
RESEARCH ARTICLE

ChatGPT as a Conversational Partner in an Undergraduate English Speaking Skills Course

Mohammad Alolayan

Associate Professor, Department of English Language and Literature, College of Languages and Humanities, Qassim University, Saudi Arabia

Corresponding Author: Mohammad Alolayan, **E-mail:** m.alolaiyan@qu.edu.sa

ABSTRACT

Undergraduate English majors in Saudi Arabia often complete speaking courses with limited opportunities for sustained, low-anxiety oral practice. This study explored the use of ChatGPT as an out-of-class conversational partner in a Speaking Skills course taught by the researcher at Qassim University during the second semester of the 2023–2024 academic year (Spring 2024). Thirty-six male undergraduate students in two intact sections of the course took part: one section ($n = 18$) completed twelve structured weekly conversation sessions with ChatGPT through its mobile application, while the other section ($n = 18$) completed the same weekly topics through conventional pair-work practice with classmates. Speaking fluency was measured before and after the intervention through monologue and dialogue tasks analysed for speech rate, mean length of run, and pause-time ratio. Willingness to Communicate (WTC) was measured with an adapted 27-item questionnaire covering general, in-class, and out-of-class WTC. Semi-structured interviews and weekly reflective journals provided qualitative data. Consistent with the exploratory aim of the study, quantitative data were analysed with descriptive statistics only (means, standard deviations, mean gains, and percentage change), and qualitative data were analysed thematically. Both sections improved, but the ChatGPT section showed considerably larger gains: speech rate rose by 38.3% compared with 11.2%, mean length of run by 56.9% compared with 16.1%, and total WTC by 39.7% compared with 14.3%. Learners described ChatGPT as a patient, non-judgemental, and always-available partner that lowered their speaking anxiety, although they continued to value human interaction. The findings suggest that structured ChatGPT practice can be a useful supplement to classroom speaking instruction in Saudi undergraduate programmes, and they offer a practical protocol that course instructors can adapt. Because the analysis is descriptive and the sample small, the results indicate trends rather than generalisable effects.

KEYWORDS

ChatGPT; speaking skills; speaking fluency; Saudi undergraduate students; mobile-assisted language learning

ARTICLE INFORMATION

ACCEPTED: 02 September 2026

PUBLISHED: 25 October 2026

DOI: 10.32996/jeltal.2026.8.11.3

1. Introduction

Speaking remains one of the most demanding skills for learners of English as a Foreign Language (EFL), particularly where learners have little contact with English outside the classroom. In Saudi Arabia, the communicative goals of the Vision 2030 education reforms have placed new emphasis on oral proficiency, yet undergraduate English majors continue to report difficulty in speaking fluently and confidently in spontaneous interaction (Aljumah, 2011; Alrabai, 2022). Speaking courses in Saudi English departments are typically constrained by class size and limited contact hours, so each student receives only a few minutes of actual speaking time per lecture (Alrasheedi, 2020).

Two constructs are central to speaking development in such settings. *Speaking fluency* refers to the ability to produce speech smoothly and at an appropriate rate, and it is commonly measured through speech rate, pause behaviour, and mean length of run (Tavakoli & Wright, 2020). *Willingness to Communicate* (WTC) refers to a learner's readiness to enter into discourse in the second language at a particular time with a specific person (MacIntyre et al., 1998). Learners who have adequate linguistic knowledge may still avoid speaking if their WTC is low, which makes WTC a practical concern for any speaking course (Lee & Lee, 2020).

Copyright: © 2026 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (<https://creativecommons.org/licenses/by/4.0/>). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

Mobile technologies have already become part of how Saudi college students learn. Alaghbari (2021) examined the perceptions and practices of both learners and teachers regarding mobile-assisted language learning (MALL) in Saudi college education and drew attention to the gap between the wide availability of mobile devices and their structured pedagogical use. Similarly, Alolayan (2025) showed that carefully designed technology-supported input—captioned authentic videos within a flipped blended environment—improved the English vocabulary of Saudi university students. Both studies suggest that technology yields learning gains when it is integrated into a course through a clear instructional design rather than left to students' informal use.

