

AI-Mediated Speaking Practice and EFL Anxiety: Linking Measured Change and Learner Experience

Budiawan^{1*}, Virga Putra Darma², Nursyifa Azzahro³

¹⁻²⁻³ Institut Prima Bangsa, Indonesia

e-mail: budiawan@ipbcirebon.ac.id *, virgaputradarma@gmail.com, nursyifa@ipbcirebon.ac.id

Abstract:

AI-mediated speaking practice may provide possibilities for practice with less interpersonal pressure, yet speaking anxiety is still a persistent challenge in EFL learning. This quasi-experimental mixed-methods study examined whether additional AI-mediated speaking practice was associated with changes in speaking anxiety among Indonesian EFL students and explored students' affective experiences during such practice. Forty-five students from two intact classes participated: 25 completed 24 additional AI-mediated speaking sessions over eight weeks, while 20 received regular speaking instruction. Speaking anxiety was measured before and after the intervention, with posttest scores analysed using ANCOVA controlling for pretest anxiety. Twelve experimental-group students were subsequently interviewed. Both groups showed small descriptive decreases in anxiety, but ANCOVA detected no between-group difference at posttest after controlling for baseline anxiety ($p = .977$, $\eta p^2 < .001$). The qualitative data indicated three patterns of affective experience: (1) students reported lower social-evaluative pressure during AI-mediated speaking practice; (2) feedback became consequential to students due to learner uptake and repeated practice; and (3) although students reported increased confidence in their speaking abilities, this was not generalised over contexts. These experiences varied across participants and were affected by students' remaining learning anxieties, variation in uptake of the AI, and varying degrees of technological frictions. Integrating the findings suggested complementarity rather than contradiction: positive affective experiences were reported during AI-mediated speaking practice, but between-group difference in speaking anxiety was not detectable. AI may offer a relatively low-pressure environment for speaking practice, but context specific comfort and confidence should not be taken as evidence of broader anxiety reduction.

Keywords: *AI-mediated speaking practice; EFL speaking anxiety; evaluative pressure; generative AI; mixed methods*

1. INTRODUCTION

Speaking anxiety remains an ongoing challenge in learning English as a Foreign Language (EFL), especially when learners must speak spontaneously, respond to teachers or peers, or perform in assessment situations. Such situations can exacerbate the fear of making mistakes, negative evaluation, and concerns about communication, thereby limiting participation and opportunities for meaningful oral practice (Dewaele & Dewaele, 2020; Horwitz et al., 1986; Al-Khotaba et al., 2019). These concerns are particularly relevant in Indonesian EFL contexts, where students often have limited opportunities to use English beyond the classroom and may experience anxiety when speaking publicly or being evaluated by others (Hanafiah et al., 2022; Noe et al., 2025).

Recent advances in generative AI have created new opportunities for speaking practice outside conventional classroom interaction. Voice-based conversational AI can provide repeated practice, immediate feedback, flexible scheduling, and interaction without a physically present interlocutor. Systematic reviews collectively identify repeated practice, feedback, self-regulation, confidence, and reduced anxiety among the potential affordances and affective benefits of AI-supported language learning (Bhar, 2026; Crompton et al., 2024; Du & Daniel, 2024; Soriano-Gonzalez & Belda-Medina, 2026; Wiboolyasarini et al., 2025). These features may be especially relevant for learners who experience human evaluation as threatening because AI-mediated practice can alter the social conditions under which speaking occurs.

However, positive experiences during AI-mediated practice should not be assumed to indicate broader reductions in speaking anxiety. Speaking anxiety is context-sensitive: learners may feel relatively comfortable in one interactional setting yet remain anxious when speaking before teachers, peers, or other human interlocutors. Aldukhayel and Namaziandost (2022), for example, reported lower anxiety in remote than face-to-face presentations, while Susoy (2026) found that the relationship between speaking anxiety and performance differed between human- and AI-facilitated assessment conditions. Such findings suggest that the interactional and evaluative conditions of speaking matter for how anxiety is experienced.

Empirical findings on AI-supported speaking are also mixed. Some studies report reduced speaking anxiety following AI-mediated practice (Qu & Chen, 2026; Wang & Wen, 2025), whereas others identify positive learner experiences without clear between-group differences in anxiety (Ballıdağ & Aydin, 2024; Y. Huang & Kakham, 2026). These differences may reflect variation in intervention design, comparison conditions, learner engagement, and how anxiety is measured. More importantly, studies reporting comfort, confidence, or reduced pressure during AI interaction do not necessarily establish that these experiences generalize to broader speaking situations involving human evaluation.

This distinction reveals a more specific research gap, which is no longer merely a matter of the lack of empirical evidence based on mixed-methods approaches. Quantitative measures can indicate whether broader speaking anxiety changes, but they provide limited insight into how learners experience evaluative pressure, feedback, repeated practice, and confidence within AI-mediated interaction. Qualitative accounts can illuminate these experiences but cannot by themselves establish broader anxiety change. What remains insufficiently understood is how specific affective experiences in a given context during interactions with AI relate to measurable changes in broader speaking anxiety, particularly when both are tested within the same intervention.

To address this gap, the present study distinguishes between proximal affective experiences during AI-mediated practice and broader changes in speaking anxiety across speaking contexts. Using an explanatory sequential mixed-methods design within a quasi-experimental framework, it first examines whether students receiving additional AI-mediated speaking practice differ from a comparison group in posttest speaking anxiety after controlling for baseline anxiety. It then explores how students describe evaluative pressure, feedback and repeated practice, and confidence across AI-mediated and human-mediated speaking contexts. Integrating these strands allows the study to examine whether positive experiences during AI practice correspond to, complement, or remain distinct from broader measured anxiety change. It addresses the following research questions:

1. Does posttest speaking anxiety differ between students who receive additional AI-mediated speaking practice and those who receive regular speaking instruction, after accounting for pretest anxiety?
2. How do students describe their affective experiences of AI-mediated speaking practice and perceived changes across speaking contexts?

