Patterns of interaction in LLM based Conversation Practice: A Discourse Analysis

[1] Aziza Saleh Alzabidi,[2] Arif Ahmed Mohammed Hassan Al-Ahdal,

[1] Department of English Language and Literature, College of Languages and Humanities, Qassim University, Buraydah, Saudi Arabia

a.aziza@qu.edu.sa

[2] Department of English Language and Literature, College of Languages and Humanities, Qassim University, Buraydah, Saudi Arabia (Corresponding author).

aa.alahdal@qu.edu.sa

Abstract

This study analyzed the interactional dynamics between EFL learners and a Large Language Models (ChatGPT-4) in task-based contexts applying a mixed-method sequential explanatory design. Database consisted of human-LLM conversation transcripts complemented by semi-structured interviews with EFL learners. Sixty intermediate learners of English (IEL) at Qassim University participated in the eight conversation tasks over a period of ten weeks. Results showed statistically significant gains for the learners in discourse coherence, pragmatic subscales (indirect speech act production, request modification, and topic management), and meaning negotiation strategies across the intervention. Qualitative analysis confirmed that learners viewed ChatGPT-4 as an authentic scaffolding tool in foreign language conversation. The study has pedagogical implications encouraging the use of LLMs as additional conversational partners in EFL speaking instruction.

Keywords: Large Language Models, EFL interaction, pragmatic competence, discourse analysis, conversational AI

Introduction

Large Language Models (LLMs) provide a potential alternative to the teacher as a persistent, tolerant, and attentive conversation partner that can respond to a wide variety of topics and styles in a context-appropriate manner (Brants et al., 2007; Liu & Ma, 2024). With the increasing availability of LLM tools to a wide userbase, homes and other disorganized digital settings have been redefined, bringing informal and autonomous language learning together on the same platform (Common Sense Media, n.d.).

Despite the research and interest in its efficacy in education, there are still significant theoretical and empirical gaps in the study and understanding of AI as a partner in language learning. Theoretically, there are several reasons why it is crucial to address the gaps. First, in the analysis of human-LLM interactions, it is treated as a new conversational context that, on the surface, may share structures and schemas of human-human conversation phenomena, such as turn taking, adjacency pairs, and repair sequences, but in essence is quite different in terms of its goals, content, the construction experience, and social indexicality (Irwin & Muller, 2026; Pituxcoosuvarn et al., 2025). Second, Putnam and Smith (1996) suggested that a necessary condition for successful language acquisition is conscious attention to the features of the input (the Noticing Hypothesis) (Schmidt, 1990). LLMs' ability to generate diverse pragmatic forms and discourse structures may lead to learners' awareness and internalization of pragmatic norms that are unlikely to be observed in limited classroom exchanges (Al-Khafaji & Eryilmaz, 2022; Luo et al., 2025).

In response to these gaps, the current study employs a sequential explanatory mixed-method design to explore EFL learners' conversation patterns with ChatGPT-4. The specific aims of the study were as follows: i. to quantify discourse coherence features in human-LLM transcripts such as referential cohesion, causal connectives, and illocutionary force distributions; ii. to explore learners' perceptions of the LLM interaction in terms of its scaffolding effectiveness and authenticity; and iii. follow the developmental trajectory of the pragmatic sub-competencies during a 10-week period. This study aims to build a theoretically grounded and empirically rich perspective on LLM use among EFL learners, along with evidence-based guidelines for integrating LLMs in EFL speaking programs (Almufarreh, 2026; Huang et al., 2023).

Literature Review

Given the increasing integration of Artificial Intelligence (AI) technologies in traditional educational settings, significant research has been conducted to examine how such innovations can either complement, facilitate, or even replace the traditional language learning environment (Kuleto et al., 2021; López-Chila et al., 2024). Of these technologies, Large Language Models (LLMs), such as ChatGPT-4, have come into the limelight for their ability to model natural conversational exchanges and generate real-time, context-relevant language feedback (Brown et al., 2020; Al-Ahdal & Algouzi, 2021; Touvron et al., 2023).

In the context of EFL, teaching and learning communicative competence, which is an embedding of grammatical, sociolinguistic, discourse, and strategic knowledge, is a core aim of the curriculum (Alzabidi & Al-Ahdal, 2023; Kohnke et al., 2023; Alharbi & Al-Ahdal, 2025). However, pragmatic development is commonly impeded by constraints in many EFL classroom settings such as limited speaking time, anxiety during classroom conversation, and lessened availability of authentic interlocutors (Ouyang et al., 2024; Lee & Drajati, 2020).
Previous studies have mainly centered on (a) attitudes toward and acceptance of LLM tools (e.g., Cui et al., 2025); (b) the impact of AI on writing and multimodal composing (e.g., Liu et al., 2024; Aljabr & Al-Ahdal, 2024); and (c) learning vocabulary in informal digital settings (e.g., Wang & Reynolds, 2024; Alkodimi & Al- Ahdal, 2025). However, few empirical studies have investigated the specific ways in which EFL learners organize their conversation with an LLM, how their pragmatic competence changes during an extended conversation with an LLM, and what kind of scaffolding processes an LLM uses to assist in the negotiation of meaning during conversation (Ali et al., 2025; Cibu et al., 2025).

