

Enhancing Speaking Skills with AI: ChatGPT-Based Dynamic Assessment for Intermediate EFL Learners

Abstract

This study examines the integration of ChatGPT within a Dynamic Assessment (DA) framework informed by Sociocultural Theory (SCT) and its effects on the speaking performance of adult intermediate EFL learners. A quasi-experimental pre-test/post-test design involved 30 participants randomly assigned to an experimental group ($n = 15$) receiving ChatGPT-mediated DA or a control group ($n = 15$) receiving traditional instruction. Speaking performance was assessed with an analytic rubric covering accuracy, fluency, lexical resource, and pronunciation. Quantitative data were analyzed via mixed between-within subjects ANOVA; qualitative data came from open-ended questionnaires completed by 18 learners and three instructors. Results showed a significant Group $\times$ Time interaction favoring the experimental group. Learners and instructors reported gains in motivation, engagement, and awareness of specific linguistic features, alongside challenges of information overload and limited contextual nuance. The study concludes that ChatGPT-mediated DA can enhance speaking skills when paired with teacher mediation and offers implications and limitations for practice and research.

Keywords: Dynamic Assessment, Sociocultural Theory, ChatGPT, Speaking Performance, Accuracy, Fluency, Technology-Enhanced Language Learning

Introduction

Proficient oral communication in English is essential for adult learners seeking global opportunities. AI-driven tools such as ChatGPT provide immediate, adaptive feedback that supports self-paced practice (Annamalai et al., 2023; Bozkurt et al., 2023). Dynamic Assessment (DA), rooted in Vygotsky's Zone of Proximal Development (ZPD), integrates assessment and instruction through mediated interaction (Lantolf & Poehner, 2004; Kozulin et al., 2003). Combined with Sociocultural Theory (SCT), DA emphasizes social mediation and collaborative knowledge construction (Vygotsky, 1978).

Prior research has shown positive effects of both interactionist and interventionist DA on speaking development (Estaji & Farahanynia, 2019; Ahmadi Safa & Beheshti, 2018) and the potential of AI tools to deliver scalable scaffolding (Chen et al., 2022; Qiao & Zhao, 2023). Yet empirical studies that deliberately embed ChatGPT within a DA framework informed by SCT for adult intermediate EFL speaking remain limited. This study addresses that gap.

Research Questions

The study addresses three research questions:

1. How does the integration of ChatGPT technology within a DA framework enriched by SCT affect the overall speaking performance (accuracy, fluency, lexical resource, and pronunciation) of adult intermediate EFL learners?
2. How do adult intermediate EFL learners perceive the integration of ChatGPT technology within the DA framework in relation to their speaking development?
3. How do instructors perceive the effectiveness of ChatGPT technology within the DA framework in facilitating personalized learning and addressing diverse learner needs?

Literature Review

DA moves beyond static testing by revealing and promoting latent potential through mediation (Barzegar & Azarizad, 2014; Alemi, 2015). Both interactionist and interventionist approaches have demonstrated benefits for oral performance (Estaji & Farahanynia, 2019). Barjesteh (2013) characterized DA as a paradigm shift toward a process-oriented philosophy that unites teaching and testing within the ZPD.

AI tools such as ChatGPT can supply graduated, contingent feedback that approximates human mediation (Shan et al., 2025; Dizon & Tang, 2020). Studies report gains in motivation, engagement, and speaking when AI is integrated thoughtfully (Song & Song, 2023; Qiao & Zhao, 2023). Barjesteh (2023) showed that technology-enhanced instruction can improve EFL vocabulary and listening comprehension, while Barjesteh (2019) underscored the continuing centrality of the teacher's mediational role—an insight directly relevant to AI-supported DA. Ethical and practical concerns, including cognitive load and teacher preparedness, remain important (Sindermann et al., 2021; Lee et al., 2023). Overall, the literature supports the theoretical synergy of DA, SCT, and AI, yet calls for empirical studies that report both quantitative outcomes and qualitative insights into how mediation operates on specific speaking sub-skills.

Method

Design and Participants

A quasi-experimental pre-test/post-test control-group design with mixed-methods data collection was employed. Thirty adult intermediate (B1 CEFR) EFL learners (aged 18–25; 11 male, 19 female) at a private language institute in Shiraz, Iran, voluntarily participated. Eligibility was confirmed via an adapted Oxford Quick Placement Test. Participants were randomly assigned to an experimental group ($n = 15$) or a control group ($n = 15$) using a computerized random-number generator. All were native Persian speakers. Baseline speaking scores and prior formal speaking-course experience were comparable across groups.

