



Enhancing Elementary School Students' Speaking Skills through Self-Recorded Video-Assisted Deep Learning

Nafisa Nurfatin^{1*}, Yunus Abidin¹, Dede Margo Irianto¹, Siti Nasuhi Sungki²

¹Universitas Pendidikan Indonesia, Indonesia

²Universiti Malaysia Sarawak, Malaysia

*Corresponding Author's Email: nafisanurfatin24@upi.edu

How to cite: Nurfatin, N., Abidin, Y., Irianto, D. M., & Singki, S. N. (2026). Enhancing Elementary School Students' Speaking Skills through Self-Recorded Video-Assisted Deep Learning. *Profesi Pendidikan Dasar*, 13(2), 158–176. <https://doi.org/10.23917/ppd.v13i2.18222>

Abstract

Keywords:

deep learning;
self-recorded video;
speaking skills;
elementary students

Article History:

Submitted: 2026-06-10
Revised: 2026-08-05
Accepted: 2026-08-22

Speaking instruction in elementary schools often emphasizes performance outcomes while providing limited opportunities for students to reflect on their speaking process. Although video-assisted learning has been widely used, limited research has integrated a Deep Learning Approach with Self-Recorded Video to support reflective speaking development in elementary education. This study aimed to analyze the effect of a Deep Learning Approach assisted by Self-Recorded Video on elementary students' Speaking Skills and explain their learning experiences. An explanatory sequential mixed-methods design employed a Nonequivalent Control Group Design involving 49 sixth-grade students, with 25 in the experimental group and 24 in the control group. Data were collected through speaking tests, Self-Recorded Video documentation, and semi-structured interviews. Quantitative data were analyzed using descriptive statistics, N-Gain, Mann-Whitney U Test, and effect size, while interviews with three students explained the quantitative findings. The experimental group improved from 9.80 to 16.68 with an N-Gain of 69.94%, while the control group improved from 9.88 to 13.29 with an N-Gain of 31.86%. The posttest difference was significant ($U = 111.500$; $Z = -3.804$; $p < 0.001$; $r = 0.543$). These findings indicate that integrating Deep Learning and Self-Recorded Video supports reflective practice and improves speaking performance.

INTRODUCTION

Background of the study

Speaking skills are one of the key competencies in Indonesian language instruction in elementary school because they relate to students' ability to express ideas, experiences, opinions, and information orally (Pituxcoosuvann et al., 2024). Mastery of speaking skills is not only related to students' confidence in speaking, but also encompasses the ability to convey messages clearly, use appropriate vocabulary, construct correct language structures, and maintain fluency when communicating. These skills are necessary when students answer questions, engage in discussions, share experiences, present their work, or express their opinions in front of others. Therefore, Indonesian language instruction must provide experiences that allow students to actively use the language in various communicative situations. Developing speaking skills starting in elementary school also supports students' ability to communicate, collaborate, and convey ideas effectively in both educational settings and daily life through oral communication (Pituxcoosuvann et al., 2024). The use of activities and media that provide

© The Author(s). 2026



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

stimulation and opportunities for practice can also support the development of speaking skills at the elementary school level (Septiani et al., 2025; Wahyudi, 2025)

Developing speaking skills requires a learning experience that gives students the opportunity to understand, apply, and reflect on the knowledge they have acquired (Fitriati et al., 2026). Speaking skills enable students to connect knowledge with experience, develop ideas, and express their thoughts through oral language (Lai, 2024). This process requires student engagement throughout learning so that students can understand communication objectives and prepare the content of their speech in a structured manner. Learning that connects the material with situations close to students' lives can also provide a more meaningful learning experience (Mandalika, 2025). In this context, the Deep Learning Approach provides a pedagogical framework through the principles of mindful learning, meaningful learning, and joyful learning, which position student engagement as an important part of the learning process (Sari & Arta, 2025). These three principles provide opportunities for students to understand the material, connect it with their experiences, use it in communication activities, and reflect on the learning process they have undertaken.

The development of learning technology provides diverse media options to support more varied learning experiences, with video offering particular potential for Indonesian language instruction because it simultaneously presents visual and auditory information (Erlitaviana & Tyas, 2025; Lavrishcheva et al., 2026). Through audiovisual presentation, students can observe examples of language use, message delivery, and communication contexts, while also reviewing specific sections to strengthen their understanding of the material (Noetel et al., 2021). In Speaking Skills instruction, these features can help students understand effective ways of conveying ideas before engaging in speaking practice. This potential can be further enhanced through Self-Recorded Video, which requires students to record their own speaking performances based on assigned tasks, shifting learning from merely observing media to actively producing content (Pertiwi, 2025). During this process, students prepare ideas, select relevant information, and practice before recording, while the resulting videos can be reviewed to evaluate pronunciation, vocabulary, language structure, and speaking fluency (Purba et al., 2024; Rohdina & Priyana, 2025). Thus, Self-Recorded Video can foster a more active learning experience by engaging students in the preparation, production, and reflection of their speaking performances.

Problem of the study

The Merdeka Curriculum directs learning toward active student engagement and provides opportunities for students to develop communication skills according to their learning needs. In Indonesian language instruction, Speaking Skills play an important role because students need to convey ideas, answer questions, participate in discussions, tell stories, and express opinions orally. However, classroom conditions indicate that students' Speaking Skills have not yet fully developed in accordance with these demands. Some students still experience difficulties when expressing ideas orally, particularly in maintaining speaking fluency, using clear pronunciation, constructing sentences, selecting vocabulary, and conveying messages so that they can be easily understood. This situation indicates a problem in students' ability to use language orally that requires attention in Indonesian language instruction.

This situation is reinforced by the results of a preliminary survey conducted with 25 students using five Speaking Skills indicators, namely fluency, pronunciation, grammar, vocabulary, and comprehensibility. The preliminary survey showed that 20% of students were in the low category, 72% were in the moderate category, 8% were in the good category, and no students reached the very good category. Thus, 92% of students were still in the low to moderate categories, indicating that most students had not demonstrated Speaking Skills at the expected level. This condition was evident when students conveyed stories or ideas, both in terms of speaking fluency, clarity of pronunciation, accuracy of grammar use, vocabulary selection, and comprehensibility of the messages conveyed. These initial findings provide empirical evidence that Speaking Skills problems were directly identified among the students who served as the research subjects.

This condition is consistent with Saputro et al. (2023), who found that 61.3% of sixth-grade elementary school students had not reached the minimum competency criterion of 77 in public

speaking, with low self-confidence identified as one contributing factor. Difficulties in Speaking Skills may limit students' participation in classroom communication, as problems in organizing ideas, vocabulary, grammar, pronunciation, and comprehensibility can hinder their ability to respond, express opinions, and convey information effectively, potentially reducing their confidence and opportunities for direct communication practice. Therefore, Speaking Skills instruction needs to provide sufficient opportunities for students to prepare ideas, practice, deliver messages, and review their performances. However, classroom constraints such as class size, limited instructional time, and differences in students' readiness may restrict repeated speaking practice, highlighting the value of media that enables more flexible preparation and practice before direct Speaking Performance. In this context, Self-Recorded Video allows students to prepare speech content, practice, record their performances, and review the resulting recordings, while Kang (2022), in a study involving 48 university students in an ESP course, linked self-recorded video-speaking tasks with self-regulated learning through recording, interaction, and reflection; although the study involved a different educational context and level, it provides a basis for incorporating video recording into speaking activities that emphasize practice and reflection.

In addition to providing opportunities for practice, the learning process also needs to provide experiences that help students understand the purpose of activities, connect the material with their experiences, and reflect on what they have done. In the educational context, the Deep Learning Approach positions mindful learning, meaningful learning, and joyful learning as components of a learning experience that encourages student engagement, meaningful understanding of the material, and reflection throughout the learning process. Nafi'ah & Faruq (2025), in their study of the Deep Learning Approach in primary education, explained that the integration of mindful, meaningful, and joyful learning is associated with the development of student engagement and contextual and reflective learning experiences. Andayanie et al. (2025) and Asadollahpour et al. (2026) also positioned Deep Learning as a pedagogical framework directed toward holistic, reflective learning experiences that are aligned with the needs of twenty-first-century learning. In Speaking Skills instruction, this framework can provide opportunities for students to understand communication objectives, prepare ideas, use language in meaningful activities, and reflect on their speaking experiences.

