Getsempena English Education Journal (GEEJ)
Volume 12, Number 2, 2025 pp. 121-138
P-ISSN: 2355-004X E-ISSN: 2502-6801
Open Access: https://ejournal.bbg.ac.id/geej

SPEAK-BOT AND GROUP DYNAMICS: EXPLORING COLLABORATIVE INTERACTION QUALITY IN AI-ASSISTED SPEAKING PEDAGOGY

Veni Nella Syahputri*[1], Nyak Mutia Ismail[2], Cut Nabilla Kesha[3]
[1]*Department of English Language and Culture, Universitas Teuku Umar, Aceh, Indoensia,*
[2]*Department of English Language, Universitas Serambi Mekkah, Aceh, Indoensia,*
[3]*Department of Public Administration, Universitas Teuku Umar, Aceh, Indoensia*

* Corresponding Author: *venninellasyahputri@utu.ac.id*

ARTICLE INFO

Article history:
Received : Sep 07, 2025
Revised : Oct 24, 2025
Accepted : Nov 25, 2025
Available online : Nov 30, 2025

Keywords:
AI-assisted pedagogy, chatbot-supported learning, collaborative interaction quality, group dynamics, SPEAK-BOT framework

ABSTRACT

This study investigates how the SPEAK-BOT framework shapes group dynamics and collaborative interaction quality in EFL speaking pedagogy. This study addresses the underexplored social dimension of language learning. Specifically, it examines how chatbot-generated prompts, when embedded in a pedagogical framework, influence turn-taking, elaboration, responsiveness, and cohesion during group discussions. A qualitative research design was employed with 20 third-semester English education students at Universitas Teuku Umar, Indonesia. The Students were organized into four small groups and engaged in structured speaking tasks that required consulting chatbot prompts as discussion starters. The instruments used in this research were an audio recorder for data collection, a thematic coding framework for discourse analysis, and a rubric-based scoring sheet to evaluate participants' performance. The data were analyzed using discourse analysis and rubric-based scoring supported by descriptive statistics. The findings revealed a clear variation across groups. One group achieved very high interaction quality, marked by equal participation, deep elaboration, and strong cohesion. Two groups performed moderately, each showing strengths in some dimensions but gaps in others. One group demonstrated weak collaboration, relying heavily on chatbot output and producing fragmented discussions. The results suggest that the SPEAK-BOT framework has the potential to foster richer collaboration when learners use AI critically, but risks weakening interaction when prompts are adopted passively. The study contributes by reframing AI not as a substitute for peer dialogue but as a pedagogical mediator that can strengthen collaborative speaking pedagogy.

INTRODUCTION

Speaking is widely recognized as one of the most essential yet difficult skills in English as a Foreign Language (EFL) learning. Unlike reading or listening, which mainly involve comprehension of input, or writing, which allows time for revision, speaking

requires learners to produce language in real time (Essien et al., 2024). Students must manage grammar, vocabulary, pronunciation, and intonation simultaneously, while also paying attention to meaning, context, and the reactions of their interlocutors. For many learners, this creates anxiety and hesitation, leading to silence or very limited contributions in class discussions. In EFL contexts where exposure to English outside the classroom is minimal, the challenge becomes even greater. Teachers and researchers have long tried to design activities that support learners in developing fluency, accuracy, and communicative competence in speaking. Traditional speaking tasks such as role play, debates, and group discussions have been widely used to provide opportunities for practice (Murtisari et al., 2020). However, these activities sometimes fail to fully engage all students. In many classrooms, participation is unequal: confident learners dominate, while weaker or shy learners remain passive. As a result, the potential of collaborative speaking tasks is not always realized. The central challenge for pedagogy is therefore not only how to help students speak more accurately or fluently, but also how to create conditions where learners can interact collaboratively (Eisenring et al., 2024).

At the same time, language education is undergoing a transformation due to the rapid development of digital technology and Artificial Intelligence (AI). Artificial Intelligence (AI) has become an important area in the field of language education. Unlike earlier technologies that relied mainly on static materials, AI systems offer dynamic, responsive, and adaptive interaction with learners (Arumugam, 2025). The most visible form of AI in education is the chatbot, a program designed to simulate human conversation. AI tools, especially conversational chatbots such as ChatGPT, Bing AI, or other educational bots, now provide new ways for learners to access authentic input, practice interaction, and receive feedback (Siregar, 2022). Unlike earlier digital tools that were limited to multiple-choice exercises or static content, AI chatbots simulate a human interlocutor, offering responses that are immediate, flexible, and context-sensitive. They can answer questions, generate prompts, suggest corrections, and even play the role of conversation partners. This technological change has attracted increasing attention in EFL research and practice. Many studies have shown that AI chatbots can support learners' fluency, helping them to speak more smoothly by providing models of natural conversation (Khalil et al., 2019). Other studies highlight improvements in accuracy, since chatbots can correct grammatical mistakes or provide examples of correct usage. Learners often report higher motivation because interaction with AI feels less threatening than

speaking to teachers or peers. Chatbots are patient, non-judgmental, and available anytime, reducing learners' anxiety (Baek et al., 2025).