2. LITERATURE REVIEW

2.1 Speaking Anxiety and Evaluative Pressure

Foreign language anxiety is an affective response associated with learning and using a foreign language, with speaking among its most common triggers. Horwitz et al. (1986) identified communication apprehension, test anxiety, and fear of negative evaluation as key components of foreign language classroom anxiety. Fear of negative evaluation is particularly relevant to speaking because learners must produce language in real time while their performance may be noticed and evaluated by lecturers, peers, or an audience. Speaking anxiety therefore reflects not only linguistic demands but also the socio-evaluative conditions in which speaking occurs.

Empirical evidence suggests that learners' anxiety levels are strongly influenced by the assessment conditions they face. For example Aldukhayel and Namaziandost (2022) found that students' anxiety tended to be lower during remote presentations compared to in-person ones, while Susoy (2026) demonstrated differences in the relationship between speaking anxiety and performance in assessments facilitated by humans versus AI. These findings suggest that anxiety varies across speaking situations. Therefore, research in this field needs to clearly distinguish between general speaking anxiety and the affective experiences that occur in the context of specific interactions; feeling comfortable when interacting with AI does not necessarily guarantee a reduction in anxiety when learners return to speaking with humans.

2.2 AI-Mediated Speaking as an Affective Practice Environment

Conversational AI changes several conditions of speaking practice. Learners can practise individually, repeat conversations, receive immediate feedback, and choose when and how often to practise with less interpersonal exposure. Systematic reviews indicate that these features are effective in supporting speaking practice, boosting self-confidence and self-regulation, and minimizing speaking anxiety (Bhar, 2026; Crompton et al., 2024; Du & Daniel, 2024; L. Huang & Liu, 2025; Wiboolyasarin et al., 2025). Bhar's (2026) synthesis of 36 empirical studies confirmed affective benefits in the form of reduced anxiety and increased willingness to communicate, while underscoring the importance of pedagogical design and implementation conditions.

One of the mechanisms underlying this phenomenon is the reduction in evaluative pressure among individuals. Aldamen et al. (2026) found that learners experienced AI-mediated speaking as psychologically safer and more supportive, with reduced concerns about interpersonal judgment. Similar patterns have been reported in Indonesia: Noe et al. (2025) described ChatGPT as a speaking partner with relatively low level pressure, while Rizal (2025) highlighted its immediate feedback, a judgement-free practice environment, increased self-confidence building, and the flexibility of self-directed practice. These findings indicate that the affective benefits of AI do not stem solely from the technology itself but from transformations in the social and interactional condition that shape how students cope with mistakes, receive feedback and engage in repetition.

2.3 Feedback, Repeated Practice, and Affordance Uptake

AI-mediated speaking also provides opportunities for immediate feedback and repeated attempts. Crompton et al. (2024) identified several affordances of AI in English language learning, including its role as a conversation partner, language coach, platform for speaking practice, and supporter of student self-regulation. On the other hand, Bhar (2026) highlighted the characteristics of AI-based feedback as being instant, repeatable, and consistent. In the Indonesian context, Rizal (2025) reported that the immediate feedback helped learners recognize errors, make corrections independently, and practice more autonomously. Thus, this feedback stimulates a continuous learning cycle: speaking → feedback → error awareness → correction → retrying.

However, the availability of these features does not mean that learners use them equally. From an affordance perspective, possibilities for action arise through relations among technological features, users, and contexts; in the case of ChatGPT, Zhong et al. (2025) show how such affordances are shaped within a broader social-ecological system. In this study, affordance uptake refers to learners' recognition and use of opportunities such as feedback and repeated practice. This perspective treats AI-mediated autonomous practice not as a treatment that automatically produces effects, but as an environment whose possibilities may be taken up differently by individual learners.

2.4 From Proximal Affective Benefits to Broader Anxiety Change

Differences in affordance uptake may help explain why evidence on AI and speaking anxiety remains inconsistent. Several experimental or quasi-experimental studies report reduced anxiety following AI-mediated speaking practice (Ding & Yusof, 2025; Ebadi et al., 2025; Qu & Chen, 2026; Wang & Wen, 2025; Yan et al., 2026). Qu and Chen (2026), for example, found a greater reduction in speaking anxiety with voice-based GenAI than with peer interaction. By contrast, Huang and Kakham (2026) and Ballıdağ and Aydın (2024) found no significant between-group differences despite participants reporting positive experiences involving feedback, repetition, confidence, or relatively non-judgmental interaction. Ekizer's (2025) meta-analysis likewise indicates heterogeneity across AI-based interventions.

This discrepancy suggests that affective changes occur at two different levels. First, proximal affective benefits, which include a reduction in evaluative pressure, an increase in comfort or perceived self-confidence, and a willingness to make mistakes while interacting with AI. Second, broader changes in anxiety, which refer to shifts detectable through broader measures of speaking anxiety and extend beyond AI interactions. These proximal benefits do not

automatically lead to broader changes, especially when learners are once again faced with human audiences, peer assessment, instructor evaluation, and the interpersonal consequences of making mistake. Therefore, research on AI-assisted speaking practice should not only focus on whether anxiety changes but also on where those changes occur and whether their effects extend to other speaking contexts.

3. RESEARCH METHODOLOGY

This study employed an explanatory sequential mixed-methods design within a quasi-experimental framework. The quantitative phase compared pretest and posttest speaking anxiety between the experimental and control groups. The posttest results then served as the basis for purposive sampling to select 12 participants from the experimental group-representing the high, medium, and low score distribution – who would be interviewed in the qualitative phase. Thus, the quantitative phase preceded and informed the qualitative sampling, with both approaches integrated during the data interpretation stage.