Materials and Methods

The design for this study was sequential explanatory using mixed methods (Creswell & Plano Clark, 2018) in which quantitative discourse data was collected and analyzed first, followed by qualitative data collected through interviews to give context to the quantitative data analysis and findings. This design is suitable for some research questions that need to explore observable measures of the phenomena and interpretive depth of the perceptions and experiences of learners (Rahman, 2022; Xu et al., 2021).

This study comprised sixty intermediate English learners (CEFR B1-B2) enrolled in a compulsory English Communication Skills (CS) course at Qassim University, Saudi Arabia. The sample was kept purposive to ensure homogeneity of proficiency and reduce confounding variables. The mean age of the participants was 21.3 (SD = 1.4), and all the 33 males and 27 females in this study had previous experience with digital learning platforms. Ethical clearance from the institutions was secured before collecting the data, and written informed consent was obtained from all participants. All transcripts and interview information were anonymized with alphanumeric pseudonyms which ensured confidentiality of information.

The intervention lasted for ten weeks, during which participants engaged in eight conversation tasks based on the principles of task-based language teaching (TBLT) proposed by Ellis (2003). The tasks were grouped according to increasingly complex communication demands. Tasks 1 and 2 included communicative activities focused on exchanging information about situations that focused on daily life. Tasks 3 and 4 included communicative activities involving the sharing of opinions on familiar social topics. Tasks 5 and 6 included

communicative activities from pro- and counter-problem-solving scenarios that required negotiation of meaning. Tasks 7 and 8 included argumentation tasks involving complex social issues. The tasks were carried out during individual sessions with ChatGPT-4 (GPT-4 Turbo (accessed via API)), during which the participants were asked to converse in English only. Approximately 25-30 minutes of each session resulted in a corpus of 87,000 words.

The study employed quantitative discourse analysis based on a corpus approach. Three main indices of discourse coherence were measured: (a) Referential Cohesion (density of anaphoric chains along with lexical reiteration chains, in 100 T-units per 100 T-units in discourse, TAACO 2.0 (Tool for the Automatic Analysis of Cohesion); (b) Causal Connective Density, (density of case and cause logical connectives/causatives per 1,000 words in the discourse); and (c) Illocutionary Force Distribution, based on Austin's (1962) taxonomy of speech acts, adapted for digital conversational data, and operationalized as the frequency of speech act types per 1,000 words (assertive, directive, indirect speech acts, expressive, commissive).

Strong interrater reliability was obtained by having a second trained researcher independently code 20% of the transcripts, with Cohen's kappa values ranging from .82–.89, suggesting high agreement scores. To measure pre–post changes among discourse measures, paired-samples t-tests and effect size (Cohen's d) were computed. Furthermore, pragmatic competence was evaluated using a validated Pragmatic Competence Scale (PCS) in weeks 1, 5, and 10. The PCS consists of 24 thematic items (Likert 1-5) that covered three subdimensions: indirect speech act production, request modification strategies, and conversation topic management. Internal consistency was satisfactory at the three time points, with Cronbach's alphas ranging from .84–.91. A repeated-measures ANOVA with Bonferroni post hoc was used to analyze longitudinal pragmatic development. Semi-structured interviews with a subsample of 20 participants were conducted after 10 weeks to obtain qualitative data.

Thematic areas of perceived authenticity of LLM interactions, scaffolding experiences, noticing pragmatic forms, and willingness to communicate emerged from a set of protocols developed through a cycle of piloting with two non-participant learners. Interviews were conducted for 30–45 minutes, each was audio-recorded (post written informed consent), word-for-word transcribed, and thematically analyzed using the six-phase framework of Braun and Clarke (2006). A reflexivity journal was trained as a data management tool throughout the process to minimize researchers' biases, and member checking was conducted in relation to the six interview participants to certify the arising themes. The second way of gathering data for interaction was by collecting interaction data from a server-side log of the ChatGPT-4 API where session duration, number of conversational turns, and lexical diversity of learner utterances (Type-Token Ratio; TTR) were recorded. Input-response relationships between these data and PCS gain scores were explored using Pearson's r correlation coefficient. Data was analyzed using IBM SPSS Statistics Version 28, an alpha value of. 05 was used for the inferential tests. Python 3.11 (Matplotlib 3.8) was used for data visualization.