Chatbots can engage learners in dialogues, generate questions, provide models of sentences, or suggest corrections. For example, learners can ask a chatbot to act as a travel agent, customer, or debate partner, creating simulated contexts that resemble authentic communication. Beyond chatbots, intelligent tutoring systems (ITS) represent another strand of AI in education. These systems are designed to diagnose learner needs and adapt tasks accordingly (Arumugam, 2025). In language learning, ITS can analyze errors, track progress, and provide individualized exercises. They are often used for grammar training, vocabulary practice, or reading comprehension. A broader term, AI-assisted pedagogy, refers to the integration of AI tools into structured teaching and learning processes (Vibhavi et al., 2024). This includes not only using AI for individual practice but embedding it into classroom activities, assessment, and collaborative tasks. AI-assisted pedagogy emphasizes the role of teachers in guiding how AI is used. Teachers are not replaced but act as mediators, ensuring that AI interaction is meaningful and pedagogically sound (Erguvan, 2021). This perspective is particularly relevant to the present study, which does not treat AI as a stand-alone tool but as a component of a larger framework for collaborative speaking practice (Safar & Anggraheni, 2024).

In relating this issue to learning, there are several related collaborative learning theories. Language learning is fundamentally a social process. Theories of collaborative learning provide a strong foundation for understanding why interaction matters in EFL classrooms. One of the most influential perspectives comes from Lev Vygotsky's Zone of Proximal Development (ZPD). According to this concept, learners can achieve higher levels of performance when working with peers or teachers than when working alone. Interaction provides scaffolding that supports learners in completing tasks slightly beyond their independent ability. When learners negotiate meaning, ask questions, and receive feedback from peers, they operate within their ZPD and gradually internalize new skills (Koriat, 2015). Social Constructivism further explains that knowledge is co-constructed through interaction. Learning is not simply the transfer of information from teacher to student, but a collaborative process in which meaning is built collectively. In EFL speaking classes, this means that students learn not only by practicing individual sentences but by engaging in discussions where ideas are shaped, contested, and refined together.

Interaction theory adds that communication itself is a driver of language acquisition (Yoo et al., 2022). Learners acquire language as they participate in conversational exchanges that require comprehension, production, and negotiation. Through interaction, they notice gaps in their language, attempt repairs, and receive input that is modified to their level. Collaborative speaking activities such as group discussions or problem-solving tasks provide opportunities for this kind of meaningful interaction (Arochman et al., 2023). Together, these theories emphasize that learning is maximized when students participate actively in group work, share turns, and co-construct meaning. They highlight why interaction quality—not only the quantity of speech—should be a central concern of pedagogy (Safar & Anggraheni, 2024).

While many studies measure speaking ability in terms of fluency and accuracy, interaction quality offers a broader lens to evaluate communication. Interaction quality refers to how well participants engage with each other during a task (Azman & Johari, 2022). Four common indicators are often used: (1) Turn-taking – Effective collaboration requires balanced participation (Alrashedi, 2020). High-quality interaction is seen when members share speaking opportunities fairly and no one dominates or remains silent; (2) Elaboration – Learners do not simply provide short answers but expand on ideas, add details, or introduce new perspectives. Elaboration shows deeper processing of language and content; (3) Responsiveness – Strong interaction quality is marked by students responding directly to what others say, building connections between contributions, and showing that they listen actively; and (4) Cohesion – Group discussions that maintain a sense of unity, support, and continuity demonstrate cohesion. Learners encourage each other, follow the same line of thought, and co-construct meaning rather than producing disconnected statements (Yoestara & Ismail, 2023). These indicators move beyond linguistic correctness and provide a way to analyze the social and collaborative dimensions of speaking tasks. They are particularly important when evaluating how AI influences group work, because AI may affect not only what students say but how they interact with each other (Ismail & Sabrina, 2023).

However, most existing research focuses primarily on individual learning outcomes such as fluency, accuracy, vocabulary development, or learner motivation. Much less attention has been paid to what happens when AI chatbots are integrated into collaborative classroom contexts. Speaking is inherently social. Real-world communication usually happens in groups, not only in one-to-one exchanges. In classrooms, group work is a central pedagogical strategy to promote peer learning,

encourage negotiation of meaning, and provide richer opportunities for practice. Yet the influence of AI on group dynamics and collaborative interaction quality is still underexplored.

Several studies have been done under the topics of using AI chatbots in language learning. Huang, et al. (2022) conducted a systematic review of 25 empirical studies on chatbot-supported language learning. Using an inductive grounded approach and Garrison's social presence framework, they examined the technological, pedagogical, and social affordances of chatbots. Findings revealed three technological strengths: timeliness, ease of use, and personalization. Five pedagogical uses were identified: chatbots as interlocutors, simulations, transmission tools, helplines, and recommendation providers. They also promoted social presence through affective, open, and coherent communication. However, challenges included technological limitations, cognitive load, and the novelty effect. The study proposed initial design principles for implementing chatbots in language education. Besides, Harisha, et al. (2024) investigated the effectiveness of chatbots in English language learning with undergraduate students at UIN Sultan Maulana Hasanuddin Banten. Using a qualitative descriptive design, they collected data through questionnaires from 30 students and conducted interviews with 3 participants. Findings emphasized the benefits of chatbots in providing speedy, tailored assistance, improving accessibility, and supporting individualized learning. However, the study also highlighted constraints such as potential errors and platform limitations. Overall, the authors concluded that chatbots have strong potential to enhance English language acquisition when used critically and effectively in instructional contexts. Last, Safar & Anggraheni (2024) investigated how AI chatbots affect language learning and cognitive load among 60 university learners through a mixed-methods study. Over a three-week period, participants' language proficiency—especially conversational fluency and vocabulary—was assessed before and after chatbot use. Results showed significant improvements in both areas. However, learners reported a moderate cognitive load, primarily due to the challenge of engaging in real-time conversation while processing feedback. Importantly, reported cognitive load varied with individual factors like prior proficiency and familiarity with AI tools. The study concludes that AI chatbots can effectively support language learning, provided cognitive load is thoughtfully managed (Lee & Hahn, 2024).

