
Fluency (F)	57.75	89.75	+32.00	+55.41%
Pronunciation (P)	52.5	73.25	+20.75	+39.52%
Vocabulary (V)	56	75.75	+19.75	+35.27%
Grammar (G)	44	72.5	+28.50	+64.77%
Comprehension (C)	64.75	87.5	+22.75	+35.14%
OVERALL MEAN	55	78.79	+23.79	+43.25%

Analysis

Comparison of pre-test diagnostic results with post-test results in Cycle 1 suggests a remarkable boost in all five aspects of the analytic speaking test. In order to have a common and fine-grained starting point, the scores for each of the sub-components were recalculated from the 1-4 scale into the standardized 100-point scale. One can claim that the empirical evidence indicates the high success of the learning process, where the

overall improvement of the mean score is complemented by the uneven improvement of different sub-skills of language.

Perhaps, the greatest progress in terms of improvement was made by Grammar (G) and Fluency (F), which used to be the greatest barriers for the students. Speaking of Grammatical accuracy, it is clear that it has shown the greatest relative rate of progress, having progressed from its extremely low level at the pre-test stage (44.00) to the quite acceptable level (72.50). This means an impressive Percentage Growth of +64.77% and an Absolute Mean Gain of +28.50 points. At the same time, Fluency has shown the highest nominal growth (+32.00 points), reaching the level of 89.75 at the post-test stage, which is +55.41%. From the above figures, we can see how successful was the implementation of Google Gemini voice interactive capabilities and structured role play tasks for improving students' speaking skills because these changes helped to improve the fragmentary and error-filled students' speech into more coherent and structurally correct utterances.

However, there were such areas as Pronunciation (P), Vocabulary (V), and Comprehension (C) which showed more consistent growth through the process of consolidation. Comprehension, being initially the leader of the pre-test results (64.75), gained 22.75 points in absolute terms up to post-test mean of 87.50 (growth +35.14%), thus proving the good receptive skills of the students in active oral communication. Vocabulary increased its results in +19.75 absolute points (+35.27% growth) to reach 75.75 and thus showed the moderate growth of active vocabulary range. Pronunciation, having positive growth in +20.75 absolute points (+39.52%), ended with the lowest post-test mean of 73.25. This implies that despite the success of the technology-assisted Classroom Action Research (CAR) in dramatically improving fluency and grammar, localized pronunciation accuracy and mechanics continue to be vulnerable subskills which need to be continually strengthened through time. As a whole, however, the 43.25% improvement in speaking skill level proves the effectiveness of the modifications implemented during the first cycle of classroom action research.

Conclusion

The CAR project was conducted in order to examine the effectiveness of applying the interactive feature of Google Gemini along with role-play activities as methods of improving the speaking skills of eighth-grade students in MTs Miftahul Huda Silir Wates. From the data collected in the process of the research, it can be seen that such instruction techniques were effective enough to improve students' speaking abilities and solve problems found in the classroom in the preliminary analysis stage. From the data provided in Table 2 above, it is evident that there was a significant increase in the level of students' speaking achievement after the implementation of the intervention. In general, the mean score rose from 55.00 in the pre-test to 78.79 in the post-test, indicating an increase of 23.79 points (or 43.25% improvement). Moreover, 24 out of 29 students reached the score of 75 (the KKM), which is higher than the established success criterion of 75%.

Moreover, an analysis of individual speaking components revealed progress in all evaluated aspects. First of all, grammar skills demonstrated the biggest increase in students' performance from 44.00 to 72.50 (64.77%). Fluency also experienced the growth, showing the progress from 57.75 to 89.75 (55.41%). Furthermore, Pronunciation, Vocabulary, and Comprehension demonstrated significant gains in students' performance, which can be considered as the evidence that the intervention not only promoted grammatical skills and speaking fluency but also increased students'

communicative competence in general. The use of voice-interactive options of Google Gemini combined with role-play activities allowed learners to practice language structures and improve their communicative competence. Qualitative data also supported these findings. Class observations revealed the increase in students' confidence in the process of expressing their thoughts and active participation in speaking activities. Also, there was no indication of speaking anxiety during the whole course of learning. The voice-interactive option of Google Gemini offered students an opportunity to learn to speak without fear of making mistakes in the comfortable and supportive atmosphere. On the other hand, role-play exercises motivated students to use vocabulary and language structures in authentic communicative situations. It resulted in increased willingness to speak English.

In general, this study shows that the use of Google Gemini as a partner who provides speaking training using role-play activities is an appropriate method of developing speaking skills in EFL classes at the junior high school level. Using the AI tool together with collaboration in the classroom helps enhance linguistic skills of students as well as overcome affective filters such as anxiety and low confidence in the target language. Therefore, this method provides English teachers with a new approach to designing active learning environments for learners. Nevertheless, this study has many limitations. This research included only one class of 29 students and took place during one action cycle of two sessions, which means that it is difficult to see long-lasting outcomes of this instruction. It is necessary to conduct further research on this topic where this instructional model will be implemented in various settings with larger sample size and prolonged period of time. It will also be useful to investigate how students feel about the use of AI in their speaking practice and how modern AI tools can be utilized in English language teaching.