Two existing (intact) speaking classes served as the experimental and control groups. To minimize bias caused by teacher-related factors, both classes were taught by the same instructor. The absence of random assignment of participants to treatment groups confirms that this study is quasi-experimental in nature.

3.1 Participants

Participants were 45 students enrolled in an English Literature study program at an Indonesian university. One intact class formed the experimental group ($n = 25$) and the other the control group ($n = 20$). Both participated in regular speaking classes, but the experimental group received an additional program consisting of AI-assisted independent speaking practice

After the posttest, the 25 experimental-group anxiety scores were ranked, and 12 participants were purposively selected from the upper, middle, and lower portions of the distribution. This strategy sought variation in experiences among participants with relatively different anxiety scores rather than creating diagnostic categories or applying predetermined cut-offs.

3.2 Instruments

3.2.1 Speaking Anxiety Questionnaire

Speaking anxiety was measured using a questionnaire adapted from the Foreign Language Classroom Anxiety Scale (FLCAS) (Horwitz et al., 1986). Of the original 33 items, 25 were selected and contextualized for English-speaking activities. DeepL provided an initial Indonesian translation, which was reviewed by the researcher, whose background is in linguistics and English language teaching, for meaning equivalence, linguistic clarity, and contextual appropriateness. The 25-item version was piloted with 18 students outside the main sample. One item with a negative corrected item-total correlation ($-.453$) was removed. The final 24-item instrument showed high internal consistency (Cronbach's $\alpha = .966$), with corrected item-total correlations ranging from .459 to .945. Five positively worded items were reverse-scored before total scores were calculated, with higher scores suggesting greater speaking anxiety.

3.2.2 Qualitative Data Collection

Qualitative data were collected through semi-structured interviews with the 12 purposively selected experimental-group participants after the posttest. Interviews were carried out in

Indonesian by research assistants who were not involved in participants' teaching or assessment and lasted approximately 15 minutes. The interview guide comprised 15 main questions addressing prior speaking experiences, AI-mediated practice, perceived changes, changes in learning practices, and reflections on the program. Follow-up probes were used for clarification or elaboration. Interviews were recorded and transcribed with transcription software by the assistants who conducted them.

3.3 AI-Mediated Autonomous Speaking Practice Intervention

The intervention lasted eight weeks and consisted of 24 scheduled sessions, each lasting 15 to 30 minutes, which were conducted at least three times a week. Participants used ChatGPT's voice feature for spoken interaction in English. ChatGPT model, account type, and device were not standardized; participants used the accounts and devices available to them. All experimental-group participants received the same prompt for each session, specifying a speaking topic or task and requesting feedback on aspects such as grammar, vocabulary, pronunciation, and fluency. The number of conversational turns was not prescribed because practice was time-based rather than turn-based. Session topics, prompts, and instructions were distributed via WhatsApp.

Participants were instructed to complete all 24 sessions and upload proof of completion to Google Drive. These tracking documents were used to verify session completion, not to measure the intensity or quality of students' spoken engagement. Although additional practice outside of scheduled sessions was permitted, such activities were not systematically recorded. During the intervention, some participants reported internet connectivity issues and occasional off-topic ChatGPT responses. On the other hand, no participant in the control group reported using ChatGPT for speaking practice during the study although such use was not independently monitored. The control group followed the same course material and regular speaking class topics without receiving the additional AI-based practice program.

3.4 Procedures and Data Analysis

Both groups completed the speaking anxiety questionnaire before and after the eight-week intervention. Following the posttest, the 12 purposively selected experimental-group participants completed semi-structured interviews. Quantitative and qualitative data were analysed separately before integration at the interpretation stage.

Before quantitative analysis, data were checked for completeness and matched pretest–posttest responses. Internal consistency was assessed using Cronbach's alpha, and descriptive statistics summarized anxiety scores. Between-group posttest differences were analysed using ANCOVA, with posttest score as the dependent variable, group as the factor, and pretest score as the covariate. Assumption checks included missing data, outliers and influential cases, residual normality, homogeneity of variance, and homogeneity of regression slopes. Results are reported using adjusted means, 95% confidence intervals, *F*, *p*-values, and partial eta squared (ηp^2), with $\alpha = .05$.

Interview data were analysed using researcher-led reflexive thematic analysis, informed by Braun and Clarke's (2006, 2019) approach. Analysis involved familiarization with the Indonesian transcripts, initial coding, development of candidate themes, review against the dataset, and theme definition and naming. Transcripts were first analysed within cases and then extended across cases, with particular attention paid to participant variation as well as disconfirming or boundary evidence.

Analysis was conducted by the lead researcher as an interpretive process; consistent with reflexive thematic analysis, intercoder agreement was not used as a criterion of analytic quality. Reflexivity was maintained by actively retaining negative experiences, minimal change, residual anxiety, and technological constraints during theme development and review, reducing the risk of overemphasizing positive experiences with AI. FNE and the concept of technology affordance were used as sensitizing concepts during interpretation process, rather than as predetermined coding categories. Because some topics were also elicited through interview probes, interpretation emphasized elaboration and cross-case patterns rather than frequency of mention. Reported quotations were initially translated using DeepL and checked against the original transcripts for meaning equivalence.

Integration occurred through connecting and merging. Quantitative posttest results informed interview sampling, and quantitative findings and qualitative themes were subsequently brought together during interpretation and in the joint display (Figure 1). Integration focused on complementarity and expansion, using individual experiences to contextualize group-level findings. The meta-inference was developed across both strands without treating positive AI-mediated experiences as evidence of broader anxiety reduction.

3.5 Ethical Considerations

The study was conducted with institutional research permission. Before data collection, participants were informed about the purpose and procedures of the study, data use and confidentiality, as well as their right to withdraw from the study without academic consequences. Participation was voluntary, and informed consent was obtained before data collection. Data were anonymized before analysis and reporting; quantitative participants were coded E1-E25 and C1-C20, and interview participants as Informant 1-Informant 12. Data were used solely for research and academic reporting.