Collectively, the reviewed studies emphasize individual learning outcomes such as fluency, accuracy, vocabulary development, learner perceptions, or cognitive load. None

of them directly investigate how chatbots function within group-based speaking activities or how they shape the quality of collaborative interaction among peers (Izzati et al., 2020). This leaves a gap in understanding the social dimension of AI-assisted pedagogy. The current study addresses this unexplored area by examining group dynamics and collaborative interaction quality—specifically turn-taking, elaboration, responsiveness, and cohesion—within the SPEAK-BOT framework, which positions chatbot prompts not as answers but as catalysts for peer discussion in EFL classrooms. Filling this gap is crucial, as collaborative speaking mirrors authentic communication and determines how effectively learners can transfer classroom interaction into real-world language use (Ma, 2023).

Building on this gap, the novelty of this study lies in its shift from the individual to the collective. Rather than asking only how AI supports accuracy or vocabulary, it investigates how AI can be embedded in group activities through the SPEAK-BOT framework. SPEAK-BOT is not a chatbot itself, but a pedagogical design that integrates chatbot interaction into structured classroom practice. Here, chatbot prompts serve as starting points for peer dialogue, positioning AI as a catalyst for collaboration rather than a substitute for human interaction. This focus is significant because successful group learning depends not on the quantity of speech alone, but on balanced participation, responsiveness, elaboration, and cohesion. By foregrounding these dimensions, the study offers a fresh lens to evaluate the impact of AI in EFL classrooms.

This study also highlights the dual role of AI in pedagogy: it can enrich discussion or, if overused, reduce interaction to mechanical exchanges. When students depend too heavily on chatbot output, collaboration weakens; but when AI prompts are framed through the SPEAK-BOT framework as material to adapt and debate, they become triggers for meaningful dialogue. This reframing is the core innovation of the study, positioning AI as a mediator of collective learning rather than a source of ready-made sentences (Premkumar et al., 2024). The study further contributes to debates on ethics and digital literacy by acknowledging risks—such as dependency, bias, and misinformation—while showing how teacher mediation and critical use can address them. By combining pedagogical design, collaborative theory, and ethical awareness, this research distinguishes itself from prior work.

Based on the rationale above, the guiding research question is:

To what extent is collaborative interaction quality reflected in student conversations stimulated by the chatbot within the SPEAK-BOT framework?

This question directs the study toward interactional processes rather than individual outcomes, shifting the focus from language products to group dynamics. In doing so, the study seeks to deepen understanding of how AI can be integrated into pedagogy to strengthen collaboration.

METHOD

This study employed an exploratory qualitative design to investigate how the SPEAK-BOT framework influences group dynamics and the quality of collaborative interaction among EFL learners. Furthermore, the study incorporated descriptive quantitative elements by scoring interaction quality using a Likert-based rubric. The participants in this study were third-semester students enrolled in the Language Education Department at Universitas Teuku Umar, Aceh, Indonesia. Data collected from the 20 students who were taking a compulsory speaking course. The selection of third-semester students was intentional. The students had already completed basic speaking courses and were familiar with pair and group activities, but they still faced challenges in fluency, confidence, and collaborative speaking. This made them an appropriate group to explore how AI-supported frameworks might enhance their interaction quality. The students were divided into small groups of five, resulting in a total of four groups for observation and analysis. Participation was voluntary, and informed consent was obtained from all students. They were assured that their performance would not affect their course grades, and data would be anonymized to protect privacy.

The instrument used for data collection was a rubric for collaborative interaction quality (Stuckey, 2015). This rubric was developed to evaluate the quality of interaction within each group. The rubric included four dimensions commonly identified in the literature: turn-taking, elaboration, responsiveness, and cohesion. Each dimension was scored on a five-point Likert scale, where 1 indicated very low quality (e.g., domination by one student, minimal responsiveness) and 5 indicated very high quality (e.g., balanced turns, deep elaboration, strong cohesion). The rubric served both as an observational tool for the researchers and as a framework for analyzing transcripts.

