Journal of Al-Akbar Pearl of Education
Volume X, Number X, Year
P-ISSN: xxxx-xxxx E-ISSN : xxxx-xxxx
Open Access: https://ejournal.mutiaraalakbar.com/jmpa

The Effect of AI Avatars in Teaching English to Enhance Interactional Speaking Skills Eighth grade SMP Tahfidz Mutiara Al-Akbar

Hikma Patika Sari[1], Kaharuddin[2], Sherlia Nur[3], Wandy Renaldy[4]
[1,2,3] UIN Alauddin Makassar
[4] IAI DDI Sidrap
Correspondence: @patikasarihikma@gmail.com

Article Info
Article history:
Received Jun 12th, 2025
Revised Aug 20th, 2025
Accepted Aug 26th, 2025

Keyword:
AI Avatar, CEFR, Interactional Speaking, Merdeka Curriculum, Speaking Function

ABSTRACT

This study investigated the effect of integrating AI Avatars on enhancing the interactional speaking skills of eighth-grade students at SMP Tahfidz Mutiara Al-Akbar. The research addressed the gap between students' current speaking proficiency and the learning outcomes of the Merdeka Curriculum, which emphasizes active communicative competence. Employing a quantitative approach with a one-group pretest–posttest design, the study involved seven students selected through purposive sampling. Data were collected through oral assessments and mapped against the Common European Framework of Reference for Languages (CEFR) benchmarks. The instructional intervention utilized AI platforms, specifically Hedra AI, Dreamina, and Canva, integrated across three instructional stages: pre-teaching, while-teaching, and post-teaching. The findings revealed an initial pre-test mean score of 2.66, placing the students at the B1 (Intermediate) level. Following the intervention, the post-test mean score increased to 3.45, indicating a statistically significant improvement, particularly in responsiveness, fluency, and sociolinguistic accuracy. The study concludes that AI Avatars provide a low-anxiety and immersive learning environment that effectively enhances student engagement and fosters communicative competence in real-life social contexts.

INTRODUCTION

The growth of technology has brought significant changes in the world of education. Based on data from Statista (2023), an online statistics portal, English is the most spoken language in the world, with more than 1.35 billion speakers. This shows that English is widely used as a means of communication in various countries. English has become a key tool for international interaction, learning, and career development around the world (Kaharuddin, 2025) In addition to being a means of communication with native speakers, English is also the most widely studied and used second language globally (Ilyosovna, 2020).

The low English proficiency of students in Indonesia is often caused by several factors. (Lowenberg, 1991, in Kaharuddin, 2018) highlights that the mismatching between the curriculum and the syllabus used is one of the main obstacles in achieving the effectiveness of English language learning. First, the curriculum that does not optimally support the development of language skills becomes a barrier for students to achieve the expected competencies. second, the crowded and noisy classroom conditions also make it difficult to create a conducive learning atmosphere. In addition, teachers' limited competence, both in terms of language mastery and teaching methodology, further worsen the situation. To address

these complex and interconnected issues, (Jain, 2023, in Kaharuddin, 2025) refines and consolidates these into four key categories that are more accessible and applicable to real-world educational contexts: (1) Curriculum Development and Evaluation, (2) Teacher Education and Professional Development, (3) Student Learning and Achievement, and (4) Educational Policies and Governance.

The curriculum has a crucial role in determining the direction and success of education (Setiawan et al., 2020). A well-designed curriculum is able to shape the character of students who are adaptive and able to function effectively in the midst of the dynamics of world change. However, many teaching materials still focus on aspects of grammar and vocabulary, without giving enough attention to contextual and relevant speaking practices (Wulandari et al., 2021). However, English learning materials in Indonesia are commonly developed without sufficient reliance on research or student-centered evaluation, as educators tend to rely on subjective views. This practice often produces content that fails to meet learners' actual needs, limiting its potential to enhance language proficiency Arafah et al. (2021). A systematically designed and goal-based curriculum has a central role in ensuring the quality of learning processes and outcomes. When the curriculum contains achievable learning objectives, relevant content, and interesting and motivating learning activities, it will have a significant impact on improving Teachers' Teaching outcomes.

In the context of English language Teaching, especially the development of speaking skills, the existence of a structured curriculum allows students to engage in a variety of communicative activities designed to increase confidence, fluency and accuracy in the language. (Kaharuddin & Abd. Hakim, 2018) Therefore, a good curriculum will ensure that students not only understand the material theoretically, but are also able to use English actively and effectively in real contexts as well as considering needs and interests so that learning objectives can be achieved well. Speaking skill is one of the important aspects in mastering English, especially in the context of English as a Foreign Language (EFL). Kaharuddin (2018) asserts that speaking mastery is not only related to the correct language structure, but also includes the functional ability to convey meaning, Richards (2008), classifies speaking based on its function into three main parts, namely speaking as Interactional, Transactional, And Performance.

Based on the results of preliminary research on the speaking ability of seventh grade students of Mutiara Al-Akbar Junior High School in pre-test taken from 8 students, it was found that there was a mismatch between the expected achievements in the learning objectives and the actual conditions of students' abilities. Student achievement had not fulfilled all the learning objectives that had been set. The learning objectives focused on mastering basic social expressions in oral form through pronunciation, intonation, and context-appropriate usage, as well as encouraging meaningful, confident, and polite interactions in English. Based on the results of the assessment of the three aspects of the speaking function, namely Interactional, Transactional, and Performance Speaking the overall average score was 2.37 on a scale of 4. This result is classified as (Basic), which indicates that students have shown initial progress in speaking skills, but have not fully achieved the learning objectives set out in the lesson plan.