Generative artificial intelligence (AI), and ChatGPT in particular, extends this possibility to speaking. ChatGPT can sustain open-ended conversation, adjust its language to the learner, and respond at any time and without visible impatience (Kohnke et al., 2023). Early research, however, has concentrated on writing and on learner perceptions (Xiao & Zhi, 2023; Yan, 2023), and studies on speaking remain few, especially in Arabic-speaking contexts and within the real constraints of an ordinary university course.

The present study responds to this gap. It reports on the integration of ChatGPT as an out-of-class conversation partner in a Speaking Skills course that the researcher taught to two sections of undergraduate English majors at Qassim University in the second semester of 2023–2024 (Spring 2024). The study is exploratory and classroom-based: it describes how the learners' fluency and WTC changed over the semester and how they experienced the tool. It is guided by three research questions:

RQ1: How did the speaking fluency of undergraduate students in a Speaking Skills course change after a semester of structured conversational practice with ChatGPT, compared with conventional pair-work practice?

RQ2: How did the students' Willingness to Communicate in English change in the two practice conditions?

RQ3: How did the students perceive ChatGPT as a conversational partner, and what factors shaped their engagement with it?

The study is significant at two levels. Empirically, it adds classroom-based evidence from an under-represented Arabic-speaking context to a literature dominated by East Asian studies. Pedagogically, it provides speaking instructors in Saudi English departments with a tested weekly protocol, including the prompt used, that can be adopted within existing course plans.

2. Literature Review

2.1 Theoretical Framework

Three complementary perspectives inform the study. Sociocultural theory holds that learning develops through interaction with a more capable other who provides scaffolding within the learner's zone of proximal development (Vygotsky, 1978). Although the theory was formulated for human interaction, it has been extended to interaction with technological tools. From this perspective, ChatGPT can act as a "more knowledgeable other" that models language, reformulates learners' utterances, and adjusts its contributions to the learner's level (Kohnke et al., 2023). Because it is available on demand, it also removes the scheduling and social constraints that limit peer interaction.

The second perspective concerns the decision to speak. MacIntyre et al. (1998) proposed a pyramid model in which WTC results from layered influences, from stable individual differences such as personality and motivation to situational variables such as the interlocutor, topic, and setting. Perceived communicative competence and communication anxiety are the most immediate predictors of the decision to speak. Lee and Lee (2020) found that EFL learners' WTC is shaped by the affective climate of the learning setting, with less threatening environments supporting greater willingness to speak. WTC also fluctuates from one situation to another within the same classroom (Zhang et al., 2022). A non-evaluative conversational agent may therefore provide the low-anxiety situation the model associates with communicative engagement.

The third perspective concerns practice itself. Swain's (2005) Output Hypothesis holds that producing language pushes learners to process syntax, test hypotheses, and notice gaps in their interlanguage. Skill acquisition theory adds that repeated, deliberate practice turns declarative knowledge into procedural knowledge and, eventually, automatic performance, which underlies fluent speech (DeKeyser, 2020). Conversational practice with ChatGPT offers frequent opportunities for both output and repetition outside class time.

2.2 Speaking Fluency and Willingness to Communicate in the Saudi Context

Lennon (1990) distinguished a broad sense of fluency, close to overall proficiency, from a narrow sense concerned with the smoothness and speed of speech. Segalowitz (2010) later separated cognitive fluency, utterance fluency, and perceived fluency. Utterance fluency is the most directly measurable, and Tavakoli and Wright (2020) identify speech rate, mean length of run, and pause-time ratio as reliable indicators. De Jong (2018) notes that these temporal measures capture both the speed and the breakdown dimensions of fluency and are therefore appropriate for classroom-based assessment. Fluency develops through sustained practice, which Saudi speaking courses struggle to provide within scheduled class time (Aljumah, 2011).

Willingness to speak is a particular concern in Saudi classrooms. Studies of Saudi learners report that communication anxiety, fear of negative evaluation, and limited prior speaking experience restrict oral participation even among learners whose reading and

writing are adequate (Aljumah, 2011; Alrasheedi, 2020). Alrabai (2022) showed that teachers can shape learners' emotions in the classroom, and that interventions addressing negative emotions can improve engagement. Activities that lower the stakes of speaking and provide experiences of successful communication are therefore likely to support WTC in this context.

