Enhancing students' speaking skills with Artificial Intelligence: Evidence from a senior high school class in Ternate, Indonesia

Syafiul Rahman*[1], Yufaini[1], Mahdi Ahmad[1]

[1]ISDIK Kie Raha Maluku Utara, Indonesia

Abstract

Speaking is a core skill for English learners, yet many senior high school students in Indonesia lack confidence, vocabulary, and opportunities to practice. This quasi-experimental study examined whether speaking practice with ChatGPT improves students' speaking skills more than traditional materials such as books, audiobooks, and videos. Ten students from SMA Negeri 3 Kota Ternate were divided into an experimental group (n = 5), which practiced with ChatGPT, and a control group (n = 5), which used traditional materials. Both groups completed a rubric-based speaking pre-test and post-test adapted from Brown (2001). Perception data were gathered through an open-ended questionnaire and post-test interviews. The experimental group's mean score rose from 64.0 to 74.4, whereas the control group's mean remained nearly unchanged (48.8 to 48.0). The difference in gain scores showed a medium effect size (Cohen's d = 0.66) but was not statistically significant ($t(8) = 1.05$, $p = .324$). Students in the ChatGPT group reported greater confidence, lower anxiety, and improved fluency and vocabulary, although they also noted accent misrecognition and unstable internet connections. The findings suggest that ChatGPT can serve as a useful supplementary tool for speaking practice, though larger studies are needed to confirm its effect.

Keywords: *English speaking skills, artificial intelligence, ChatGPT, EFL learners, senior high school*

1 Introduction

Artificial intelligence (AI) is increasingly used to support teaching and learning. In education, AI can provide interactive materials, personalized practice, and immediate feedback (Holmes et al., 2022; Zawacki-Richter et al., 2021). In language education, these features are especially relevant to speaking, a skill that involves grammar, vocabulary, pronunciation, and fluency, as well as the confidence to communicate in real time (Derakhshan et al., 2021).

English is widely used in education, business, and daily life, which makes speaking ability central for learners in Indonesia. Many students nevertheless struggle. Shyness, nervousness, and fear of making mistakes often keep them from taking part in speaking activities (MacIntyre et al., 2022), and limited vocabulary and pronunciation problems add to the difficulty. In large

*Corresponding Author: syafiurrahmansehan@gmail.com

classes with limited time, teachers may not be able to give each student enough practice or individual feedback (Bao, 2020), which can slow progress and reduce motivation.

ChatGPT, a conversational AI tool, has been proposed as one way to address these constraints. It can hold interactive dialogues, correct errors, and offer flexible practice that students can repeat at any time, without the pressure of speaking in front of peers (Kohnke et al., 2023; Li & Zhao, 2021). Such low-pressure practice may be especially valuable because motivation and confidence strongly influence willingness to communicate (Harmer, 2021; Schunk et al., 2020). Ryan and Deci (2020) also note that autonomy can support intrinsic motivation, and independent practice with a conversational tool gives learners some control over their own pace.

Recent studies generally report positive effects of AI tools on speaking. Qiao and Zhao (2023) compared AI-based and traditional instruction among 93 Chinese EFL students and found greater speaking improvement in the AI group. Xiao and Zhi (2023) reported that students viewed ChatGPT as a helpful learning companion that increased their motivation to practice, while Wang (2025) and Yuan (2023) found favorable attitudes toward AI-supported interaction and improved English ability. Kadaruddin (2023) observed significant speaking gains among university English students who used several AI applications in a one-group pre-test and post-test design. Fewer studies have compared ChatGPT practice against traditional materials with a control group among high school learners.

The present study addresses this gap by examining whether ChatGPT-assisted speaking practice is more effective than traditional learning materials for senior high school students. Three research questions guided the study: (1) *Is there a significant difference in speaking skill improvement between students who practice with ChatGPT and students who use traditional materials?* (2) *What are students' perceptions of using ChatGPT to support their English speaking practice?* And (3) *What benefits and challenges do students experience when using ChatGPT for speaking practice?* It was hypothesized that ChatGPT-assisted learning would improve speaking skills more than traditional instruction alone. For students, the findings may point to a low-pressure way to practice. For teachers, they may suggest a supplementary tool for situations where class time is limited.

2 Methods

Design This study used a quasi-experimental pre-test and post-test design with two groups. Quantitative scores were combined with qualitative perception data in a mixed-methods approach (Creswell & Plano Clark, 2018). The study used one existing class rather than a randomly selected sample from a larger population, which is typical of classroom-based research.

