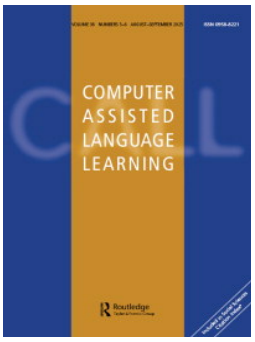


Routledge
Taylor & Francis Group

Computer Assisted Language Learning

ISSN: 0958-8221 (Print) 1744-3210 (Online) Journal homepage: www.tandfonline.com/journals/ncal20

Exploring interaction process, L2 motivation and intended effort in chatbot-mediated interactions: evidence from Latent Dirichlet Allocation (LDA) topic modeling

Yunfei Du & Barry Lee Reynolds

COMPUTER ASSISTED LANGUAGE LEARNING
https://doi.org/10.1080/09588221.2025.2552116

Routledge
Taylor & Francis Group

Check for updates

Exploring interaction process, L2 motivation and intended effort in chatbot-mediated interactions: evidence from Latent Dirichlet Allocation (LDA) topic modeling

Yunfei Du[a,b,c] and Barry Lee Reynolds[b,c]

[a]Department of English, School of Languages and Communication Studies, Chongqing University of Technology, Chongqing, China; [b]Faculty of Education, University of Macau, Taipa, Macau SAR, China; [c]Centre for Data Science, University of Macau, Taipa, Macau SAR, China

ABSTRACT
Interaction shapes the learning opportunities and learners' motivation through exposure, feedback and practice. However, despite the growing interest of incorporating chatbots in second language (L2) speaking, few studies have examined how the dynamics of chatbot-mediated interactions trigger motivational changes over time. To remedy the gap, the present study empirically examined the impact of chatbot-mediated interactions on learners' L2 motivational self system (L2MSS) and intended effort, as well as unpacking the interaction process as evidence for these motivational changes. In this study, 232 undergraduate students were recruited and assigned to either the experimental group (EG-CHAT) or the control group (CG). Doubao, a chatbot built on large language models (LLMs), was introduced to deliver chatbot-mediated interactions with LLMs-powered speaking activities for the EG-CHAT group, while FiF Speaking Practice was employed for the CG to offer structured practices such as video watching, repeating and shadowing. After the five-week quasi-experiment, two-way mixed-design ANOVA results revealed significant interaction effects for both the ideal L2 self and intended effort, indicating that the EG-CHAT group exhibited greater motivational changes across the intervention period than the CG. However, while both groups demonstrated significant improvements in the L2 learning experience, no group outperformed the other, and no significant change was observed in the ought-to L2 self from pre- to post-test. The Latent Dirichlet Allocation (LDA) topic modeling results of the dialogic content provide further insight into these motivational changes and the interaction process, as well as identifying insufficient learner behaviors that call for motivational scaffolding and cognitive intervention from teachers.

ARTICLE HISTORY
Received 17 December 2024
Accepted 20 August 2025

KEYWORDS
Chatbot; interaction process; L2 motivational self system; intended effort; LDA topic modeling

CONTACT Barry Lee Reynolds barryreynolds@um.edu.mo E33 Room 1014 Faculty of Education, University of Macau, Taipa, Macau SAR, China

1. Introduction

Interaction is a vital component of language acquisition, which both shapes the learning opportunities (Illeris, 2016) and learners' motivation through exposure, feedback, and practice (Al-Hoorie, 2018). Beyond its general role, specific features of interaction—such as dialogic content, interactive styles, and negotiation of meaning—significantly influence learners' L2 motivational self system (L2MSS) and intended effort. Although the importance of interaction has been long recognized, students often face limited opportunities and input-poor environments for practicing speaking in second language (L2) contexts (Gan, 2012). The lack of exposure to meaningful interaction and authentic language use may increase learners' speaking anxiety and weaken their language proficiency and fluency (Fathi et al., 2024; Tai & Chen, 2023). To foster a more supportive environment and mitigate affective barriers, chatbots have been integrated in lectures, tutorials and out-of-class activities to enrich conversation practices (Ebadi & Amini, 2022).

Chatbots, such as Google Assistant and ChatGPT, are conversational agents designed to interact with humans through text or speech, featured by their human-like conversations and adaptive responses (Ji et al., 2023). Among these, ChatGPT, which is built upon large language models (LLMs), has significantly enhanced the immersive and life-like quality of interactions between learners and chatbots (Kasneci et al., 2023). Leveraging their interactive and generative features, chatbots are widely used in L2 speaking contexts to design speaking activities, improve learners' communicative skills and boost their motivation and willingness to communicate (Tai, 2024). For instance, the English chatbot Andy was employed to design interactive speaking activities, such as role-play, scenario-based conversation, and targeted vocabulary practice (Fathi et al., 2024). Ye et al. (2022) reported that when EFL learners participated in chatbot-mediated interactions, Microsoft Xiaoying offered timely feedback and corrected their pronunciation errors during conversational tasks, which led to a significant improvement in their pronunciation accuracy. In addition, Kim and Su (2024) found that students learning Korean as a foreign language were highly motivated and possessed stronger willingness to communicate when engaging with the chatbot on Danbee AI.

In light of previous studies, the pedagogical and motivational impact of chatbot-mediated interactions have garnered significant interest (Huang & Mizumoto, 2024; Um et al., 2023). While interaction is widely considered important in shaping learners' motivation (Al-Hoorie, 2018), few studies have examined how the dynamics of chatbot-mediated interactions trigger motivational changes over time. Without unpacking the detailed processes of these interactions, teachers may struggle to gauge learners' cognitive and motivational engagement, thereby hindering their ability to provide timely and adaptive interventions. Although some studies have employed interviews and self-report questionnaires to identify interaction behaviors and strategies (e.g., corrective feedback seeking, gamified interaction, and self-directed learning) in chatbot-mediated activities (Li et al., 2024; Tai & Chen, 2023; Ye et al., 2022), these methods capture only retrospective reflections rather than the real-time dialogic content of the interaction process. These gaps call for fresh perspectives and unbiased, data-driven research to bridge motivational changes with real-time chatbot-mediated interactions.

In response, we aim to examine how learners' L2MSS and intended effort change during chatbot-mediated interactions, while also exploring the real-time interaction process between learners and a chatbot. To achieve these objectives, we intend to analyze motivational changes through two-way mixed-design analyses of variance (ANOVA) and process dialogic content generated during learner-chatbot interactions using Latent Dirichlet Allocation (LDA) topic modeling. This unsupervised machine learning approach enables the discovery of hidden thematic structures in interaction data, offering a data-driven perspective that transcends the scope of interviews and self-report questionnaires. As the interaction data is clustered and interpreted, it could provide evidence to the motivational changes, and evaluate the routines, preferences, and behaviors that L2 learners adopt during chatbot-mediated interactions. The insights gained from unpacking these interactions and motivational components can further help teachers design effective speaking activities, sustain high-quality language exposure, and satisfy individual differences in the interactive environment.

2. Literature review

2.1. Chatbot-mediated interactions in the L2 speaking context

Interaction research holds a central and enduring position in second language acquisition (SLA), recognized for its unparalleled volume, longevity, and profound impact on follow-up studies (Plonsky & Gass, 2011). Chatbot-mediated interaction builds on this legacy by offering learners the opportunity to engage in meaningful, goal-directed conversations in a low-pressure environment. Early chatbots such as ELIZA and Google Assistant, though limited to scripted rule-based responses, laid the foundation for integrating conversational AI into language learning. These systems were found to support vocabulary recall, oral rehearsal, and reduced learner anxiety (Ahn & Lee, 2016; Tai, 2022). In contrast, LLMs-powered chatbots such as ChatGPT represent a paradigm shift: they enable spontaneous, multi-turn, and personalized dialogue, facilitating more authentic and student-directed interaction (Gao et al., 2024; Kasneci et al., 2023).

As chatbots provide L2 learners with opportunities to practice the targeted language in meaningful ways, research has uncovered interactive styles and learner behaviors that might shape the dynamics and effectiveness of chatbot-mediated interactions (Engwall et al., 2021; Wu et al., 2024). Interactive styles, such as interlocutor, narrator, interviewer, and facilitator (Engwall et al., 2021), affect the ways that chatbots interacted with learners. While some studies have conflated interactive styles with broader pedagogical or technological affordances, including simulation, resource provider, and teachable and motivational agent (Huang et al., 2022; Kuhail et al., 2023), the distinctive interactive nature of chatbots serves as the anchor of designing speaking activities in chatbot-mediated interactions.