Table 1. collaborative interaction quality ((Bruner, 2006), (Leahey, 2012), and (Vygotsky, 1978))

Dimension	1 = Very Low Quality	2 = Low Quality	3 = Moderate Quality	4 = High Quality	5 = Very High Quality
Turn-taking	One student dominates;	Uneven participation;	Some sharing of	Turns are fairly well	All members

Dimension	1 = Very Low Quality	2 = Low Quality	3 = Moderate Quality	4 = High Quality	5 = Very High Quality
	others mostly silent.	limited contribution from weaker members.	turns, but still unbalanced.	distributed, most members contribute.	share speaking time equally and actively.
Elaboration	Responses are minimal; no expansion of ideas.	Occasional elaboration but mostly short or surface-level answers.	Moderate elaboration with some details, but not sustained.	Frequent elaboration; students expand and add relevant details.	Consistent, in-depth elaboration; ideas are developed, extended, and enriched.
Responsiveness	Little or no response to peers; comments are isolated.	Limited acknowledgment of peers' contributions.	Some responsive moves but often disconnected.	Regularly responds to peers, shows understanding, and connects ideas.	Highly responsive; peers build on each other's input seamlessly.
Cohesion	Discussion feels fragmented; no sense of group unity.	Weak cohesion; ideas are disconnected and lack continuity.	Moderate cohesion; some linking but also gaps in flow.	Strong cohesion; ideas linked clearly and supportive atmosphere.	Very strong cohesion; group shows unity, mutual support, and co-construction of meaning.

The rubric was developed by adapting insights from Long's Interaction Hypothesis (1996), which emphasizes negotiation of meaning during conversational exchanges, Vygotsky's Sociocultural Theory (1978), which highlights the importance of collaboration and scaffolding within the Zone of Proximal Development, and interaction analysis frameworks such as Flanders (1970), which provide systematic categories for evaluating participation and discourse patterns in classroom interaction.

The procedure of collecting data involve some phases. First, students were introduced to the SPEAK-BOT framework, including how to use chatbot prompts responsibly as triggers for discussion. They were reminded that the chatbot output was not the "answer" but material to be adapted, debated, or elaborated in their groups. Second, students were assigned a discussion task relevant to their course content, such as planning an English Day event or debating solutions to an environmental issue. Groups were instructed to begin by consulting SPEAK-BOT for initial prompts or example sentences. After receiving these prompts, the groups continued the discussion among

themselves, integrating, adapting, or critiquing the AI-generated material. At the end of the session, students provided short written reflections on how they used chatbot prompts and how they felt about the group collaboration. These reflections supported the interpretation of data found in rubric.

The data were analyzed using a thematic coding and quantitative scoring. Initially, all audio recordings were transcribed verbatim. The transcripts included not only the words spoken but also pauses, overlaps, and other markers of interaction. Discourse analysis was applied to examine how students managed their group discussions. This involved identifying sequences of turn-taking, examples of elaboration, and instances of responsiveness or lack thereof. Particular attention was given to how chatbot prompts entered the flow of conversation—whether they were simply repeated, adapted, or used as springboards for further dialogue. After the initial discourse analysis, transcripts were subjected to thematic coding. Codes were developed both deductively (based on the four dimensions of interaction quality) and inductively (emerging from the data). For example, deductive codes included categories such as "balanced turn-taking" or "surface-level elaboration," while inductive codes included themes like "over-reliance on chatbot output" or "peer encouragement through humor." Thematic coding allowed the researchers to identify recurring patterns and contrasts across different groups. To complete the qualitative findings, each group's interaction was scored using the rubric for collaborative interaction quality. Two raters independently evaluated the recordings and transcripts, assigning scores for each dimension on the 1–5 Likert scale. Inter-rater reliability was calculated to ensure consistency. The resulting scores were analyzed descriptively, producing averages and ranges for each dimension. These descriptive statistics provided a general picture of interaction quality across the eight groups and helped highlight which aspects were strongest or weakest.

Triangulation was applied by combining multiple data sources to ensure the trustworthiness: transcripts, rubric scores, and student reflections. Inter-rater reliability between the two scorers strengthened the credibility of the quantitative rubric results. Additionally, this study adhered to ethical research standards. Participants were informed about the purpose of the research and their right to withdraw at any time. Anonymity was guaranteed by replacing names with codes in transcripts and reports. Since chatbot interaction involved no personal data storage, there was no risk of privacy violation from the AI tool. Still, students were reminded to avoid sharing sensitive information during sessions.

RESULTS AND DISCUSSION

Results

The results of this study provide insight into how the SPEAK-BOT framework influenced collaborative interaction quality among third-semester EFL learners. By examining turn-taking, elaboration, responsiveness, and cohesion across four groups, the findings reveal distinct patterns of group dynamics, ranging from highly collaborative exchanges to fragmented discussions. These variations highlight both the potential and the limitations of AI-supported pedagogy in speaking classes. In the following section, the results are presented in detail, followed by a discussion that interprets them in light of collaborative learning theories, interaction frameworks, and previous studies on AI in language education.

Table 2. Results of Collaborative Interaction Quality (Groups 1–4)

Dimension	Group 1	Group 2	Group 3	Group 4
Turn-taking	4 – Turns fairly distributed, most members contributed.	3 – Some sharing of turns, but one stronger member often dominated.	5 – All members shared time equally and actively.	2 – Uneven participation; weaker members rarely spoke.
Elaboration	3 – Some elaboration but not always sustained.	4 – Frequent elaboration with added details, though uneven across members.	5 – Consistent, in-depth elaboration; ideas developed fully.	2 – Mostly short answers with minimal development.
Responsiveness	4 – Regular responses and connections, some minor gaps.	3 – Some responses but often disconnected; limited follow-up.	5 – Highly responsive; peers built seamlessly on each other's input.	2 – Limited acknowledgment of peers; ideas often isolated.
Cohesion	3 – Moderate cohesion; some linking but occasional gaps.	4 – Strong cohesion, ideas linked clearly with supportive atmosphere.	5 – Very strong cohesion; unity and co-construction evident.	2 – Weak cohesion; fragmented discussion without flow.