To bridge this gap, integrating Artificial Intelligence (AI) through AI Avatars offers a strategic and innovative solution. AI Avatars provide an interactive and immersive environment that allows students to practice speaking in a low-anxiety setting, which is essential for developing interactional competence. This technology-integrated approach aligns with the principles of *Curriculum Merdeka*, which promotes active participation and personalized learning. Furthermore, utilizing such modern tools is not only academically relevant but also resonates with the progressive values of Islamic education at SMP Tahfidz Mutiara Al-Akbar.

Consequently, this study aims to investigate The Effect of AI Avatars in Teaching English to Enhance Interactional Speaking Skills: A Study of Eighth Grade Students at SMP

Tahfidz Mutiara Al-Akbar," this research seeks to provide empirical evidence on how AI-driven tools can stimulate engagement and substantially improve the ability of students to communicate effectively in real-life social contexts.

RESEARCH METHODS

This study employs a quantitative approach with a pre-experimental design, specifically the one-group pretest-posttest design. This design is utilized to evaluate the effectiveness of an instructional intervention by comparing students' speaking proficiency before and after the treatment (Sugiyono, 2013). While this design lacks a control group, it remains a highly relevant and practical method for assessing educational interventions in real classroom settings (Ary et al., 2010). The participants of this study were seven students from the Eighth-grade class at SMP Tahfidz Mutiara Al-Akbar.

Data collection

Data collection is focused on measuring students' speaking baseline and progress through three key speaking functions identified by Richards (2008): interactional, transactional, and performance-based speaking.

1. Interactional Speaking: Assessed through pair or small-group interviews on social topics. Criteria (adapted from CEFR, 2020) include responsiveness, turn-taking, fluency, pronunciation, and sociolinguistic accuracy.
2. Transactional Speaking: Evaluated through role-play activities focusing on the exchange of information. Criteria (adapted from Brown, 2004) include information clarity, grammar accuracy, and vocabulary use.
3. Performance Speaking: Assessed via picture description, oral presentations, and storytelling. Criteria (adapted from Hughes, 2003) emphasize structure, content relevance, and overall delivery.

Instructional Treatment

The intervention follows a three-stage instructional model—Pre-Teaching, While-Teaching, and Post-Teaching as adapted from Kaharuddin (2018). AI Avatar technology (e.g Hedra Ai, Dremina Ai and Canva) serves as the primary medium:

1. Pre-Teaching: Teachers set learning objectives and introduce the AI Avatar tools, ensuring students understand how to interact with the technology.
2. While-Teaching: The core phase where students engage in active tasks. Interactional skills are practiced through simulated conversations with AI Avatars to lower anxiety (Kaplan-Rakowski et al., 2024). Transactional skills are honed through role-plays and AI-based video assignments. Finally, performance skills are developed through AI-generated picture descriptions and storytelling based on the *English for Nusantara* curriculum.
3. Post-Teaching: A reflection phase where the teacher and students review the learning process and provide feedback on the AI integration.

Data Analysis and Assessment Criteria

To ensure high reliability, an analytic scoring rubric with a four-level scale (1–4) was implemented. While the initial rubric was designed to address specific local classroom needs, the resulting scores were subsequently mapped to the Common European Framework of Reference for Languages (CEFR) to align with international benchmarks (Little, 2006). This mapping ensures that the learners' speaking abilities are interpreted through both a contextual and global lens. To maintain objectivity and minimize potential bias, the assessment process

was conducted in collaboration with an external evaluator from the American English School (Miss Rima), who served as an independent rater.

The conversion of the analytic scores into proficiency levels and qualitative descriptions is detailed in Table 1 below:

Table 1 Guideline for Evaluating the Students speaking skill

Range Score	Level of Proficiency	CEFR Level	Qualitative Description
3.76-3.90	Advanced	C1	Very good
3.51-3.75	Upper-Intermediate	B2	Good
2.51-3.50	Intermediate	B1	Fair
1.51-2.50	Basic/Elementary	A2	Poor
1.00-1.50	Very low/ Beginner	A1	Very Poor

Table 1.1 Guideline for Evaluating the Students speaking skill

Source: Adapted from CEFR (2001) for Proficiency Levels and Kaharuddin & Yassi (2018) for Qualitative Descriptions.

The assessment framework presented in Table 1 serves as the standardized for evaluating the students' progress throughout the study. By integrating the scoring system with the CEFR benchmarks, the researcher can precisely categorize each student's speaking ability, moving from "Very Poor" (A1) to "Very Good" (C1).

In this study, this classification is crucial for identifying the specific gap between the students' baseline performance in the pre-test and their potential growth after the AI Avatar intervention. Furthermore, the qualitative descriptions ranging from Basic to Advanced provide a clear narrative of the students' communicative competence, ensuring that the results are not only statistically measurable but also pedagogically meaningful in an international context.