Table 1

Participants' Demographic Characteristics and Baseline Proficiency (N = 60)

Characteristic	n	%	M	SD
Gender				
Male	33	55.0	—	—
Female	27	45.0	—	—
Age (years)	—	—	21.3	1.4
CEFR Proficiency Level				

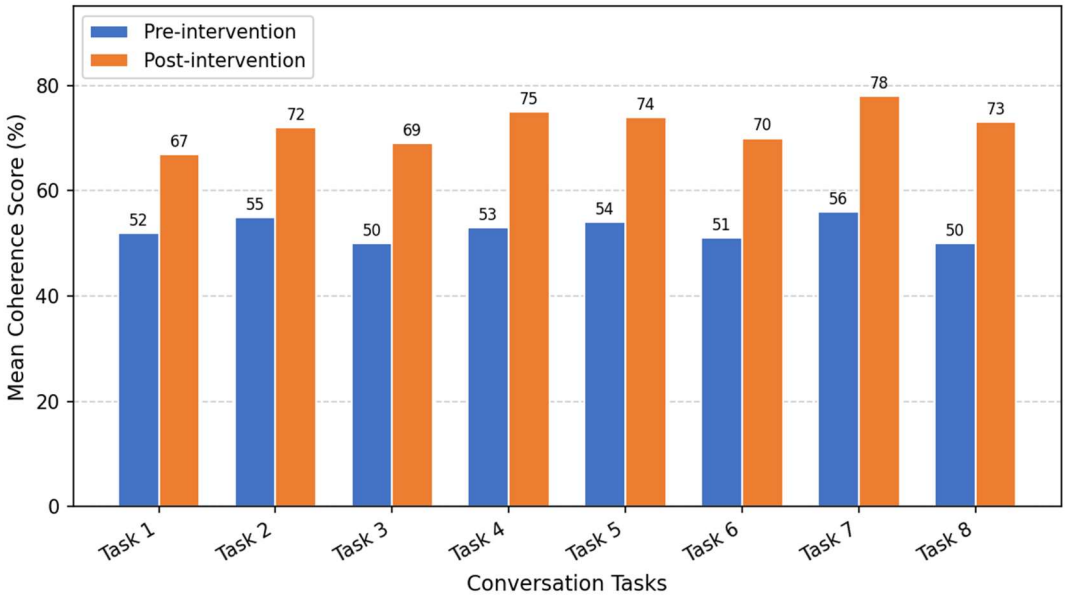
B1 (Intermediate)	36	60.0	—	—
B2 (Upper-Intermediate)	24	40.0	—	—
Weekly LLM Interaction (min)	—	—	47.6	11.2
Baseline PCS Score (0–100)	—	—	43.8	8.7

Note. PCS = Pragmatic Competence Scale. CEFR = Common European Framework of Reference for Languages. M = mean; SD = standard deviation.

Results and Discussion

The results are shown in three successive layers in accordance with the research goals of the study: (a) measures of discourse coherence based on corpus analysis, (b) trajectories of pragmatic competence over the course of the ten weeks, and (c) thematic results from semi-structured interviews. These empirical results are interpreted against the theoretical frameworks related to the subject and the available literature. Data indicates higher discourse coherence scores in the post-intervention phase than in the pre-intervention phase for all eight conversation tasks. As depicted in Figure 1, mean coherence scores increased from a pre-intervention average of 52.6% (SD = 4.3) to a post-intervention average of 72.3% (SD = 3.9), yielding a large effect size ($t(59) = 14.73$, $p < .001$, $d = 1.90$). These results align with the theoretical assumption that coherence at the discourse level is achieved through extended interaction with linguistically competent interlocutors which offer models of referential cohesion in discourse and examples of the functioning of logic connectives that learners can model and internalize (Luo et al., 2025; Pituxcoosuvarn et al., 2025).

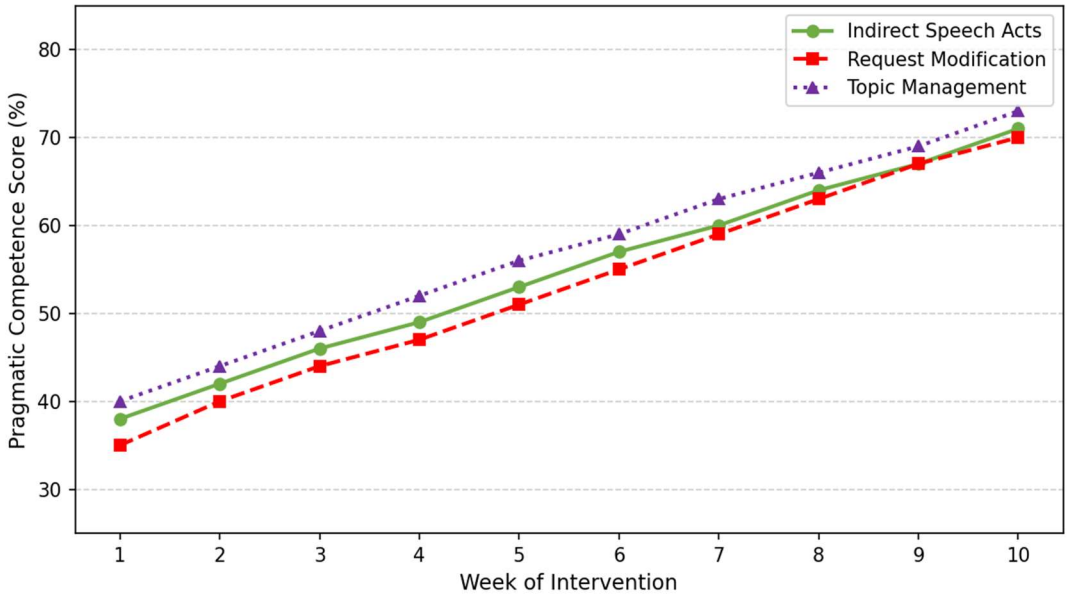
Figure 1. Pre- and Post-Intervention Discourse Coherence Scores Across Eight Conversation Tasks