4. RESULTS

4.1 Quantitative Findings

4.1.1 Descriptive Statistics and Instrument Reliability

The speaking anxiety questionnaire showed good internal consistency at both measurement points. Cronbach's alpha for the control group was .947 for the pretest and .948 for the posttest, while for the experimental group, the values were .855 and .934, respectively. Descriptive statistics showed that the mean anxiety scores decreased slightly in both groups. In the control group, the mean score decreased from 83.30 (SD = 19.11) to 82.00 (SD = 19.42), while in the experimental group, it decreased from 81.72 (SD = 10.71) to 80.64 (SD = 15.80).

Table 1. Instrument reliability and descriptive statistical analysis of anxiety levels

Group	Time	<i>n</i>	Number of items	Cronbach's α	M	SD	ΔM
Control	Pretest	20	24	.947	83.30	19.11	
Control	Posttest	20	24	.948	82.00	19.42	-1.30
Experimental	Pretest	25	24	.855	81.72	10.71	
Experimental	Posttest	25	24	.934	80.64	15.80	-1.08

Note: ΔM = posttest – pretest; a negative value indicates a decrease in anxiety scores. Higher scores indicate higher levels of speaking anxiety.

4.1.2 ANCOVA Results

ANCOVA was used to compare the posttest anxiety scores of the two groups after controlling for pretest scores. The assumption checks revealed no significant violations. The interaction between pretest scores and group was not significant ($p = .532$), so the assumption of homogeneity of regression slopes was met. The residuals also showed no significant deviations from normality (Shapiro-Wilk, $p = .412$) or homogeneity of variances (Levene's test, $p = .289$). No outliers or cases that exerted an excessive influence on the model were found.

After controlling for pretest scores, the adjusted posttest mean was 81.19 (95% CI [76.30, 86.08]) in the control group and 81.29 (95% CI [76.91, 85.66]) in the experimental group. ANCOVA indicated that the difference in adjusted posttest scores between the two groups was not significant, $F(1, 42) < .001$, $p = .977$, $\eta^2 < .001$. The group coefficient estimate also showed a very small difference: $B = 0.095$, $SE = 3.254$, $t = 0.029$, 95% CI [-6.47, 6.66]. Thus, after accounting for differences in pretest scores, no evidence of a difference in posttest anxiety between the two groups was found.

Table 2. ANCOVA Results for Posttest Anxiety Scores

Group	Adjusted M	95% CI
Control	81.19	[76.30, 86.08]
Experimental	81.29	[76.91, 85.66]

Group effect: $F(1, 42) < .001$, $p = .977$, $\eta^2 < .001$; $B = 0.095$, $SE = 3.254$, 95% CI [-6.47, 6.66].

4.2 Qualitative Findings

The interviews provided a more contextual understanding of the participants' affective experiences during AI-mediated speaking practice. Cross-case analysis yielded three interrelated themes: (1) lower social-evaluative pressure during AI-mediated practice, (2) feedback becoming consequential through learner uptake and repeated practice, and (3) confidence gains that were situated rather than uniformly generalized across speaking contexts. Factors such as individual variability, residual anxiety, exercise intensity, and technical barriers were analysed as interrelated cross-cutting conditions, rather than as independent themes. Given that several dimensions – such as feedback, self-confidence, and anxiety – were also explored through follow-up questions during the interviews, the interpretation process focused on participants' elaborations, consistency across questions and cases, and evidence that either limited or clarified dominant patterns, rather than relying solely on the frequency of statements.

Theme 1: Lower social-evaluative pressure in AI-mediated practice

The first theme suggests that one of the most significant differences between practicing with AI and specific human-mediated speaking situation is the reduction in interpersonal evaluative pressure. Factors such as errors in language structure or accent, anxiety about being evaluated by instructors and peers, feelings of awkwardness, and mental stress caused by being observed by others were identified as sources of speaking discomfort. Interestingly, these social barriers diminished during interaction sessions with AI. These findings suggest that these psychological changes stem from a shift in how learners interpret their own spoken errors. Informant 6, for example, described the AI's response to grammar errors as: "It's okay ... *yang penting kamu sudah berusaha*." ("It's okay... the important thing is that you tried.")

That response was part of his explanation of why he gradually began to enjoy the exercises more. A similar sense of comfort was expressed by Informant 7, who felt free to interact with the AI because corrections were provided without any sense of judgment. In addition, the absence of the interlocutor's physical presence also had a positive effect; one participant noted that direct eye contact during in-person communication often triggers nervousness that disrupts concentration—an obstacle they did not encounter when interacting with AI technology.

Nevertheless, a reduction in evaluative pressure does not necessarily eliminate all sources of anxiety. Concerns about linguistic accuracy still linger for some participants, and the psychological comfort they experience while practicing with AI often does not carry over when they face oral exams, peer criticism, or presentations in front of professors. This cross-case phenomenon demonstrates that AI technology plays a greater role in reducing the perception of interpersonal social pressure for some individuals, rather than acting as a space that completely eliminates speaking anxiety.

Theme 2: Feedback Became Consequential through Learner Uptake and Repeated Practice

The second theme shows that feedback is not only available as an AI feature but, in some cases, is followed by action on the part of the learner. Participants described corrections to grammar, vocabulary, pronunciation, and word choice as part of the exercises, but the ways in which they used that feedback varied. In some cases, feedback was followed by checking, correction, retrying, or additional practice, while in other cases its use was more limited.

The clearest example was demonstrated by a participant who critically double-checked the AI's corrections against a grammar book, rather than accepting them at face value. This case demonstrates that technological feedback is not merely the final answer in a dialogue, but rather a stimulus that encourages students to engage in ongoing independent academic exploration.