RESULTS AND DISCUSSION

The Effects of an Avatar AI-Based Lesson Plan in Enhancing speaking skill

In this study, the term 'Effect' is analyzed through established linguistic definitions. The Cambridge Dictionary (n.d.) defines it as the outcome of a specific cause, while the Oxford Learner's Dictionary (n.d.) describes it as a change which is a result or consequence of an action or other cause. Based on these definitions, it can be concluded that an 'Effect' is the tangible result or specific change that occurs as a direct consequence of a particular action. In the context of this research, the 'effect' refers to the measurable improvement in student performance resulting from a specific teaching intervention. Consequently, this subchapter addresses the first research question by examining the effects of implementing an Avatar AI-based lesson plan on seventh-grade students' speaking skills.

To begin with, the first section discusses the planning stage, focusing on the Planning integration of AI Avatars into English Language Teaching (ELT). Following this, the second section explains the implementation of speaking instruction based on specific speaking

functions, ranging from Interactional and Transactional to Performance. Ultimately, the final section presents the evaluation of the teaching process conducted throughout the implementation of AI Avatar-mediated instructional implementation in teaching English.

1. Planning the Integration of AI Avatar in ELT

This subchapter explains the planning of AI Avatar integration in English Language Teaching (ELT) for speaking instruction. It is further divided into the following sub-sections: (a) Lesson plans based on AI Avatars, (b) speaking teaching materials based on AI Avatars, and (c) Evaluation of Teaching after the Use of AI Avatars in Classroom Speaking Functions

a) Lesson plans based on AI Avatars

The lesson plan plays a crucial role in structuring the instructional process by guiding teachers in determining learning outcomes and organizing instructional activities (Kaharuddin & Yassi, 2018). A well-designed lesson plan ensures that teaching is systematic, measurable, and pedagogically sound through essential components such as goals, objectives, materials, procedures, evaluation and extra-class work which collectively ensure that the teaching process is systematic, measurable (Brown, 2001).

Grounded in these principles, the lesson plans in this study were designed in alignment with Indonesia's national curriculum, namely the Merdeka Curriculum, in accordance with the Ministry of Education, Culture, Research, and Technology Regulation No. 12 of 2024. This alignment is further reinforced by the Ministry of Primary and Secondary Education Regulation No. 13 of 2025, which mandates the implementation of deep learning approaches starting from the 2025/2026 academic year to move beyond surface learning and foster meaningful understanding and essential competencies.

Deep learning emphasizes conceptual understanding, reflective thinking, and higher-order cognitive skills through Meaningful, Mindful, and Joyful Learning. This approach also fosters key global competencies character, citizenship, collaboration, communication, creativity, and critical thinking as outlined by Fullan et al. (2017), which are essential for 21st-century learning and align with Indonesia's ongoing educational transformation.

b. Speaking teaching materials based on AI Avatars

This study adapted the three-stage instructional framework consisting of Lesson A (Skill Getting), Lesson B (Skill Using), and Lesson C (Review) proposed by Kaharuddin and Yassi (2018).

1) Lesson A : (Skill Getting)

This study focuses on equipping learners with essential linguistic elements required for effective communication. Students are introduced to vocabulary, sentence patterns, grammatical features, and pronunciation drills that align with the topic of the unit. After being exposed to these linguistic forms, students engage in pair practice designed to help them connect the forms with their communicative functions. They then proceed to speaking practice relevant to the current topic to apply the target language in context. In this study, the speaking practice is enhanced through the use of an AI Avatar, which assists students in distinguishing sentence structures, grammar usage, vocabulary selection, and pronunciation more accurately and interactively.

2) Lesson B : (Skill Using)

This is designed to help students transition from merely knowing linguistic forms to being able to use them effectively in spoken communication. This stage consists of two main sub-

activities: functional communication and social interaction. The functional communication activity demonstrates how the vocabulary and expressions learned earlier operate within authentic conversational contexts. It provides learners with realistic situations that incorporate specific language structures, vocabulary items, and pronunciation features that they may naturally encounter when discussing a given topic. Meanwhile, the social interaction activity offers students the opportunity to apply their linguistic knowledge through a semi-scripted role-play. By engaging in this interactive task, learners practice participating in social exchanges related to the topic, using the linguistic forms introduced in the previous lesson

3) Lesson C : (Review)

The final phase of each instructional unit, aimed at determining the extent to which students have internalized the material covered in the preceding lessons. In this study, the review stage incorporates two main activities: language game, intended to assess students' linguistic creativity in an enjoyable, interactive manner while enabling them to practice speaking in authentic communicative situations, and extra-class work, which offers additional exercises either connected to the current topic or serving as preparation for the upcoming lesson. Within the established framework, the review phase typically includes language games and supplementary assignments, all designed to gauge how well students can apply the language features and skills taught earlier. These activities encourage learners to communicate spontaneously and naturally, thereby consolidating their understanding and enhancing their overall communicative ability.

Simulation

As defined by the Oxford Learner's Dictionary (n.d.), simulation is a situation where specific conditions are artificially created to study or experience something that exists in the real world. Furthermore, Carson (2005) expands this definition by viewing simulation as an instrument for measuring the effectiveness of both existing and proposed systems across various configurations over extended periods. In this study, the simulation is conducted to determine whether the teaching module and instructional materials are appropriately aligned and effective for actual classroom implementation.