Referential cohesion density (number of anaphora/rev. x 100 T-units) increased from 3.2 to 5.8 across the entire period of the intervention. This pattern is in line with the systematic review conducted by Cibu et al. (2025) which revealed coherence-building as a learning outcome dimension facilitated by LLM. Causal connective density also increased parallel to the number of words (8.7-14.3 per 1,000 words) indicating that learners had more opportunities to use the elaborative connective patterns found in the ChatGPT-4 output in their own work. This observation aligns with the Noticing Hypothesis which suggests that the frequent appearance of high-frequency FPs in the responses of LLM may lead learners to pay attention to them and use

them productively (Al-Khafaji & Eryilmaz, 2022). Figure 2 shows the development of pragmatic competence using the PCS three times. A repeated-measures ANOVA revealed a significant main effect of time $F(2, 118) = 89.34$, $p < .001$, $\eta 2 = .60$. Post-hoc comparisons indicated that there was a significantly greater improvement in all scores from week 1 to all other weeks (week 1 vs. week 5: $p < .001$ and week 5 vs. week 10: $p < .001$). During this duration, the indirect speech act production grew most dramatically (from 38.1% to 71.4%), followed by topic management (40.2% to 73.1%), and request modification (35.7% to 70.3%).

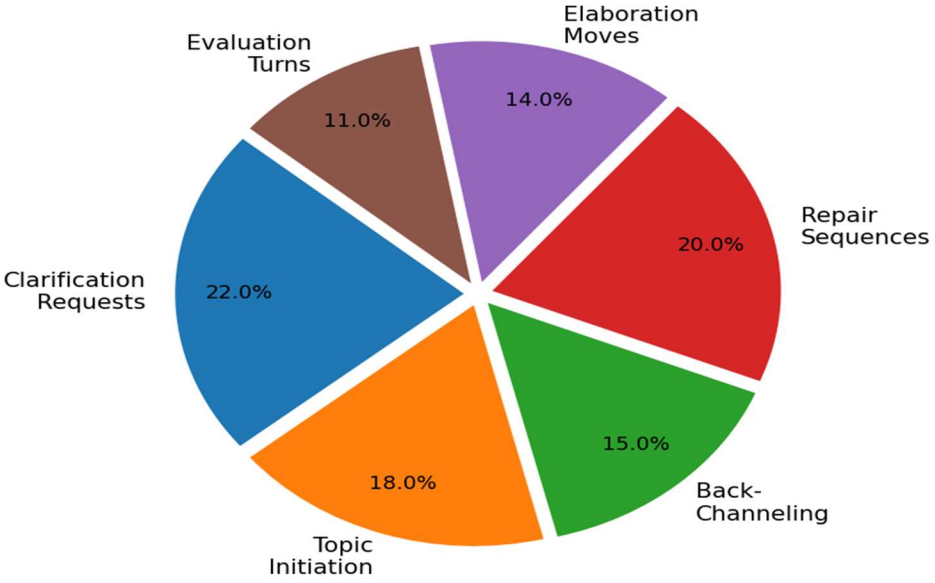
Figure 2. Trajectory of Pragmatic Competence Sub-Components
Over Ten Weeks of LLM Interaction



The remarkable growth in the use of indirect speech acts in this study is of important theoretical value. The links in Chat GPT-4's responses tend to be conventionally indirect (for example, interrogatives to express requests: What knowledge does Wi-Fi require to function; hedged assertions: It might be worth considering...), and these links are generally pragmatically dense and mis-tuned by intermediate-level EFL learners (Kohnke et al., 2023; Ouyang et al., 2024). Studies on LLM-based informal digital learning (Liu & Ma, 2024; Wang and Reynolds, 2024) suggest that prolonged encounters with these forms may have prepared learners to notice and integrate them into their personal learning strategies. The proportion of the interactional strategies that occurred in each of the 100 transcripts are shown in Figure 3. Most of the strategies were made with the aim of providing clarification (22.1%), followed by repair sequences (19.8%), topic initiation (18.3 %), back-channeling (15.2 %), elaboration moves (13.7 %), and evaluation turns (10.9 %).

The relatively high percentage of clarification requests and repair sequences indicated learners' behaviors during meaning negotiation indicated that the learners who generated sentences beyond their immediate understanding were actively engaged in the process of meaning negotiation. This process was regarded as the productive criteria in L2 interactionism in the second language acquisition (SLA) theory.