2.3 Mobile Learning and Generative AI in Language Learning

Saudi research on MALL has moved from questions of access to questions of pedagogy. Alaghbari (2021) investigated both learners' and teachers' perceptions and practices of MALL in Saudi college education and showed that mobile devices are already embedded in students' everyday routines; the study's main implication is that their educational value depends on how deliberately teachers integrate them into instruction. Alolayan (2025) provides an example of such integration: in a quasi-experimental study, Saudi university students who watched captioned authentic videos in a flipped blended environment outperformed a control group on vocabulary measures. The present study builds on both works by moving from receptive vocabulary to productive speaking, and from video input to interactive AI conversation delivered through a mobile application.

Generative AI adds an interactive dimension to mobile learning. Kohnke et al. (2023) identified several affordances of ChatGPT for language learning, including personalised feedback, models of contextually appropriate language, on-demand availability, and extended interaction, while cautioning that these depend on how the tool is integrated and on learners' digital literacy. Xiao and Zhi (2023) and Yan (2023) reported positive but mixed experiences in writing contexts, including concerns about over-reliance. Mohamed (2023) found that EFL teachers welcomed ChatGPT's potential while worrying about accuracy and academic integrity.

Evidence on speaking is emerging. Tai and Chen (2023) found that interaction with an intelligent personal assistant increased adolescent EFL learners' WTC, and Fathi et al. (2024) reported that AI-mediated speaking activities improved university learners' speaking skills and WTC more than face-to-face peer activities. These studies support the potential of conversational AI for speaking but leave open how such tools work inside a regular speaking course in an Arabic-speaking university.

This literature leaves four gaps that motivate the present study. First, research on ChatGPT has focused mainly on writing. Second, few studies come from Saudi or other Arabic-speaking settings. Third, few studies are embedded in an ordinary course and conducted by the course instructor over a full semester, which is the setting in which most teachers would actually use the tool. Fourth, the affective experience of speaking with an AI partner has received limited attention. This study addresses these gaps through a semester-long classroom study that combines temporal fluency measures, a WTC questionnaire, and qualitative data.

3. Methodology

3.1 Design, Context, and Participants

The study used an exploratory, classroom-based mixed-methods design with a quasi-experimental component: a pretest–posttest comparison of two intact sections of the same course. Random assignment of individual students was not possible because students register in sections through the university system. The type of out-of-class speaking practice (ChatGPT conversation versus conventional pair work) was the condition of interest; speaking fluency and WTC were the outcome measures. In line with the exploratory purpose and the size of the two sections, the quantitative data were analysed descriptively, and the study does not claim statistical generalisation (see Section 3.4).

The study took place in the Department of English Language and Literature, College of Languages and Humanities, Qassim University, during the second semester of the 2023–2024 academic year (Spring 2024). The setting was the department's Speaking Skills course ENG 328, a required course in the BA programme whose outcomes include speaking fluently on familiar and academic topics, sustaining conversation, and expressing and defending opinions. The course meets for two contact hours per week over a 15-week semester and combines class discussion, presentations, and role-play. The researcher was the instructor of record for both sections, so curriculum, materials, contact hours, and assessment were identical across them.

Participants were 36 male Saudi undergraduate English majors enrolled in the two sections of the course. All were native speakers of Arabic who had studied English at school for at least nine years and had passed the programme's prerequisite language courses. They were aged 19 to 22 ($M = 20.4$, $SD = 0.8$). One section was designated as the ChatGPT group (CHG, $n = 18$) and the other as the comparison group (CMG, $n = 18$) by a coin toss. Participation required enrolment in the course, a smartphone able to run the ChatGPT application, no prior systematic use of ChatGPT for language learning. All 36 students in the two sections met these conditions and completed all measures. Pretest scores (Tables 1 and 3) show that the two sections started from similar levels.

Because the researcher also taught the course, several steps were taken to limit researcher bias. Journals and interview responses carried no weight in course grades. The fluency measures were computed from audio recordings using acoustic software rather than from the instructor's impressionistic ratings, and a second trained rater checked the pause annotation of 20% of the recordings (see Section 3.2). The end-of-semester interviews were conducted by a departmental colleague who did not teach the course, and the researcher analysed the data only after final course grades had been submitted. At the same time, the researcher's familiarity with the learners and the course is acknowledged as a source of insight in interpreting the qualitative data.