Research State of the Art

Speaking Skills are an important ability in elementary language instruction because through these skills students learn to convey ideas, experiences, information, and opinions orally. This ability is not only related to students' willingness to speak but also to the quality of message delivery during communication. Speaking Skills are understood through five aspects, namely fluency, pronunciation, grammar, vocabulary, and comprehensibility (Bayu et al., 2026; Brown, 2004). Fluency refers to students' ability to convey ideas smoothly, pronunciation concerns clarity of articulation, while grammar concerns the accuracy of the language structures used. Meanwhile, vocabulary concerns the ability to select appropriate words, and comprehensibility concerns the extent to which messages are understood by listeners. These five aspects provide a more comprehensive picture of the quality of students' oral performance; therefore, the development of Speaking Skills should be directed not only toward the frequency with which students speak but also toward the quality of the language used during communication.

The development of Speaking Skills in recent years has increasingly been associated with the use of digital technology, which provides stimuli, communication models, and opportunities for students to produce language orally. Ramalingam et al. (2022) examined the use of digital technology in learning related to communication skills, while Santhanasamy and Yunus (2022) demonstrated the use of flipped learning and Blendspace in supporting learners' Speaking Skills. Nair and Yunus (2021) also examined the use of digital technology in language and communication learning activities. More recently, Almulla et al. (2026) examined a digital storytelling intervention to improve elementary-level students' speaking abilities. Findings from these studies indicate a tendency for digital technology to function not only as a source of learning materials but also as an environment that allows students to obtain examples, develop ideas, and use language in communication activities. Although the contexts, educational levels, and types of technology used differ, these studies provide a basis for positioning digital media more closely within students' language production and speaking performance.

Video supports speaking instruction by simultaneously presenting visual and auditory information, enabling students to observe language use, message delivery, communication contexts, and opportunities for repeated practice. Dzakiah et al. (2020) found that smartphone video recording was more effective than in-class monologue presentations in improving EFL students' speaking performance, while students also showed high interest in using video recording for speaking instruction, suggesting that recording technology can function not only as a medium for documenting learning outcomes but also as a tool for preparing and demonstrating speaking abilities. Similarly, Dang et al. (2022), in a study of 26 university students divided into experimental and control groups, found that students who used video recording achieved higher speaking performance and expressed positive attitudes toward completing speaking tasks through video recording. Further evidence was reported by Belmekki (2023), whose quasi-experimental pretest-posttest study found that self-video recording contributed 6.43% to the development of learners' speaking ability, providing empirical support for its use as a learning practice. Moreover, Belmekki (2023) also found that learners held positive attitudes toward using self-video recording in speaking instruction.

The use of Self-Recorded Video at an educational level closer to the context of elementary education has also begun to receive attention. Beltran-Palanques (2024) specifically examined Self-Recorded Video in Promoting Students' Speaking Ability among seventh-grade students. The study involved 28 students and used a one-group pretest-posttest design. The study aimed to measure the effectiveness of Self-Recorded Video as a learning medium in improving students' speaking ability. This study is important because it demonstrates that the use of Self-Recorded Video is not limited to higher education contexts but can also be applied to learners at primary and secondary educational levels. Based on these studies, the function of video in speaking instruction has developed. Initially, video recording was used primarily as a means of practicing and documenting speaking performance. Parmaxi (2023) showed that recordings produced by students could be used to observe developments in pronunciation, vocabulary, grammar, and fluency. Subsequently, Dang et al. (2022) demonstrated that the use of video recording could support speaking achievement and received positive responses from learners. Belmekki (2023) then provided more specific evidence regarding the effect of self-video recording on speaking ability, while Beltran-Palanques (2024) extended its application to seventh-grade students. Thus, previous studies have provided a basis for the potential of recording speaking performance to support learners' oral development.

Gap Study and Objective

Previous studies have shown that video technology has been used in language instruction to support speaking activities through various forms of intervention. Santhanasamy and Yunus (2022) used Blendspace in a flipped learning approach and demonstrated improvements in Speaking Skills among elementary school students. Almulla et al. (2026) also examined the use of digital storytelling to develop elementary school students' speaking abilities in English language instruction. Both studies demonstrate that digital technology has been used in speaking instruction at the elementary level, indicating that elementary school students are not an entirely unexplored context. At higher educational levels, Ferng and Hwang (2026) used student-produced videos in Japanese language learning and associated them with oral performance, self-efficacy, and learning satisfaction. Al-Hawamleh et al. (2022), also used video recording and peer observation in university students' oral presentation instruction. These findings indicate that video has an empirical basis for supporting the practice, documentation, and evaluation of speaking performance across various learning contexts.

Although research on video and speaking ability among elementary school students has been conducted, the forms and functions of video used remain diverse. Santhanasamy and Yunus (2022) positioned digital technology within a flipped learning approach, whereas Almulla et al. (2026) used digital storytelling as a speaking learning activity. Both studies provide evidence regarding the use of digital technology to support elementary school students' speaking abilities but did not specifically position student-produced recordings as part of a cycle of speaking practice, review, reflection, and performance improvement. Ferng and Hwang (2026) provide a context closer to student-produced video, but the study was conducted with university students learning Japanese. Al-Hawamleh et al. (2022), also demonstrated the potential of video recording to help learners observe oral performance,

but the context involved university-level presentation instruction. These differences indicate that the use of video in speaking instruction has different mechanisms depending on its function, educational level, and language-learning context.

Research on Self-Recorded Video provides a more specific basis for the use of self-recording in speaking instruction. Parmaxi (2023) examined the use of Self-Recorded Video in English language learning by considering pronunciation, vocabulary, grammar, and fluency. The study provides relevant evidence that self-recording can be used as part of speaking practice, but its participants were in higher education. Belmekki (2023) also examined the use of self-video recording in speaking instruction and demonstrated positive learner responses to the activity, but the research focused on attitudes toward the use of self-video recording. Ferng and Hwang (2026) expanded research on student-produced video through Japanese language learning among university students and reported its association with oral performance. These studies reinforce the potential of student-produced video, but the application of Self-Recorded Video among elementary school students in Indonesian language instruction remains limited and has not been extensively examined through the Deep Learning Approach as a pedagogical framework.

Based on the research gap, this study aimed to analyze the effect of the Deep Learning Approach assisted by Self-Recorded Video on elementary school students' Speaking Skills in Indonesian language instruction. Specifically, this study analyzed students' speaking ability based on five aspects: fluency, pronunciation, grammar, vocabulary, and comprehensibility. This focus was directed toward obtaining empirical evidence of changes in the quality of speaking performance after students participated in learning that integrated the Deep Learning Approach with Self-Recorded Video. The learning process included idea preparation, speaking practice, performance recording, recording review, reflection, and performance improvement. Through this sequence, the study was expected to provide empirical evidence regarding the implementation of an integrated pedagogical approach and self-recording in Speaking Skills instruction among elementary school students.

METHOD

Type and Design

This study employed a mixed-methods approach using an explanatory sequential design, in which quantitative and qualitative stages were conducted sequentially: the quantitative stage measured changes in students' Speaking Skills through pretests and posttests, while the subsequent qualitative stage explored students' experiences to explain the quantitative findings, consistent with the characteristics of explanatory sequential research (Draucker et al., 2020; McCrudden & McTigue, 2019). The quantitative stage used a quasi-experimental Nonequivalent Control Group Design involving an experimental group of 25 sixth-grade students from School A and a control group of 24 sixth-grade students from School B. The two schools were selected based on comparable characteristics, including students' grade level, general learning conditions, curriculum implementation, and availability of Indonesian language instruction, while both schools also demonstrated sufficient readiness to participate in the study. School A was assigned as the experimental group because it met the predetermined criteria and was available to implement the Deep Learning Approach assisted by Self-Recorded Video, whereas School B was assigned as the control group because it provided comparable learning conditions and continued the existing instructional approach without Self-Recorded Video. Both groups completed pretests and posttests without individual randomization. This design enabled comparison of changes in students' speaking ability between the treatment and comparison groups and is consistent with its use in educational research where participant randomization is not conducted (Ardhiati et al., 2021).

