BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg

METACOGNITIVE PLANNING STRATEGIES IN AI-DRIVEN SPEAKING PRACTICE: ENHANCING EFL LEARNERS' FLUENCY AND SPEECH PREP

Ibodullayeva Diyora Orifjon qizi
Samarkand State Institute of Foreign Languages
diyoraibodullayeva50@gmail.com

Abstract. This study investigates the impact of pre-task metacognitive planning strategies on English as a Foreign Language (EFL) learners' oral fluency during interactive encounters with conversational Voice Artificial Intelligence (AI). While voice-based AI applications offer accessible, low-anxiety speaking practice, learners frequently experience cognitive overload, resulting in prolonged hesitations and fragmented speech when interacting without strategic preparation. Utilizing a quasi-experimental pre-test/post-test control group design, intermediate-level university students were divided into an experimental group, which employed a structured Metacognitive Planning Sheet prior to AI voice sessions, and a control group, which engaged in unguided AI conversations over a multi-week intervention period. Temporal fluency parameters – specifically speech rate and silent pause frequency – were measured alongside qualitative assessments of perceived cognitive load and speaking confidence. The findings revealed that learners utilizing pre-task metacognitive planning achieved a statistically significant increase in speech rate and a substantial reduction in unfilled silent pauses, markedly outperforming the control group. Qualitative data indicated that pre-task goal-setting and conversational anticipation reduced real-time cognitive processing strain, allowing learners to reallocate mental resources toward speech execution and discourse continuity. The paper concludes with pedagogical implications for integrating metacognitively scaffolded frameworks into AI-enhanced language learning environments.

Keywords. Metacognitive planning, EFL oral fluency, Voice AI interaction, speech rate, silent pause reduction, cognitive load theory, Task-Based language teaching, learner autonomy, speech preparation strategies, Artificial Intelligence in Pedagogy

Introduction. In recent years, the integration of Artificial Intelligence (AI) into foreign language pedagogy has transformed the landscape of English as a Foreign Language (EFL) instruction. Among various AI innovations, conversational Voice AI tools – such as advanced Large Language Model (LLM) speech interfaces, speech-recognition applications, and interactive virtual agents – have emerged as viable solutions to one of language teaching's most persistent challenges: providing personalized, scalable, and low-anxiety speaking practice. Oral fluency, characterized by the smooth, continuous, and efficient production of spoken speech with minimal disruptive hesitation, remains a critical benchmark of communicative competence. However, achieving fluency in a traditional EFL classroom is often constrained by limited individual speaking time, large class sizes, and heightened affective barriers like performance anxiety. Voice AI tools address these limitations by offering on-demand, non-judgmental, and interactive speaking environments where learners can engage in real-time oral discourse.

Despite the proliferation of Voice AI technologies, access to intelligent conversational agents alone does not automatically guarantee improvements in oral fluency. A common issue in AI-mediated speaking tasks is that learners frequently experience cognitive overload when tasked with

generating dynamic, real-time responses. When interacting with voice assistants without structured guidance, EFL learners often struggle with lexical retrieval, syntactic structuring, and idea generation simultaneously. Consequently, their oral output becomes characterized by prolonged silent pauses, frequent filled pauses (e.g., um, uh), self-corrections, and fragmented utterances. Rather than fostering fluid communication, unguided AI interaction can inadvertently reinforce passive, reactive speech patterns. This discrepancy underscores the necessity of combining technological tools with sound pedagogical frameworks that scaffold the cognitive processes underlying speech production.

To mitigate cognitive load and enhance fluency during interactive speaking, metacognitive strategy instruction offers a robust theoretical framework. Grounded in Flavell's (1979) theory of metacognition and Levelt's (1989) Speech Production Model, metacognitive strategies involve higher-order cognitive executive control – specifically planning, monitoring, and evaluating one's own learning processes. In the context of oral fluency, pre-task metacognitive planning serves as an essential cognitive scaffold. By engaging in explicit goal-setting, predicting potential conversational trajectories, organizing conceptual frameworks, and pre-selecting relevant lexical units prior to interacting with the AI, learners reduce the real-time processing demands imposed during live conversation. As a result, cognitive resources can be reallocated toward speech pacing, temporal execution, and discourse continuity rather than dynamic vocabulary retrieval.