3.2 Instruments

Fluency was assessed at pretest (Week 1) and posttest (Week 14) through two tasks: (a) a four-minute monologue ("Describe a place that is important to you and explain why"), and (b) a five-minute paired dialogue in which each student discussed a debatable statement with a trained interlocutor. Both tasks were audio-recorded. Three measures were computed in Praat (Boersma & Weenink, 2023), a freely licensed acoustic analysis program: (1) speech rate, the number of syllables per minute of total speaking time; (2) mean length of run, the average number of syllables between silent pauses of 250 milliseconds or longer; and (3) pause-time ratio, the proportion of speaking time occupied by silent pauses (Tavakoli & Wright, 2020). Scores from the two tasks were averaged. The second rater's pause boundaries agreed with the researcher's in 93% of cases, and disagreements were resolved by discussion.

WTC was measured with a 27-item questionnaire adapted from MacIntyre et al. (2001), covering general WTC in English (10 items), WTC inside the classroom (9 items), and WTC outside the classroom (8 items). Items were rated on a five-point scale from 1 (*almost never willing*) to 5 (*almost always willing*). The Arabic version was reviewed by two bilingual applied linguists for conceptual equivalence. Internal consistency (Cronbach's alpha) was .92 at pretest and .94 at posttest.

Qualitative data came from two sources. After the posttest, all 18 CHG students and six CMG students were interviewed individually in Arabic. The protocol had 12 open questions on experiences with the practice activity, perceived benefits and limitations, emotional responses, and suggestions. Interviews lasted 25–40 minutes and were recorded and transcribed.

In addition, CHG students kept weekly reflective journals describing their sessions, feelings, perceived progress, and difficulties. Entries were submitted electronically each week.

3.3 Procedures

From Week 2 to Week 13, CHG students completed one structured 30-minute session per week with ChatGPT, using the free version of the official mobile application (GPT-3.5), giving twelve sessions or about six hours of practice. Each session followed four stages: warm-up and small talk (5 minutes); guided conversation on the week's course topic (15 minutes); learner-led free conversation (5 minutes); and a short written reflection (5 minutes). Topics followed the course plan and included personal experiences, local culture, campus life, current events, and ethical dilemmas.

Students opened every session with the same prompt, refined in two pilot sessions: *"I am a Saudi undergraduate student practising English speaking. Please act as a friendly conversation partner. Talk with me naturally about [topic]. Use clear vocabulary, ask me follow-up questions, and gently correct my major errors. Be patient and encouraging."* All sessions were conducted in the application's two-way voice conversation mode, which had been available to free users since late 2023 (OpenAI, 2023), so students spoke and listened throughout without typing. Students uploaded a screenshot of each conversation history to the course page so that completion could be monitored. Students were advised not to enter personal information into their conversations.

The comparison group attended the same course, taught by the same instructor, with identical in-class activities. Instead of ChatGPT sessions, they completed weekly 30-minute pair-work conversations with assigned classmates outside class, using the same weekly topics and the same four-stage structure, and submitted a short audio excerpt of each conversation.

Overall, in Week 1 both groups completed the fluency tasks and the WTC questionnaire. The practice sessions ran from Week 2 to Week 13. The posttests were administered in Week 14, and the interviews took place in Weeks 15 and 16. Journals were collected throughout the practice period.

3.4 Data Analysis

Quantitative data were analysed descriptively in Microsoft Excel. For each measure and group, the analysis reports the mean (*M*) and standard deviation (*SD*) at pretest and posttest, the mean gain (posttest minus pretest), and the percentage change relative to the pretest mean. Comparison between the groups is based on the size and direction of these descriptive differences. No inferential tests were conducted: the study was designed as an exploratory classroom inquiry with two intact sections of 18 students, and the aim was to describe change within a real course rather than to estimate population effects.

Qualitative data from interviews and journals were analysed through reflexive thematic analysis (Braun & Clarke, 2019): familiarisation, initial coding, theme generation, review, definition, and reporting. A second coder independently coded 25% of the data; the two coders agreed on 88% of code assignments, and differences were resolved through discussion. Five participants reviewed a summary of the themes (member checking). The number of participants who expressed each theme is reported to show how widely it was shared.

4. Results

4.1 RQ1: Speaking Fluency

Table 1 presents the descriptive statistics for the three fluency measures.