The quantitative stage followed a pretest–treatment–posttest sequence, with both groups completing one-hour pretests offline; the experimental group then received online instruction over five days through task instructions, practice, recording, and reflection, followed by an online posttest, while the control group completed a one-hour offline posttest. The posttest used a picture sequence with a different theme but comparable task demands to the pretest to minimize the possibility of students simply reproducing the initial story. After the quantitative results were analyzed, the qualitative stage involved three experimental-group students selected according to low, moderate,

and high N-Gain categories to represent different levels of improvement. Interviews explored students' experiences in understanding the tasks, practicing, using Self-Recorded Video, and reflecting on their speaking performance, with the findings subsequently used to explain patterns and differences identified in the quantitative results. This integration, in which quantitative findings guide participant selection for the subsequent qualitative stage, is consistent with the connecting process characteristic of explanatory sequential mixed-methods research (Draucker et al., 2020).

Data and Data Sources

The research involved 49 sixth-grade students, comprising 25 students in the experimental group (12 male and 13 female) and 24 students in the control group (11 male and 13 female), with a total of 23 male and 26 female participants aged 11–12 years from public and private schools; the experimental and control schools were anonymized as School A and School B, respectively, and participants were not individually randomized. The primary quantitative data consisted of students' Speaking Skills scores from pretests and posttests, assessing five aspects—fluency, pronunciation, grammar, vocabulary, and comprehensibility—using a 1–4 scale based on the study's assessment rubric. Qualitative data were obtained through interviews with three experimental-group students selected after quantitative analysis based on low, moderate, and high N-Gain categories to explore their experiences in understanding tasks, practicing speaking, using Self-Recorded Video, and reflecting on their performances. Supporting data consisted of Self-Recorded Video documentation produced during the treatment, which demonstrated students' speaking activities and supported reflection on strengths and areas for improvement but did not replace the pretest and posttest scores as the primary measure of speaking ability. Thus, the data sources comprised 49 student participants, pretest and posttest scores as the primary quantitative data, interviews with three selected students as qualitative data, and Self-Recorded Video documentation as supporting data.

Data Collection Techniques

Data collection was conducted sequentially through Speaking Skills tests, Self-Recorded Video documentation, and semi-structured interviews. The speaking tests provided pretest and posttest data for both groups, with the experimental group completing a one-hour offline pretest followed by a five-day online treatment and an online posttest, while the control group completed both tests offline for one hour each. The tests used storytelling tasks based on picture sequences, in which students orally constructed stories from the sequence of events; the same assessment procedure was applied to both groups, while the posttest used a different but comparable theme to minimize the possibility of recalling or repeating the pretest story. Speaking Skills were assessed in five aspects: fluency, pronunciation, grammar, vocabulary, and comprehensibility, each scored on a 1–4 scale according to the research rubric, with higher scores indicating better performance.

After the pretest, the experimental group received Indonesian language instruction using the Deep Learning Approach assisted by Self-Recorded Video through a five-day online learning sequence consisting of four stages: task instruction, practice, recording, and reflection. Students first received instructions about the speaking task, then understood the picture sequence, organized ideas, and practiced storytelling before recording their own speaking performance using Self-Recorded Video. They subsequently reviewed their recordings to identify strengths and areas requiring improvement, creating a learning sequence that connected task understanding, practice, performance, and reflection. In contrast, the control group received comparative instruction without Self-Recorded Video, completing a one-hour offline pretest and posttest. The posttest used a different but comparable picture-sequence theme to minimize recall of the pretest story, and students' performances in both groups were assessed using the same five Speaking Skills aspects: fluency, pronunciation, grammar, vocabulary, and comprehensibility.

Semi-structured interviews were conducted after the quantitative data had been collected and analyzed. Three students from the experimental group were selected based on N-Gain categories, namely students with low, moderate, and high improvement. The interviews focused on students' experiences in understanding the task or material, practicing, using Self-Recorded Video, and reflecting on the recording results. The interview questions were used to obtain explanations of learning

experiences that could help explain differences in speaking ability improvement identified in the quantitative data. With this sequence, data collection proceeded from the pretest, treatment and instruction, posttest, quantitative analysis, and interviews as the qualitative stage.

Data Analysis

Quantitative data were analyzed to describe initial ability, changes in scores, and improvement in Speaking Skills in the experimental and control groups. Descriptive analysis was used to present the characteristics of pretest and posttest scores in both groups. The descriptive statistics included the mean, standard deviation, minimum score, and maximum score according to the needs of the research data presentation. Subsequently, speaking ability improvement was calculated using N-Gain based on the pretest score, posttest score, and maximum possible score. N-Gain values were used to classify students' levels of improvement into low, moderate, and high categories and served as the basis for selecting three interview participants. The use of N-Gain as a measure of changes in learning outcomes has also been employed in educational research to describe improvement in ability following treatment (Rofii & Syarifah, 2024).

Before inferential analysis was conducted, the data distribution was examined to determine the appropriate statistical techniques. A normality test was used to determine whether the data met the assumption of a normal distribution, while a homogeneity test was used to examine the equality of variances between groups when parametric analysis was used. Because the study involved experimental and control groups that were not formed through randomization, comparisons of changes in ability were analyzed by considering scores from both groups. When the data met parametric assumptions, differences in improvement between groups were analyzed using an appropriate parametric test, whereas data that did not meet parametric assumptions were analyzed using an appropriate nonparametric alternative. The selection of inferential techniques was based on the two-group data structure and measurements obtained before and after instruction rather than on a one-group design.

Interview data were analyzed descriptively by grouping information according to students' experiences during the learning process. The analysis focused on students' experiences in understanding the task or material, applying their understanding through speaking practice, using Self-Recorded Video, and reflecting on the recording results. Information from the three students was compared to identify similarities and differences in experiences across low, moderate, and high improvement categories. The interview findings were then connected with N-Gain results and quantitative findings from the experimental group. Thus, qualitative data were used to explain quantitative patterns, particularly learning experiences associated with variations in Speaking Skills improvement. The use of quantitative data to determine interview participants and the use of interview findings to clarify quantitative results constitute an integration process consistent with the explanatory sequential design.

RESULTS

Description of Students' Speaking Ability

Descriptive analysis was conducted to provide an initial overview of students' speaking ability in the experimental and control groups before and after instruction. The data presented included the number of students (N), mean, median, mode, standard deviation, minimum value, and maximum value. The results of the descriptive analysis are presented in Table 1.

Table 1. Descriptive Statistics of Students' Speaking Ability

Group	Test	N	Mean	SD	Min.	Max.
Experimental	Pretest	25	9.80	2.986	6	16
Experimental	Posttest	25	16.68	2.155	14	20
Control	Pretest	24	9.88	1.801	7	13
Control	Posttest	24	13.29	2.851	8	20

Based on Table 1, the mean pretest score of the experimental group was 9.80, while that of the control group was 9.88. These results indicate that the initial abilities of the two groups were relatively similar. After instruction, the mean score of the experimental group increased to 16.68, whereas that of the control group increased to 13.29. Thus, the increase in the mean score was greater in the experimental group than in the control group. In the experimental group, the minimum and maximum posttest scores were 14 and 20, respectively, whereas those of the control group were 8 and 20.