While existing literature has extensively examined the general benefits of AI in language learning, empirical research specifically investigating how pre-task metacognitive planning influences oral fluency parameters during Voice AI interactions remains sparse. To address this gap, this study aims to evaluate the effectiveness of integrating a structured Metacognitive Planning Framework into AI-driven oral practice for university-level EFL learners. Specifically, the study addresses the following research questions:

RQ1: To what extent does pre-task metacognitive planning impact EFL learners' temporal speaking fluency (measured by speech rate/words per minute and silent pause frequency) during interactive AI voice tasks?

RQ2: How does structured metacognitive preparation affect learners' perceived cognitive load and speaking confidence when interacting with Voice AI in real time?

Methodology. This study adopted a quasi-experimental pre-test/post-test control group design to investigate the impact of pre-task metacognitive planning on EFL learners' oral fluency during interactive AI voice sessions. Over a six-week intervention period, both groups engaged in regular conversational tasks with an AI voice agent, but only the experimental group received structured instruction and scaffolding in metacognitive preparation prior to each speaking session. The participant pool comprised 60 undergraduate students enrolled in an English for Academic Purposes (EAP) course at a regional university. All participants were native speakers of Uzbek or Russian with an intermediate level of English proficiency, verified as CEFR B1 through an institutional placement test administered at the start of the semester. Participants were randomly assigned by intact class groups into either the experimental group (n = 30) or the control group (n = 30). To maintain consistency, students who reported prior daily usage of voice-based AI applications for language learning were excluded from the final analysis, ensuring that baseline familiarity with the technology was roughly equal across both cohorts.

Two main instruments were deployed in this study: the Voice AI interface used for speaking practice and the Metacognitive Planning Sheet (MPS) designed for pre-task preparation.

Voice AI Platform: Both groups interacted with an advanced LLM-powered conversational voice interface set to a friendly, conversational tone with standard American English pronunciation. The

prompts were calibrated to generate open-ended questions across six real-world academic and social topics (e.g., environmental sustainability, digital media habits, dynamic problem-solving in everyday scenarios).

Metacognitive Planning Sheet (MPS): Developed specifically for the experimental group, the MPS guided learners through three distinct pre-task mental operations before triggering the AI voice agent:

Goal Formulation: Identifying target vocabulary, transitions, and communication goals.

Conversational Anticipation: Predicting prospective follow-up questions from the AI and structuring appropriate conversational turns.

Self-Monitoring Strategies: Setting personal guidelines for managing hesitation without breaking the flow of speech (e.g., using filler phrases intentionally rather than pausing silently).

Oral fluency was assessed at two intervals: immediately before the six-week intervention (pre-test) and right after its completion (post-test). Both tests required participants to complete a five-minute conversational task with the AI agent on an unfamiliar topic.

Audio recordings from both tests were transcribed and analyzed using speech analysis software (Praat) to extract two key objective temporal parameters of oral fluency:

Speech Rate: Calculated as the total number of articulated words divided by the total time taken (expressed as Words Per Minute, or WPM).

Silent Pause Frequency: Measured as the number of unfilled silent pauses exceeding 1.5 seconds per minute of speech.

In addition to quantitative audio metrics, a 5-point Likert scale post-survey accompanied by short semi-structured interviews was administered to gather qualitative insights into the experimental group's perceived cognitive load and confidence levels during speech production.

Results. To evaluate the impact of pre-task metacognitive planning on students' oral speech production, descriptive and inferential statistical analyses were performed on two primary temporal metrics: speech rate (measured in Words Per Minute, WPM) and silent pause frequency (unfilled pauses exceeding 1.5 seconds per minute of speech).

At baseline, an independent samples t-test indicated no statistically significant difference between the experimental group ($M = 86.4$, $SD = 10.2$) and the control group ($M = 87.1$, $SD = 9.8$) regarding speech rate ($t(58) = 0.27$, $p = 0.788$). This established that both cohorts possessed equivalent baseline oral fluency prior to the intervention. Similarly, baseline pause frequency was roughly identical across both groups, with the experimental cohort averaging 6.8 silent pauses per minute ($SD = 1.4$) and the control group averaging 6.6 pauses per minute ($SD = 1.5$).