Table 1

Descriptive Statistics for Speaking Fluency Measures

Measure	CHG Pre M (SD)	CHG Post M (SD)	CMG Pre M (SD)	CMG Post M (SD)
Speech rate (syllables/min)	118.4 (14.2)	163.7 (15.8)	120.1 (13.7)	133.5 (14.6)
Mean length of run (syllables)	4.32 (0.81)	6.78 (0.92)	4.41 (0.78)	5.12 (0.86)
Pause-time ratio	0.41 (0.07)	0.27 (0.06)	0.40 (0.08)	0.35 (0.07)

Note. CHG = ChatGPT group ($n = 18$); CMG = comparison group ($n = 18$). Lower pause-time ratio indicates less pausing.

At pretest the two groups were close on every measure; the largest difference (1.7 syllables per minute in speech rate) was about one-eighth of a standard deviation. Table 2 summarises the change in each group.

Table 2

Mean Gains and Percentage Change in Speaking Fluency

Measure	CHG gain	CHG change	CMG gain	CMG change
Speech rate	+45.3	+38.3%	+13.4	+11.2%
Mean length of run	+2.46	+56.9%	+0.71	+16.1%
Pause-time ratio	-0.14	-34.1%	-0.05	-12.5%

Note. Gain = posttest mean – pretest mean; change = gain as a percentage of the pretest mean.

Both groups improved on all three measures, but the gains of the ChatGPT group were roughly three times as large. Its speech rate rose by 45.3 syllables per minute (38.3%), compared with 13.4 (11.2%) in the comparison group. Mean length of run increased by 2.46 syllables (56.9%) against 0.71 (16.1%), and pause-time ratio fell by 0.14 (34.1%) against 0.05 (12.5%). At posttest, the difference between the group means in speech rate (30.2 syllables per minute) was about twice the size of either group's standard deviation, indicating that the two distributions overlapped only slightly.

4.2 RQ2: Willingness to Communicate

Table 3 presents the WTC results by dimension, and Table 4 the corresponding gains.

Table 3

Descriptive Statistics for Willingness to Communicate

Dimension	CHG Pre M (SD)	CHG Post M (SD)	CMG Pre M (SD)	CMG Post M (SD)
General WTC	2.84 (0.61)	3.92 (0.54)	2.78 (0.58)	3.21 (0.62)
WTC inside class	2.71 (0.66)	3.85 (0.59)	2.75 (0.64)	3.08 (0.67)
WTC outside class	2.92 (0.71)	4.04 (0.61)	2.88 (0.69)	3.31 (0.65)
Total WTC	2.82 (0.59)	3.94 (0.51)	2.80 (0.57)	3.20 (0.60)

Note. Scale range 1–5; higher scores indicate greater willingness to communicate.

Table 4

Mean Gains and Percentage Change in Willingness to Communicate

Dimension	CHG gain	CHG change	CMG gain	CMG change
General WTC	+1.08	+38.0%	+0.43	+15.5%
WTC inside class	+1.14	+42.1%	+0.33	+12.0%
WTC outside class	+1.12	+38.4%	+0.43	+14.9%
Total WTC	+1.12	+39.7%	+0.40	+14.3%

Both groups began with moderate-to-low WTC (means around 2.8) and both improved. The ChatGPT group's total WTC rose by 1.12 points (39.7%), from below to near the "often willing" level, while the comparison group's rose by 0.40 points (14.3%). The largest difference between the groups was in WTC inside the classroom, where the ChatGPT group gained 1.14 points (42.1%) compared with 0.33 (12.0%). This pattern suggests that confidence built in private practice carried over into face-to-face classroom participation, a reading supported by the interview data below.

4.3 RQ3: Learners' Perceptions

Thematic analysis of the CHG interviews and journals produced five themes, four positive and one critical. Table 5 shows how many of the 18 students expressed each of the first four themes.