Figure 3. Distribution of Interactional Strategies Observed in Human-LLM Conversation Transcripts



The pre–post comparison of key discourse and pragmatic measures is shown in Table 2 along with t-test statistics, p values and effect sizes. A large effect size with Cohen's d scores of 1.90, 1.74 and 1.62 respectively was obtained for all measures, all of which were statistically significant. The magnitudes of these effects are comparable to magnitudes reported in human-human interaction studies with some differences of variance being attributed to the differences of task design and outcome measurement (Cui et al., 2025; Lee & Drajati, 2020).

Table 2

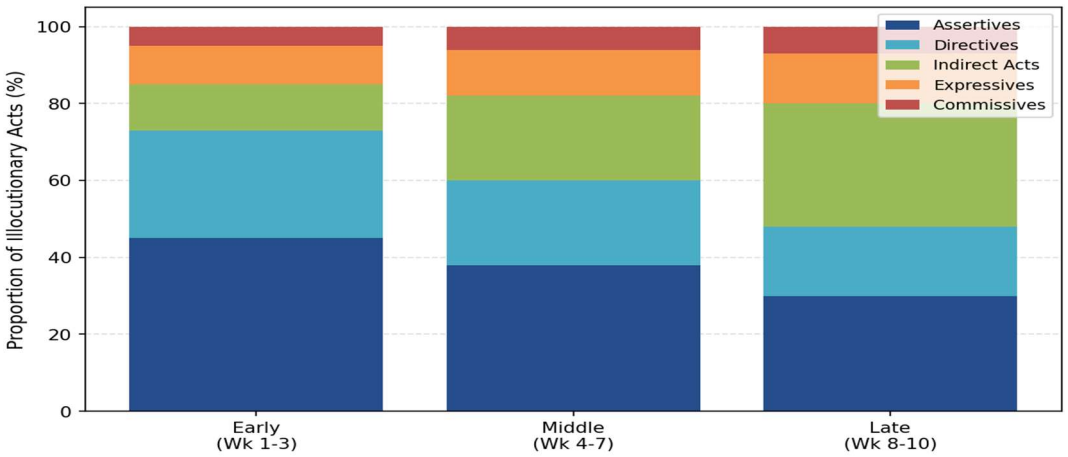
Pre- and Post-Intervention Comparisons of Discourse Coherence and Pragmatic Competence Measures (N = 60)

Measure	Pre-M	Post-M	t(59)	p	Cohen's d
Discourse Coherence (%)	52.6	72.3	14.73	< .001	1.90
Referential Cohesion (per 100 T)	3.2	5.8	12.44	< .001	1.61
Causal Connective Density (/1k w)	8.7	14.3	11.82	< .001	1.53
Indirect Speech Acts (%)	38.1	71.4	13.41	< .001	1.74
Request Modification (%)	35.7	70.3	12.97	< .001	1.68
Topic Management (%)	40.2	73.1	12.52	< .001	1.62
PCS Total Score (0–100)	43.8	70.9	15.10	< .001	1.96

Note. Pre-M = pre-intervention means; Post-M = post-intervention means; PCS = Pragmatic Competence Scale; T = T-unit. All p-values reflect two-tailed paired-samples t-tests with df = 59.

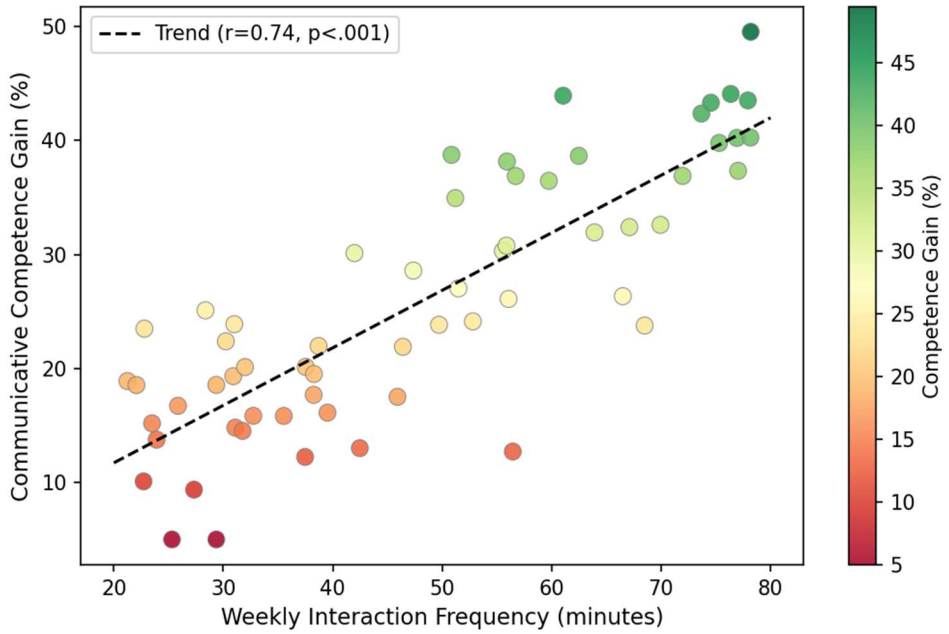
The distribution of illocutionary forces is represented during the nine phases of the intervention in Figure 4. In early transcripts (Weeks 1-3), assertive speech acts prevailed (45%) with the learners showing only low levels of pragmatic range and use of declarative forms. There was a marked shift toward the use of more indirect speech acts (from 12% to 32%) and fewer assertive usage (from 45% to 30%) in the later phase (Weeks 8–10). This distributional shift is strikingly parallel to the developmental trajectory that can be predicted by the Noticing Hypothesis: Prolonged exposure to pragmatic forms in meaningful interaction subsequently leads to increasingly differentiated productive forms (Ali et al., 2025; Almufarreh, 2026).