Prior the actual in the classroom implementation, simulations were conducted to ensure that the teaching process could be implemented effectively and to adjust the time allocation in the lesson plans and teaching materials to fit the research schedule. Two simulation stages were carried out. Simulation 1 was conducted on 18 November 2025 with one student at Perumahan Citra Pandang-Pandang from 16:00 to 18:00 WITA see Apenddix . Simulation 2 was conducted on 22 November 2025 with six students at Puri Diva Istanbul from 15:30 to 17:00 WITA. See appendix

The simulations aimed to test the feasibility of the materials, identify time constraints, and refine the sequencing of activities. Based on the results of the first simulation, adjustments were made to improve time management and instructional flow, which were then applied in the second simulation. The findings indicate that the revisions were effective, as the classroom implementation ran smoothly and remained within the planned time frame. These results suggest that the lesson plans and teaching materials are practical and appropriate for classroom use

***c.* Evaluation of Teaching after the Use of AI Avatars in Classroom Speaking Functions**

The speaking assessments were designed based on the three speaking functions proposed by Richards (2008): Interactional, Transactional, and Performance speaking. Interactional speaking was evaluated through student interviews to explore communicative interaction and learning experiences (Jailani, 2023). Transactional speaking was assessed using role-play activities that simulate real-life communication and promote active language use (Astalini et al., 2021; Marlia et al., 2018). Performance speaking was measured through picture description tasks utilizing visual stimuli (Farih & Dewi, 2021), oral presentations to assess organization, grammar, vocabulary, and fluency (Ati & Parmawati, 2022; Tsang, 2020), and storytelling activities to enhance fluency, creativity, and learner engagement (Ikrammuddin, 2017).

To address the first research question, the evaluation was conducted during the teaching process by applying the Learning level of Kirkpatrick's (1967) evaluation model through pre-test and post-test assessments. The results were analyzed using the CEFR standard, as presented (see Appendix). The pre-test and post-test were fully conducted by an external course institution, and the researcher was not involved in the assessment process. The pre-test took place on Tuesday, 24 November 2025, and the post-test was conducted on 28 November 2025, from 10:00 to end WITA, in the Grade 9 classroom of SMP Tahfidz Mutiara Al-Akbar (see Appendix 6).

2. Implementing AI-based Teaching Materials

The implementation stage, which coincided with the data collection process, was conducted over a one-week period from 24 to 28 November 2025. Within this timeframe, the core instructional sessions were delivered using AI Avatar as a medium for English speaking instruction. These instructional treatments specifically took place over three planned sessions on 25, 26, and 27 November 2025. By utilizing AI Avatar-based materials, three core topics were presented: greeting and saying goodbye, introducing my self and others, and talking about hobbies and interests. Each topic was meticulously designed to address three key speaking functions: interactional, transactional, and performance speaking.

The first and second treatments (25–26 November) were delivered exclusively by the researcher to seven students in Grade 8 class from 09:30 to 11:00 WITA. The third treatment (27 November) involved a collaborative approach between the researcher and the English teacher in Grade 9 class , with five students present from 13:30 to 15:00 WITA. In this final session, the researcher initiated the lesson, followed by the teacher who provided instruction on daily life vocabulary and hobbies using supplementary materials .The teacher's involvement was essential to address the second research question, specifically to identify instructional factors that require teacher presence when integrating AI Avatar as a medium in the classroom. The selected topics were aligned with the Merdeka Curriculum, ensuring they were suitable for the students' linguistic levels and encouraged meaningful language application both inside and outside the classroom

3. Evaluating the Outcomes of Teaching AI Avatar

This subchapter details the evaluation of teaching outcomes following the implementation of AI Avatar as a medium for instruction to ensure the objectivity and professional standards of the data, the assessment process was conducted by an external examiner from the American English School Course her name is Miss Rima ST. The evaluation specifically focuses on the comparative analysis of the results from the Pre-test, administered on November 24, 2025, and the Post-test, conducted on November 28, 2025. The collected

quantitative data were rigorously analyzed using SPSS to provide a clear measure of student improvement in interactional, transactional, and performance speaking.

The subsequent sections present the results of the pre-speaking test, the post-speaking test, and a comparative analysis to determine the students' overall gain. To provide a comprehensive interpretation, the results are categorized based on a synthesized guideline. The proficiency levels and standardized scales are adapted from the CEFR proficiency standards (2001), while the qualitative descriptions (Excellent to Very Poor) are adopted from the classification framework proposed by Kaharuddin and Yassi (2018). This integration, as illustrated in Table 1.1, allows for both a global and a descriptive measure of the students' speaking achievement.

Table 2 Guideline for Evaluating the Students speaking skill

Range Score	Level of Proficiency	CEFR Level	Qualitative Description
3.76-3.90	Advanced	C1	Very good
3.51-3.75	Upper-Intermediate	B2	Good
2.51-3.50	Intermediate	B1	Fair
1.51-2.50	Basic/Elementary	A2	Poor
1.00-1.50	Very low/ Beginner	A1	Very Poor

Source: Adapted from CEFR (2001) for Proficiency Levels and Kaharuddin & Yassi (2018) for Qualitative Descriptions.

1. The Result of the Pre-Speaking Test

In describing the pre-test results, several critical parameters must be clarified in accordance with the framework adopted from Kaharuddin & Yassi (2018**)**. These essential items include the number of participants, the precise schedule, the location, the involvement of reviewers, and the total duration of the assessment.