Following the six-week intervention period, the experimental group demonstrated a substantial increase in overall speech rate, gaining an average of 26.4 WPM to reach a post-test mean of 112.8 WPM ($SD = 11.5$). A paired samples t-test confirmed that this intra-group improvement was statistically significant with a large effect size ($t(29) = 8.92$, $p < 0.001$, Cohen's\ $d = 1.18$). In contrast, while the control group also showed a slight increase in speech rate—rising from 87.1 WPM to 94.2 WPM ($SD = 10.6$)—their gain of 7.1 WPM yielded a much smaller effect size ($t(29) = 2.68$, $p = 0.012$, Cohen's\ $d = 0.35$). An independent samples t-test comparing post-intervention gain scores between the two cohorts confirmed that the experimental group's speech rate improvement significantly exceeded that of the control group ($t(58) = 6.45$, $p < 0.001$).

A similar trajectory was observed regarding silent pause frequency. Post-intervention data revealed a dramatic reduction in prolonged silent hesitations for the experimental group, dropping by nearly 57% from 6.8 pauses per minute down to 2.9 pauses per minute ($SD = 0.9$). This decline was statistically significant ($t(29) = -12.41$, $p < 0.001$, Cohen's\ $d = 1.62$). Meanwhile, the control

group exhibited a more modest reduction, moving from 6.6 to 5.2 silent pauses per minute (SD = 1.2). These comparative metrics indicate that structured pre-task planning enabled students to maintain a noticeably smoother and more continuous speech flow during dynamic turn-taking with the AI voice assistant.

The quantitative findings were complemented by qualitative data gathered through post-intervention Likert-scale survey items and semi-structured interview sessions with participants in the experimental group.

Analysis of the survey data demonstrated a strong consensus among learners regarding the cognitive benefits of the Metacognitive Planning Sheet (MPS). A overwhelming majority of participants (86.7%) agreed or strongly agreed that completing the pre-task preparation helped them anticipate prospective conversational turns, which prevented them from "freezing" when receiving unexpected responses from the AI interface. Furthermore, 80.0% of respondents noted that pre-selecting key lexical structures and transition phrases significantly reduced the time spent actively searching for vocabulary mid-sentence, thereby lowering their real-time cognitive burden.

In qualitative interviews, students consistently highlighted how structured planning mitigated the affective pressure associated with instantaneous AI interactions. One participant noted that before using the planning sheet, they experienced continuous hesitation because they attempted to formulate ideas, select vocabulary, and manage grammar simultaneously.

Discussion. The empirical results of this study demonstrate that integrating structured pre-task metacognitive planning into Voice AI practice significantly enhances EFL learners' oral fluency. While conversational AI applications provide an accessible, low-anxiety medium for target language exposure, mere interaction without strategic scaffolding yields only incremental gains. The marked improvements observed in the experimental group – specifically a gains differential of +19.3 Words Per Minute and a 57% reduction in unfilled silent pauses compared to the control group – highlight the transformative role of explicit cognitive preparation in speech production.

These findings can be theoretically anchored in Levelt's (1989) Speech Production Model and Sweller's (1988) Cognitive Load Theory. Levelt posits that oral speech involves three interconnected stages: Conceptualization (determining what to say), Formulation (selecting words and applying grammatical rules), and Articulation (the physical execution of speech). In real-time, unguided interactions with AI agents, novice or intermediate learners often experience cognitive bottlenecking when attempting to execute conceptualization and formulation simultaneously. By introducing the Metacognitive Planning Sheet (MPS), learners offload the heavy cognitive burden of conceptualizing ideas and retrieving baseline vocabulary before initiating the call. Consequently, working memory capacity during live interaction is freed up, allowing learners to focus primarily on articulation, speech pacing, and active listening.

The outcomes of this research align closely with previous studies on metacognitive scaffolding in traditional language classrooms (e.g., Skehan, 1998; Foster & Skehan, 1996), which emphasized that pre-task planning consistently leads to higher fluency and structural complexity. However, this study extends these classic findings to the modern context of AI-mediated language learning.