Table 5*Frequency of Themes Among ChatGPT Group Participants (n = 18)*

Theme	n	%
Reduced speaking anxiety	16	88.9
Always-available practice partner	18	100.0
Patient and adaptive interaction	15	83.3
Increased speaking confidence	14	77.8

The most widely shared theme was the reduction of speaking anxiety. Sixteen students said that speaking with ChatGPT made them less anxious, mainly because it did not judge them:

"When I speak with ChatGPT, I don't feel embarrassed when I make mistakes. With a human, even my friends, I always worry about how they will judge my English." (CHG-P7)

Closely related was the value of an always-available partner. All 18 students appreciated being able to practise at any hour, which addressed the scarcity of English interlocutors outside class:

"I practise with ChatGPT whenever I have free time, sometimes late at night. There is no other way I could find an English conversation partner at 1 a.m." (CHG-P12)

Students also described the interaction as patient and adaptive. Fifteen of them noted that ChatGPT simplified its language on request and never showed impatience:

"If I don't understand something, I can ask ChatGPT to use simpler words, and it does so without any irritation. With my classmates, I sometimes feel they are losing patience with me." (CHG-P3)

As a result, fourteen students reported growing confidence, which several linked to more frequent participation in class:

"After practising with ChatGPT every week, I started raising my hand in class to answer in English. Before this, I would never volunteer to speak." (CHG-P9)

At the same time, students were aware of the tool's limitations: occasional unsuitable vocabulary, limited understanding of Saudi cultural references, occasional speech-recognition errors, and, above all, the absence of a human connection:

"ChatGPT is a great practice partner, but at the end of the day it is not a real person. I still need to speak with people." (CHG-P15)

The six CMG students interviewed found pair work useful but said it was limited by their partners' availability and motivation and by worry about classmates' judgement. Several asked to try ChatGPT practice after hearing about the other section's experience.

5. Discussion

This study described how undergraduate students in a Speaking Skills course responded to a semester of structured ChatGPT conversation practice. The ChatGPT section showed markedly larger descriptive gains than the comparison section in all fluency measures and in all WTC dimensions, and the learners' own accounts are consistent with these numbers. Because both sections were taught by the same instructor with the same materials, the main difference between them lay in the type of out-of-class practice. Nevertheless, given the descriptive design and the small intact groups, the findings are best read as indicative trends.

With regard to fluency, the larger gains in the ChatGPT group are consistent with skill acquisition theory (DeKeyser, 2020): about six hours of extra, focused speaking practice gave learners repeated opportunities to proceduralise their knowledge. Pair work in the comparison group offered a similar nominal amount of time, but interviews suggest that it was often shortened or interrupted by scheduling problems and by partners switching to Arabic, so the effective amount of English output was probably lower. The results agree with those of Fathi et al. (2024), who found AI-mediated speaking activities more effective than peer interaction, and they extend that finding to a Saudi undergraduate course.

The most notable result, however, is the larger gain in in-class WTC. In terms of MacIntyre et al.'s (1998) model, private practice with a non-evaluative partner appears to have raised learners' perceived competence and lowered their anxiety, which then affected their willingness to speak in front of classmates and the instructor. This mirrors Tai and Chen's (2023) finding with an intelligent personal assistant and supports Lee and Lee's (2020) emphasis on a non-threatening affective climate. For Saudi learners, for whom fear of negative evaluation is a recurring barrier (Aljumah, 2011; Alrasheedi, 2020), such a rehearsal space may be particularly valuable.

More broadly, the results support the argument that technology produces learning gains when it is built into course design. Alaghbari (2021) highlighted the gap between Saudi college students' everyday mobile use and its structured pedagogical use, and Alolayan (2025) showed that a deliberately designed captioned-video environment improved vocabulary. In the present study,

ChatGPT was not simply recommended to students: it was tied to weekly course topics, framed by a common prompt, organised into fixed stages, and monitored. This design probably explains much of the benefit, and it is the element instructors can most easily transfer.

These findings should be read in light of several limitations. First, the sample comprised 36 male students in one course at one university, so the findings cannot be generalised. Second, the analysis was descriptive by design; differences between the groups were not tested statistically, and no causal claim is made. Third, the groups were intact sections and, despite similar pretest scores, may have differed in unmeasured ways. Fourth, the researcher taught both sections; although safeguards were used (Section 3.1), instructor expectations may have influenced the learners. Finally, no delayed posttest was given, so it is not known whether the gains were retained.

Despite these limitations, the study has practical value. For instructors of speaking courses in Saudi English departments, it suggests that (a) ChatGPT practice can be assigned as a structured weekly out-of-class component linked to course topics; (b) a common, level-appropriate prompt helps keep the conversations pedagogically useful; (c) short reflections and simple monitoring (such as uploading conversation histories) sustain engagement; and (d) AI practice should supplement rather than replace human interaction, which learners continue to value.