Another participant described corrections as something that can be accepted without embarrassment, so that mistakes can be corrected and practice can continue. Repeated practice was also evident in the way some participants used their practice opportunities. Informant 1 described a change from having almost never practiced before:

“Aku tuh nggak pernah latihan ya, nggak pernah latihan.” (“I never practice, you know—I never practice.”). After participating in the program, the participant reported practising beyond the scheduled sessions and, at times, even daily because they had begun to *“ngerasain impact-nya”* (“feel the impact”).

Some participants took the initiative to explore new topics, practice tasks repeatedly, and interact with AI outside of program hours. However, this trend was not uniform, as some other participants tended to be passive, merely meeting the minimum attendance requirements for mandatory sessions without increasing the intensity of their practice. Thus, the qualitative evidence distinguishes between the availability of feedback and practice opportunities and participants' actions regarding those opportunities. Feedback becomes more consequential when participants use it for checking, correction, retrying, or further practice. However, because the frequency, duration, and quality of student engagement are not systematically measured, this pattern does not indicate that repeated use of a feature or practice automatically leads to changes in anxiety levels.

Theme 3: Confidence Gains Were Situated Rather Than Uniformly Generalized

The third theme indicates that participants' perceived confidence varies depending on the speaking setting. Some participants felt more confident, particularly when speaking with AI; some reported that this change began to be evident in classroom interactions; while others still experienced nervousness when facing presentations or human audiences.

Informant 11 made a clear distinction between the two situations, stating that they felt more at ease conversing with an AI than with other people. Informant 5 showed a similar pattern: hesitancy persisted in class because of worries about grammatical faults, while comfort and confidence rapidly increased when engaging with an AI. In contrast, Informants 6 and 8 reported that the changes they felt began to manifest in human-mediated contexts—for example, becoming bolder in responding in class or interacting with instructors and peers. However, these experiences represent self-reported changes, not objective measures of confidence transfer.

Most importantly, Informant 12 provided a deviant case from the generally more positive pattern, stating: “*nggak terlalu banyak membantu ... cuma ada beberapa vocabulary yang aku tahu aja*” (“It didn't help much ... there were only a few vocabulary items that I learned”). This case suggests that access to a practice environment with relatively low evaluative pressure and available feedback is not necessarily accompanied by increased perceived confidence.

Overall, the cross-case patterns suggest that confidence is best understood as situated and variable across speaking contexts. Positive experiences with AI-mediated practice among some participants remained primarily tied to the conditions of AI interaction; for others, these experiences were reported to have begun extending into the classroom; while other participants showed only partial or minimal changes. These accounts do not establish that confidence gained during AI-mediated practice transferred to human-mediated speaking, as cross-context changes were not independently measured.

These variations were also evident in how participants utilized the practice opportunities and in the challenges they encountered during their interactions with the AI. Some participants extended or added to their practice sessions, while others primarily attended the required sessions. Some participants also reported interactional or technological friction, such as interrupted conversations, AI responses that strayed from the topic, or difficulty conveying their intentions to the AI. Thus, lower evaluative pressure, feedback, and opportunities for repeated practice were not experienced or utilized uniformly. These variations are treated as boundary conditions that cut across all three themes, rather than as separate themes.

Overall, the qualitative results show that the positive affective experiences of the participants were mostly context-dependent and proximal. Some participants felt that AI-mediated practice offered less social-evaluative pressure, and when they took advantage of these chances, feedback and repeated practice gained significance. However, lingering worry, limited uptake, little perceived improvement, and technological friction were still noted, and confidence did not always extend to human-mediated speaking. Thus, the interviews primarily describe changes in the conditions and experiences of speaking practice, rather than evidence that broader speaking anxiety has generally decreased.

Figure 1. Integration of Quantitative and Qualitative Findings and Meta-Inference



Note: The qualitative themes represent participants' reported experiences of AI-mediated speaking practice rather than independently measured changes across speaking contexts. Boundary conditions cut across the three themes and are not treated as a separate theme.

Figure 1 integrates quantitative and qualitative findings during the interpretation phase. Quantitative analysis showed that no between-group differences in speaking anxiety were detected on the posttest after controlling for baseline anxiety, whereas interviews revealed lower social-evaluative pressure, variations in the use of feedback and opportunities for repeated practice, and situational self-confidence during or after speaking practice with AI. These two findings are complementary, not contradictory, as they capture different levels of affective experience. The questionnaire measures broader speaking anxiety, while the interviews capture participants' more specific experiences regarding the context within and across speaking situations. The resulting meta-inference is that proximal positive experiences during practice with AI can coexist with the absence of detectable differences in broader speaking anxiety between the two groups.

5. DISCUSSION

The findings indicate that the relationship between AI-mediated speaking practice and speaking anxiety is more complex than the assumption that AI use directly reduces anxiety. After baseline anxiety was controlled through ANCOVA, no between-group difference in posttest anxiety was detected ($p = .977$, $\eta^2 < .001$). Yet the interviews revealed more positive affective experiences during AI-mediated practice, particularly lower social-evaluative pressure, opportunities to use feedback and practise repeatedly, and greater comfort and confidence. These experiences varied across participants and did not consistently extend to human-mediated speaking situations. The central finding, therefore, is not that AI succeeded or failed in reducing speaking anxiety, but that positive affective experiences during AI-mediated practice can coexist with no detectable between-group difference in broader speaking anxiety.

5.1 No Detectable Between-Group Difference in Speaking Anxiety

The quantitative findings showed no between-group difference in posttest speaking anxiety after baseline anxiety was controlled. This differs from studies reporting anxiety reductions following AI-mediated speaking interventions (Ding & Yusof, 2025; Ebadi et al., 2025; Qu & Chen, 2026; Wang & Wen, 2025; Yan et al., 2026). Qu and Chen (2026), for example, found a greater reduction in speaking anxiety with voice-based GenAI than with peer interaction. The present

findings are closer to Huang and Kakham (2026) and Ballıdağ and Aydın (2024) who found no significant between-group differences despite participants reporting positive experiences with AI. Together, these findings suggest that affective outcomes of AI-mediated speaking vary across pedagogical and implementation conditions.