Specifically, chatbots as interlocutors immerse learners in predefined or flexible conversational exchanges that mimic real-life dialogues, providing immediate responses, phonetic modeling, and creative input (Lin & Mubarok, 2021; Wang et al., 2022). As young learners participated in multiple rounds of interaction with the chatbot

MSLIPA, they became more active and demonstrated a stronger willingness to communicate (Wu et al., 2024). As narrators, chatbots provide story-based or knowledge-rich content, often through one-directional but dialogically framed explanations that support the learning of grammar, collocations, and sentence patterns (Bailey et al., 2021; Zhang et al., 2023). For example, the AI chatbot Huang delivered substantive and elaborated responses beyond simple definitions, leading to high levels of learner engagement during an eight-week experiment where students asked over 20 questions per week (Zhang & Huang, 2024). In interviewer or facilitator roles, chatbots can either elicit structured learner responses or guide learners through goal-oriented tasks such as simulated job interviews, decision-making activities, or problem-solving exercises (Ghafouri, 2024; Tai & Chen, 2022). For instance, Timpe-Laughlin et al. (2023) conducted a role-play task in an interviewer-style format, where 47 primary-level EFL learners were asked to make requests to their boss who was performed by an automated AI agent and a human interlocutor in succession. Although learners experienced significantly more dialogic rounds with the human boss, they engaged in slightly longer conversations during the automated interactions. This difference may be attributed to the additional time spent on building social rapport, clarifying requests, and using backchanneling strategies (Timpe-Laughlin et al., 2024).

In addition to system design, learner strategies, such as repeating, rephrasing, and seeking clarification, emerge in chatbot-mediated interactions as ways to manage breakdowns (Moussalli & Cardoso, 2020; Wu et al., 2024). Strategy use often varies with proficiency: higher-proficiency learners favor rephrasing, while lower-proficiency learners rely on repetition or avoidance (Dizon & Tang, 2020). Implicit feedback from the chatbot, such as questions, silences, or incorrect responses, often prompts learners to negotiate meaning and refine their output (Wu et al., 2024). Compared to peer interactions, chatbot-mediated conversations elicit richer and more adaptive strategies, as learners face less social pressure and fear of embarrassment (Rahman & Tomy, 2023). The less-threatening nature may lessen learners' affective filter, encouraging persistence and attempts to overcome communication challenges (Moussalli & Cardoso, 2020). However, less attention has been paid to the real-time interaction process. Although Wu et al. (2024) made groundbreaking progress by analyzing the amount, content and strategies of interactions between primary school students and a chatbot, further research is needed to establish a stronger connection between chatbot-mediated interactions and learners' motivation.

2.2. L2 motivational self system and intended effort

Motivation is 'the state of cognitive and emotional arousal which leads to a conscious decision to act' (Williams & Burden, 1997, p. 120). It might be shaped by intrinsic factors such as personality and intelligence, as well as extrinsic ones including societal influences, family, school and technology (Dörnyei & Ushioda, 2013). To better understand the complexities of motivation in L2 learning, Dörnyei (2005) developed the L2MSS, a theoretical framework grounded in self and identity theories.

The L2MSS, as a tripartite model, is composed of three interrelated components: the ideal L2 self, the ought-to L2 self, and the L2 learning experience (Dörnyei, 2005, 2009). Sitting in the center of the three dimensions, the ideal L2 self refers to the ideal state that L2 learners expect to achieve, relating to personal goals and future-oriented motivations. The ought-to L2 self reflects external expectations and obligations, driving learners to meet societal or interpersonal demands or to avoid negative outcomes. The L2 learning experience encompasses situation-specific factors related to the immediate learning environment, such as the influence of teachers, peers, and AI tools (Taguchi et al., 2009). Unlike the future-oriented selves, the L2 learning experience focuses on learners' attitudes toward their current learning context and has been highlighted as a significant factor in shaping motivation through authentic engagement (Dörnyei, 2019).

Intended effort, a key outcome of motivation, emerges from the discrepancy between a learner's current self and their ideal or ought-to selves (Henry & Cliffordson, 2017). When learners perceive a gap between their present abilities and their aspirations or external expectations, they are often motivated to exert effort to close this gap (Al-Hoorie, 2018). Research suggested that the ideal L2 self tends to have a stronger and more consistent positive correlation with intended effort than the ought-to L2 self, which may sometimes demotivate learners due to its obligation-driven nature (Dörnyei & Chan, 2013). The L2 learning experience also shows moderate correlations with intended effort, although its direct relationship with achievement is weaker (Al-Hoorie, 2018).

Motivation-related studies in chatbot-mediated learning environments yielded mixed results. For instance, Chien et al. (2022) found that while extrinsic motivation increased during a four-week English conversation exercise using the LINE ChatBot, intrinsic motivation only rose when competition was introduced. Conversely, Fryer et al. (2017, 2019) identified the novelty effect as crucial for sustaining motivation, noting that learners' motivation often declined as the novelty of chatbots faded. Bräuer and Mazarakis (2022) confirmed this pattern, which reported that long-term chatbot use produced weaker motivational outcomes compared to short-term studies. These earlier studies, however, primarily involved rule-based or retrieval-based chatbots with limited conversational depth and motivational adaptability.

Extending earlier attempts, a few recent studies have turned their attention to LLMs-powered chatbots, which are capable of more natural and adaptive interaction. So far, this line of inquiry has largely focused on writing contexts. Huang and Mizumoto (2024), for example, examined the integration of ChatGPT in EFL writing classes and found a slight increase in students' ideal L2 self, but no significant effects on ought-to L2 self or L2 learning experience. Despite these developments, few studies have directly linked motivation, particularly L2MSS constructs, to the chatbot-mediated interaction process. Given that learners' interactions with teachers, peers and chatbots are integral components of L2 learning experience (Kuhail et al., 2023), it is necessary to explore how dialogic engagement with LLMs-powered chatbots influences these motivational elements. Building upon the previous literature and our objectives of this study, three research questions are proposed below.

RQ1. How do learners' ideal L2 self, ought-to L2 self, and L2 learning experience change from pre- to post-test during chatbot-mediated interactions, and how do these changes differ from those observed in structured speaking activities?

RQ2. How does learners' intended effort change from pre- to post-test during chatbot-mediated interactions, and how does this change compare to that in structured speaking activities?

RQ3. What are the thematic structures and characteristics of dialogic content in the chatbot-mediated interaction process, and how might the identified topics and behaviors relate to motivational changes?

3. Methodology

3.1. Participants

We adopted a mixed-methods research design to measure the interaction process, L2MSS, and intended effort in a quasi-experimental setting. A total of 232 undergraduate students were recruited from four intact classes at a public university in Southwest China. All participants were second-year non-English majors enrolled in the same compulsory College English course, which is part of the university's general education curriculum. As such, participant selection followed a convenience sampling approach, based on existing class assignments rather than individual randomization.

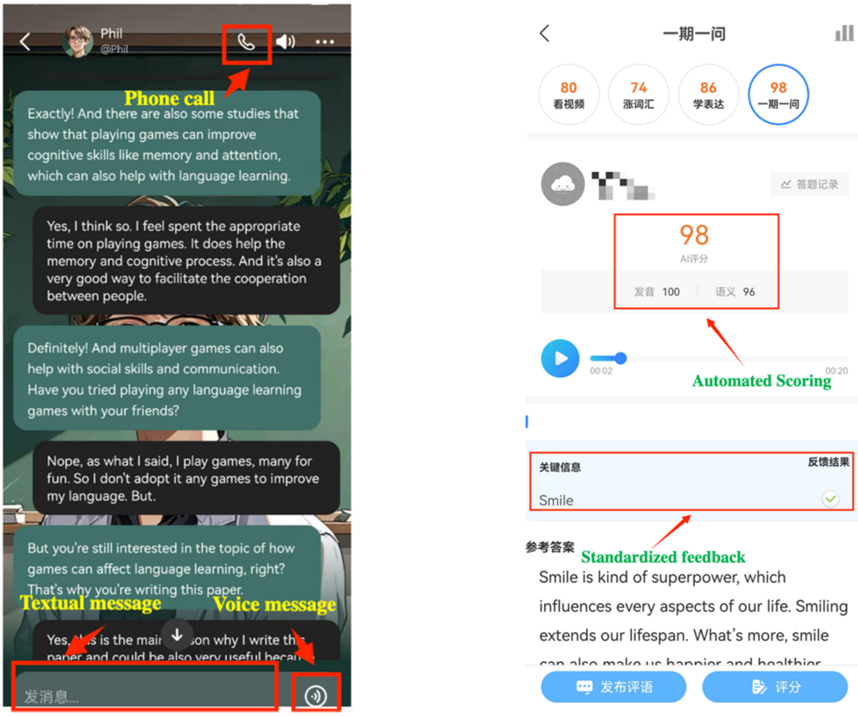
Students' English proficiency levels were inferred based on their performance in the College English Test (CET) Band 4 and Band 6, which are widely used high-stakes proficiency tests in mainland China. These tests are aligned with the Common European Framework of Reference for Languages (CEFR), with CET-4 corresponding approximately to CEFR B1 and CET-6 to B2 levels (Zhao et al., 2017). Accordingly, participants in this study were estimated to fall within the A2-B1 range on the CEFR scale.

For research purposes, these four classes were randomly labeled as Class A, B, C, and D. Class A and B were assigned to the experimental group (EG-CHAT), consisting of 118 students (83 males and 35 females), while Class C and D formed the control group (CG), with 114 students (78 males and 36 females). The four classes were considered comparable in terms of gender distribution, academic year, and English proficiency levels, as reflected in their CET scores. All students received instruction from the same experienced English teacher using a shared syllabus, textbook, and weekly schedule. The only difference lay in their assigned post-class speaking activities: the EG-CHAT completed LLMs-powered speaking activities through chatbot-mediated interactions, while the CG used the FiF Speaking Practice platform, which involved structured speaking activities, such as video watching, repeating and shadowing.