Following the establishment of the evaluation criteria, the students' initial speaking abilities were measured to determine their baseline proficiency levels. This pre-speaking test was administered on November 24, 2025, from 11:00 a.m. to 12:00 p.m. in the Grade 9 classroom, involving eight students from Grade 7 at SMP Tahfidz Mutiara Al-Akbar Makassar. Conducted by the American English School Course Institution, this assessment took place prior to the implementation of the AI Avatar-based teaching materials to ensure an accurate measure of the students' starting competence.

The pre-speaking test assessed students' abilities across three speaking functions: interactional (conducted through an interview), transactional (executed via a role-play scenario), and performance speaking (evaluated through storytelling, oral presentation, and picture description). To ensure a detailed evaluation, the interviews were conducted with the teacher on a rotational basis. This process generated comprehensive information regarding the students' initial speaking skills, which is subsequently presented in the table below.

Interactional speaking

Analyzing the shift between these scores allows for a definitive evaluation of the training's success in enhancing student competence. The comparison of the gain scores is illustrated in the section below

No	Respondent Number	Interactional Speaking					Total
		Interactive & Responsive	Trun-taking skills	Fluency	Pronunciation	Sociolinguistics accuracy	
1	001	3	2	3	2	2	12
2	002	3	3	3	2	3	14
3	003	3	2	3	2	2	12
4	004	3	3	3	3	2	14
5	005	3	3	3	3	3	15
6	006	3	3	3	3	3	15
7	007	3	2	2	2	2	12
	Average	3,0	2,57	2,86	2,43	2,43	**13,29**

*Source: Primary data processing: pre-test Interactional Speaking***.**

Based on the analysis of the Pre-test data for the Interactional Speaking component, the total scores achieved by the seven (7) respondents ranged from a minimum of eleven (11**)** (achieved by respondent 007) to a maximum of fifteen (15) (achieved by respondents 005 and 006). Overall, the students attained an average score of **13.29** out of a possible maximum of 15. This average score indicates that the students generally possess overall is **Fair** to the commencement of the AI Avatar-based learning intervention, as detailed in the table below.

Table 1.3: The Results of Students' Interactional Speaking Scores and Proficiency Levels

No	Respondent Number	Total Score	Average Score	CEFR Level
1	001	12	2,40	A2
2	002	14	2,80	B1
3	003	12	2,40	A2
4	004	14	2,80	B1
5	005	15	3,00	B1
6	006	15	3,00	B1
7	007	11	2,20	A2
	Average	13,29	**2,66**	**B1**

Source: Primary data processing: pre-test Interactional based on CEFR Level

As shown in the detailed results, the overall average component score of **2.66** places the student group at the B1 (Intermediate) proficiency level. This interpretation is further supported by the individual scores, where a majority of students (4 out of 7) achieved the B1 level, while the remaining three (3) were categorized as A2.

This B1 average suggests that the students are generally able to engage in basic social interactions, respond to routine questions, and maintain simple conversations, thereby demonstrating a strong foundational competence prior to the commencement of the AI Avatar-based learning intervention. However, the varying individual scores indicate that the intervention should also address the specific limitations observed in components such as turn-taking management, pronunciation clarity, and sociolinguistic appropriateness to ensure consistent B1 mastery across the entire group.

The pre-test Interactional speaking test results were analyzed through SPSS, and the analysis is shown in the descriptive statistics table below.

Statistics

Pre-Test Interactional Speaking

N	Valid	7
	Missing	0
Mean		13.29
Median		14.00
Std. Deviation		1.604
Variance		2.571
Minimum		11
Maximum		15

the descriptive statistics of the pre-test Interactional speaking test were further illustrated using a statistical histogram

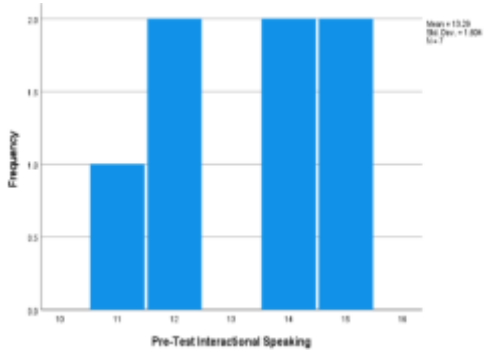

Chart 1. The Pre-Test Score Distribution and Frequency

Post test; Interactional Speaking

Table 2. The Result of Post-test Interactional speaking

No	Respondent Number	Interactional Speaking					Total
		Interactive & Responsive	Trun-taking	Fluency	Pronunciation	Sociolinguistics accuracy	
1	001	4	4	4	3	3	18
2	002	4	3	3	3	3	16
3	003	4	4	4	3	3	18
4	004	4	4	3	3	3	17
5	005	4	4	3	4	4	19
6	006	4	4	4	3	3	18
7	007	3	3	3	3	3	15
	Average	3,86	3,71	3,43	3,14	3,14	**17,29**

Source: Primary data processing: post-test Interactional Speaking.