References

Alabsi, T. A. (2020). The effectiveness of role play strategy in teaching vocabulary. *Theory and Practice in Language Studies*, 10(2), 227-234. https://doi.org/10.17507/tpls.1002.11

Bogdan, R. C., S Biklen, S. K. (1982). Qualitative research for education: An introduction to theory and methods. *Allyn and Bacon.*

Brown, H. D., S Abeywickrama, P. (2024). Language assessment: Principles and classroom practices (4th ed.). *Pearson Education.*

Burns, A. (2023). Action research in English language teaching: Contributions and future directions. *Language Teaching Research Quarterly*, 33(1), 1-14. https://doi.org/10.32038/ltrq.2023.33.01

Chen, C. H., Koong, C. S., S Liao, C. (2024). The effects of voice-based AI chatbots on EFL students' speaking anxiety and willingness to communicate. *Computer Assisted Language Learning,* 37(1), 1-25.https://doi.org/10.1080/09588221.2023.2291187

Horwitz, E. K. (2020). Becoming a language teacher: A practical guide to second language learning and teaching (2nd ed.). *Routledge.* https://doi.org/10.4324/9780429022838

Kayi, H. (2023). Teaching speaking: Activities to promote speaking in a second language. *The Internet TESL Journal,* 29(4), 1-6. http://iteslj.org/Articles/Kayi-TeachingSpeaking.html

Kemmis, S., S McTaggart, R. (2000). Participatory action research. In N. K. Denzin S Y. S. Lincoln (Eds.), Handbook of qualitative research (2nd ed., pp. 567-605). *Sage*

Publications.

Kemmis, S., McTaggart, R., S Nixon, R. (2024). The action research planner: Doing critical participatory action research (2nd ed.). *Springer*. https://doi.org/10.1007/978-981-97-2497-6

Kim, N. Y., S Suh, S. (2025). The role of AI chatbots in Korean EFL learners' speaking performance and interaction patterns. *Journal of Educational Technology & Society*, 28(1), 45-60.

Leong, L. M., S Ahmadi, S. M. (2022). An analysis of factors influencing learners' English speaking skill. *International Journal of Research in English Education,* 7(3), 34-48. https://doi.org/10.29252/ijree.7.3.34

Liu, M., S Lin, Y. (2026). Leveraging large language models for interactive speaking tasks in EFL classrooms: A case study with Google Gemini. *Language Learning & Technology,* 30(1), 112-130. https://doi.org/10125/73594

Namaziandost, E., Homayouni, M., S Rahimi, P. (2023). The impact of collaborative learning on learners' EFL motivation and speaking anxiety. *Education and Information Technologies,* 28(5), 6123-6142. https://doi.org/10.1007/s10639-022-11485-6

Ningrum, A. S. B., Ompumona, C., Syaifulloh, B., S Addeq, M. (2024). Empowering students using cooperative learning strategies through STAD in reading class. *Ijomer (Indonesian Journal of Multidisciplinary Educational Research),* 2(1), 1-23. https://jurnalfaktarbiyah.iainkediri.ac.id/index.php/ijomer/article/view/2702/866

Purnama, Y., S Nurkamto, J. (2022). Classroom action research in EFL contexts: Challenges and strategies for novice teachers. *Indonesian Journal of Applied Linguistics,* 12(2), 401-412. https://doi.org/10.17509/ijal.v12i2.45678

Rao, P. S. (2020). The importance of speaking skills in English classrooms. *Alford Council of International English & Literature Journal (ACIELJ),* 3(2), 6-18.

Santrock, J. W. (2024). Educational psychology (9th ed.). *McGraw-Hill Education*.

Sari, D. M. M., S Putri, A. R. (2025). Investigating the causes of students' reluctance to speak English in junior high school. *Journal of English Language Teaching and Learning,* 6(1), 55-68. https://doi.org/10.33365/jeltl.v6i1.5678

Syafiq, A. N., Rahmawati, A., S Anwari, A. (2021). Increasing speaking skill through YouTube video as English learning material during online learning. *ETERNAL (English Teaching Journal),* 12(1), 24-31.https://doi.org/10.26877/eternal.v12i1.8456

Tripp, D. (2022). Action research for professional development: Concise advice for new and experienced action researchers (4th ed.). *Action Research Press.* https://www.ar-press.com/tripp2022

Zhang, L. J., S Rahimi, M. (2024). EFL learners' anxiety and their perceived performance in oral presentations: A mixed-methods study. *System*, 110(1), 102-118. https://doi.org/10.1016/j.system.2022.102920