The findings should also be considered in relation to the context-sensitive nature of speaking anxiety. Foreign language anxiety has been conceptualized as situation-specific, with fear of negative evaluation reflecting concern about others' judgments (Horwitz et al., 1986). Aldukhayel and Namaziandost (2022) also discovered that anxiety varied depending on whether the presentation was done in person or remotely. Therefore, lower anxiety in more general speaking settings does not necessarily follow from more comfort in one interactional environment. At the same time, the absence of a between-group difference should not be interpreted as evidence that AI offered no affective benefit. Given the relatively small sample, the ANCOVA indicates only that this study detected no posttest difference between conditions after controlling for baseline anxiety.

Moreover, the experimental group received additional AI-mediated speaking practice whereas the control group received regular instruction. The design therefore cannot separate AI-specific effects from additional time-on-task, repeated task exposure, individualized feedback, scheduling flexibility, novelty, or expectancy effects. These alternative explanations must be considered when interpreting participants' positive experiences.

5.2 Lower Social-Evaluative Pressure as a Situated Experience

The first qualitative theme emphasizes a feature that set AI-mediated practice apart from some human interaction methods: less social-evaluative pressure. Concerns about errors, peer or lecturer criticism, embarrassment, and, in certain situations, the audience's presence or gaze were all mentioned by participants. Some of these social cues were diminished by AI-mediated practice. This pattern can be explained by fear of negative evaluation (Horwitz et al., 1986). Certain forms of socially evaluative anxiety may also decrease when a human interlocutor and the potential for interpersonal judgment become less prominent.

This interpretation is consistent with previous research. Aldamen et al. (2026) found that learners experienced AI-mediated speaking as psychologically safer and more supportive, with reduced concerns about interpersonal judgment. In Indonesia, Noe et al. (2025) similarly described AI as a relatively low-pressure speaking partner, while Rizal (2025) identified non-judgmental practice, feedback, and confidence among learners' valued experiences. AI's affective value may therefore lie partly in how it changes the social and interactional conditions surrounding errors, feedback, and repetition rather than in the technology alone.

The present findings, however, qualify this interpretation. Lower evaluative pressure during AI interaction was not equivalent to lower anxiety across speaking contexts. Some participants remained nervous during presentations or when speaking before lecturers or human audiences. AI is therefore better understood here as providing a lower-stakes interactional condition for some learners, rather than as an environment that inherently removes anxiety. This distinction helps explain how positive interview accounts could coexist with no detectable between-group difference in broader speaking anxiety.

5.3 Feedback, Repeated Practice, and the Attribution Problem

The second theme demonstrates that possibilities for repetition and feedback were more than just AI features; some participants actively utilized them by asking for corrections, attempting again, practicing more, or cross-referencing feedback with other sources. From the standpoint of technology affordance, affordances are opportunities for action that result from relationships between users, situations, and technological features. In this study, affordance uptake occurred when learners acted on such possibilities through correction, retry, or additional independent practice. In the context of ChatGPT, Zhong et al. (2025) similarly show how affordances are shaped within a broader social-ecological system.

This interpretation is consistent with research identifying feedback and repetition as important affordances of AI-supported language learning. Bhar (2026) emphasizes the promptness, consistency, and repeatability of AI-mediated feedback, while Crompton et al. (2024) list conversational partnership, language coaching, speaking practice, and self-regulation help. According to Rizal (2025), prompt feedback can facilitate error recognition, correction, and more autonomous practice in Indonesia. The current research adds to this body of work by differentiating between availability and uptake: learners did not use feedback and repetition in the same manner or to the same level just because they had access to the same AI capabilities.

These patterns do not establish that affordance uptake reduced anxiety. Practice intensity, feedback quality, and the relationship between engagement and anxiety change were not systematically measured, and some participants practised beyond the scheduled sessions. Perceived benefits may therefore reflect AI-mediated feedback alongside additional time-on-task, repeated exposure, or self-directed practice. Future studies could examine these relationships using anxiety measures alongside usage logs, practice frequency and duration, and evidence of how learners act on feedback.

5.4 Context-Dependent Confidence and the Limits of Generalization

The third theme indicates that participants' perceived confidence gains were situated rather than uniformly generalized. Some explicitly felt more confident speaking with AI than with humans, while others reported that greater comfort had begun to extend to classroom participation. Still others remained nervous during presentations or audience-facing situations, and some reported minimal change. This variation is consistent with speaking anxiety as sensitive to interactional and evaluative conditions rather than stable across situations.

Previous research supports this interpretation. Aldukhayel and Namaziandost (2022) showed that anxiety can differ across presentation conditions, while Susoy (2026) found that the relationship between speaking anxiety and performance differed between human- and AI-facilitated assessment conditions. Qu and Chen (2026), despite reporting anxiety reduction in a GenAI condition, also noted limitations in AI's representation of some paralinguistic cues present in human communication. Social-evaluative features reduced during AI-mediated practice may therefore reappear when learners return to human interaction.

These patterns support a distinction between a proximal affective experiences during AI-mediated practice and their possible generalization across speaking context. While the latter focuses on whether such feelings persist when interactional contexts change, the former refers to comfort, reduced evaluative pressure, or confidence felt during AI-mediated practice. This disparity does not suggest that transfer or generalization were examined separately in this study.

Instead, it clarifies why there was no detectable between-groups-difference in posttest speaking anxiety despite individuals reporting favorable AI-mediated experiences. Therefore, without specific cross-context testing, confidence during AI engagement should not be taken to represent generalized speaking confidence.