Participants and groups The participants were 10 students from SMA Negeri 3 Kota Ternate, a class identified as having low English speaking skills. These students commonly showed low confidence, nervousness, difficulty expressing ideas, and limited vocabulary. The 10 students were divided into two groups of five. The experimental group used ChatGPT as its main tool for speaking practice. The control group practiced with traditional materials such as books, audiobooks, and videos. Both groups received instruction on how to use their assigned resource before

the practice period, and both went through the same practice period so that the comparison was fair.

Instruments Three instruments were used: a speaking assessment rubric, an open-ended perception questionnaire, and interview protocols delivered as spoken dialogues. The rubric was adapted from Brown (2001) and rates five criteria: grammar, vocabulary, comprehension, fluency, and pronunciation. Each criterion is scored from 1 to 5, giving a maximum of 25 points. Raw scores are converted to a 0–100 scale using the formula *Score = (Student's points ÷ 25) × 100*. Converted scores are classified as Excellent (86–100), Good (71–85), Fair (56–70), Poor (41–55), or Very Poor (40 or below).

Data collection and analysis Data were collected in four steps. We administered a speaking pre-test to all students and gave an open-ended questionnaire at the same stage to capture initial perceptions. After the practice period, we administered a speaking post-test, followed by a brief spoken interview about each student's experience, including confidence, nervousness, and motivation. The interviews were transcribed. For quantitative analysis, gain scores (post-test minus pre-test) were compared between groups using an independent-samples t-test. The questionnaire and interview data were reviewed thematically to identify recurring perceptions, benefits, and challenges.

3 Results

3.1 Speaking Performance

Table 1 summarizes the pre-test, post-test, and gain scores for both groups. The experimental group's mean rose from 64.0 to 74.4, a mean gain of +10.4, which moved its average classification from Fair to Good. Four of the five experimental students improved. The control group's mean was virtually unchanged, from 48.8 to 48.0 (mean gain of −0.8), and remained in the Poor range. Three control students improved, while two declined.

Table 1: Speaking Test Scores and Gain Scores by Group (0–100 scale)

Group	n	Pre-test M	Post-test M	Mean gain	SD of gain	Classification (pre → post)
Experimental (ChatGPT)	5	64.0	74.4	+10.4	17.34	Fair → Good
Control (traditional materials)	5	48.8	48.0	−0.8	16.35	Poor → Poor

Note. SD = standard deviation. Classifications follow the rubric score categories described in the Methods section.

An independent-samples t-test on the gain scores showed a difference in the expected direction, $t(8) = 1.05$, $p = .324$ (two-tailed). This did not reach significance at = .05. The effect size was medium (Cohen's $d = 0.66$), indicating a meaningful practical difference despite the non-significant result. With only five students per group, statistical power was limited, so the test may not detect a real difference of this size.

3.2 Individual Variation

Not every student improved. One experimental student declined by 20 points, giving markedly shorter post-test answers that relied more on Indonesian than the pre-test answers. Two control students also declined (by 16 and 20 points), while one gained 16 points. These drops may reflect interview engagement, question fatigue, or the question format. The experimental group's post-test asked narrow questions about opinions of ChatGPT, such as "What do you think about ChatGPT?" The pre-test and the control group's post-test used broader prompts about daily routines, hobbies, and personal relationships. This asymmetry in question difficulty is acknowledged as a limitation of the instrument design.

3.3 Students' Perceptions, Benefits, and Challenges

Most experimental students described ChatGPT positively, calling it very helpful (*sangat membantu*), easy to understand, and useful for practicing without pressure. Students said that daily practice built their confidence, that voice interaction helped their speaking more than books did, and that practicing with AI made them more ready to talk with others. Fluency, vocabulary, and pronunciation were the areas students mentioned most often as improving, followed by grammar. Control-group students also acknowledged benefits of AI in general, even though they had not used ChatGPT intensively.

Students reported several benefits: increased confidence and reduced anxiety in a private, non-judgmental setting; immediate correction of mistakes; the flexibility to practice anytime; learning new vocabulary and more natural expressions; and a greater willingness to speak in longer sessions. The reported challenges were occasional misrecognition of Indonesian-accented speech, AI vocabulary that sometimes felt too advanced, internet and signal problems, and difficulty knowing how to start a conversation. Students who engaged more elaborately with ChatGPT appeared to show clearer gains, which suggests that the quality of engagement matters.

4 Discussion

The findings offer partial support for the hypothesis that ChatGPT-assisted practice improves speaking more than traditional materials alone. The experimental group showed a larger mean gain than the control group, and the medium effect size suggests a practically relevant difference. However, the result was not statistically significant, so chance cannot be ruled out given the very small sample. The results should therefore be read as preliminary and suggestive rather than conclusive.