3.2. Instruments

3.2.1. Doubao and FiF speaking practice

Two learning systems, Doubao and FiF Speaking Practice were selected to assign activities in the quasi-experiment. Doubao, a representative LLMs-powered chatbot developed by ByteDance, was employed to deliver self-regulated and flexible speaking activities for EG-CHAT. The chatbot features extensive knowledge, emergent capability, and multifaceted functions for handling communication tasks, information search, and even music generation. Learners can interact with Doubao through textual messages, voice messages, or even simulated phone calls, all of which are



a. Modes of Interaction in Doubao b. Interface in FiF Speaking Practice

Figure 1. Flexible conversation on Doubao and structured practice on FiF Speaking Practice.

transcribed into text during interaction (see Figure 1a). Given its performance comparable to GPT-4o and Claude 3.5, as well as its easy access on mobile phones in China, it was chosen as the favorable chatbot for supporting the LLMs-powered speaking activities in this study.

FiF Speaking Practice is a commercial English learning system that functions without the support of LLMs. It provides a structured bank of speaking activities drawn from IELTS tasks, TED talks, and authentic global news content related to economics and technology (Du & Gao, 2022). This system integrates automated scoring and standardized feedback, offering learners numerical evaluations on key speaking dimensions and example responses for comparison (see Figure 1b). Due to its reliability and ease of use, FiF Speaking Practice has been integrated into College English programs in China for nearly a decade (Zhang & Li, 2019). This was used by the CG as a comparison to Doubao.

3.2.2. LLMs-powered speaking activities in chatbot-mediated interactions

The optimal use of chatbots requires well-designed interaction activities and call-to-action prompting. By following the four interactive styles including interlocutor, narrator, interviewer and facilitator proposed by Engwall et al. (2021), we prepared seven LLMs-powered speaking activities (see Table 1). The first category, reflects the interlocutor style, aiming to create open-ended, interest-driven conversations that facilitate spontaneous expression (Bibauw et al., 2019). The second category,

Table 1. LLMs-powered speaking activities categorized by interaction styles.

Styles	Activities	Description	Sample Prompt
Interlocutor	Free talk	Starting a flexible topic.	Let's talk about your favorite game.
	Debate	Practicing communicative and critical thinking skills through discussion.	I don't agree with you. Smiling is not only a facial expression…
Narrator	Storytelling	Asking Doubao to tell a story in English.	Can you tell me a story about the Little Mermaid?
	Knowledge delivery	Learning vocabulary, grammar, or syntax through interaction.	What's meaning of the word timid? How can I use the word in practice?
Interviewer	Role-play	Simulating real-life oral tasks like ordering food or job interviews.	Let's have a role-play about ordering a take-out in a restaurant.
	Language proficiency test	Taking mock English tests like IELTS or College English Test with Doubao.	Can you work as an IELTS examiner to start a speaking test?
Facilitator	Problem solving	Asking Doubao for advice or support with a specific issue.	I am struggling with time management. Can you give me some advice?

which centers on storytelling and knowledge delivery, corresponds to the narrator role. Leveraging Doubao's extensive knowledge base, these activities help students improve listening comprehension while acquiring target language forms. The third category, simulation-based activities, adopts the interviewer style by incorporating real-life or business-like scenarios into structured dialogues that simulate practical communicative settings (Lee & Lim, 2023; Tai & Chen, 2022). Finally, the fourth category follows the facilitator style, where Doubao functions as an expert, engaging students to solve specific problems or explore complex issues by offering tailored advice. Students in the EG-CHAT were encouraged to talk with Doubao for at least 15 min per week, using but not limited to the seven recommended activities.

In contrast, the CG participated in structured practices like video watching, repeating and shadowing, also for a minimum of 15 min each week. In this context, repeating referred to students listening to a model sentence or passage and repeating it verbatim, while shadowing involved listening and simultaneously speaking along with the audio to improve fluency and pronunciation. Both were delivered through the FiF Speaking Practice, which also provided automated scores and standardized feedback based on key content words.

3.2.3. L2 attitude/motivation scale

In order to examine the impact on L2MSS and intended effort between chatbot-mediated interactions and regular exercises *via* FiF Speaking Practice, we adopted the L2 attitude/motivation scale designed by Taguchi et al. (2009). The 40 items in the questionnaire were categorized into four aspects, including the ideal L2 self, the ought-to L2 self, the L2 learning experience, and intended effort (see Appendix A). The overall reliability of the questionnaire is excellent (Cronbach $\alpha = 0.935$), referring the high internal consistency of the scale. Participants were asked to respond to the items on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

3.2.4. Latent Dirichlet Allocation (LDA) topic modeling

Instead of using a semi-structured interview that falls into the regular research design, we employed Latent Dirichlet Allocation (LDA) topic modeling to identify and cluster textual characteristics for the dialogue logs of the EG-CHAT. LDA is a widely-used topic modeling algorithm that reveals thematic structure by estimating the distribution of words within each topic (Blei et al., 2003). The parameters for LDA, including the Dirichlet priors ($\alpha = 0.1$ and $\beta = 0.01$), were empirically set to control the topic and word distributions (Griffiths & Steyvers, 2004; Wallach et al., 2009). The optimal number of topics was determined using perplexity, which is commonly used to assess the LDA model's quality. Though the lower perplexity predicts the better fit, adding more topics beyond a certain point might result in overfitting for the model when the perplexity curve starts to level off (Chang et al., 2009). Coherence was not used in this study because the conversations between learners and Doubao were scattered across diverse dimensions rather than focusing on a single broad topic. As suggested by Aletras and Stevenson (2013), coherence metrics are more effective for structured and content-rich datasets, whereas they can be unreliable for fragmented or unstructured contexts. This adoption slightly differs from Wu et al. (2024), who employed a combination of perplexity and coherence metrics to examine interactions between primary school students and the self-designed chatbot MSLIPA. While Wu et al. (2024) focused on structured in-class online tasks, our study encouraged students to engage in flexible, conversational interactions outside of class.

3.3. Data collection

As shown in Figure 2, we followed quasi-experimental procedures to collect interaction and questionnaire data. Before the experiment, written informed consent was obtained from all participants. The pre-test was performed with the participants in both groups to measure their motivation and intended effort levels with the L2 attitude/motivation scale. The participants in both groups were also informed how to appropriately use the two systems, respectively.

During the five-week experiment, participants in both EG-CHAT and CG received three hours of in-class instruction per week. All four classes were taught by the same experienced teacher, who was also one of the researchers. This ensured consistency in syllabus, materials, and in-class learning activities, which covered topics such as 'Laypeople's understanding of radiation and radioactivity', 'The Art of Reputation', and 'The NBA Data Scientist'. The chatbot-mediated and structured speaking activities were assigned as post-class tasks, completed by students independently. The teacher's involvement in these tasks was limited to initial guidance and technical support to ensure proper use of each platform.

At the end of the experiment, participants in the two groups spent 20 min to complete the L2 attitude/motivation scale as post-test. We collected the questionnaire results with a pencil-and-paper survey while students in the EG-CHAT sent their screenshots of dialogue logs to the teacher *via* e-mail.

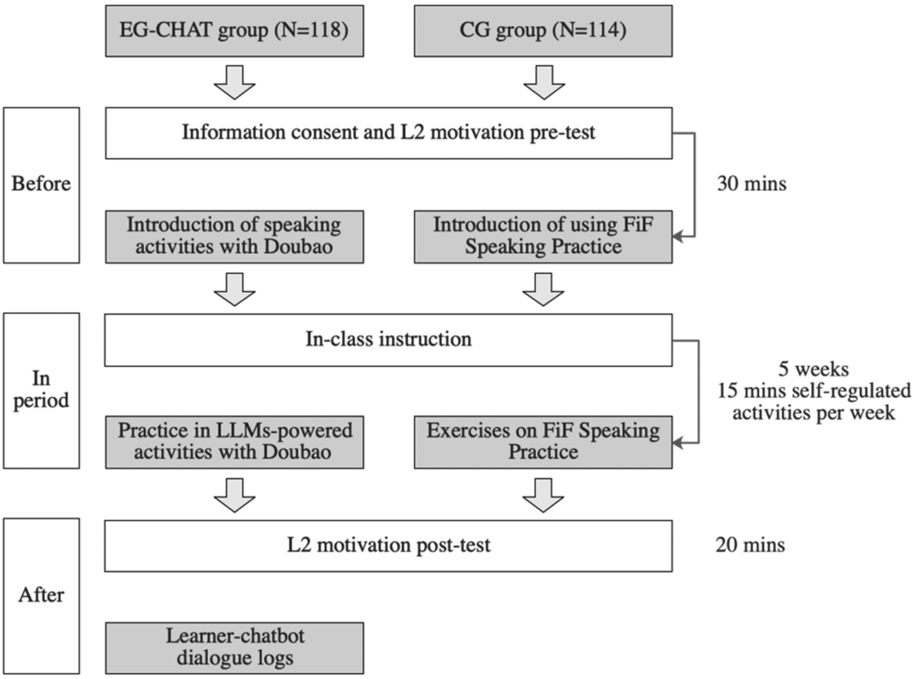


Figure 2. Procedures of the quasi-experiment.