The analysis of the post-test results for the Interactional Speaking component demonstrates a significant improvement in the students' overall proficiency following the intervention. The students' total scores ranged from a minimum of fifteen (15) (achieved by Respondent 007) to a maximum of nineteen (19) (achieved by Respondent 005), out of a maximum possible score of 20. The group collectively achieved a high average total score of **17.29.** as the total average score in the post test which means that the quality of the students' speaking skill afther getting treatment using the teaching material is **Fair**

gh score indicates a substantial mastery of the interactional skills, providing evidence of the effectiveness of the AI Avatar-based learning material, as detailed in the table below

Table 2.1 The Results of Students' Interactional Speaking Scores and Proficiency Levels

No	Respondent Number	Total Score	Average Score	CEFR Level
1	001	18	3,60	B2
2	002	16	3,20	A2
3	003	18	3,60	A2
4	004	17	3,40	A2
5	005	19	3,80	B1
6	006	18	3,60	A2
7	007	15	3,00	A2
	Average	**17,29**	**3,45**	**B1**

Source: Primary data processing: post-test Interactional e based on CEFR Level

As shown in the detailed results, the overall average component score of **3.45** for the Post-Test places the student group at the **B1 (Intermediate) proficiency level**, corresponding to a "Fair" qualitative description. This interpretation is supported by the individual scores, where a majority of students (**5 out of 7**) were categorized as **A2**, while

the remaining two (**2**) achieved **B1** or **B2**. This B1 average suggests that students are now generally able to handle complex interactions, respond appropriately, and maintain conversation flow with greater confidence compared to the pre-test, thus demonstrating a significant improvement following the intervention. However, the varying individual scores especially the fact that five students remain categorized as A2 indicate that the intervention still needs to ensure consistent mastery across the entire group to solidify the B1 level and move students toward the B2 level.

Post-test scores for interactional speaking, gathered from the Interview task, were analyzed with SPSS, and the outcomes are shown in the table of descriptive statistics below

Statistics

Post-Test Interactional Speaking

N	Valid	7
	Missing	0
Mean		17.29
Median		18.00
Std. Deviation		1.380
Variance		1.905
Minimum		15
Maximum		19

The post-test interactional speaking scores collected from the interview activity were further depicted using a statistical histogram.

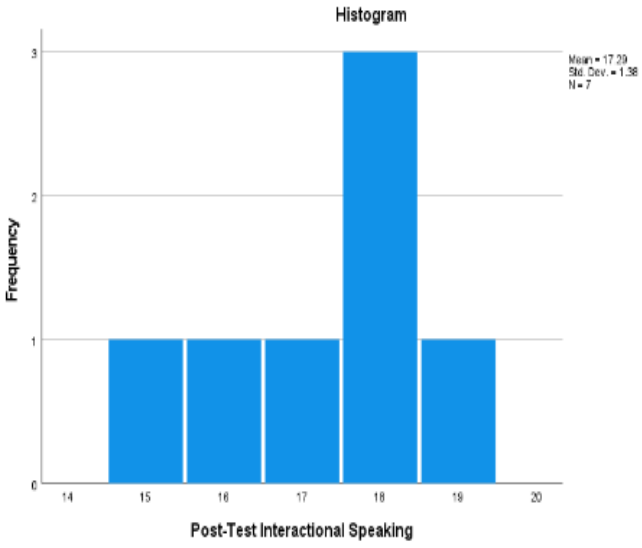

Chart 2. The Pre-Test Score Distribution and Frequency

2. The Overall Results of Pre-Speaking Test and Post-Speaking Test

The overall results of the speaking test provide a comprehensive overview of the comparisons between the various assessment phases. These pre-tests and post-tests are specifically designed to measure the knowledge gained through the intervention of

instructional materials developed from the initial needs analysis. In alignment with the framework proposed by Richards (2008), the evaluation covers three fundamental functions of speaking: interactional speaking, transactional speaking and performance speaking , While the pre-test establishes the participants' baseline proficiency across these categories before the treatment, the post-test evaluates their progress upon completion of the course.

1. **Interactional speaking**

Analyzing the shift between these scores allows for a definitive evaluation of the training's success in enhancing student competence. The comparison of the gain scores is illustrated in the section below

Table 1.9 Test Scores Comparison Between Pre-Test And Post-Test of Interactional speaking

No	Respondent Number	Interactional Speaking Component	
		Pre-Test	Post-Test
1	001	12	18
2	002	14	16
3	003	12	18
4	004	14	17
5	005	15	19
6	006	15	18
7	007	11	15
	Average	2,66	3,45
	Description	B1	B1

(Source: Primary data processing Interactional Speaking)

The table shows that a significant improvement takes place in the students' scores before and after treatment, which means that the students achieved a measurable enhancement in their speaking skills after learning with the teaching materials developed in this study. The improvement can be observed by looking at the minimum and the maximum scores in the pre-test and the post-test, in which **eleven (11)** is the minimum score in the pre-test while **fifteen (15)** is the minimum score in the post-test; besides, **fifteen (15)** is the maximum score in the pre-test, while **nineteen (19)** is the maximum score in the post-test. In addition, the students' progress in learning can also be measured by comparing the students' average component score in the pre-test of **2.66** and the post-test score of **3.45**. If these two scores are taken into account regarding the students' speaking quality before and after treatment, it can be stated that the **quality of the students' speaking skills significantly**

increases from "Fair" to "Good," proving that the intervention successfully improved their interactional competence

The data presented above were further processed using SPSS to generate box plots, which illustrate the distribution of scores. These box plots provide a clear visualization of the significant improvement in students' speaking proficiency by comparing their performance before (pre-test) and after the intervention (post-test)