5.5 Mixed-Methods Interpretation: Complementarity Across Levels of Experience

Mixed-methods integration is best understood in terms of complementarity and expansion, rather than convergence or contradiction. Quantitatively, no between-group difference in posttest speaking anxiety was detected after controlling for baseline anxiety. Qualitatively, participants described lower evaluative pressure, differential uptake of feedback and repeated practice, and confidence that varied across contexts. The two strands therefore capture different but complementary levels of the phenomenon: broader group-level anxiety and context-specific affective experience.

This interpretation also helps contextualize apparently inconsistent findings in previous research. Qu and Chen (2026) reported measurable anxiety reduction alongside positive GenAI experiences, whereas Huang and Kakham (2026) and Ballıdağ and Aydın (2024) found positive AI-related experiences without significant between-group differences in overall speaking anxiety. The present study provides an empirical illustration of why proximal affective experiences and broader anxiety change should be distinguished.

The resulting meta-inference is that positive proximal experiences during AI-mediated speaking practice are not equivalent to generalized reductions in speaking anxiety. The qualitative findings expand rather than overturn the quantitative result: interviews illuminate what participants experienced within AI-mediated practice, while ANCOVA indicates that these experiences did not coincide with a detectable between-group difference in broader speaking anxiety. The contribution is therefore not to classify AI as successful or unsuccessful in reducing anxiety, but to distinguish the affective conditions experienced during AI-mediated practice from broader affective change across speaking contexts.

Pedagogically, these findings support positioning AI as a complement rather than a replacement for human speaking interaction. Lower-stakes AI practice may provide opportunities for rehearsal, experimentation, feedback, and repetition before learners encounter peer, small-group, or more evaluative speaking situations. This sequencing remains a pedagogical implication rather than an effect tested in the present study.

5.6 Limitations and Directions for Future Research

There are a number of limitations to this study. First, the experimental condition contained extra AI-mediated speaking practice that was not matched in the control group for practice time, repeated task exposure, feedback, autonomy, or schedule flexibility. Additionally, the two intact classes were not assigned at random. As a result, the design is unable to separate AI-specific impacts from other intervention elements or further practice. Additionally, the sample was recruited from a single institution and was somewhat small, which limited statistical accuracy and generalizability. Because the sample was chosen based on available intact classes, no a priori power analysis was performed; as a result, the study may not have been able to identify small between-group effects. Larger samples and comparative circumstances that are matched for practice time and instructional exposure should be used in future research.

Second, the questionnaire measured speaking anxiety relatively broadly, whereas the interviews revealed variation across AI-mediated and human-mediated speaking contexts. Context-specific affective experiences during AI practice may therefore not be fully reflected in overall anxiety scores. Future research could combine broader anxiety measures with situational measures across AI-mediated, peer, classroom presentation, and other human-evaluative speaking conditions.

Third, interviews were conducted only after the intervention, preventing direct examination of how participants' affective experiences developed over time. Session records verified completion but did not capture the intensity or quality of engagement. Longitudinal designs combining repeated interviews, behavioral logs, and repeated affective measures could examine how affordance uptake relates to anxiety and confidence over time. Although research data were anonymized, the study did not implement or evaluate AI-specific protocols governing how participants' interaction data were processed or stored by the platform. Future studies should address platform-level privacy and data governance more explicitly.

Fourth, the cross-language adaptation of the anxiety instrument did not include back-translation or formal expert-panel review, and the pilot sample was insufficient for examining its construct structure. Further research should therefore evaluate the psychometric validity of the adapted version in larger samples.

Finally, the eight-week intervention did not include a delayed posttest. The study therefore cannot determine whether participants' positive affective experiences during AI-mediated practice persisted after the intervention or generalized more broadly to human-mediated speaking. Future studies could incorporate delayed measures and designs that progressively connect AI rehearsal with peer, small-group, and classroom speaking to examine retention and cross-context affective generalization more directly.

6. CONCLUSION

This study examined whether eight weeks of additional AI-mediated speaking practice was associated with differences in speaking anxiety relative to regular instruction, while also exploring learners' affective experiences during AI-mediated practice. After adjusting for baseline anxiety, ANCOVA found no between-group difference in posttest speaking anxiety, despite a small decline in anxiety levels in both groups. However, the interviews revealed that participants reported confidence increases that varied depending on speaking circumstances, frequently felt less social-evaluative pressure, and interacted with feedback and repeated practice in diverse ways.

Taken together, these findings suggest that positive affective experiences during AI-mediated practice do not necessarily correspond to broader reductions in speaking anxiety. By separating proximal affective sensations from more general measurable anxiety change and demonstrating how the two can coexist without being contradictory, the work advances research on AI-mediated language acquisition. Therefore, AI-mediated practice may serve as a lower-pressure setting for preparing learners for human-mediated speaking while broader affective changes need to be assessed across speaking contexts and over time.

Acknowledgments

This study was funded by the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia (Kemdiktisaintek) through the Penelitian Dosen Pemula (PDP) Research Grant for the 2026 fiscal year. The research assistants' assistance with the interviews and the students' voluntary involvement are both appreciated by the writers.

Conflict of Interest

Regarding the conduct, authorship, and publishing of this work, the authors disclose no conflicts of interest.

Ethical Considerations and Informed Consent

After learning about the goals and methods of the study, participants voluntarily participated and provided informed consent prior to the start of data collection. Throughout the study, their privacy and identities were maintained, even when the results of the interviews were made public.