To ensure comparability, both the EG-CHAT and CG were recommended to engage in post-class speaking activities for a minimum of 15 min per week over the five-week period. The time-on-task was communicated as a weekly requirement and monitored through dialogic screenshots (for EG-CHAT) and learning logs recorded in the FiF system (for CG). Although the input formats varied—interactive dialogue with Doubao versus structured speaking exercises *via* FiF Speaking Practice—the time and frequency of engagement were designed to be equivalent across groups. Given this fundamental difference in interactivity and depth of feedback, the learning logs, scores, and standardized feedback recorded on FiF Speaking Practice were not used for the comparison of interaction processes. Instead, only the EG-CHAT's dialogic content was included for topic modeling, as it directly reflected spontaneous responses and thematic structure within chatbot-mediated interactions.

3.4. Data analysis

To evaluate the differences in L2MSS and intended effort, independent-samples t tests were conducted between the EG-CHAT and CG prior to the quasi-experiment. Table 2 shows that there was no significant difference between the two groups in the ideal L2 self ($t = -0.663$, $p = 0.508$), ought-to L2 self ($t = -0.032$, $p = 0.974$), L2 learning experience ($t = 0.846$, $p = 0.398$), and intended effort ($t = 0.625$, $p = 0.533$). The two-way mixed design ANOVA was subsequently applied to post-test scores to measure the effects of the two types of speaking activities (i.e., LLMs-powered and pre-arranged practices) on these variables. The selection of the two-way mixed design ANOVA allows for gauging within-subject effects from pre- to post-test,

Table 2. Descriptive statistics of pre-test of L2MSS and intended effort.

	Group	*N*	Mean	*SD*	*t*	*df*	*p*
ideal L2 self	EG-LLMs	118	3.169	0.9389	−0.663	230.000	0.508
	CG	114	3.252	0.9513			
ought-to L2 self	EG-LLMs	118	2.710	0.9864	−0.032	230.000	0.974
	CG	114	2.714	0.8323			
L2 learning experience	EG-LLMs	118	3.515	0.9921	0.846	230.000	0.398
	CG	114	3.413	0.8358			
Intended effort	EG-LLMs	118	3.194	0.7669	0.625	230.000	0.533
	CG	114	3.131	0.7776			

$^*p < 0.05$, $^{**}p < 0.01$.

between-subject effects between the experimental and control groups, and the interaction effects between time and group (Şendağ & Odabaşı, 2009).

As for the analysis of the interaction process, the dialogue logs of the EG-CHAT were processed and analyzed using Python through the following steps: Optical Character Recognition (OCR) transcription, data preprocessing, and LDA topic modeling. In the first step, the acquired screenshots of dialogue logs were transcribed into text using OCR. Out of 118 PDF or RAR packages of screenshots, 87 were eligible for transcription. Files were excluded for some reasons such as poor image quality, the use of Chinese language, and bullet-points instead of full and natural responses. In the second step, data cleaning, tokenization, and lemmatization were carried out using the NLTK library. During the process, special characters, emojis, non-informative symbols, and stopwords (e.g. 'you', 'it', 'my', 'wouldn't') were removed, followed by breaking the text into tokens and converting them to their base forms using WordNetLemmatizer. The preprocessing step produced 7,007 cleaned sentences, with a total word count of 127,234 words and 4,931 unique terms. In the final step, the cleaned data were subjected to LDA topic modeling using the Gensim library, with perplexity scores and topic clusters reported in the results section below.

4. Results

4.1. Changes of L2MSS between the EG-CHAT and CG groups

By means of questionnaire survey before and after the experiment, we gauged the impact of the two types of speaking activities on students' L2 motivational system for both the EG-CHAT and CG. Descriptive statistics (see Table 3) showed that the two groups exhibited different changes in their motivational scores across the intervention period. For the ideal L2 self, the EG-CHAT experienced an increase from 3.169 ($SD = 0.939$) to 3.490 ($SD = 0.814$), whereas the CG showed a minor decrease from 3.252 ($SD = 0.951$) to 3.191 ($SD = 0.906$). The ought-to L2 self mean scores for the EG-CHAT slightly decreased from 2.710 ($SD = 0.986$) to 2.619 ($SD = 0.912$), while the CG remained relatively unchanged. Both groups demonstrated gains in L2 learning experience, with the EG-CHAT increasing from 3.515 ($SD = 0.992$) to 3.897 ($SD = 0.643$), and the CG from 3.413 ($SD = 0.836$) to 3.712 ($SD = 0.738$).

The two-way mixed design ANOVA was conducted to statistically assess whether these observed motivational changes differed significantly between the EG-CHAT and CG from the pre- to post-test. As shown in Table 4, although both groups

Table 3. Descriptive statistics of pre- and post L2 motivational system.

		Pretest			Posttest		
	Group	Mean	*SD*	*N*	Mean	*SD*	*N*
ideal L2 self	EG-CHAT	3.169	0.939	118	3.490	0.814	118
	CG	3.252	0.951	114	3.191	0.906	114
ought-to L2 self	EG-CHAT	2.710	0.986	118	2.619	0.912	118
	CG	2.714	0.832	114	2.715	0.963	114
L2 learning experience	EG-CHAT	3.515	0.992	118	3.897	0.643	118
	CG	3.413	0.836	114	3.712	0.738	114

Table 4. Summary of two-way mixed design ANOVA with ideal L2 self.

Source	SS	df	MS	*F*	*p*	η^2
Between groups	4976.961	1	4976.961	5655.301	0.000**	0.961
Group (EG-CHAT vs CG)	1.357	1	1.357	1.542	0.216	0.007
Error (between)	202.412	230	0.880			
Within groups						
Time (pre vs post)	1.957	1	1.957	2.598	0.108	0.011
Time × Group	4.205	1	4.205	5.584	0.019*	0.024
Error (time)	173.222	230	0.753			

*$p < 0.05$, **$p < 0.01$.

Table 5. Summary of two-way mixed design ANOVA with ought-to L2 self.

Source	SS	df	MS	*F*	*p*	η^2
Between groups	3355.684	1	3355.684	2991.498	0.000**	0.929
Group (EG-CHAT vs CG)	0.286	1	0.286	0.255	0.614	0.001
Error (between)	258.000	230	1.122			
Within groups						
Time (pre vs post)	0.234	1	0.234	0.395	0.530	0.002
Time × Group	0.243	1	0.243	0.410	0.522	0.002
Error (time)	136.255	230	0.592			

*$p < 0.05$, **$p < 0.01$.

began with comparable levels of ideal L2 self, a significant interaction effect between time and group ($F_{(1,230)}$ = 5.584, p < .001, η^2 = .024) indicated that the two groups followed divergent changes across measurements: the EG-CHAT showed a significant increase, while the CG experienced a slight decline. However, the main effect of time was not significant ($F_{(1,230)}$ = 2.598, $p = 0.108$), nor was the main effect of group ($F_{(1,230)}$ = 1.542, $p = 0.216$), suggesting no overall difference across groups or time points.

As demonstrated in Table 5, neither the main effect of time ($F_{(1,230)}$ = 0.395, $p = 0.530$) nor the main effect of group ($F_{(1,230)}$ = 0.255, $p = 0.614$) was significant for the ought-to L2 self. Additionally, there was no significant interaction effect between time and group ($F_{(1,230)}$ = 0.410, $p = 0.522$). That is to say, the intervention did not produce a differential effect in the ought-to L2 self scores between the EG-CHAT and CG from pre- to post-test.

Despite the very small effect sizes, Table 6 shows a significant main effect of time ($F_{(1,230)}$ = 22.347, $p < 0.01$, η^2 = 0.089) while no significant effect of group ($F_{(1,230)}$ = 3.325, $p = 0.070$) across the measurements. It indicates that all participants,

Table 6. Summary of two-way mixed design ANOVA with L2 learning experience.

Source	SS	df	MS	*F*	*p*	η^2
Between groups	6127.555	1	6127.555	8515.451	0.000**	0.974
Group (EG-CHAT vs CG)	2.393	1	2.393	3.325	0.070	0.014
Error (between)	165.504	230	0.720			
Within groups						
Time (pre vs post)	13.458	1.000	13.458	22.347	0.000**	0.089
Time × Group	0.200	1.000	0.200	0.332	0.565	0.001
Error (time)	138.511	230.000	0.602			

$^*p < 0.05$, $^{**}p < 0.01$.

Table 7. Descriptive statistics of pre- and post intended effort.

	Pretest			Posttest		
Group	Mean	*SD*	*N*	Mean	*SD*	*N*
EG-CHAT	3.194	0.767	118	3.647	0.548	118
CG	3.131	0.778	114	3.258	0.764	114

Table 8. Summary of two-way mixed design ANOVA with intended effort.