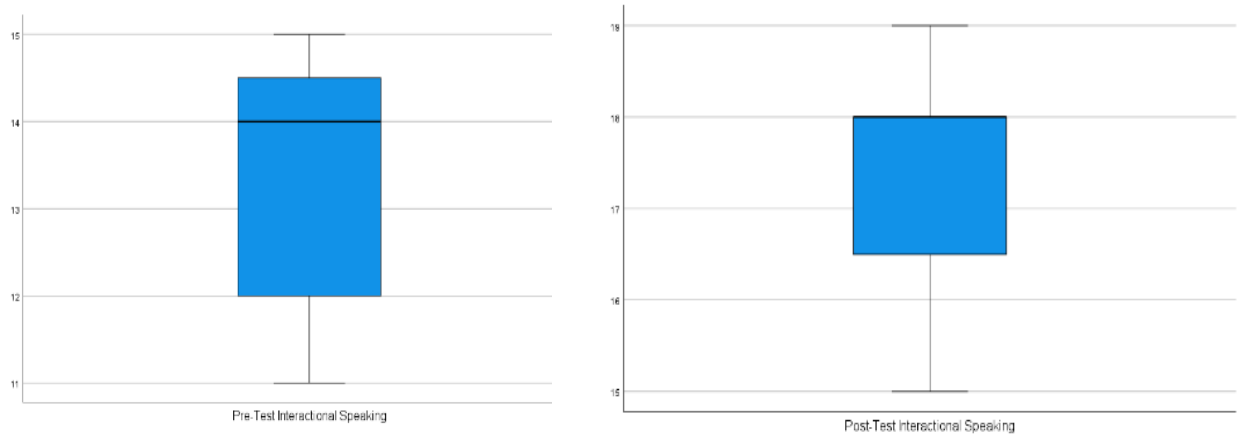

Chart 3.1 The comparison between pre-test and post test scores distribution

Assumption Test

Before conducting the hypothesis testing, it is essential to evaluate the sample distribution to determine whether the data is normal or non-normal. This procedure is commonly referred to as a **normality test**. These tests are designed to validate the assumptions required for hypothesis testing by comparing the null hypothesis (H_0) against the alternative hypothesis (H_1). The decision-making criteria for the normality test are defined as follows:

H_0 = (The Null Hypothesis)

The sample follows a normal distribution if the significance value is greater than 0.05 ($p > 0.05$).

H_1 = (The Alternative Hypothesis)

The sample is not normally distributed if the significance value is less than 0.05 ($p < 0.05$)

To assess the normality of the sample distribution, the Kolmogorov-Smirnov statistical test was employed in this study as follows

Table 1.10 One-sample Kolmogorov-semirnov Test Interactional Speaking

One-Sample Kolmogorov-Smirnov Test

			Pre-Test Interactional Speaking	Post-Test Interactional Speaking
N			7	7
Normal Parameters[a,b]	Mean		13.29	17.29
	Std. Deviation		1.604	1.380
Most Extreme Differences	Absolute		.243	.269
	Positive		.217	.160
	Negative		-.243	-.269
Test Statistic			.243	.269
Asymp. Sig. (2-tailed)[c]			.200[d]	.135
Monte Carlo Sig. (2-tailed)[e]	Sig.		.246	.131
	99% Confidence Interval	Lower Bound	.235	.122
		Upper Bound	.257	.139

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

d. This is a lower bound of the true significance.

e. Lilliefors' method based on 10000 Monte Carlo samples with starting seed 2000000.

Based on the statistical analysis presented in Table 29, it is evident that both the pre-test and post-test data follow a normal distribution. Specifically, the **pre-test** yielded a significance value of **0.246**, while the **post-test** resulted in a significance value of **0.131**. Since both values are greater than the 0.05 threshold (**Sig. > 0.05**), there is insufficient evidence to reject the null hypothesis, confirming that the sample data is **normally distributed**. Consequently, because the normality assumption has been fulfilled, the data is eligible for further parametric statistical testing to evaluate the research hypothesis

Paired Sample T-Test

After finding out that the samples of the pre-test and post-test are normally distributed, an Independent T-Test is administered to evaluate the impact of the developed instructional materials on the students' speaking proficiency. This statistical analysis aims to test the research hypothesis as follows:

H_0 (The Null Hypothesis)

There is no significant improvement in the speaking skills of students who utilized the 'Speaking Course One' instructional materials at a significance level of $\alpha = 0.05$, indicating that the treatment has no significant influence on their achievement.

H_1 (The Alternative Hypothesis)

There is a significant improvement in the speaking skills of students who utilized the 'Speaking Course One' instructional materials at a significance level of $\alpha = 0.05$, demonstrating that the treatment has a significant influence on their achievement

The null hypothesis is rejected if the p-value falls below 0.05, confirming that the treatment has a significant impact on students' speaking performance. This determination is made using an independent t-test, which serves as the statistical basis for accepting or rejecting the research hypothesis

Table 5.3. Paired Samples T-Test of Speaking Interactional speaking

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Pre-Test Interactional Speaking - Post-Test Interactional Speaking	-4.000	1.528	.577	-5.413	-2.587	-6.928	6	<,001

With regard to the table above, the SPSS output for the paired samples t-test shows a t-value of **-6.928** v of **< .001**. This output reveals that the significance value is much lower than the standard alpha level ($p < 0.$

This result implies that the treatment has a highly significant effect on the students' speaking skills achiev less than 0.05, the **null hypothesis is therefore rejected**, and the alternative hypothesis is accepted. Th statistically significant improvement in speaking performance following the intervention.