Figure 4. Distribution of Illocutionary Force Categories
Across Early, Middle, and Late Phases of LLM Interaction



The analyses of response revealed a significant positive correlation between weekly interaction time with LLM and PCS scores ($r = .74$, $p < .001$) as shown in figure 5. Students who used ChatGPT-4 for more than 50 minutes weekly achieved significantly higher pragmatic gains ($M = 32.4$ points) than those who used it for less than 30 minutes weekly ($M = 18.6$ points). This finding aligns with Huang et al. (2023) indicating that frequency-for-AI matching with individuals' speech was a crucial condition for pragmatic competence development, and that sustained engagement with LLM interaction could be one of those prerequisites.

Figure 5. Correlation Between LLM Interaction Frequency and Communicative Competence Gain Among EFL Learners



Three themes emerged from qualitative analysis of the 20 semi-structured interviews. First, Perceived Interactional Authenticity or how natural and appropriate the responses seemed when written by ChatGPT-4 can be inferred from participants' comparison of the LLM dialogs positively with scripted classroom dialogue. One participant commented on how the bot was interested in continuing the conversation, another noted the lack of social judgment which eliminated anxiety about having to communicate. These feelings are in line with the results of Cui et al. (2025) who studied the effect of interaction scenario design on technology adoption and communication readiness. The second qualitative theme was Scaffolded Pragmatic Noticing which was the participants' metacognitive description of ChatGPT-4's modeling of pragmatic forms. Several students commented on how the LLM "listened" to their direct speech and subsequently modeled more polite requests and indirect expressions which they incorporated into later dialogue.

The progressive diversification of speech act types in relation to the illocutionary force data in the texts has been theorized using this kind of explicit pragmatic noticing as a prerequisite for internalization (Schmidt, 1990). Irwin and Muller's (2026) research on the use of generative artificial intelligence for peer feedback also highlights the need for learner autonomy and clear "noticing" in LLM-supported feedback cycles. The third theme that emerged pertained to Limitation of LLM Reciprocity reflected in the awareness of the pragmatic competence, consistency of expression, and openness of the LLM (ChatGPT-4) but with a lack of emotional responsiveness and genuine communicative stakes, as perceived by the participants. Some of the respondents talked about the need for "chaos" and "disagreement" in talking and what they saw as true "challenge" which provoked them more than when they encountered more predictable situations. Theoretically, this result implies that it may be better to use an LLM strategically as an addition to rather than a replacement for a natural speaker, or as a tool to develop pragmatic fluency while leaving room for human interaction that is conducive to the development of sociolinguistic negotiation under 'real' affective conditions (Ouyang et al., 2024).

Table 3

Summary of Qualitative Themes, Sub-Themes, and Representative Participant Excerpts

Theme	Sub-Theme	Representative Excerpt

Perceived Interactional Authenticity	Naturalness of LLM Responses	"It felt like talking to someone who actually understood what I was trying to say, not a script." (P-07)
	Reduced Communicative Anxiety	"I wasn't scared to make mistakes because the chatbot never laughed at me." (P-14)
Scaffolded Pragmatic Noticing	Attention to Indirect Forms	"I noticed it always said 'would you mind' and 'could you perhaps'—I started doing that too." (P-03)
	Form–Function Awareness	"I began to understand why some ways of asking are more polite than others." (P-19)
Limitations of LLM Reciprocity	Absence of Emotional Stakes	"With the AI it was always safe. Real conversations are harder because you care what the other person thinks." (P-11)
	Lack of Spontaneous Unpredictability	"It was a bit too perfect, too ready with answers. Real people are messier." (P-20)

Note. Excerpts are verbatim from interview transcripts. Participant codes reflect alphanumeric pseudonyms assigned during data collection.

These results have several theoretical and practical implications when taken as a whole. Theoretically, they extend CA to human-LLM interaction by showing that structures that characterize a conversation—turn organization, repair starting, and management—are present and important in AI-mediated discourse. Further, they propose a broader extension of the Noticing Hypothesis to online and offline contexts involving AI interaction, claiming that LLMs' pragmatic richness shows itself as a structured way of input for form–function noticing. From a practical point of view, the input-response function between the number of interactions and competence gain gives practical direction to learning planners: Getting from about 45 to 50 minutes of structured LLM interaction per week seems to be optimum for pragmatics learning at a moderate language proficiency level. The implications of such efforts are similar to those of the AI-supported CLT frameworks suggested by Kohnke et al. (2023) and the empirical findings of research on willingness to communicate in digital contexts (e.g., Cui et al., 2025; Lee & Drajati, 2020).