Source	SS	df	MS	*F*	*p*	η^2
Between groups	5073.889	1	5073.889	8889.357	0.000**	0.975
Group (EG-CHAT vs CG)	5.925	1	5.925	10.381	0.001**	0.043
Error (between)	131.280	230	0.571			
Within groups						
Time (pre vs post)	9.744	1	9.744	20.964	0.000**	0.084
Time × Group	3.069	1	3.069	6.603	0.011**	0.028
Error (time)	106.900	230	0.465			

$^*p < 0.05$, $^{**}p < 0.01$.

regardless of group, experienced a similar increase in their L2 learning experience from pre-test to post-test. However, no significant interaction effect was found between time and group ($F_{(1,230)}$ = 0.332, $p = 0.565$). It could suggest that while the L2 learning experience scores for both groups increased, neither group outperformed the other across the intervention period.

4.2. Changes of intended effort between the EG-CHAT and CG groups

To examine how learners' intended effort developed across the intervention period, we analyzed both descriptive statistics and the ANOVA results (see Tables 7 and 8). As shown in Table 7, the EG-CHAT's mean scores increased from 3.194 ($SD = 0.767$) to 3.647 ($SD = 0.548$), whereas the CG's scores rose from 3.131 ($SD = 0.778$) to 3.258 ($SD = 0.764$).

As shown in Table 8, the two-way mixed-design ANOVA identified a statistically significant main effect of time ($F_{(1,230)}$ = 20.964, $p < 0.01$, η^2 = 0.084) and of group ($F_{(1,230)}$ = 10.381, $p < 0.01$, η^2 = 0.043). More importantly, a significant interaction effect between time and group ($F_{(1,230)}$ = 6.603, $p < 0.01$, η^2 = 0.028) was found. In this regard, although both groups reported increased intended effort from pre- to post-test, the EG-CHAT demonstrated a significantly greater increase than the CG.

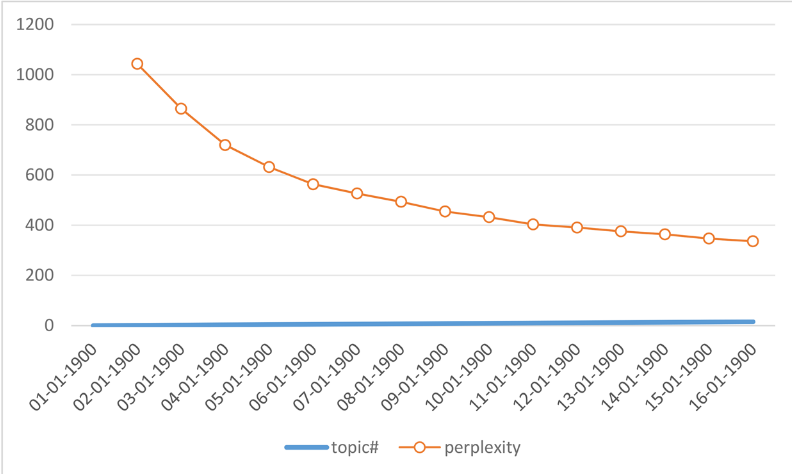


Figure 3. The perplexity curve of LDA modeling.

4.3. Interaction process in chatbot-mediated interactions

In order to analyze the interaction process, the dialogue logs between Doubao and the 87 students in the EG-CHAT were clustered using LDA topic modeling. According to Figure 3, the perplexity curve shows a sharp decline in perplexity as the number of topic increases, with a noticeable leveling off around 10 to 12 topics. This pattern indicates that adding more topics beyond this range yields diminishing returns in model fit. When reviewing the topic clusters produced by the 10-, 11-, and 12-topic solutions, we observed that several topics in the 11- and 12-topic models exhibited semantic overlap and thematic redundancy. To ensure interpretability and clarity, two researchers independently examined the outputs and reached a consensus that the 10-topic model offered the most distinct and coherent structure. A collaborative labeling of the 10 topics was then conducted using a consensus coding approach between the two researchers.

To better interpret the interaction process, the 10 topics were further grouped into three overarching dimensions: learning-related, class-related, and life-related. Table 9 presents each topic, the number of clustered sentences, and the top 20 characteristic words generated by the LDA model.

4.3.1. Learning-related conversations

The learning-related dimension revolves around issues related to learning, including topics such as English learning, tasks and scheduling, and academic development. On the central topic of English learning, students emphasized pronunciation, vocabulary, grammar, listening, speaking, and other related skills. They frequently asked questions like '*Do you have any tips for remembering new words*', '*What are some effective ways to learn English grammar*', and '*How do I improve my listening and comprehension skills*'. In addition to seeking advice, students shared useful tips and resources with Doubao. English books, movies, music, TED talks, and famous speeches like *The King's Speech* as helpful materials frequently appeared in their

Table 9. Clustering results of LDA topic modeling.

Categories	Topics	Sentences	Top 20 characteristic words
Learning-related	English learning	1052	English, question, learn, skill, practice, keep, movie, talk, language, word, free, learning, topic, book, ask, listening, grammar, conversation, course, pronunciation
	Tasks & scheduling	616	Time, say, focus, start, mean, thank, form, sense, little, task, break, down, exam, situation, point, okay, first, advice, run, schedule
	Academic development	557	Study, idea, future, support, part, goal, sleep, plant, research, mind, set, attention, call, test, night, class, performance, phone, smell, habit
Class-related	Sports & games	613	Watch, play, player, game, team, show, future, world, football, hard, action, sport, character, basketball, ability, live, game, movie, career, nba
	Arts, music & cultures	689	Work, favorite, day, sound, fun, song, love, story, music, kind, plan, artist, nice, today, activity, famous, chongqing, place, art, beautiful
	Big data	590	Datum, example, big, student, problem, benefit, area, education, transportation, system, security, ensure, teacher, field, concern, college, method, progress, privacy, decision
	Radiation	460	Radiation, health, include, nuclear, risk, high, lead, effect, level, increase, source, human, cause, potential, exposure, safety, medical, change, travel, factor
Life-related	Leisure time	532	Food, look, one, process, sodium, exercise, type, weather, enhance, option, body, amount, check, app, choice, home, intake, fish, diet, gain
	Societal challenges	628	People, new, challenge, waste, material, while, information, offer, development, issue, energy, deal, technology, resource, aspect, internet, opinion, china, society, celebrity
	Mood & emotions	544	Help, important, thing, right, life, friend, experience, smile, relationship, doctor, patient, quality, share, stress, answer, mood, physical, communication, job, knowledge

conversations. Also, when learners perceived Doubao's responses as too lengthy, they sought ways to negotiate, saying things like, '*Why are there so many words in your answer? I don't need such long conversations*'.

Regarding tasks and scheduling, students often expressed concerns about time management, final exams, and organizing their daily schedules. Treating Doubao as a friend, they shared personal reflections like '*I lack time to spend on learning*' and asked for advice on how to '*arrange my time in the period of final exam review*'. Some students also shared their end-of-day routines, commenting '*making a cup of hot coffee and enjoying the review in the rain*' as a way to unwind.

As for the topic of academic development, the concerned issues are composed of future goals, academic performance and career aspirations. Many discussions centered on assignments, tests, and grades, with students expressing curiosity about their life after graduation. Since students spent their spare time to talk with Doubao in the evening, some of them wrapped up conversations with interesting closings like '*I'm going to sleep now to prepare for tomorrows exam*' and '*Man I'm going to sleep*'.

4.3.2. Class-related conversations

The class-related dimension focuses on topics closely related to in-class listening and reading materials, including sports and games, arts and music, big data, and radiation. Notably, students naturally connected their personal interests to classroom content, as they engaged in open-topic speaking activities with Doubao, rather than discussing pre-defined topics. The first two topics—sports and games, and arts and music—established a strong connection between students' interests and the in-class content. After studying materials like '*the NBA Data Scientist*' and '*the Art of Reputation*' in class, influential names such as '*Curry*', '*SGA*', '*Cristiano Ronaldo*', '*Rembrandt*', '*Vermeer*', and '*Taylor Swift*' appeared frequently in their conversations. What students learned in English class recalled memories of splendid moments from the NBA playoff and European cup, as well as masterpieces by great artists that had been tucked away in their minds. In addition to sharing their favorite players, artists and singers, students also discussed their experiences of visiting and appreciating different cities like Chongqing and Xi'an. As they asked and commented, '*Do you think Chongqing is a city worth visiting*', '*What do you know about the cultural customs of Chongqing*', and '*Chongqing is a very unique and vibrant city*'.

Regarding the topics of big data and radiation, the knowledge students gained in class raised concerns and fear. Although they attempted to balance the benefits and risks of big data, students posed questions about '*What are the problems in education*', '*How can we protect our privacy and security when using self-tracking technologies*', and '*What are the best practices for protecting student data security in schools, colleges and universities*'. As for radiation, students shared in-depth understandings of the types and causes of radiation. Some reflected, '*ionizing radiation is harmful to living cells and non-ionizing is less harmful*', and '*radiation made by nuclear*', while others held unusual but intriguing thoughts like, '*if there are enough levels of radiation, will it cause a person to travel through space and time*'.

4.3.3. Life-related conversations

The life-related category includes topics that touch on students' daily lives and broader social issues, such as leisure time, societal challenges, and mood and emotions. In the topic of leisure time, students discussed activities like hanging out to eat, exercising, and maintaining a healthy lifestyle. Following students' sharing like '*Today I had a nice day I enjoyed hot pot with my roommate*', '*How often should I exercise to lose weight effectively*', and '*How can I limit my sodium intake*', Doubao replied actively and gave advice and support.