Description of Speaking Ability Scores by Indicator

Descriptive statistical analysis was conducted to provide an overview of students' speaking ability in the experimental and control groups before and after instruction. Speaking ability was assessed based on five indicators: fluency, pronunciation, grammar, comprehensibility, and vocabulary. The mean and standard deviation of each indicator are presented to describe the condition of speaking ability in each group.

Table 2. Descriptive Statistics of Students' Speaking Ability by Indicator

Indicator	Experimental	Experimental	Control	Control
	Pretest	Posttest	Pretest	Posttest
	M	SD	M	SD
Fluency	2.00	.82	3.16	.55
Pronunciation	2.04	.84	3.32	.48
Grammar	1.76	.66	3.28	.54
Comprehensibility	1.96	.74	3.40	.50

Based on Table 2, the experimental group showed an increase in the mean scores of all speaking ability indicators after instruction. The mean fluency score increased from 2.00 at pretest to 3.16 at posttest. Pronunciation increased from 2.04 to 3.32, grammar from 1.76 to 3.28, comprehensibility from 1.96 to 3.40, and vocabulary from 2.04 to 3.52. In the control group, increases were also observed across all indicators. The mean fluency score increased from 2.17 to 2.42, pronunciation from 1.88 to 2.75, grammar from 1.75 to 2.58, comprehensibility from 1.92 to 2.88, and vocabulary from 2.17 to 2.67. Descriptively, the experimental group had higher posttest mean scores than the control group across all speaking ability indicators.

Normality Test

The normality test was conducted to determine the distribution of students' speaking ability data in the experimental and control groups before further analysis. The Shapiro–Wilk test was used because the number of participants in each group was fewer than 50 students. The results of the normality test are presented in Table 3.

Table 3. Results of the Normality Test for Speaking Ability Data

Group	Pretest	Interpretation	Posttest	Interpretation
Experimental	.061	Normal	.003	Non-normal
Control	.235	Normal	.835	Normal

Based on Table 3, the significance value for the experimental group was .061 for the pretest and .003 for the posttest. In the control group, the significance values were .235 for the pretest and .835 for the posttest. Data were considered normally distributed when the significance value was greater than .05. Thus, the experimental group's pretest and the control group's pretest and posttest were normally distributed, whereas the experimental group's posttest was not normally distributed.

Wilcoxon Test Results for the Experimental Group

The Wilcoxon Signed-Rank Test was conducted to determine changes in students' speaking ability in the experimental group between the pretest and posttest results. This test was used because

the normality test indicated that the experimental group's posttest data were not normally distributed. The Wilcoxon test results are presented in Table 4.

Table 4. Wilcoxon Test Results for the Experimental Group

Comparison	N	Z	Sig. (2-tailed)
Pretest and Posttest	25	-4.386	< .001

The Wilcoxon test results showed a Z value of -4.386 with a significance value of $p < .001$. The significance value was lower than .05, indicating a significant difference between the pretest and posttest speaking ability scores of students in the experimental group. Descriptively, the mean speaking ability score increased from 9.80 at pretest to 16.68 at posttest. These results indicate a change in students' speaking ability after participating in instruction in the experimental group.

Wilcoxon Test Results for the Control Group

The Wilcoxon Signed-Rank Test was conducted to determine differences in students' speaking ability in the control group before and after instruction. The analysis results are presented in Table 5.

Table 5. Wilcoxon Test Results for the Control Group

Comparison	N	Z	Sig. (2-tailed)
Pretest-Posttest	24	-3.66	< .001

Based on Table 5, the Z value was -3.66 and $p < .001$. Because $p < .05$, there was a significant difference between the pretest and posttest speaking ability scores of students in the control group. The mean score increased from 9.88 to 13.29, indicating that students' speaking ability in the control group improved after instruction.

N-Gain Test

N-Gain analysis was conducted to describe the level of improvement in students' speaking ability based on pretest and posttest scores in the experimental and control groups. In addition to the mean N-Gain value, category distributions were used to examine patterns of speaking ability improvement in each group. The results are presented in Table 6.

Table 6. Results of N-Gain Analysis and Distribution of Speaking Ability Improvement Categories

Group	N	Mean N-Gain (%)	Category	Low	Moderate	High
Experimental	25	69.94	Moderate	0 (0%)	16 (64%)	9 (36%)
Control	24	31.86	Low	10 (41.7%)	13 (54.2%)	1 (4.2%)

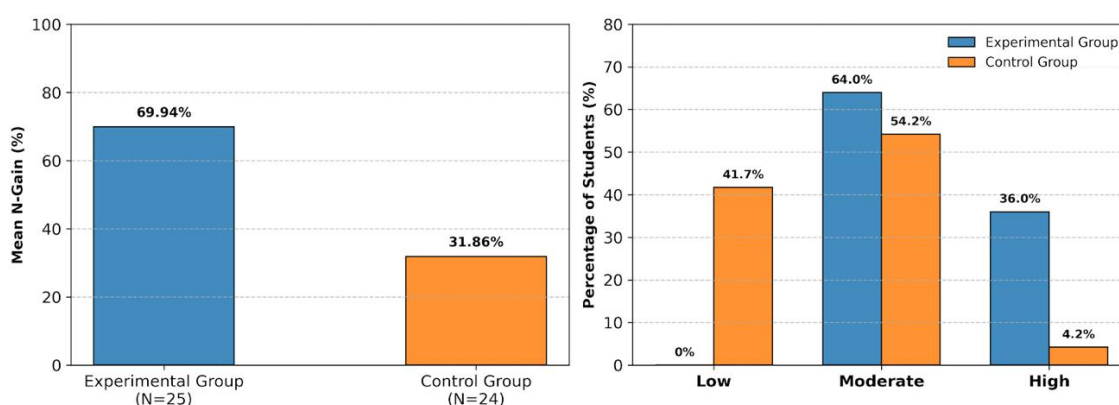


Figure 1. N-Gain Analysis and Distribution of Speaking Ability Improvement

Based on Table 6 and Figure 1, the experimental group obtained a mean N-Gain of 69.94%, which was categorized as moderate, whereas the control group obtained a mean N-Gain of 31.86%, which

was categorized as low. Descriptively, these results indicate that speaking ability improvement in the experimental group was higher than that in the control group. The category distribution showed that in the experimental group, most students were in the moderate category (64%), followed by the high category (36%), and no students were in the low category. In contrast, 41.7% of students in the control group were in the low category, 54.2% were in the moderate category, and only 4.2% were in the high category. This pattern indicates that the experimental group had a better level of speaking ability improvement, both in terms of the mean N-Gain and its category distribution.

Test of Differences in Speaking Ability between Groups

After analyzing changes in speaking ability within each group, the next analysis was conducted to determine differences in speaking ability outcomes between the experimental and control groups. The Mann–Whitney U Test was conducted on posttest scores because the data did not fully meet the normality assumption. The significance level used was .05.

Table 7. Mann–Whitney Test Results for Speaking Ability Posttest Scores

Group	N	Mean Rank	Mann–Whitney U	Z	Asymp. Sig. (2-tailed)
Experimental	25	32.54	111.500	-3.804	< .001
Control	24	17.15	—	—	—

Based on Table 7, the experimental group had a Mean Rank of 32.54, whereas the control group had a Mean Rank of 17.15. The Mann–Whitney test showed $U = 111.500$, $Z = -3.804$, and $\text{Asymp. Sig. (2-tailed)} < 0.001$. The significance value was lower than 0.05, indicating a significant difference in speaking ability posttest scores between the experimental and control groups. The higher Mean Rank of the experimental group than that of the control group indicates that students in the experimental group tended to have higher speaking ability posttest scores. This finding demonstrates a difference in speaking ability outcomes between the two groups after the instruction was provided.

Effect Size

To complement the results of the difference test between the experimental and control groups, an effect size analysis was conducted. This analysis was used to determine the magnitude of the difference in speaking ability posttest scores between the two groups so that the research findings would not only indicate statistical significance but also provide an overview of the strength of the observed difference. The effect size was calculated based on the Z value from the Mann–Whitney test.