7. REFERENCES

- Aldamen, H., Almashour, M., Al-Deaibes, M., & Alsharefeen, R. (2026). AI as a reflective partner in EFL speaking: a qualitative study informed by cyborg psychology. *Frontiers in Psychology*, 17. <https://doi.org/10.3389/fpsyg.2026.1786617>
- Aldukhayel, D., & Namaziandost, E. (2022). Remote Presentations: Making L2 Presentations Less Stressful. *Hindawi: Education Research International*, 2022(Cmc). <https://doi.org/10.1155/2022/5353819>
- Ballıdağ, M & Aydin. (2024). A comparison of the effects of AI-based chatbots and peer interactions on speaking anxiety among EFL learners. *Wiley*, 2025 (3)(September), 224–238. <https://doi.org/DOI: 10.1002/fer3.70005>
- Bhar, S. K. (2026). Artificial Intelligence in EFL Speaking Instruction: A Systematic Review of Pedagogical Design, Affective Conditions and Instructional Input. *Encyclopedia*, 6(4). <https://doi.org/10.3390/encyclopedia6040074>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2019). *Reflecting on reflexive thematic analysis*. *Qualitative research in sport, exercise and health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, 55(6), 2503–2529. <https://doi.org/10.1111/bjet.13460>
- Dewaele, J., & Dewaele, L. (2020). *Are foreign language learners' enjoyment and anxiety specific to the teacher? An investigation into the dynamics of learners' classroom emotions*. 10(1), 45–65. <http://dx.doi.org/10.14746/ssllt.2020.10.1.3>
- Ding, D., & Yusof, A. M. B. (2025). Speaking Anxiety : a Mixed Methods Study. *Humanities and Social Sciences Communications*, 12(2025), 1–16. <https://doi.org/10.1057/s41599-025-05550-z>
- Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and*

- Education: Artificial Intelligence*, 6(April), 100230.
<https://doi.org/10.1016/j.caeai.2024.100230>
- Ebadi, S., Velayati, S., Ramezanzadeh, A., & Rawdhan Salman, A. (2025). Exploring the impact of AI-powered speaking tasks on EFL learners' speaking performance and anxiety: An activity theory study. *Acta Psychologica*, 259(April), 105391.
<https://doi.org/10.1016/j.actpsy.2025.105391>
- Ekizer, F. N. (2025). Exploring the impact of artificial intelligence on English language teaching: A meta-analysis. *Acta Psychologica*, 260(2025), 105649.
<https://doi.org/10.1016/j.actpsy.2025.105649>
- Hanafiah, W., Aswad, M., Sahib, H., Yassi, A. H., & Mousavi, M. S. (2022). The Impact of CALL on Vocabulary Learning, Speaking Skill, and Foreign Language Speaking Anxiety: The Case Study of Indonesian EFL Learners. *Education Research International*, 2022.
<https://doi.org/10.1155/2022/5500077>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign Language Classroom Anxiety. *The Modern Language Journal*, 70(2), 125–132. <https://doi.org/10.2307/327317>
- Huang, L., & Liu, M. (2025). The impact of technology on foreign language anxiety: A systematic review of empirical studies from 2004 to 2024. *Humanities and Social Sciences Communications*, 12(1), 1–19. <https://doi.org/10.1057/s41599-025-05921-6>
- Huang, Y., & Kakham, P. (2026). The impacts of AI conversational agents on EFL learners' oral proficiency and foreign language speaking anxiety. *Frontiers in Education*, 11(April), 1–16.
<https://doi.org/10.3389/educ.2026.1799269>
- Hussein Hammodeh Ahmad Al-Khotaba, Eid Hamoudeh Ahmed Alkhataba, Samsiah Abdul-Hamid, I. B. (2019). Foreign Language Speaking Anxiety: A Psycholinguistic Barrier Affecting Speaking Achievement of Saudi EFL Learners. *Arab World English Journal (AWEJ) Volume*, 10(4), 313–329. <https://dx.doi.org/10.24093/awej/vol10no4.23>
- Noe, S., Yusuf, F. N., & Wirza, Y. (2025). ChatGPT as a Speaking Partner: Investigating Its Impact on Students' Speaking Anxiety. *JL3T (Journal of Linguistics, Literature and Language Teaching)*, 11(2), 190–199. <https://doi.org/10.32505/jl3t.v11i2.12646>
- Qu, Z., & Chen, L. (2026). Simulating the “cultural other”: the impact of GenAI interlocutors on EFL learners' intercultural communicative competence, speaking anxiety, and learning engagement. *Frontiers in Psychology*, 17(March).
<https://doi.org/10.3389/fpsyg.2026.1799695>
- Rizal, R. P. (2025). Students' Perceptions of AI-Assisted Feedback in Speaking Practice of Higher Education Students. *IJEAL (International Journal of English and Applied Linguistics)*, 5(3), 490–496. <https://doi.org/10.47709/ijeal.v5i3.7290>
- Soriano-Gonzalez, L., & Belda-Medina, J. (2026). Chatbots for oral proficiency in English language learning: A systematic review. *Telematics and Informatics Reports*, 23(2026), 100346. <https://doi.org/10.1016/j.teler.2026.100346>
- Susoy, Z. (2026). Reducing anxiety and enhancing performance: the impact of AI chatbots versus human facilitation on EFL speaking assessment outcomes. *Frontiers in Psychology*, 16(January), 1–19. <https://doi.org/10.3389/fpsyg.2025.1745942>
- Wang, X., & Wen, Y. (2025). The role of artificial intelligence tools on Chinese EFL learners' speaking skills, speaking anxiety, and motivation. *Acta Psychologica*, 261(August), 105957. <https://doi.org/10.1016/j.actpsy.2025.105957>
- Wiboolyasarini, W., Wiboolyasarini, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and

- pedagogical implications. *Ampersand*, 14(May), 100224.
<https://doi.org/10.1016/j.amper.2025.100224>
- Yan, H., Kaur, M., & Singh, S. (2026). The impact of AI-mediated instruction on speaking proficiency, enjoyment, anxiety, and emotional engagement: a mixed-methods approach. *Humanities and Social Sciences Communications*, 13(568). <https://doi.org/10.1057/s41599-026-06705-2>
- Zhong, Y., Lian, J., Huang, H., & Deng, H. (2025). Affordances, constraints, and implications of ChatGPT in education from a social-ecological perspective: A data mining approach. In *Education and Information Technologies* (Vol. 30, Number 12).
<https://doi.org/10.1007/s10639-024-13237-2>