From the table above, it is learned that collaborative interaction quality across the four groups revealed clear variation in how students engaged with each other when using the SPEAK-BOT framework. Overall, the findings show that while some groups were able to use chatbot prompts as effective triggers for collaboration, others struggled to move beyond surface-level responses and displayed weaker patterns of group interaction. Group 1 achieved an average quality score of 3.5, which falls between moderate and high. The group demonstrated balanced participation, with most members contributing and

turns fairly well distributed. Responsiveness was also strong, as members often acknowledged and built on peers' comments. However, elaboration remained limited; ideas were expanded occasionally but not consistently sustained. Cohesion was moderate, as the discussion sometimes lacked continuity despite general group effort. Group 2, also with an average score of 3.5, performed similarly to Group 1 but with different strengths and weaknesses. The group showed frequent elaboration and strong cohesion, suggesting that ideas were often extended and the group maintained a supportive atmosphere. Yet, turn-taking was less balanced, as one or two students dominated the interaction, while weaker members contributed less. Responsiveness was also uneven, with several comments not fully connected to previous turns.

Group 3 emerged as the strongest group, with an average score of 5 across all dimensions. This group achieved highly collaborative interaction where all members shared speaking turns equally, elaborated ideas in depth, and responded seamlessly to each other's contributions. Cohesion was very strong, and the discussion reflected a collective effort in meaning-making. This suggests that SPEAK-BOT prompts were effectively used as springboards for co-construction of knowledge. Group 4, in contrast, obtained the lowest average score of 2. The group displayed weak interaction quality, with one member dominating and others remaining largely silent. Responses were short and isolated, elaboration was minimal, and cohesion was lacking, resulting in fragmented discussion. This indicates that the group relied on chatbot input without effectively engaging in peer collaboration. In sum, the results show that the SPEAK-BOT framework can foster high-quality interaction, but its impact depends on how groups manage participation, responsiveness, and the use of AI prompts.

Below is the chart showing collaborative interaction quality across the four groups. It compares scores for Turn-taking, Elaboration, Responsiveness, and Cohesion on the 1–5 Likert scale.

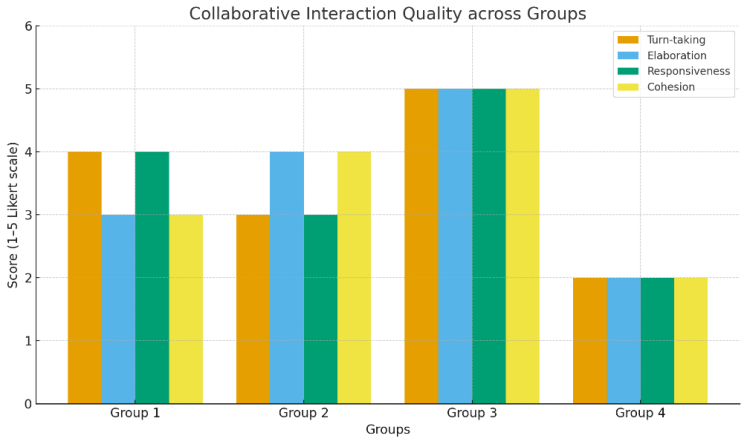

Figure 1. Collaborative interaction quality

In addition, Figure 2 below illustrates the average scores of collaborative interaction quality across the four groups. The chart provides a clear overview of how each group performed in terms of turn-taking, elaboration, responsiveness, and cohesion, summarized into a single mean value. By presenting the results in this way, it becomes easier to identify patterns of group dynamics and compare relative strengths and weaknesses. The visualization highlights the overall contrast between high-performing and low-performing groups, offering an accessible representation of interaction quality under the SPEAK-BOT framework.

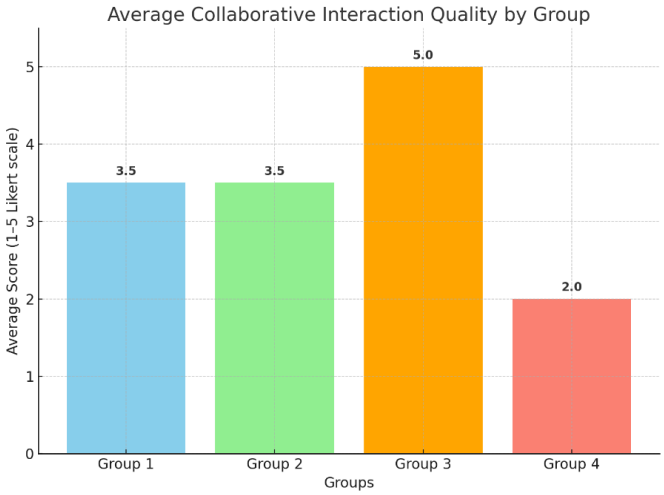

Figure 2. Average collaborative interaction quality across the four groups

The average scores of collaborative interaction quality highlight clear contrasts among the four groups. Group 3 stood out with a perfect mean of 5.0, indicating very high-quality collaboration across all dimensions. Members of this group distributed turns equally, elaborated ideas in depth, responded seamlessly to one another, and maintained strong cohesion throughout their discussion. In contrast, Group 4 obtained the lowest

mean score of 2.0, reflecting weak collaboration. Their discussion was fragmented, with uneven participation, minimal elaboration, and limited responsiveness, showing a tendency to rely heavily on chatbot prompts without deeper peer engagement. Both Group 1 and Group 2 achieved mid-level averages of 3.5, but with different strengths and weaknesses: Group 1 showed balanced turn-taking and responsiveness but lacked consistent elaboration and cohesion, while Group 2 demonstrated stronger elaboration and cohesion yet suffered from uneven participation. These variations suggest that while the SPEAK-BOT framework has the potential to stimulate collaborative interaction, its effectiveness depends on how groups negotiate participation and use AI-generated prompts as triggers for dialogue rather than final answers.