Table 8. Effect Size of the Difference in Speaking Ability Posttest Scores

Analysis	Z	N	Effect Size (r)	Interpretation
Mann–Whitney U	-3.804	49	.543	Large

Based on Table 8, the effect size calculation, an r value of .543 was obtained. This value indicates that the difference in speaking ability posttest scores between the experimental and control groups had a large effect. Thus, the difference identified through the Mann–Whitney test was not only statistically significant but also indicated a relatively strong magnitude of difference.

Interviews

Interviews were conducted to obtain an overview of students' experiences after participating in Deep Learning-based instruction with self-recorded video. The interview analysis focused on three elements: understanding, applying, and reflecting. These three elements were used to examine how students understood the story content, addressed difficulties while constructing the story, and reassessed their performance. The interview findings were subsequently grouped according to low,

moderate, and high N-Gain categories. This grouping was used to provide qualitative evidence supporting the speaking ability improvement identified through N-Gain analysis.

For the student with low N-Gain, understanding of the story content remained limited to objects and events visible in the pictures. The student could mention several parts of the story but was not yet able to explain the sequence completely and coherently. When experiencing difficulty in selecting words, the student tended to stop and ask for the recording to be restarted from the beginning. The ability to reflect on the video results was also limited because the student was not yet able to independently determine which parts needed improvement. This was reflected in the student's statement: "I don't know, Ma'am. My mother said it was already okay, so it was immediately sent to Drive. But when I spoke, I still said 'um... um...' a lot." (Student, low N-Gain). This finding indicates that students in the low category still required assistance in understanding the story sequence, finding ways to overcome speaking difficulties, and evaluating their performance.

For the student with moderate N-Gain, understanding of the story content had developed further. The student was able to explain the main sequence of the pictures and convey the moral message using simple sentences. When experiencing difficulty in determining words, the student still needed time to find appropriate expressions but was able to continue the story using simpler word choices. In reflection, the student began to identify aspects that had been performed well and aspects that still needed improvement, such as voice volume and speaking rate. This condition indicates that students in the moderate category had begun to use simple strategies to overcome speaking difficulties. In addition, the ability to evaluate performance had begun to emerge, although it had not yet been conducted deeply and independently.

For the student with high N-Gain, understanding of the story sequence was more complete and coherent. The student was able to explain the relationship between each picture and convey a moral message appropriate to the story content. When experiencing difficulty in selecting words, the student could review the visual information and replace inappropriate words with more suitable expressions. The student also demonstrated the ability to use conjunctions and more varied vocabulary when developing the story. Reflective ability was evident in the student's ability to identify the strengths and weaknesses of the performance based on the task requirements. One student stated, "My body position from my head to my waist was already right. My fluency was good and I managed to convey the moral message. But the timing was a little too fast, only 45 seconds, not yet one full minute." (Student, high N-Gain). This statement indicates that the student was able to evaluate the performance based on the established criteria and identify aspects that still needed improvement.

Overall, the interview findings showed differences in the characteristics of learning experiences across the three N-Gain categories. Students with low N-Gain still experienced difficulties in understanding the story sequence, finding solutions when encountering difficulties, and evaluating their performance. Students with moderate N-Gain were able to understand the story and convey the main ideas, although limitations remained in fluency and word selection. Meanwhile, students with high N-Gain demonstrated better abilities in understanding the story content, selecting strategies when encountering difficulties, and evaluating their performance. These findings provide supporting evidence for the variation in speaking ability improvement observed in the N-Gain analysis. Thus, the interview data helped explain the characteristics of students' abilities at different levels of improvement without replacing the statistical analysis results.

DISCUSSION

This study showed that the implementation of the Deep Learning Approach assisted by Self-Recorded Video resulted in greater Speaking Skills improvement in the experimental group than in the control group. The mean score of the experimental group increased from 9.80 at pretest to 16.68 at posttest, representing an increase of 6.88 points. Meanwhile, the control group increased from 9.88 to 13.29, representing an increase of 3.41 points. The difference was also evident in the N-Gain results, namely 69.94% in the experimental group, categorized as moderate, and 31.86% in the control group, categorized as low. The Mann-Whitney U Test showed a significant posttest difference between the two groups ($U = 111.500$; $Z = -3.804$; $p < .001$), with a mean rank of 32.54 for the experimental group and 17.15 for the control group. The effect size of $r = .543$ indicates a relatively large practical effect.

These findings provide evidence that integrating the Deep Learning Approach and Self-Recorded Video is associated with better Speaking Skills outcomes than comparative instruction.

The difference in outcomes cannot be attributed solely to the presence of video, but rather to the way in which Self-Recorded Video was integrated into the learning process through task instruction, practice, recording, review, and reflection, enabling students to produce and evaluate their own speaking performances and to identify areas for improvement. Video thus functioned as a learning artefact rather than merely a medium for delivering instructional content. This finding is supported by Dang et al. (2022), who reported better speaking outcomes and more positive learner responses among students using video recording than in a comparison group; although their study involved university students learning English, it likewise demonstrated that recording affords additional opportunities for speaking practice and performance. The finding is also consistent with Ferng and Hwang (2026), who demonstrated the benefits of student-produced video for oral performance in higher-education Japanese language learning. Differences in educational and linguistic context, however, suggest that the contribution of video is closely related to its pedagogical design; the present study therefore extends previous findings by demonstrating how Self-Recorded Video can be integrated into the Deep Learning Approach as a learning cycle that connects task understanding, practice, performance production, and reflection.

At the elementary school level, the findings are consistent with those of Ramalingam et al. (2022), who reported that digital storytelling supported students' comprehension, vocabulary, and fluency by providing digital stimuli through which to organise ideas and communicate orally; the present study nonetheless extends this approach by requiring students to produce and evaluate their own performances through Self-Recorded Video, rather than simply engaging with digital stories. Similarly, Nair and Yunus (2022) found that digital storytelling supported speaking ability, engagement, and confidence, while Santhanasamy and Yunus (2022) showed that flipped learning supported by Blendspace provided opportunities for speaking practice both inside and outside the classroom; these studies share the present study's emphasis on flexible learning and increased opportunities for practice, but the present intervention further incorporated the recording and reviewing of performances. Furthermore, Zanah et al. (2023) demonstrated that interactive video combined with Discovery Learning improved elementary students' Indonesian speaking skills, while Cahyanti and Nuroh (2023) reported improvements through digital storytelling across pronunciation, grammar, vocabulary, fluency, and story content. Taken together, these findings support the use of digital media combined with pedagogical activities to develop elementary students' speaking abilities; the present study, in turn, offers a distinct intervention by integrating student-produced Self-Recorded Video with the principles of mindful, meaningful, and joyful learning.

Support for the use of video as a means of speaking production can also be seen in Sukrutrit (2023). The study compared speaking in front of the class with speaking through self-recorded video and found that learners tended to prefer self-recording because they felt more confident, more relaxed, enjoyed the activity more, and had greater opportunities to regulate their learning process. This finding is relevant to the present study because self-recording provides students with an opportunity to prepare their performance before submission. Thus, Self-Recorded Video can reduce the pressure of direct performance while providing additional time to organize ideas. These findings are also supported by Belmekki (2023), who examined the use of self-video recording in speaking instruction. However, the context of the learners in that study differed from that of the present study. The similarity is that self-recording provides learners with an opportunity to review their oral production. Yaşar and Mede also positioned the use of video technology in language learning as an activity that can support oral performance development. Differences in educational level and language of instruction should be considered, but they also indicate that the mechanism of observing one's own performance has potential for application across different learning contexts.