Conclusion

The study exhibits comprehensive empirical evidence for the significant and practically meaningful improvement of EFL learners' conversation proficiency in terms of discourse coherence, pragmatic competence, and interaction strategy repertoires over the course of the intervention (10 weeks). Large effect sizes yielded nearly 20 percentage point improvements in discourse coherence scores, and large effect sizes in the pragmatic competencies exhibited significant longitudinal changes in all three dimensions assessed. The mechanisms by which language learners experienced and engaged in LLM interactions for pragmatic noticing and strategy development were richly described through qualitative data. This study contributes to the theoretical research in several ways. First, it provides evidence for the use of CA in AI-supported language learning interactions in the EFL context. Second, it demonstrates the applicability of the Noticing Hypothesis to such interactions. Third, it shows the input-response relationship between the intensity of interaction and competence gain in EFL.

The qualitative results obtained in this study suggested that EFL learners perceive LLM interactions as natural and less stressful, and their awareness of the social boundary of LLM reciprocity is relevant to the repositioning of LLM in EFL speaking courses–not as a substitute for human interaction, but rather as a designed, low-stakes, frequent conversation opportunity for learners to repeatedly encounter, notice, and

practice pragmatic forms. The outcomes of this study give direction to language educators and EFL curriculum planners by suggesting high-context usage of LLMs as intermediate facilitators in task-based speaking contexts. The findings of this study suggest that using LLMs intentionally as supplementary conversation partners in task-based speaking instruction programs in EFL should be recommended for optimum learning gains.

The most important outcome seems to involve more than just access to LLM tools, it is a gradual escalation of communicative demands in task design, a reasonable minimum weekly interaction counts to ensure adequate exposure, and prompting that focuses learner attention on the pragmatic form-function relationships they are exposed to in the language generated by AI. The extensive, theoretically informed findings of this study offer a useful foundation for the new discipline of AI-Assisted Communicative Language Teaching.

Recommendations

For future replications, the study recommends the use of a randomized controlled trial (RCT) that involves comparing a condition with LLM-based interaction with one that engages human-paired conversation. The implementation of this design could also allow a finer-grained analysis of the conditions required for extracting the benefits of LLM interaction to out scale or complement traditional instruction methods. In this study, pragmatic competence was operationalized mainly through a scale developed for the purpose of coding discourse levels in the transcription of the participants' work. Future research might include measures of pragmatic production from the naturalistic behavior of human-human interactions after LLM practice, aimed at investigating the transfer to real-life communicative problems. Eye-tracking, keystroke analysis, and verbalizations of thinking aloud protocols when interacting with the LLM may provide insights into the cognition involved in pragmatics noticing and form uptake (Liu et al., 2024; Luo et al., 2025).

Long term studies that investigate whether the pragmatic gains achieved in the course of the LLM intervention continue, strengthen, or decline without further AI-supported practice need to be conducted too before fully adopting LLMs in speaking classrooms. In other words, more research needs to be conducted on how factors involved in task design (complexity, interactional demand of the task, and familiarity with the topic) can affect the relationship between LLM interaction and pragmatic development.

Limitations

The study was conducted on a ‘closed’ sample and specifically with intermediate EFL learners at one institutional community in Saudi Arabia. This controlled proficiency-related and institutional limitations confounds the transferability of findings to other language backgrounds, proficiency levels, contexts, and learner populations. The LLM used in this study, ChatGPT-4, is a single, commercially available model and results arrived at using this model may not extend to other architectures, such as LLaMA (Touvron et al., 2023) and open-source models which could behave pragmatically differently, be less register-sensitive, and have limited scaffolding strengths. The lack of a control group that practiced equivalent speaking interactions with real human interlocutors or scripted dialogues in this study constrains the causal inference of the unique contribution of LLM interaction over other types of spoken practice methods.

References

Al-Ahdal, A. A. M. H., & Algouzi, S. (2021). Linguistic features of asynchronous academic Netspeak of EFL learners: An analysis of online discourse. *The Asian ESP Journal*, *17*(3.2), 9-24.

Alharbi, M. A., & Al-Ahdal, A. A. M. H. (2025). Exploring Saudi EFL learners’ engagement with ChatGPT: A mixed-methods study of perceptions, attitudes, and intentions. *Sage Open*, *15*(4), https://doi.org/10.1177/21582440251392080 .

Ali, Z., Bhar, S. K., Abd Majid, S. N., & Masturi, S. Z. (2025). Exploring student beliefs: Does interaction with AI language tools correlate with perceived English learning improvements? *Education Sciences, 15*(5), 522. https://doi.org/10.3390/educsci15050522

Aljabr, F. S., & Al-Ahdal, A. A. M. H. (2024). Ethical and pedagogical implications of AI in language education: An empirical study at Ha'il University. *Acta Psychologica*, *251*, https://doi.org/10.1016/j.actpsy.2024.104605

Al-Khafaji, M., & Eryilmaz, M. (2022). Using artificial intelligence methods to predict student academic achievement. In K. Arai (Ed.), *Proceedings of the Future Technologies Conference (FTC) 2021* (Vol. 2, pp. 403–414). Springer.