Discussion

The primary purpose of this study was to investigate how the SPEAK-BOT framework influences group dynamics and the quality of collaborative interaction in EFL speaking classrooms. While previous research on AI in education has often highlighted individual benefits such as improved fluency, accuracy, and vocabulary growth, this study contributes by focusing on the collective dimension of speaking pedagogy. Specifically, it emphasizes how chatbot-generated prompts, when integrated through a pedagogical framework, can stimulate group discussion and enhance or hinder collaboration. By directing attention to interaction quality—measured through turn-taking, elaboration, responsiveness, and cohesion—this study provides a novel contribution to the growing field of AI-assisted pedagogy. The findings extend the discussion from individual practice with AI toward its role as a mediator of peer collaboration in classroom contexts. The study employed an exploratory qualitative design with descriptive quantitative support. Participants were third-semester EFL students at Universitas Teuku Umar, divided into four small groups. Each group engaged in speaking tasks that required them to consult chatbot prompts within the SPEAK-BOT framework before developing their own discussions. Group interactions were audio-recorded, transcribed, and analyzed using three approaches: discourse analysis to examine conversational features, thematic coding to identify patterns of interaction, and rubric-based scoring to provide measurable indicators of collaborative quality. The rubric, developed from established theories of interaction and sociocultural learning, assessed four dimensions—turn-taking, elaboration, responsiveness, and cohesion—on a five-point Likert scale.

The results showed significant variation among the four groups. Group 3 achieved the highest performance, scoring consistently at the maximum level across all dimensions. Their interaction was marked by equal participation, in-depth elaboration of ideas, strong responsiveness, and high cohesion. Group 4, in contrast, performed weakest, with low scores on every dimension, demonstrating uneven participation, minimal elaboration, poor responsiveness, and fragmented cohesion. Groups 1 and 2 fell in the mid-range with average scores of 3.5, though their strengths differed. Group 1 demonstrated balanced turn-taking and strong responsiveness but weaker elaboration and cohesion, whereas Group 2 showed strong elaboration and cohesion but uneven participation and weaker responsiveness. These results suggest that while the framework has potential to foster collaborative interaction, its impact varies depending on how groups manage participation and use chatbot prompts. The findings align with key insights from Vygotsky's Sociocultural Theory (1978) and the Zone of Proximal Development (ZPD)(Vygotsky, 1978), which emphasize that learners achieve more when supported through collaboration and scaffolding. Group 3 exemplified this principle, as students worked within each other's ZPD, building ideas collectively and achieving higher-quality interaction. In contrast, Group 4 showed limited scaffolding and peer support, resulting in weak collaboration.

The results also support Premkumar's et al. (2024) finding which argues that negotiation of meaning during interaction promotes language learning. Groups that responded to each other's contributions and elaborated on ideas (especially Group 3) demonstrated precisely the kind of negotiation Long describes. Conversely, Group 4's lack of responsiveness reduced opportunities for negotiation, limiting learning potential. Furthermore, the study connects with Social Constructivism, which frames learning as a co-construction of knowledge through social interaction. The varying levels of cohesion across groups illustrate how the social context strongly shapes learning outcomes (Belda-Medina & Calvo-Ferrer, 2022). Cohesive groups were able to co-construct meaning, while fragmented groups failed to achieve the same level of depth.

Several explanations can account for the patterns observed. First, the strong performance of Group 3 may be attributed to their ability to use chatbot prompts as springboards rather than final answers. By adapting and debating AI output, they maintained active engagement with peers, which strengthened all dimensions of interaction quality. This reflects findings from previous studies, such as those highlighting how chatbot prompts can enhance motivation and stimulate dialogue when used

critically. Second, the weaker performance of Group 4 suggests the risk of over-reliance on AI, which has also been reported in earlier research. Some studies note that learners tend to copy chatbot responses without deeper elaboration, leading to superficial interaction. The same pattern emerged here, where students leaned too heavily on AI-generated sentences, resulting in fragmented discussions and weak cohesion, as found by (Lau et al., 2024).