6. Conclusion

This classroom study explored ChatGPT as a conversational partner in a Speaking Skills course for Saudi undergraduate English majors. Over one semester, the section that practised with ChatGPT showed considerably larger descriptive gains in speaking fluency and Willingness to Communicate than the section that practised in pairs, and learners described the tool as patient, available, and non-judgemental, while still valuing human conversation. The study contributes classroom-based evidence from a Saudi context and a ready-to-use weekly protocol for speaking instructors. Future research should involve larger and mixed-gender samples across several institutions, apply inferential analyses to confirm the trends observed here, include delayed posttests, and examine how AI practice can best be combined with human interaction in speaking courses.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

ORCID iD: <https://orcid.org/0009-0003-5787-3216>

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Alaghbari, G. S. (2021). Mobile-assisted language learning in Saudi college education: Exploring learners and teachers' perceptions and practices. *Asian ESP Journal*, 17(3.1), 8–31.
- [2] Aljumah, F. H. (2011). Developing Saudi EFL students' oral skills: An integrative approach. *English Language Teaching*, 4(3), 84–89. <https://doi.org/10.5539/elt.v4n3p84>
- [3] Alolayan, M. (2025). The effect of captions in a flipped blended learning environment on developing EFL vocabulary among Saudi university students. *Journal of Arabic Sciences and Humanities*, 18(4), 392–416. <https://doi.org/10.71422/0198-018-004-013>
- [4] Alrabai, F. (2022). The role of mixed emotions in language learning and teaching: A positive psychology teacher intervention. *System*, 107, Article 102821. <https://doi.org/10.1016/j.system.2022.102821>
- [5] Alrasheedi, S. (2020). Investigation of factors influencing speaking performance of Saudi EFL learners. *Arab World English Journal*, 11(4), 66–77.
- [6] Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* [Computer software]. University of Amsterdam. <https://www.praat.org>
- [7] 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>
- [8] De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, 15(3), 237–254. <https://doi.org/10.1080/15434303.2018.1477780>
- [9] DeKeyser, R. (2020). Skill acquisition theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (3rd ed., pp. 83–104). Routledge. <https://doi.org/10.4324/9780429503986-5>

- [10] Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, Article 103254. <https://doi.org/10.1016/j.system.2024.103254>
- [11] Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- [12] Lee, J. S., & Lee, K. (2020). Affective factors, virtual intercultural experiences, and L2 willingness to communicate in in-class, out-of-class, and digital settings. *Language Teaching Research*, 24(6), 813–833. <https://doi.org/10.1177/1362168819831408>
- [13] Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 387–417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- [14] MacIntyre, P. D., Baker, S. C., Clément, R., & Conrod, S. (2001). Willingness to communicate, social support, and language-learning orientations of immersion students. *Studies in Second Language Acquisition*, 23(3), 369–388. <https://doi.org/10.1017/S0272263101003035>
- [15] MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545–562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- [16] Mohamed, A. M. (2023). Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies*, 29, 3195–3217. <https://doi.org/10.1007/s10639-023-11917-z>
- [17] OpenAI. (2023, September 25). *ChatGPT can now see, hear, and speak*. <https://openai.com/index/chatgpt-can-now-see-hear-and-speak/>
- [18] Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge. <https://doi.org/10.4324/9780203851357>
- [19] Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Lawrence Erlbaum.
- [20] Tai, T.-Y., & Chen, H. H.-J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, 31(3), 1485–1502. <https://doi.org/10.1080/10494820.2020.1841801>
- [21] Tavakoli, P., & Wright, C. (2020). *Second language speech fluency: From research to practice*. Cambridge University Press. <https://doi.org/10.1017/9781108589109>
- [22] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- [23] Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3), Article 212. <https://doi.org/10.3390/languages8030212>
- [24] Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28, 13943–13967. <https://doi.org/10.1007/s10639-023-11742-4>
- [25] Zhang, J., Beckmann, N., & Beckmann, J. F. (2022). One situation doesn't fit all: Variability and stability of state willingness to communicate in a Chinese college English classroom. *Language Teaching Research*, 26(3), 504–529. <https://doi.org/10.1177/1362168819891330>