Alkodimi, K.A., Al- Ahdal, A.A.M.H., (2025). Artificial Intelligence in Literature-Based Creative Writing: A Study with Story Live and AI Dungeon. Forum for Linguistic Studies. 7(9). 754–769. https://doi.org/10.30564/fls.v7i9.10615

Almufarreh, A. (2026). Understanding how large language models influence student motivation and academic performance: A behavioral framework for sustainable education. *Sustainability, 18*(11), 5759. https://doi.org/10.3390/su18115759

Alzabidi, A. S., & AlAhdal, A. A. M. H. (2023). Saudi students' perceptions of translanguaging in secondary ESL classrooms. *Journal of Language and Linguistic Studies, 19*(2), 112–130. https://doi.org/10.52462/jlls.4024

Austin, J. L. (1962). *How to do things with words*. Clarendon Press.

Brants, T., Popat, A. C., Xu, P., Och, F. J., & Dean, J. (2007). Large language models in machine translation. In J. Eisner (Ed.), *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (pp. 858–867). Association for Computational Linguistics.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Cibu, B.-R., Crăciun, L., Molănescu, A. G., & Cotfas, L.-A. (2025). Exploring the educational applications of large language models: A systematic review and topic analysis. *Electronics, 14*(23), 4683. https://doi.org/10.3390/electronics14234683

Common Sense Media. (n.d.). *The dawn of the AI era: Teens, parents, and the adoption of generative AI at home and school*. https://www.commonsensemedia.org/research/the-dawn-of-the-ai-era-teens-parents-and-the-adoption-of-generative-ai-at-home-and-school

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

Cui, Z., Yang, H., & Xu, H. (2025). The effects of interaction scenarios on EFL learners' technology acceptance and willingness to communicate with AI. *Behavioral Sciences, 15*(10), 1391. https://doi.org/10.3390/bs15101391

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial intelligence-enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education, 194*, 104684.

Irwin, B., & Muller, T. (2026). Positioning generative AI in EFL peer feedback: Training feedback literacy and enabling uptake in speaking classes. *Education Sciences, 16*(4), 544. https://doi.org/10.3390/educsci16040544

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550.

Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O. M. D., Păun, D., & Mihoreanu, L. (2021). Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. *Sustainability, 13*(18), 10424.

Lee, J. S., & Drajati, N. (2020). Willingness to communicate in digital and non-digital EFL contexts: Scale development and psychometric testing. *Computer Assisted Language Learning, 33*(5–6), 560–583.

Liu, G., & Ma, C. (2024). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching, 18*(2), 125–138.

Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education, 211*, 104977.

López-Chila, R., Llerena-Izquierdo, J., Sumba-Nacipucha, N., & Cueva-Estrada, J. (2024). Artificial intelligence in higher education: An analysis of existing bibliometrics. *Education Sciences, 14*(1), 47.

Luo, Y. T., Liu, T., Pang, P. C.-I., Wang, Z., & Chan, K. I. (2025). Exploring information interaction preferences in an LLM-assisted learning environment with a topic modeling framework. *Applied Sciences, 15*(13), 7515. https://doi.org/10.3390/app15137515

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies, 28*(4), 4221–4241.

Ouyang, Z., Jiang, Y., & Liu, H. (2024). The effects of Duolingo, an AI-integrated technology, on EFL learners' willingness to communicate and engagement in online classes. *International Review of Research in Open and Distributed Learning, 25*(3), 97–115.

Pituxcoosuvarn, M., Tanimura, M., Murakami, Y., & White, J. S. (2025). Enhancing EFL speaking skills with AI-powered word guessing: A comparison of human and AI partners. *Information, 16*(6), 427. https://doi.org/10.3390/info16060427

Rahman, A. (2022). Mapping the efficacy of artificial intelligence-based online proctored examination (OPE) in higher education during COVID-19: Evidence from Assam, India. *International Journal of Learning, Teaching and Educational Research, 21*(2), 76–94.

Richard, C. (2017, March 24). Artificial intelligence to grow 47.5% in education over next 4 years. *The Journal*. https://thejournal.com/articles/2017/03/24/ai-market-to-grow-47-5-percent-over-next-four-years.aspx

Rubinger, L., Gazendam, A., Ekhtiari, S., & Bhandari, M. (2023). Machine learning and artificial intelligence in research and healthcare. *Injury, 54*(Suppl. 3), S69–S73.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., … Lample, G. (2023). *LLaMA: Open and efficient foundation language models* (arXiv:2302.13971). https://arxiv.org/abs/2302.13971

Wang, X., & Reynolds, B. L. (2024). Beyond the books: Exploring factors shaping Chinese English learners' engagement with large language models for vocabulary learning. *Education Sciences, 14*(5), 496. https://doi.org/10.3390/educsci14050496

Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., ... Qiu, C. W. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation, 2*(4), 100179.