Comparing these findings with earlier studies reveals both similarities and differences. Similar to previous research, this study confirms that chatbots can reduce anxiety and encourage participation, particularly for less confident learners. Evidence from Group 1 supports this, as weaker members contributed more confidently when supported by prompts. However, unlike many earlier studies that primarily assessed individual fluency and accuracy, this research demonstrates that the collective dimension—interaction quality—may not improve automatically with chatbot integration, as mentioned by Zhang & Zheng. 2024) and Wahid et al. (2024). The contrasting performances of Groups 3 and 4 show that AI alone does not guarantee collaborative learning; rather, its effectiveness depends on how it is framed pedagogically and how learners choose to engage with it. Another difference lies in the explicit focus on group dynamics. While previous studies often highlight learner–chatbot interaction, this study positions chatbot use within a framework of peer discussion. This positioning is crucial, because it shows that AI can be intentionally designed to stimulate collaboration rather than replace it. By structuring chatbot prompts as triggers for peer talk, the SPEAK-BOT framework addresses the gap identified in prior literature, which often overlooked the social dimension of classroom interaction. Finally, the differences among groups illustrate the context-dependent nature of AI integration. Cultural factors, group composition, and student personalities may all influence how chatbot prompts are used. For instance, a dominant student in Group 2 affected turn-taking, while supportive humor in Group 3 enhanced cohesion. These findings suggest that the success of AI-assisted frameworks is not only technological but also deeply social, shaped by the interaction of tools, tasks, and participants.

The findings of this study imply that the integration of AI into speaking pedagogy must go beyond individual practice. The contrast between highly collaborative groups and weaker ones shows that AI can either enhance or undermine peer interaction. This means teachers and institutions need to treat AI not simply as a language tool but as part of a pedagogical design that shapes how students learn together. For teachers, the

implication is clear: SPEAK-BOT or similar frameworks should be used to stimulate discussion rather than supply answers. Teachers should train learners to adapt, critique, and extend chatbot prompts so that collaboration is preserved. For students, the results imply that effective group work requires active participation, responsiveness, and cohesion, not passive reliance on technology. At the broader level, the study implies that policymakers and curriculum designers can recognize AI as a partner in education. When structured carefully, frameworks like SPEAK-BOT offer a pathway to prepare learners not only for linguistic competence but also for the collaborative skills demanded in real communication.

CONCLUSION

This study examined how the SPEAK-BOT framework influences group dynamics and collaborative interaction quality in EFL speaking classrooms. Unlike much existing research that has focused mainly on fluency or accuracy, this study emphasized the social dimension of speaking, analyzing how students shared turns, elaborated ideas, responded to peers, and built cohesion when chatbot prompts were incorporated into group tasks. The findings revealed important variations across the four groups. Group 3 demonstrated very high-quality interaction, with balanced participation, deep elaboration, and strong cohesion. Groups 1 and 2 performed at a moderate level, each showing particular strengths but also weaknesses in responsiveness or turn distribution. Group 4, by contrast, scored lowest, relying heavily on chatbot output without sustaining meaningful dialogue. These results suggest that the SPEAK-BOT framework is effective when learners use chatbot prompts critically and collaboratively, but its impact is diminished when students treat AI responses passively. These findings reinforce sociocultural and interactionist perspectives on language learning. Group 3 illustrated how scaffolding and negotiation of meaning can enrich learning, while Group 4 reflected the risks of dependency and lack of peer support. The implication is that AI must be integrated thoughtfully: teachers should guide students in adapting and extending chatbot prompts, and assessments should include interaction quality alongside language accuracy.

The study is not without limitations. The small sample size, limited to third-semester students from one university, restricts the generalizability of the results. Data were collected from short-term discussions only, and the focus on audio excluded non-verbal aspects such as gestures or eye contact, which also influence collaboration. Future studies should therefore involve larger and more diverse groups, employ longitudinal

designs, and capture multimodal data to provide a fuller understanding of collaborative interaction.

Based on these insights, several recommendations can be made. Teachers should employ frameworks like SPEAK-BOT to ensure chatbot prompts are used as stimuli for dialogue, not as final answers, while monitoring group dynamics to balance participation. Students should be encouraged to treat chatbot responses critically, adapting and debating them to sustain collaboration. Curriculum designers should incorporate measures of interaction quality into assessment rubrics, recognizing collaboration as a learning goal. Finally, future research should broaden scope and method, testing AI-supported frameworks across contexts and over time.

REFERENCES

Alrashedi, N. (2020). A new decade for social changes. *Technium Social Sciences Journal*, *9*(July), 101–105.

Arochman, T., Jabbar, S. A., Hutabarat, P. S. P., & Pitaloka, N. S. A. (2023). Infographics as a Tool to Facilitate English Learning Activities: Student's Perceptions. *ENGLISH FRANCA : Academic Journal of English Language and Education*, *7*(2), 369. https://doi.org/10.29240/ef.v7i2.7508

Arumugam, S. (2025). Deep Learning-Based Smart Invigilation System for Enhanced Exam Integrity. *Proceedings of Engineering and Technology Innovation*, *29*, 99–115. https://doi.org/10.46604/peti.2024.14105

Azman, A., & Johari, M. (2022). Investigating the effectiveness of videos designed using cognitive load theory on Biology students' students' academic achievement. *Jurnal Pendidikan IPA Indonesia*, *11*(3), 461–468. https://doi.org/10.15294/jpii.v11i3.37324

Baek, G., Cha, C., & Han, J. H. (2025). AI Chatbots for Psychological Health for Health Professionals: Scoping Review. *JMIR Human Factors*, *12*, 1–15. https://doi.org/10.2196/67682

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using Chatbots as AI Conversational Partners in Language Learning. *Applied Sciences (Switzerland)*, *12*(17), 1–16. https://doi.org/10.3390/app12178427

Bruner, J. (2006). *Theory of Cognitive Development*. Routledge.