Beyond personal stories, students also engaged in conversations about global and societal issues, expressing concerns about '*China-US relations*', '*new energy vehicles*', and '*the future development trends in artificial intelligence*'. When discussing food emergencies, they showed empathy by supporting actions like '*donate food and funds to local food banks*'.

Lastly, emotional topics related to smiling, happiness, and handling bad emotions were frequently discussed. A noteworthy point is, some emotional responses either from Doubao or the students were clustered within this topic. Doubao often provided positive and affirmative responses like '*That's an interesting question*', '*You have a point there*', and '*Wow, you have some amazing people in your life*'. Students appreciated these supportive interactions, returning with messages such as '*I'm happy to talk to you*', '*Thank you I'm in a better mood*', and '*It's been a long day, I want to talk with you*'.

5. Discussion

5.1. Improvement of ideal L2 self and L2 learning experience

The two-way mixed-design ANOVA results revealed a statistically significant interaction effect between time and group for the ideal L2 self ($F_{(1,230)} = 5.584$, $p < .001$, $\eta^2 = .024$), suggesting that the two groups exhibited different changes across the intervention period. Specifically, the EG-CHAT's ideal L2 self scores increased from 3.169 to 3.490, while the CG experienced a slight decline (from 3.252 to 3.191). These results provide quantitative evidence that chatbot-mediated speaking activities may have positively influenced learners' self-perception as future L2 users. However, for both groups, whether they engaged in chatbot-mediated interactions or practiced on FiF Speaking Practice, there were no significant change on the ought-to L2 self from pre- to post test. Regarding the L2 learning experience, both groups showed significant improvements, though no group outperformed the other. These findings diverge significantly from those reported by Huang and Mizumoto (2024). In their study, despite a slight increase in students' ideal L2 self, this change was not statistically significant. Plus, no significant increases were observed in their ought-to L2 self or L2 learning experience after students collaborated with and received feedback from ChatGPT in an English writing class. The discrepancy may stem from the different contexts of the two studies. While Huang and Mizumoto (2024) focused on writing tasks, our study centers on spoken interactions, which may elicit different motivational responses. Another potential reason is that chatbot-mediated interactions in our study were observed as more natural and human-like, which may have contributed to the significant motivational changes. These findings reinforce the importance of interaction in language acquisition (Illeris, 2016), whether through human-human conversations or learner-chatbot interactions.

The LDA topic modeling results offer further insight into the significant increase in the ideal L2 self for the EG-CHAT. The content-rich interactions between learners and Doubao reflect an implicit language acquisition process, wherein learners build tacit knowledge that enables them to communicate effectively, even if they cannot explicitly articulate the rules governing their language use (Rebuschat & Williams, 2012). Through incidental exposure in chatbot-mediated interactions, learners access to and mimic lexical uses and syntactic structures. As Loewen (2020, p. 25) noted, effective communication relies on implicit knowledge, which is essential for learners to speak fluently and effortlessly. These perceived opportunities for incidental learning align with Hwang et al. (2022), who emphasized that the free talk in chatbot-mediated interactions can motive learners to apply words, phrases, and sentences acquired in the classroom, thereby sparking their willingness to sustain dialogue. Upon examining the dialogic content, we also found that learners often treated Doubao as a peer or a fellow English speaker, rather than as a teacher or authority figure. Learners tended to share their favorite NBA stars, talk about struggles they faced before exams, and even say farewell to Doubao, following a similar interactive pattern reported by Wang et al. (2022). When the chatbot is perceived as human-like, learners, as noted by Chaves and Gerosa (2021), may engage in more personal interactions. The ability to practice speaking at any time without the pressure of a classroom setting allows learners to feel more autonomous and emotionally supportive, which can contribute to a stronger ideal self-image as competent language users.

The ought-to L2 self is shaped by external factors, such as societal norms and expectations from teachers (Dörnyei, 2009). Since neither the chatbot-mediated interactions nor the regular speaking activities on FiF Speaking Practice directly align with these external demands, no significant changes were observed in the ought-to L2 self for either the EG-CHAT or CG. Also, a rather short period of five weeks may not provide enough opportunities for learners to internalize the chatbot-mediated interaction as a habit or motivating force, as the teacher intended to encourage.

As for the L2 learning experience, the increase in the EG-CHAT may be attributed to the personalized feedback and high-quality language exposure provided by the chatbot, which created an anxiety-free environment and offered user-friendly opportunities for conversational practice. The finding was similarly found in a wide range of studies (Moussalli & Cardoso, 2020; Rahman & Tomy, 2023; Wan & Moorhouse, 2024). However, speaking activities on FiF Speaking Practice, such as watching videos, repeating, and shadowing, also allowed learners to hear authentic language use, improve pronunciation, and understand contextual usage. Although these activities are rather fixed and passive compared to chatbot-mediated interactions, they still positively contributed to learners' language experience. Moreover, the similar improvement in both groups might also related to the same supportive classroom setting, which includes access to resources, encouragement from the teacher, and opportunities for peer interaction. These elements likely contributed to comparable benefits in L2 learning experiences across both groups.

5.2. Higher investment on chatbot-mediated interactions

According to the ANOVA results, both groups followed a similar increase on intended effort. But the EG-CHAT perceived greater efforts in speaking activities than the CG, as evidenced by a significant interaction effect between time and group ($F_{(1,230)} = 6.603$, $p < 0.01$, $\eta^2 = 0.028$). Despite the increased perceived effort, the EG-CHAT might have experienced a sense of satisfaction, autonomy, and relatedness, which could have contributed to their behavioral engagement with the chatbot over time. As Mercer (2019) emphasized, these self-determined factors can drive learners to invest more time and effort into language learning. Aside from the utilitarian purposes of completing tasks to meet teacher expectations, learners in the EG-CHAT appeared to develop a more self-motivated approach. This change echoes the significant increase of the ideal L2 self, where students are driven by a sense of personal achievement rather than external obligations.

Combined with the LDA results, the conversations between students and Doubao were heavily influenced by interaction styles and emotional connections. These styles such as interlocutor and facilitator enabled students to explore a wide range of topics related to learning, class, and daily life, which liberate speaking practices from the 'form-meaning' dichotomy and provide explicit support for cognition, interaction, and collaboration (Tyler, 2008). The frequent occurrence of emotional words, including affirmative and caring feedback, suggests a relatively strong connection between learners and the chatbot. When learners feel that their emotions are acknowledged and mirrored, it fosters a sense of authenticity and immersion (Lang et al., 2022), which in turn transformed motives into practical behaviors. These cognitive and affective supports, aligned with the core mission of intelligent systems (Bodnar et al., 2016), compensate for the absence of a human teacher, who is supposed but unable to provide motivational scaffolding and manage frustration for every student in large-class settings.

5.3. Insufficient learner behaviors which deserve teachers' intervention

Although we did not collect self-report data on learner behaviors through questionnaires or interviews, the dialogic content offers an intuitive reflection of the diverse strategies students employed during chatbot-mediated interactions. Because learners were informed in advance how to communicate with the chatbot with various activities, they developed self-awareness to seek speaking opportunities and ways to improve vocabulary and grammar. This explains why the topic related to English learning appeared more frequently compared to the others. Besides, unlike interactions with traditional chatbots such as MSLIPA (Wu et al., 2024), where primary school students often relied on rephrasing or explicitly stating 'I don't know', university-level learners interacting with Doubao demonstrated a different approach. When confronted with Doubao's lengthy responses and excessive questioning in a single round of dialogue, learners in our study actively negotiated meaning and sought solutions. These behaviors may be partially attributed to age differences, while Doubao's favorable capabilities in understanding and generating content likely played a more significant role in encouraging learners to adopt a proactive problem-solving approach.

Despite these positive behaviors, the exchanges between students and the chatbot were generally broad and superficial, with students rarely demonstrating in-depth and analytical perspectives. The LDA results reveal that, even in topics they were most familiar with, such as sports, arts, and music, conversations were limited to sharing names of prominent figures or simple facts. This shallow level of interaction might be due to relatively low English proficiency and weak communicative skills among learners, while the absence of scaffolding in free communication may serve as a more important contributing factor (Tong et al., 2024). As Ushida (2005) argued, it is teachers who shape the culture of the classroom, decide on the use of course materials, and address learners' needs in diverse ways. When obstacles arise in the learner-chatbot interaction, a third party—whether a teacher or a peer—can be introduced to improve the conversation complexity and sustain engagement. Moreover, another reason behind the shallow communication is related to insufficient prior knowledge. It is evident as in-depth discussions emerged in those less common topics of big data and radiation. The prior knowledge acquired from the class triggered meaningful conversations about personal privacy protection and the causes and classifications of radiation. It confirms the findings of Dao et al. (2021) and Chen et al. (2023) that found the familiarity with prior knowledge significantly impacted learner engagement levels during tasks. This finding emphasizes the importance of in-class materials and activities, which serve as a catalyst for analytical and critical interactions between learners and the chatbot.