The development of the five Speaking Skills indicators shows a more specific pattern. In the experimental group, fluency increased from 2.00 to 3.16, pronunciation from 2.04 to 3.32, grammar from 1.76 to 3.28, vocabulary from 2.04 to 3.52, and comprehensibility from 1.96 to 3.40. All indicators improved, but vocabulary obtained the highest posttest score, while fluency remained one of the aspects requiring further strengthening. This pattern indicates that the intervention did not produce

uniform changes across all speaking components. Students' ability to select and use vocabulary appeared to develop more rapidly than their ability to maintain speaking fluency consistently. Vocabulary development can be associated with the use of picture sequences as stimuli. Pictures provide a context that helps students connect objects, actions, and events with the words needed to construct a story. Students also have opportunities to prepare ideas before recording. When the recordings are reviewed, students can recognize inappropriate word choices and attempt to use alternative words in subsequent performances. This finding is consistent with Wang et al. (2024), who reported that audiovisual media can support the development of speaking ability, vocabulary, pronunciation, and comprehension among elementary-level learners. Thus, the increase in vocabulary in this study may not be related solely to video use but to the combination of visual stimuli, idea preparation, oral practice, and performance evaluation.

In pronunciation, scores increased from 2.04 to 3.32. Recording allowed students to listen again to how they pronounced words and delivered sentences. This opportunity may help students identify unclear parts and improve them in subsequent attempts. This finding is consistent with Wang et al. (2024) and Nair and Yunus (2022), who demonstrated development in oral aspects through audiovisual activities and digital storytelling. The difference in the present study is that the recordings were not only used as final products but also as sources of self-evaluation during the learning process. In grammar, scores increased from 1.76 to 3.28. Although the increase was substantial, the score remained lower than vocabulary. This may be understood because the oral use of language structures requires students to coordinate content, vocabulary, grammar, pronunciation, and fluency simultaneously. Fan & Chen (2023) demonstrated that digital storytelling can support elementary school students' speaking performance. However, the findings of the present study indicate that the development of language structures may require a longer practice period than simply providing technology-based speaking experiences. Thus, Self-Recorded Video can serve as a means of providing practice and evaluation but does not automatically guarantee optimal grammatical mastery.

Comprehensibility increased from 1.96 to 3.40. This increase indicates that students became increasingly able to convey messages that could be understood by listeners. The picture sequence helped students maintain connections between ideas so that the story did not merely consist of separate descriptions of objects. This finding can be compared with Zhussupova & Shadiev (2023), who showed that digital storytelling can support idea organization and communication in a multimodal learning environment. Although their study involved different learners, the similarity lies in the use of visual elements to support oral communication production. Thus, visual stimuli in the present study may have functioned as a scaffold in the process of organizing messages. Colognesi et al. (2023) provide another relevant comparison because their study also used video recording in the development of elementary school students' oral communication. Their findings showed that video-based learning can provide benefits for several components of communication, but video does not automatically improve all aspects of performance. This comparison is important for the present study because it indicates that the advantage of the experimental group should not be understood as resulting from video technology alone. The improvement identified is more reasonably understood as the result of interaction among speaking tasks, practice, recording, opportunities to review performance, and reflection.

This pattern becomes clearer when compared with the control group. The control group also showed development across all indicators, namely fluency from 2.17 to 2.42, pronunciation from 1.88 to 2.75, grammar from 1.75 to 2.58, vocabulary from 2.17 to 2.67, and comprehensibility from 1.92 to 2.88. This means that comparative instruction also contributed to the development of speaking ability. However, the improvement in the experimental group was greater across all indicators. Therefore, the findings do not support the claim that speaking ability can only be developed through video. Instead, the findings indicate a relative advantage when Self-Recorded Video is integrated with a particular pedagogical structure.

The mechanism underlying the intervention can be summarised as a cycle of task instruction → practice → Self-Recorded Video → review → reflection → improvement, in which students grasp the performance requirements, organise and rehearse their ideas, produce concrete evidence of their speaking performance, review their recordings, and identify areas for improvement, thereby becoming both performers and observers of their own learning process. Amgott (2024) highlighted the

importance of reflective video in developing awareness of language use and speech organisation, while Dang et al. (2022) demonstrated that video-recording activities can contribute to speaking performance; together, these studies support the view that video becomes more meaningful when embedded within a structured pedagogical cycle. In the Indonesian context, Beltran-Palanques (2024) found that Self-Recorded Video supported the speaking ability of seventh-grade students through a one-group pretest–posttest design, whereas the present study extends this evidence by means of a quasi-experimental comparison of experimental and control groups and its integration with the Deep Learning Approach. Similarly, Rohdina and Priyana (2025) reported improvements in speaking ability through Self-Recorded Video in a classroom action research study involving secondary school students, in which self-recording served as a space for speaking practice; the present study extends this approach to elementary education using a quasi-experimental design. Fauziah (2025) further demonstrated that vlog-based video production could be integrated into speaking instruction among university EFL students, whereas the present study adapts the concept of student-produced video into a simpler Self-Recorded Video activity suited to the characteristics of elementary school students. In this study, implementation documentation is presented in Figure 1 to illustrate the differences in learning contexts between the experimental and control groups.



Figure 2. Documentation of Learning Implementation in the Experimental and Control Groups
(a) Self-Recorded Video activity in the Experimental Group; (b) Classroom Learning Activity in the Control Group

Figure 2(a) shows students in the experimental group participating in learning using the Deep Learning Approach assisted by Self-Recorded Video. Student activities included completing speaking tasks, practicing, recording performances, and reviewing the resulting recordings. The documentation shows that Self-Recorded Video was used as part of the learning activity rather than merely as a medium for storing students' work. Meanwhile, Figure 2(b) shows the learning activity of the control group, which received comparative instruction without the integration of Self-Recorded Video. The two forms of documentation provide visual context for the difference in treatment used in the Nonequivalent Control Group Design. Thus, Figure 1 supports the interpretation that differences in the outcomes of the two groups were associated with differences in the learning activities provided during the intervention.

Interview data from the explanatory sequential design also helped explain variations in the quantitative results. Interviews with students in the low, moderate, and high N-Gain categories indicated that students with low N-Gain scores still experienced difficulties in understanding the storyline and identifying aspects that needed improvement, whereas students in the moderate category began to recognize weaknesses such as speaking volume and pace. Students with high N-Gain scores demonstrated better abilities to understand the relationships between images, select strategies when encountering difficulties, and evaluate their performance based on the task criteria. These findings support the interpretation that improvements in speaking ability are associated with the development of students' abilities to understand, apply, and reflect on their learning experiences. The

interview data did not replace the statistical results but rather provided an explanation of variations in students' experiences across different levels of improvement.

From a state-of-the-art perspective, previous studies have established the use of video technology in various forms, including digital storytelling, flipped learning, video recording, vlogging, and student-produced videos, with evidence that video can support different aspects of oral communication (Amgott, 2024; Colognesi et al., 2023; Dang et al., 2022; Fan & Chen, 2023; Ferng & Hwang, 2026; Nair & Yunus, 2022; Ramalingam et al., 2022; Santhanasamy & Yunus, 2022; Sukrutrit, 2023). Research in Indonesia has likewise demonstrated the growing application of digital media in speaking instruction (Beltran-Palanques, 2024; Cahyanti & Nuroh, 2023; Fauziah et al., 2025; Rohdina & Priyana, 2025; Zanah et al., 2023), indicating that the use of video is not in itself novel. In practical terms, Self-Recorded Video can be integrated into elementary speaking instruction through clear and straightforward instructions, sufficient time for practice, and guided review of students' recordings, without requiring complex technology, since existing devices can be adapted to students' circumstances. The feasibility of self-recording is supported by Beltran-Palanques (2024) and by Rohdina and Priyana (2025), whereas the present study further demonstrates its integration with Deep Learning principles in elementary education.

Overall, this study provides evidence that the integration of the Deep Learning Approach assisted by Self-Recorded Video resulted in higher Speaking Skills outcomes than those of the comparison group. This advantage was reflected in a mean increase of 6.88 points, an N-Gain of 69.94%, and a significant posttest difference with an effect size of $r = .543$. All speaking ability indicators showed improvement, with vocabulary achieving the highest posttest score, while fluency and grammar still require further practice. Interview data also indicated that the abilities to understand, apply, and reflect on learning experiences developed differently among students with different levels of N-Gain. Thus, the contribution of this study lies not merely in the use of video but in the integration of instruction → practice → Self-Recorded Video → review → reflection → improvement within the framework of mindful, meaningful, and joyful learning. This positioning strengthens previous research while also providing a new contribution to Indonesian language learning at the elementary school level.