Eisenring, M., Jamiluddin, J., Hairul, M., & Putri, D. (2024). The use of chatbots in the English language teaching to promote modern language learning: A literature review. *International Journal of Indonesian Education and Teaching*, *8*(1), 127–139.

Erguvan, I. D. (2021). The rise of contract cheating during the COVID-19 pandemic: a qualitative study through the eyes of academics in Kuwait. *Language Testing in Asia*, *11*(34), 1–21. https://doi.org/10.1186/s40468-021-00149-y

Essien, A., Bukoye, O. T., O'Dea, X., & Kremantzis, M. (2024). The influence of AI text generators on critical thinking skills in UK business schools. *Studies in Higher Education*, *49*(5), 865–882. https://doi.org/10.1080/03075079.2024.2316881

Harisha, N., Suminih, S., & Oktaviana, F. (2024). the Power of Chatbots in English Language Learning: a New Age of Learning. *Lingua*, *20*(1), 19–31. https://doi.org/10.34005/lingua.v20i1.3492

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal*

of Computer Assisted Learning, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Ismail, N. M., & Sabrina, S. (2023). Virtual Learning and Memory Dissonance. *Jurnla Ilmiah Peuradeun*, *11*(3), 925–948.

Izzati, I. D. C., Tentama, F., & Suyono, H. (2020). Academic stress scale: A psychometric study for academic stress in senior high school. *European Journal of Education Studies*, *7*(7), 153–168. https://doi.org/10.46827/ejes.v7i7.3161

Khalil, M., Prinsloo, P., & Slade, S. (2019). In the nexus of integrity and surveillance: Proctoring (re)considered. *Sustainability (Switzerland)*, *11*(1), 1–14. http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI

Koriat, A. (2015). When two heads are better than one and when they can be worse: The amplification hypothesis. *Journal of Experimental Psychology: General*, *144*(5), 934–950. https://doi.org/10.1037/xge0000092

Lau, K. H. C., Yun, B., Saruba, S., Bozkir, E., & Kasneci, E. (2024). *Wrapped in Anansi's Web: Unweaving the Impacts of Generative-AI Personalization and VR Immersion in Oral Storytelling* (Vol. 1, Issue 1). Association for Computing Machinery. http://arxiv.org/abs/2409.16894

Leahey, T. (2012). *Cognition and Learning*. Wiley Publisher.

Lee, I., & Hahn, S. (2024). On the relationship between mind perception and social support of chatbots. *Frontiers in Psychology*, *15*(March), 1–10. https://doi.org/10.3389/fpsyg.2024.1282036

Ma, Z. (2023). The Study on the Influence of Academic Pressure on Academic Performance. *Journal of Education and Educational Research*, *3*(2), 106–109. https://doi.org/10.54097/jeer.v3i2.9045

Murtisari, E. T., Salvadora, L., & Hastuti, G. (2020). Isolated and Integrated Grammar Teaching in Tertiary Efl Context: Indonesian Teachers' Beliefs. *SAGA: Journal of English Language Teaching and Applied Linguistics*, *1*(1), 17–30. https://doi.org/10.21460/saga.2020.11.9

Premkumar, P., Yatigammana, M., & Kannangara, S. (2024). Impact of generative AI on critical thinking skills in undergraduates: A systematic review. *The Journal of Desk Research Review and Analysis*, *2*(1), 199–215.

Safar, M., & Anggraheni, D. (2024). Language Learning through AI Chatbots : Effectiveness and Cognitive Load Analysis. *Journal of Social Science Utilizing Technology*, *2*(3), 430–445.

Siregar, I. (2022). The pressure of academic stress and self-efficacy among student. *Jurnal Konseling Dan Pendidikan*, *10*(3), 394. https://doi.org/10.29210/156700

Stuckey, H. (2015). Three types of interviews: Qualitative research methods in social health. *Journal of Social Health and Diabetes*, *01*(02), 056–059. https://doi.org/10.4103/2321-0656.115294

Vibhavi, R., Swetha, A., Shruthi, B., Shravya, S., Balal, R., & Narayan, R. (2024). AI-Driven Chatbots in Mental Health : Enhancing Emotional Support and AI-Driven Chatbots in Mental Health : Enhancing Emotional Support and Therapy Through Cognitive Behavioural Approaches. *Gradiva Review Journal*, *10*(9), 323–328.

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. MIT Press.

Wahid, M. E., Ain, R., Abdul, M. L., Arief, D., & Shamsul, H. (2024). Digital Folk Literature as a Need for Educational Transformation : A Literary Narrative. *International Journal of Academic Research in Progressive Education and Development*, *13*(4), 3439–3460. https://doi.org/10.6007/IJARPED/v13-i4/24338

Yoestara, M., & Ismail, N. M. (2023). Employing video-editing skill in designing materials for Speaking class. *EnJourMe (English Journal of Merdeka) : Culture, Language, and Teaching of English*, *7*(2), 241–253. https://doi.org/10.26905/enjourme.v7i2.8817

Yoo, H., Jang, J., Oh, H., & Park, I. (2022). The potentials and trends of holography in education: A scoping review. *Computers and Education*, *104533*.

Zhang, Y., & Zheng, B. (2024). Transforming Narratives : The Influence OF Digital Storytelling on Traditional Narratives in English Literature. *Pacific International Journal, 7*(749), 51–60. https://doi.org/10.55014/pij.v7iSpecial