The qualitative data are consistent with the numerical trend. Students' reports of greater confidence, improved fluency and vocabulary, and preference for AI over books match earlier studies showing that AI tools can reduce anxiety and increase motivation (Qiao & Zhao, 2023; Xiao & Zhi, 2023). The challenges students reported, such as accent recognition, advanced vocabulary, and connectivity, are practical issues that teachers should plan for when integrating ChatGPT into speaking practice.

The comparison with Kadaruddin (2023) helps place these findings in context. That study, which involved university English-major students and several AI applications, reported statistically significant gains and a "Good" perception score. The present study, with younger high school learners and a control group, shows a similar direction of effect but lacks statistical power. Together, the two studies suggest that AI tools may support speaking development across educational levels, although larger samples and carefully matched instruments are needed for stronger claims.

Limitations Three limitations should be noted. First, the sample of five students per group is small, which limits statistical power and the generalizability of the t-test result. Second, the experimental group's post-test questions differed in scope and difficulty from its pre-test and from the control group's post-test, which may have affected the comparability of scores. Third, pronunciation scores were estimated from written transcripts rather than from audio recordings, which is a less precise method.

5 Conclusion and Recommendation

This study examined whether ChatGPT-assisted speaking practice improves EFL students' speaking skills more than traditional learning materials. Using a quasi-experimental pre-test and post-test design with 10 students at SMA Negeri 3 Kota Ternate, the experimental group gained an average of 10.4 points, while the control group showed almost no change.

Regarding the first research question, there is a numerical, medium-sized advantage for the ChatGPT group, but this advantage was not statistically confirmed at this sample size. Regarding the second, students held positive perceptions of ChatGPT and described it as helpful, confidence-building, and preferable to traditional materials for speaking practice. Regarding the third, the benefits included increased confidence, improved fluency and vocabulary, flexible practice, and reduced anxiety, while the challenges included accent misrecognition, advanced vocabulary, connectivity problems, and difficulty starting conversations. Taken together, the results suggest a positive trend for ChatGPT as a supplementary resource for speaking practice, in line with related research, though not a replacement for classroom instruction.

For teachers, ChatGPT can supplement regular classroom teaching by giving students low-pressure speaking practice, but teachers should guide students on how to start conversations and how to handle advanced vocabulary. For students, regular and sustained practice matters more than any single tool, and practicing in a quiet place with a stable internet connection is recommended. For future researchers, the study should be replicated with a larger sample and with pre-test and post-test instruments of matched difficulty and format. Pronunciation should be scored from audio recordings where possible. Future work could also compare ChatGPT with other AI speaking tools and examine whether the gains are retained over time.

References

Bao, W. (2020). Covid-19 and online teaching in higher education: A case study of peking university. *Human Behavior and Emerging Technologies*, *2*(2), 113–115. https://doi.org/10.1002/hbe2.191

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy* (2nd). Longman.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd). SAGE Publications.

Derakhshan, A., Khalili, A. N., & Beheshti, F. (2021). Developing efl learners' speaking ability, accuracy, and fluency. *Journal of Applied Linguistics and Language Research*, *8*(2), 1–16.

Harmer, J. (2021). *How to teach english* (New). Pearson Education.

Holmes, W., Bialik, M., & Fadel, C. (2022). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Kadaruddin. (2023). *Improving students' speaking skills by using artificial intelligence on the second semester of english program at iain parepare* [Undergraduate thesis]. State Islamic Institute (IAIN) Parepare.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Chatgpt for language teaching and learning. *RELC Journal*. https://doi.org/10.1177/00336882231162868

Li, J., & Zhao, Y. (2021). The role of feedback in improving speaking skills in efl learning. *Computer Assisted Language Learning*.

MacIntyre, P. D., Gregersen, T., & Mercer, S. (2022). Language anxiety and learner psychology. *Annual Review of Applied Linguistics*, *42*, 1–20.

Qiao, Y., & Zhao, X. (2023). The effect of artificial intelligence on efl learners' speaking skills. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1182390

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Schunk, D. H., Meece, J. L., & Pintrich, P. R. (2020). *Motivation in education: Theory, research, and applications* (4th). Pearson.

Wang, Y. (2025). The impact of chatgpt-4 on english communication skills. *SAGE Open*.

Xiao, Y., & Zhi, Y. (2023). Students' perceptions of chatgpt in language learning. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11886-z

Yuan, M. (2023). Chatbots in language learning: A review of recent studies. *Computers & Education*.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2021). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, *18*(39). https://doi.org/10.1186/s41239-021-00242-3