6. Conclusion

The momentum of AI development shows no signs of slowing, as if the pause button were never installed. The present study empirically examined the significantly positive impact of chatbot-mediated interactions on learners' L2MSS and intended effort, as well as unpacking the interaction process as evidence for these motivational changes. A significant contribution of this research is to introduce LDA topic modeling as a text mining approach, which allows for the direct tracking of interactions between learners and chatbots. This interdisciplinary approach effectively bridges the gap between quantitative measures and real-time dialogic content. In addition to that, the typical perspective on the interactive nature of chatbots offers a nuanced and dynamic understanding of motivational changes in chatbot-mediated environments. Such insights are instrumental for researchers and teachers to manage scaffolding and interventions. Furthermore, the LLMs-powered speaking activities and the collaborative workflow between learners and chatbots offer practical guidelines for the appropriate and responsible use of chatbots in L2 speaking contexts.

Admittedly, this study possessed some limitations. First, the CG, which engaged in FiF Speaking Practice, did not produce any dialogic data. Therefore, we were unable to conduct a direct comparison of the amount and types of dialogic content between the two groups. Second, while post-class practice time and curricular content were standardized across groups, the identified topics may have partly reflected students' concurrent academic concerns rather than entirely emergent interaction patterns, as the interactions occurred during an ongoing academic semester. Third, due to the nature of interaction data, we were only able to observe learner behaviors instead of accessing their thoughts and perceptions. To mitigate these issues in future research, it might be a constructive solution to incorporate more controlled prompts or retrospective interviews alongside topic modeling, thereby capturing both the spontaneous and context-driven aspects of learner engagement. By adopting this multi-layered design, subsequent studies can achieve a more favorable balance between depth and accuracy. Lastly, this study serves as an exploratory work with a relatively short duration. To understand the continuous influences as the novelty effect faded, it calls for longer-term experiments to follow up.

Acknowledgements

The authors would like to thank the reviewers and editors for their efforts and comments, which greatly improved the manuscript's quality.

Author contribution statement

Both authors contributed extensively to the work presented in this paper and approved the final manuscript. D.Y.F contributed to the study design, data collection and analysis, writing and revision. B.R worked on data analysis and revision. Both authors have approved the published manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to polish the language. After using this tool, the authors rigorously reviewed and edited the content as needed and take full responsibility for the content in the publication.

Disclosure statement

The authors declare that they have no competing interests.

Ethical approval

Ethical approval was obtained from the Ethics Committee for Research in School of Foreign Languages of Chongqing University of Technology [reference approval: CQUTSFL 2024/103].

Funding

This work was funded by the [Chongqing Municipal Education Commission] under grant number of [23SKGH250].

Notes on contributors

Yunfei Du is an Associate Professor in the School of Languages and Communication Studies at Chongqing University of Technology, Chongqing, and a PhD student in the Faculty of Education at University of Macau, China. His research interests include AI-assisted language learning (AILL), second language acquisition, and English for specific purposes.

Barry Lee Reynolds is an Associate Professor in the Faculty of Education at University of Macau, China. His research interests include computer-assisted language learning (CALL), L1/L2 vocabulary acquisition, written corrective feedback for L2 writing development, digital game-based learning, and eye tracking.

ORCID

Yunfei Du http://orcid.org/0000-0002-3213-2011
Barry Lee Reynolds http://orcid.org/0000-0002-3984-2059

Data availability statement

The datasets used and/or analyzed during the current study are available from the first author on reasonable request.

References

Ahn, T. Y., & Lee, S. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, *47*(4), 778–786. https://doi.org/10.1111/bjet.12354

Aletras, N., & Stevenson, M. (2013, March). Evaluating topic coherence using distributional semantics. In *Proceedings of the 10th International conference on computational semantics (IWCS 2013)—Long papers* (pp. 13–22).

Al-Hoorie, A. H. (2018). The L2 motivational self system: A meta-analysis. *Studies in Second Language Learning and Teaching*, *8*(4), 721–754. https://doi.org/10.14746/ssllt.2018.8.4.2

Bailey, D., Southam, A., & Costley, J. (2021). Digital storytelling with chatbots: Mapping L2 participation and perception patterns. *Interactive Technology and Smart Education*, *18*(1), 85–103. https://doi.org/10.1108/ITSE-08-2020-0170

Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*, *32*(8), 827–877. https://doi.org/10.1080/09588221.2018.1535508

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bodnar, S., Cucchiarini, C., Strik, H., & van Hout, R. (2016). Evaluating the motivational impact of CALL systems: Current practices and future directions. *Computer Assisted Language Learning*, *29*(1), 186–212. https://doi.org/10.1080/09588221.2014.927365

Bräuer, P., & Mazarakis, A. (2022). How to design audio-gamification for language learning with Amazon Alexa?—A long-term field experiment. *International Journal of Human–Computer Interaction*, *40*(9), 2343–2360. https://doi.org/10.1080/10447318.2022.2160228.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: how humans interpret topic models, In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C. and Culotta, A. (Eds), Advances in Neural Information Processing Systems, Vol. 22, Curran Associates.

Chaves, A. P., & Gerosa, M. A. (2021). How should my Chatbot interact? A survey on human-Chatbot interaction design. *International Journal of Human–Computer Interaction*, *37*(8), 729–758. https://doi.org/10.1080/10447318.2020.1841438

Chen, W., Liu, D., & Lin, C. (2023). Collaborative peer feedback in L2 writing: Affective, behavioral, cognitive, and social engagement. *Frontiers in Psychology*, *14*, 1078141. https://doi.org/10.3389/fpsyg.2023.1078141

Chien, Y.-C., Wu, T.-T., Lai, C.-H., & Huang, Y.-M. (2022). Investigation of the influence of Artificial Intelligence markup language-based LINE ChatBot in contextual English learning. *Frontiers in Psychology*, *13*, 785752. https://doi.org/10.3389/fpsyg.2022.785752

Dao, P., Nguyen, M. X. N. C., Duong, P. T., & Tran–Thanh, V. U. (2021). Learners' engagement in L2 computer-mediated interaction: Chat mode, interlocutor familiarity, and text quality. *The Modern Language Journal*, *105*(4), 767–791. https://doi.org/10.1111/modl.12737

Dizon, G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning: An investigation of Alexa. *JALT CALL Journal*, *16*(2), 107–120. https://doi.org/10.29140/jaltcall.v16n2.273

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum.
Dörnyei, Z. (2009). The L2 motivational self system. *Motivation, Language Identity and the L2 Self*, *36*(3), 9–11.
Dörnyei, Z. (2019). Towards a better understanding of the L2 learning experience, the Cinderella of the L2 motivational self system. *Studies in Second Language Learning and Teaching*, *9*(1), 19–30. https://doi.org/10.14746/ssllt.2019.9.1.2
Dörnyei, Z., & Chan, L. (2013). Motivation and vision: An analysis of future L2 self images, sensory styles, and imagery capacity across two target languages. *Language Learning*, *63*(3), 437–462. https://doi.org/10.1111/lang.12005
Dörnyei, Z., & Ushioda, E. (2013). *Teaching and researching: Motivation*. Routledge.
Du, Y., & Gao, H. (2022). Determinants affecting teachers' adoption of AI-based applications in EFL context: An analysis of analytic hierarchy process. *Education and Information Technologies*, *27*(7), 9357–9384. https://doi.org/10.1007/s10639-022-11001-y
Ebadi, S., & Amini, A. (2022). Examining the roles of social presence and human-likeness on Iranian EFL learners' motivation using artificial intelligence technology: A case of CSIEC chatbot. *Interactive Learning Environments*, *32*(2), 655–673. https://doi.org/10.1080/10494820.2022.2096638
Engwall, O., Lopes, J., & Åhlund, A. (2021). Robot interaction styles for conversation practice in second language learning. *International Journal of Social Robotics*, *13*(2), 251–276. https://doi.org/10.1007/s12369-020-00635-y
Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, *121*, 103254. https://doi.org/10.1016/j.system.2024.103254
Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and human task partners. *Computers in Human Behavior*, *75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045
Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, *93*, 279–289. https://doi.org/10.1016/j.chb.2018.12.023
Gan, Z. (2012). Understanding L2 speaking problems: Implications for ESL curriculum development in a teacher training institution in Hong Kong. *Australian Journal of Teacher Education*, *37*(1), 43–59 (Online). https://doi.org/10.1016/j.compedu.2009.01.008
Gao, Y., Wang, Q., & Wang, X. (2024). Exploring EFL university teachers' beliefs in integrating ChatGPT and other large language models in language education: A study in China. *Asia Pacific Journal of Education*, *44*(1), 29–44. https://doi.org/10.1080/02188791.2024.2305173
Ghafouri, M. (2024). ChatGPT: The catalyst for teacher-student rapport and grit development in L2 class. *System*, *120*, 103209. https://doi.org/10.1016/j.system.2023.103209
Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(Suppl. 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101
Henry, A., & Cliffordson, C. (2017). The impact of out-of-school factors on motivation to learn English: Self-discrepancies, beliefs, and experiences of self-authenticity. *Applied Linguistics*, *38*(5), 713–736. https://doi.org/10.1093/applin/amv060
Huang, J., & Mizumoto, A. (2024). The effects of generative AI usage in EFL classrooms on the L2 motivational self system. *Education and Information Technologies, 30*, 6435–6454. https://doi.org/10.1007/s10639-024-13071-6
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
Hwang, W.-Y., Guo, B.-C., Hoang, A., Chang, C.-C., & Wu, N.-T. (2022). Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Computer Assisted Language Learning*, *37*(7), 1632–1658. https://doi.org/10.1080/09588221.2022.2095406