CONCLUSION

The novelty of this study lies in the integration of the Deep Learning Approach with Self-Recorded Video to support elementary school students' speaking skills. In this study, video became an integral part of the learning sequence, comprising task instruction, practice, recording, review, and reflection. This study also provides an empirical contribution through the measurement of speaking skills based on five aspects, namely fluency, pronunciation, grammar, vocabulary, and comprehensibility, as well as through a comparison of the results between the experimental and control groups. This integration demonstrates that the principles of mindful learning, meaningful learning, and joyful learning can be connected with student-produced video activities to create a more active, reflective, and contextual speaking learning experience. This study has several limitations, including the use of a Nonequivalent Control Group design without individual randomisation and the selection of participants from different schools, which may have introduced differences in student and school characteristics that could not be fully controlled. In addition, the experimental and control groups completed the posttest online and offline respectively, which may have influenced performance, while the relatively small sample limits the generalisability of the findings. Future studies should involve larger and more diverse samples, apply consistent posttest procedures, extend the intervention period in order to examine the sustainability of improvements in fluency and grammar, and deepen the qualitative analysis of students at different levels of improvement in order to better understand the mechanisms of self-recording and reflection. Despite these limitations, Self-Recorded Video may be considered a viable strategy for speaking instruction in elementary schools when systematically integrated into the learning process through clear task instructions, adequate time for preparation and practice, and guided review and reflection. Its implementation should emphasise pedagogy rather than technology, positioning video as a tool for practice, self-monitoring, and reflection, an orientation that aligns with the Merdeka Curriculum's emphasis on student independence, meaningful engagement, and reflection. Future research may further integrate digital

feedback or artificial intelligence to provide individualised support, while preserving teachers' roles in guiding and evaluating learning.

ACKNOWLEDGMENT

The authors would like to thank the principals, teachers, and sixth-grade students of the two elementary schools that took part in this study for their cooperation and enthusiasm throughout the pretest, learning, and posttest stages. Special thanks go to the three students who shared their learning experiences in the interviews, and to the parents and guardians who supported the students in completing and submitting their self-recorded videos at home. The authors are also grateful to Universitas Pendidikan Indonesia and Universiti Malaysia Sarawak for the academic support provided during this research.

AUTHOR'S DECLARATIONS

Authors' Contributions

NN: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft. YA: Conceptualization, Supervision, Writing – review & editing. DMI: Supervision, Validation, Writing – review & editing. SNS: Methodology, Resources, Writing – review & editing. All authors read and approved the final manuscript.

Funding Statement

This research received no external funding.

Availability of Data and Materials

The datasets generated and/or analyzed during the current study are stored by the authors and are available from the corresponding author, RPS, upon reasonable request.

Competing Interests

The authors declare that they have no competing interests.

REFERENCES

- Al-Hawamleh, M. S., Alazemi, A. F., & Al-Jamal, D. A. (2022). Digital portfolio and self-regulation in speaking tasks. *Asian-Pacific Journal of Second and Foreign Language Education*, 7(1), 1–19. <https://doi.org/10.1186/s40862-022-00141-w>
- Almulla, A. A., Khasawneh, M. A., Tashtoush, M. A., & Alsayed, A. (2026). Digital storytelling interventions to enhance the speaking abilities of elementary-level students with English language difficulties. *Frontiers in Psychology*, 17, 1825184. <https://doi.org/10.3389/fpsyg.2026.1825184>
- Amgott, N. (2024). “Je suis youtubeur”: Multilingual multimodal composing abroad. *Journal of Language, Identity & Education*, 23(5), 825–841. <https://doi.org/10.1080/15348458.2022.2057992>
- Andayanie, L. M., Adhantoro, M. S., Purnomo, E., & Kurniaji, G. T. (2025). Implementation of deep learning in education: Towards mindful, meaningful, and joyful learning experiences. *Journal of Deep Learning*, 1(1), 47–56. <https://doi.org/10.23917/jdl.v1i1.11157>
- Ardhiati, N. T., Kartono, K., & Rochmad, R. (2021). The effectiveness of stop motion video assisted discovery learning model on mathematics problem solving ability in elementary school students. *Educational Management*, 10(3), 211–220. <https://journal.unnes.ac.id/sju/eduman/article/view/56071>
- Asadollahpour, F., Mazaheri, S., Shahouzaei, N., & Nakhaei, M. A. (2026). Prevalence of dyslexia and its relationship with decoding and comprehension in persian-speaking elementary students. *Indian Journal of Psychological Medicine*, 48(2), 180-186. <https://doi.org/10.1177/02537176251389498>
- Bayu, G. W., Tantri, A. A. S., Kresnawati, L. P. D., Antara, I. G. W. S., & Phillips, M. (2026). A vocabulary building system based on the words stage model to improve elementary school English productive skills. *International Journal of Language Education*, 10(1), 190-201. <https://doi.org/10.26858/ijole.v10i1.87761>

- Belmekki, M. (2023). The effect of self-video recording on the development of students' speaking skill in higher education. *The Journal of Quality in Education*, 13(21), 21–31. <https://doi.org/10.37870/joqie.v13i21.345>
- Beltran-Palanques, V. (2024). Assessing video game narratives: Implications for the assessment of multimodal literacy in ESP. *Assessing Writing*, 60, Article 100809. <https://doi.org/10.1016/j.asw.2024.100809>
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Longman.
- Cahyanti, N. N., & Nuroh, E. Z. (2023). Digital storytelling media to improve students' speaking skills in elementary school. *Journal of Education Technology*, 7(2), 261–268. <https://doi.org/10.23887/jet.v7i2.58679>
- Colognesi, S., Coppe, T., & Lucchini, S. (2023). Improving the oral language skills of elementary school students through video-recorded performances. *Teaching and Teacher Education*, 128, 104141. <https://doi.org/10.1016/j.tate.2023.104141>
- Dang, T. N. A., Nguyen, V. T., & Nga, P. T. (2022). Utilizing video recording to develop EFL student's speaking skills. *International Journal of Linguistics, Literature and Translation*, 5(5), 63–71. <https://doi.org/10.32996/ijlt.2022.5.5.8>
- Draucker, C. B., Rawl, S. M., Vode, E., & Carter-Harris, L. (2020). Integration through connecting in explanatory sequential mixed method studies. *Western Journal of Nursing Research*, 42(12), 1137–1147. <https://doi.org/10.1177/0193945920914647>
- Dzakiah, D., Assiddiq, M. A., & Mustari, S. H. (2020). Smartphone video recording as a learning tool to improve Indonesian EFL students' speaking performance. *Indonesian EFL Journal*, 6(2), 157–164. <https://doi.org/10.25134/iefli.v6i2.3384>
- Erlitaviana, D., & Tyas, D. N. (2025). The relationship between the use of audio-visual learning media and learning motivation with learning outcomes in elementary school. *Jurnal Penelitian Pendidikan IPA*, 11(3), 163-170. <https://doi.org/10.29303/jppipa.v11i3.10375>
- Fan, T.-Y., & Chen, H.-L. (2023). Developing cooperative learning in a content and language integrated learning context to enhance elementary school students' digital storytelling performance, English speaking proficiency, and financial knowledge. *Journal of Computer Assisted Learning*, 39(4), 1354–1367. <https://doi.org/10.1111/jcal.12804>
- Fauziah, I. (2025). Implementasi metode show and tell dalam meningkatkan keterampilan berbicara pada peserta didik kelas V dalam pembelajaran Bahasa Indonesia [Implementation of the show and tell method in improving speaking skills in fifth grade students in Indonesian language learning]. *Jurnal Muria Research Guidance and Counseling*, 4(1), 87–95. <https://jurnal.umk.ac.id/index.php/mrgc/article/view/14694>
- Ferng, B.-J., & Hwang, G.-J. (2026). Enhancing Japanese as a foreign language learners' oral performance with digital content creation: A student-produced mobile video-based learning approach. *British Journal of Educational Technology*, 57(1), 33–54. <https://doi.org/10.1111/bjet.70008>
- Fitriati, S. W., Urip, S. R., Farida, A. N., Hapsari, C. T., Rahmawati, Y., & Willian, A. (2026). Habitual speaking berbasis deep learning untuk meningkatkan keterampilan berbicara bahasa Inggris di SMP Negeri 24 Semarang [Habitual speaking based on deep learning to improve English speaking skills at SMP Negeri 24 Semarang]. *Lamahu: Jurnal Pengabdian Masyarakat Terintegrasi*, 5(2), 445–453. <https://doi.org/10.37905/lipmt.v5i2.40366>
- Kang, N. (2022). Promoting EFL learners' self-regulated learning using a self-recorded video-speaking task in an online ESP course. *Journal of Asia TEFL*, 19(4), 1141–1162. <https://doi.org/10.18823/asiatefl.2022.19.4.1.1141>
- Lai, C. J. (2024). Enhancing multimodal output in CLIL education: The impact of VR games on fourth-grade students' English poster designs and presentations in Taiwan. *Humanities and Social Sciences Communications*, 11(1), Article 1465. <https://doi.org/10.1057/s41599-024-03999-y>
- Lavrishcheva, E. V., Panarina, G. I., Pozdniakova, M. N., & Tregubova, Y. A. (2026). Role-play games as a means of enhancing foreign language acquisition of primary school students. *Перспективы науки и образования*, 2(80), 394-408. <https://cyberleninka.ru/article/n/role-play-games-as-a-means-of-enhancing-foreign-language-acquisition-of-primary-school-students>
- Mandalika, K. S. B. (2025). Analisis penerapan model pembelajaran kontekstual berbasis meaningful