Earlier research on voice-based AI tools often treated technological interactivity as a standalone solution for speaking hesitation. This study demonstrates that technological tools alone are insufficient without pedagogical frameworks. By combining metacognitive self-prompting with dynamic AI feedback, learners transition from reactive conversationalists to autonomous speakers capable of maintaining discourse continuity under real-time conditions.

Several limitations should be considered when interpreting these findings:

Sample Scope: The study was conducted with a cohort of 60 undergraduate EFL learners at the intermediate (CEFR B1) level over a six-week duration. The long-term retention of these fluency gains across different proficiency levels (e.g., A2 or B2+) requires further longitudinal investigation.

Focus on Temporal Fluency: Fluency was evaluated primarily through temporal parameters (speech rate and silent pauses). While qualitative feedback was gathered, future studies could expanded to evaluate how pre-task planning concurrently influences grammatical accuracy and lexical diversity.

Conclusion. This study investigated the synergistic potential of pairing pre-task metacognitive planning with interactive Voice AI practice to improve EFL learners' oral fluency. The quantitative findings confirmed that learners who utilized structured planning sheets prior to AI voice sessions achieved significantly higher speech rates and exhibited fewer disruptive pauses than those engaging in unguided interactions. Qualitative data further revealed that metacognitive scaffolding reduced cognitive overload, lowered performance anxiety, and enhanced overall communicative confidence. Ultimately, the study underscores that AI conversational agents reach their maximum pedagogical potential when embedded within structured, metacognitively guided learning cycles.

Based on these findings, several recommendations are proposed for language educators and curriculum developers:

Adopt a Three-Stage AI Pedagogy: Instructors should structure AI-based speaking activities into a distinct cycle: Pre-task Planning (goal-setting, lexical anticipation), In-task Interaction (fluid conversation with AI), and Post-task Reflection (self-evaluating performance against initial goals).

Train Learners in Metacognitive Strategies: Language programs should explicitly teach self-monitoring and goal-setting strategies rather than assuming students naturally possess these executive control skills when using educational technology.

Integrate Structured Scaffolding in AI Tools: EdTech developers should consider building built-in metacognitive prompts (e.g., pre-conversation planning modules) directly within Voice AI applications to optimize learning outcomes.

References:

1. Baddeley, A. (2000). The episodic buffer: A new component of working memory? Trends in Cognitive Sciences, 4(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2
2. Derakhshan, A., & Shirvan, M. E. (2023). Harnessing artificial intelligence in foreign language learning: Opportunities and cognitive challenges. Computers & Education, 198, 104765. https://doi.org/10.1016/j.compedu.2023.104765
3. Ellis, R. (2009). The differential effects of pre-task planning and on-line planning on L2 speaking performance. Applied Linguistics, 30(4), 474–509. https://doi.org/10.1093/applin/amp012
4. Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American Psychologist, 34(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906
5. Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. Studies in Second Language Acquisition, 18(3), 299–323. https://doi.org/10.1017/S0272263100015048
6. Goh, C. C., & Burns, A. (2012). Teaching speaking: A holistic approach. Cambridge University Press.

7. Jeon, J. (2022). Exploring the pedagogical potential of AI voice assistants in developing EFL learners' oral fluency. Language Learning & Technology, 26(2), 45–63.
8. Levelt, W. J. (1989). Speaking: From intention to articulation. MIT Press.
9. Li, J., & Kim, L. (2024). Overcoming speech hesitation through conversational AI: Cognitive load and fluency outcomes in higher education. System, 121, 103210. https://doi.org/10.1016/j.system.2024.103210
10. Oxbrow, G. (2020). Metacognitive awareness and strategy instruction in oral production. International Journal of Applied Linguistics, 30(1), 112–129. https://doi.org/10.1111/ijal.12264
11. Skehan, P. (1998). A cognitive approach to language learning. Oxford University Press.
12. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
13. Vandergrift, L., & Goh, C. C. (2012). Teaching and learning second language listening: Metacognitive awareness and strategy development. Routledge.
14. Wang, Y., & Choi, S. (2023). Generative AI and language education: A systematic review of voice-based interactive tools for oral communication. Educational Technology Research and Development, 71(4), 1823–1845. https://doi.org/10.1007/s11423-023-10255-w
15. Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2