CONCLUSION

The findings show that AI Avatar-based instruction has a positive effect on students' interactional speaking skills. The average pre-test score was 2.66 and increased to 3.45 in the post-test, indicating a meaningful improvement after the implementation of AI Avatars. This result demonstrates that the use of AI Avatars helped students perform better in interactional communication. Although the overall CEFR level remained at B1, students showed clear qualitative improvement in their interactional speaking performance. They became more confident, fluent, responsive, and more accurate in using language appropriately in social interactions, suggesting that AI Avatars function effectively as a supportive medium for developing interactional speaking skills in EFL classrooms.

REFERENCES

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. AW Longman.

Cambridge University Press & Assessment. (n.d.). Effect. In *Cambridge Dictionary*. Retrieved January 3, 2026, from https://dictionary.cambridge.org/dictionary/english/effect

Carson, J. S. (2005, December). Introduction to modeling and simulation. In *Proceedings of the Winter Simulation Conference, 2005* (pp. 8-pp). IEEE.

Chiu, T. K. (2021). A holistic approach to the design of artificial intelligence (AI) education for K-12 schools. *TechTrends*, *65*(5), 796-807.

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, *210*, 104967.

Fullan, M., Quinn, J., & McEachen, J. (2017). *Deep learning: Engage the world change the world*. Corwin Press.

Ghafoor, A., & Younus, N. (2025). Unconscious applications of NLP in ELT: An exploratory study of Pakistani teachers' beliefs and practices. *International Premier Journal of Languages & Literature*, *3*(4), 291-308.

Huang, M. (2025). Student engagement and speaking performance in AI-assisted learning environments: A mixed-methods study from Chinese middle schools. *Education and Information Technologies*, *30*(6), 7143-7165.

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, *7*(4), 396.

Kaharuddin, A., & Yassi, A. (2018). *Syllabus design for English language teaching*. Indonesia Trust Media Publishing.

Kang, H. (2022). Effects of Artificial Intelligence (AI) and Native Speaker Interlocutors on ESL Learners' Speaking Ability and Affective Aspects. *Multimedia-Assisted Language Learning*, *25*(2).

Kirkpatrick, D. L., Craig, R. L., & Bittel, L. R. (1967). Evaluation of training. *Evaluation of Short-Term Training in Rehabilitation*, 35.

Krashen, S. (1982). *Principles and practice in second language acquisition*.

Larsen-Freeman, D. (2000). *Techniques and principles in language teaching*. Oxford University Press.

Levy, M. (2009). Technologies in use for second language learning. *The Modern Language Journal*, *93*, 769-782.

Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press.

Met, M. (1991). Learning language through content: Learning content through language. *Foreign Language Annals*, *24*(4), 281-295.

Nunan, D. (2003). *Practical English language teaching*. McGraw-Hill.

Oxford University Press. (n.d.). Effect. In *Oxford Learner's Dictionaries*. Retrieved January 3, 2026, from https://www.oxfordlearnersdictionaries.com/definition/english/effect_1

Oxford University Press. (n.d.). Factor. In *Oxford Learner's Dictionaries*. Retrieved January 3, 2026, from https://www.oxfordlearnersdictionaries.com/definition/english/factor_1

Oxford University Press. (n.d.). Simulation. In *Oxford Learner's Dictionaries*. Retrieved January 3, 2026, from https://www.oxfordlearnersdictionaries.com/definition/english/simulation

Richards, J. C. (2008). *Teaching listening and speaking* (Vol. 35). Cambridge University Press.

Richards, J. C., & Farrell, T. S. (2005). *Professional development for language teachers: Strategies for teacher learning*. Cambridge University Press.

Romundza, F., Afrida, A., Asmiyunda, A., Hendra, H., Siregar, D. R., & Ulfah, F. (2025). Implementasi Synthesia AI dalam pengembangan video pembelajaran animasi untuk meningkatkan kompetensi literasi digital guru. *Jurnal Pengabdian Masyarakat Khatulistiwa*, *8*(2), 170-179.

Shoei, Y., Shuhei, N., Peeraya, S., & Hiroaki, S. (2025). Effects of Avatar Identity (AI or Human) and Conversation on Personal Space in Virtual Reality. *International Journal of Asia Digital Art and Design*, *29*(S1), 29-37.

Underhill, N. (1995). *Testing spoken language: A handbook of oral testing techniques*. Klett Ernst/Schulbuch.

Wiederhold, B. K. (2023). Treading carefully in the metaverse: The evolution of AI avatars. *Cyberpsychology, Behavior, and Social Networking*, *26*(5), 321-322.

Wright, A. (1989). *Pictures for language learning*. Cambridge University Press.

Zou, D., Huang, Y., & Xie, H. (2021). Digital game-based vocabulary learning: Where are we and where are we going? *Computer Assisted Language Learning*, *34*(5-6), 751-777.

Zuhriyah, M. (2017). Storytelling to improve students' speaking skill. *English Education: Jurnal Tadris Bahasa Inggris*, *10*(1), 119-134.