- learning di sekolah dasar [Analysis of the implementation of contextual learning models based on meaningful learning in elementary schools]. *Jurnal Dinamika*, 6(2), 131–152. <https://doi.org/10.18326/dinamika.v6i2.131-152>
- McCrudden, M. T., & McTigue, E. M. (2019). Implementing integration in an explanatory sequential mixed methods study of belief bias about climate change with high school students. *Journal of Mixed Methods Research*, 13(3), 381–400. <https://doi.org/10.1177/1558689818762576>
- Nafi'ah, J., & Faruq, D. J. (2025). Conceptualizing deep learning approach in primary education: Integrating mindful, meaningful, and joyful. *Journal of Educational Research and Practice*, 3(2), 225–237. <https://doi.org/10.70376/jerp.v3i2.384>
- Nair, V., & Yunus, M. M. (2021). A systematic review of digital storytelling in improving speaking skills. *Sustainability*, 13(17), Article 9829. <https://doi.org/10.3390/su13179829>
- Noetel, M., Griffith, S., Delaney, O., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. (2021). Video improves learning in higher education: A systematic review. *Review of educational research*, 91(2), 204–236. <https://doi.org/10.3102/0034654321990713>
- Parmaxi, A. (2023). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments*, 31(1), 172–184. <https://doi.org/10.1080/10494820.2020.1765392>
- Pertiwi, A. (2025). *Students' perspective on the effectiveness of video recording in oral communication: A case study at a vocational high school in Pekanbaru* [Undergraduate thesis, State Islamic University of Sultan Syarif Kasim Riau]. UIN Sultan Syarif Kasim Riau Repository. <https://repository.uin-suska.ac.id/88054/>
- Pituxcoosuvann, M., Radhapuram, S. C. T., & Murakami, Y. (2024). Taboo talks: Enhancing ESL speaking skills through language model integration in interactive games. *Procedia Computer Science*, 246, 3674–3683. <https://doi.org/10.1016/j.procs.2024.09.189>
- Purba, T. D. C., Silalahi, M., & Purba, R. (2024). The effectiveness of using self-recording video to improve student's speaking ability. *Bilingual: Jurnal Pendidikan Bahasa Inggris*, 6(2), 154–158. <https://doi.org/10.36985/24kcj302>
- Ramalingam, K., Jiar, Y. K., & Mathiyazhagan, S. (2022). Speaking skills enhancement through digital storytelling among primary school students in Malaysia. *International Journal of Learning, Teaching and Educational Research*, 21(3), 22–35. <https://www.ijlter.net/index.php/ijlter/article/view/1269>
- Rofii, A., & Syarifah, E. F. (2024). The use of innovative technology in teaching speaking skills to elementary school teacher education students. *Jurnal Cakrawala Pendas*, 10(3), 458–470. <https://doi.org/10.31949/jcp.v10i3.9263>
- Rohdina, N. F., & Priyana, J. (2025). From fear to fluency: Self-recorded video use to improve EFL speaking proficiency. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 13(2), 8011–8027. <https://jeet.fkdp.or.id/index.php/jeet/article/view/160>
- Santhanasamy, C., & Yunus, M. M. (2022). The flipped learning and blendspace to improve pupils' speaking skills. *Frontiers in Psychology*, 13, 866270. <https://doi.org/10.3389/fpsyg.2022.866270>
- Saputro, U. G., Basuki, I. A., & Anggraini, A. E. (2023). Analysis of elementary students' public speaking skills. *Primary: Jurnal Pendidikan Guru Sekolah Dasar*, 12(3), 576–583. <https://primary.ejournal.unri.ac.id/index.php/JPFKIP/article/view/272>
- Sari, A. W., & Arta, D. J. (2025). Implementasi deep learning: Suatu inovasi Pendidikan [Implementation of deep learning: An educational innovation]. *WASPADA (Jurnal Wawasan Pengembangan Pendidikan)*, 13(1), 121–126. <https://ejournal.undaris.ac.id/index.php/waspada/article/view/727/546>
- Septiani, S., Khusaini, F., & Fatimah, N. (2025). Effectiveness of role-playing method and interactive video media on students' speaking ability in class IV elementary school. *Journal of Educational Sciences*, 9(4), 2996–3006. <https://jes.ejournal.unri.ac.id/index.php/JES/article/view/970>
- Sukrutrit, P. (2023). Students' perceptions of speaking English in front of the class versus speaking English via self-recorded videos posted on a private Facebook group. *LEARN Journal: Language Education and Acquisition Research Network*, 16(1), 272–295. <https://so04.tci-thaijo.org/index.php/LEARN/article/view/263442>

- Wahyudi, E. (2025). Pengaruh media video animasi terhadap keterampilan menyimak dan berbicara siswa kelas IV sekolah dasar [The influence of animated video media on the listening and speaking skills of fourth grade elementary school students]. *Jurnal Binagogik*, 12(2), 51–57. <https://ejournal.uncm.ac.id/index.php/pgsd/article/view/1728>
- Wang, R., Do Dange, M., & Izadpanah, S. (2024). The effect of animated movies on speaking skills among motivated English foreign language learners: Elementary level. *European Journal of Education*, 59(3), 12665. <https://doi.org/10.1111/ejed.12665>
- Zanah, D. N., Putri, C., & Melati, I. I. (2023). Penggunaan media video interaktif untuk meningkatkan keterampilan berbicara bahasa Indonesia dengan model discovery learning [The use of interactive video media to improve Indonesian speaking skills with the discovery learning model]. *Journal of Education Research*, 4(2), 592–598. <https://doi.org/10.37985/jer.v4i2.213>
- Zhussupova, R., & Shadiev, R. (2023). Digital storytelling to facilitate academic public speaking skills: Case study in culturally diverse multilingual classroom. *Journal of Computers in Education*, 10(3), 499–526. <https://doi.org/10.1007/s40692-023-00259-x>