Illeris, K. (2016). *How we learn: Learning and non-learning in school and beyond*. Routledge.
Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, *55*(1), 48–63. https://doi.org/10.1080/15391523.2022.2142873
Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274
Kim, A., & Su, Y. (2024). How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. *System*, *122*, 103256. https://doi.org/10.1016/j.system.2024.103256
Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3
Lang, Y., Xie, K., Gong, S., Wang, Y., & Cao, Y. (2022). The impact of emotional feedback and elaborated feedback of a pedagogical agent on multimedia learning. *Frontiers in Psychology*, *13*, 810194. https://doi.org/10.3389/fpsyg.2022.810194
Lee, K.-A., & Lim, S.-B. (2023). Designing a leveled conversational teachable agent for English language learners. *Applied Sciences*, *13*(11), 6541. https://doi.org/10.3390/app13116541
Li, B., Bonk, C. J., Wang, C., & Kou, X. (2024). Reconceptualizing self-directed learning in the era of Generative AI: An exploratory analysis of language learning. *IEEE Transactions on Learning Technologies*, *17*, 1515–1529. https://doi.org/10.1109/TLT.2024.3386098
Lin, C. J., & Mubarok, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society*, *24*(4), 16–35.
Loewen, S. (2020). *Introduction to instructed second language acquisition*. Routledge.
Mercer, S. (2019). Language learner engagement: Setting the scene. In: Gao, X. (eds), *Second handbook of English language teaching* (pp. 643–660). Springer, Cham: Springer International Handbooks of Education.
Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, *33*(8), 865–890. https://doi.org/10.1080/09588221.2019.1595664
Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning*, *61*(2), 325–366. https://doi.org/10.1111/j.1467-9922.2011.00640.x
Rahman, A., & Tomy, P. (2023). Intelligent personal assistant—An interlocutor to mollify Foreign language speaking anxiety. *Interactive Learning Environments*, *32*(8), 4569–4586. https://doi.org/10.1080/10494820.2023.2204324
Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, *33*(4), 829–856. https://doi.org/10.1017/S0142716411000580
Şendağ, S., & Odabaşı, H. F. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & Education*, *53*(1), 132–141. https://doi.org/10.1016/j.compedu.2009.01.008
Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. *Motivation, Language Identity and the L2 Self*, *36*, 66–97. https://doi.org/10.21832/9781847691293
Tai, T.-Y. (2022). Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*, *2022*, 1–30. https://doi.org/10.1080/09588221.2022.2075013
Tai, T.-Y. (2024). Comparing the effects of intelligent personal assistant-human and human-human interactions on EFL learners' willingness to communicate beyond the classroom. *Computers & Education*, *210*, 104965. https://doi.org/10.1016/j.compedu.2023.104965

Tai, T.-Y., & Chen, H. H.-J. (2022). The impact of intelligent personal assistants on adolescent EFL learners' speaking proficiency. *Computer Assisted Language Learning*, 37(5–6), 1224–1251. https://doi.org/10.1080/09588221.2022.2070219

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

Timpe-Laughlin, V., Dombi, J., Sydorenko, T., & Sasayama, S. (2023). L2 learners' pragmatic output in a face-to-face vs. a computer-guided role-play task: Implications for TBLT. *Language Teaching Research*. https://doi.org/10.1177/13621688231188310

Timpe-Laughlin, V., Sydorenko, T., & Dombi, J. (2024). Human versus machine: Investigating L2 learner output in face-to-face versus fully automated role-plays. *Computer Assisted Language Learning*, *37*(1–2), 149–178. https://doi.org/10.1080/09588221.2022.2032184

Tong, P., Yin, Z., & Tsung, L. (2024). Student engagement and authentic language use on WeChat for learning Chinese as a foreign language. *Computer Assisted Language Learning*, *37*(4), 687–719. https://doi.org/10.1080/09588221.2022.2052906

Tyler, A. (2008). Cognitive linguistics and second language instruction. In *Handbook of cognitive linguistics and second language acquisition* (pp. 466–498). Routledge.

Um, H., Kim, H., Choi, D., & Oh, H. (2023). An AI-based English education platform during the COVID-19 pandemic. *Universal Access in the Information Society, 23*, 1233–1248. https://doi.org/10.1007/s10209-023-01046-2

Ushida, E. (2005). The role of students' attitudes and motivation in second language learning in online language courses. *CALICO Journal*, *23*(1), 49–78. https://doi.org/10.1558/cj.v23i1.49-78

Wallach, H. M., Murray, I., Salakhutdinov, R., Mimno, D. (2009, June). Evaluation methods for topic models. In *Proceedings of the 26th annual international conference on machine learning* (pp. 1105—1112). https://doi.org/10.1145/1553374.1553515

Wan, Y., & Moorhouse, B. L. (2024). Using Call Annie as a generative artificial intelligence speaking partner for language learners. *RELC Journal*, *56*(2), 489–498. https://doi.org/10.1177/00336882231224813

Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*, *37*(4), 814–840. https://doi.org/10.1080/09588221.2022.2056203

Williams, M., & Burden, R. (1997). *Psychology for language teachers* (Vol. 88). Cambridge University Press.

Wu, J., Li, Y., Zhou, J., & Chen, S. (2024). The impact of intelligent personal assistants on Mandarin second language learners: Interaction process, acquisition of listening and speaking ability. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2024.2317849

Ye, Y., Deng, J., Liang, Q., & Liu, X. (2022). Using a smartphone-based chatbot in EFL learners' oral tasks. *International Journal of Mobile and Blended Learning (IJMBL)*, *14*(1), 1–17. https://doi.org/10.4018/IJMBL.299405

Zhang, R., Zou, D., & Cheng, G. (2023). Chatbot-based learning of logical fallacies in EFL writing: Perceived effectiveness in improving target knowledge and learner motivation. *Interactive Learning Environments*, *32*(9), 5552–5569. https://doi.org/10.1080/10494820.2023.2220374

Zhang, Y. H., & Li, L. (2019). Research on the Influence of "FIF Oral Training System" on Non-English Majors' Oral English Learning in the perspective of mobile-assisted language learning. *Modern Linguistics*, *7*(5), 781–791. https://doi.org/10.12677/ml.2019.75104

Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. *Heliyon*, *10*(3), e25370. https://doi.org/10.1016/j.heliyon.2024.e25370

Zhao, W., Wang, B., Coniam, D., & Xie, B. (2017). Calibrating the CEFR against the China Standards of English for College English vocabulary education in China. *Language Testing in Asia*, *7*, 1–18. https://doi.org/10.1186/s40468-017-0036-1

Appendix A. The L2 attitude/motivation scale designed by Taguchi et al. (2009)

Variables	Description	Questions
Intended effort	Learners' intended efforts toward learning a second language	1. If an English course was offered at university or somewhere else in the future, I would like to take it. 2. If an English course was offered in the future, I would like to take it. 3. I am working hard at learning English. 4. I am prepared to expend a lot of effort in learning English. 5. I think that I am doing my best to learn English. 6. I would like to spend lots of time studying English. 7. I would like to concentrate on studying English more than any other topic. 8. Compared to my classmates, I think I study English relatively hard. 9. If my teacher would give the class an optional assignment, I would certainly volunteer to do it. 10. I would like to study English even if I were not required.
Ideal L2 Self	The ideal state that L2 learners expect to achieve	11. I can imagine myself living abroad and having a discussion in English. 12. I can imagine myself living abroad and using English effectively for communicating with the locals. 13. I can imagine a situation where I am speaking English with foreigners. 14. I can imagine myself speaking English with international friends or colleagues. 15. I imagine myself as someone who is able to speak English. 16. I can imagine myself speaking English as if I were a native speaker of English. 17. Whenever I think of my future career, I imagine myself using English. 18. The things I want to do in the future require me to use English. 19. I can imagine myself studying in a university where all my courses are taught in English. 20. I can imagine myself writing English e-mails fluently.
Ought-to L2 Self	The attributes that an individual believes one ought to possess	21. I study English because close friends of mine think it is important. 22. I have to study English, because, if I do not study it, I think my parents will be disappointed with me. 23. Learning English is necessary because people surrounding me expect me to do so. 24. My parents believe that I must study English to be an educated person. 25. I consider learning English important because the people I respect think that I should do it. 26. Studying English is important to me in order to gain the approval of my peers/teachers/family/boss. 27. It will have a negative impact on my life if I don't learn English. 28. Studying English is important to me because an educated person is supposed to be able to speak English. 29. Studying English is important to me because other people will respect me more if I have a knowledge of English. 30. If I fail to learn English, I'll be letting other people down.
L2 learning experience	Situation-specific motives connected to the immediate learning environment and experience	31. I like the atmosphere of my English classes. 32. Do you like the atmosphere of your English classes? 33. I find learning English really interesting. 34. Do you find learning English really interesting? 35. I always look forward to English classes. 36. Do you always look forward to English classes? 37. I really enjoy learning English. 38. Do you really enjoy learning English? 39. Would you like to have more English lessons at school? 40. Do you think time passes faster while studying English?