Instruments

Speaking performance was measured with pre- and post-tests consisting of two structured prompts (descriptive and opinion-based), each lasting 3–5 minutes. Three trained raters, blind to group assignment, scored performances using an analytic rubric adapted from IELTS Speaking Band Descriptors. Four criteria were rated on a 0–10 scale: accuracy (grammatical correctness and control), fluency (smoothness, hesitation, and repair), lexical resource (range and appropriacy), and pronunciation (intelligibility and articulation). The reported score is the mean of the four subscale scores. Inter-rater reliability was excellent ($ICC = 0.81$).

In the experimental group, ChatGPT functioned as a mediational tool within an interventionist DA framework. Learners submitted weekly speaking samples as audio recordings accompanied by transcripts. Structured prompts enacted graduated prompting, scaffolding, and contingent feedback targeting grammatical accuracy, lexical choice, fluency markers visible in transcripts (fillers, self-repairs), and pronunciation guidance via phonetic notes and common

error patterns. The control group completed identical tasks but received only conventional teacher feedback.

Qualitative data were gathered through open-ended written questionnaires completed by 18 learners from the experimental group and three instructors. Questions addressed perceived impact on speaking development (accuracy, fluency, lexical resource, pronunciation), motivation, usefulness of feedback, and limitations.

Procedure and Analysis

After pre-testing, a six-week intervention was conducted. Both groups completed identical weekly speaking tasks; only the experimental group received individualized ChatGPT-mediated feedback. Post-testing used the same prompts and raters. Questionnaires were administered immediately after the post-test.

Quantitative data were analyzed with a mixed between-within subjects ANOVA (Group $\times$ Time) on mean speaking scores. Assumptions of normality and homogeneity were checked; effect sizes (partial η^2 and Cohen's d) were reported. Qualitative data were analyzed via thematic analysis following Braun and Clarke's (2006) six-phase framework. Two independent coders performed inductive coding; inter-coder agreement exceeded 85% after reconciliation. Themes were linked to the four speaking criteria and to the quantitative results.

Results

Quantitative Findings

Descriptive statistics are presented in Table 1. The control group showed a modest mean increase from 6.07 (SD = 1.03) to 6.40 (SD = 0.99). The experimental group improved from 6.33 (SD = 1.23) to 7.33 (SD = 1.05).

Table 1. Descriptive Statistics for Speaking Scores

Group	Time	Mean	SD
Control (n = 15)	Pre	6.07	1.03
	Post	6.40	0.99
Experimental (n = 15)	Pre	6.33	1.23
	Post	7.33	1.05

A mixed ANOVA revealed a significant Group $\times$ Time interaction, $F(1, 28) = 5.12$, $p = .031$, partial $\eta^2 = .155$. Simple-effects analysis indicated significant improvement for the experimental group ($p = .034$, $d \approx 0.61$) but not for the control group ($p = .371$). The between-groups difference at post-test was also significant ($p = .018$). These results indicate that ChatGPT-mediated DA produced statistically and practically meaningful gains in overall speaking performance relative to traditional instruction.

Qualitative Findings

Thematic analysis of the 18 student and three teacher questionnaires yielded four themes. Coding was inductive; representative quotations explicitly address accuracy, fluency, lexical resource, and pronunciation.

Theme 1: Enhanced Engagement and Motivation. Learners and teachers described increased willingness to practice. “ChatGPT made it feel like I was talking to a real partner. I practiced more because I got feedback right away” (S4). “Students stayed more engaged; they expected immediate comments and kept trying to improve fluency” (T1).

Theme 2: Perceived Quality and Specificity of Feedback. Participants linked feedback to sub-skill gains. “I finally understood the grammar mistakes I always repeat—accuracy improved because the feedback was step-by-step” (S10). “It pointed out fillers and long pauses; after a few weeks I spoke more smoothly” (S7). “Suggestions for alternative vocabulary made my speech richer” (S13). “The phonetic notes helped me notice and correct pronunciation problems” (S2). Teachers noted differentiation: “It gives just enough support to lower-performing students without doing the work for them” (T2).

Theme 3: Cognitive Overload and Limitations. Some learners found the volume of feedback overwhelming: “Sometimes there was too much info. I didn’t know what to focus on first” (S9). Teachers observed occasional misinterpretation of intended meaning (T3). Pronunciation feedback, while useful for common patterns, was constrained by the text-based nature of the interaction after transcription.

Theme 4: Pedagogical Integration. Instructors emphasized the need for teacher preparation: “Planning speaking tasks that work well with ChatGPT takes time and training; the AI is a powerful mediator but does not replace the teacher’s role in interpreting nuanced performance” (T1).

These qualitative insights corroborate the quantitative gains: learners explicitly attributed improvements in accuracy, fluency, lexical resource, and pronunciation to the contingent, graduated feedback provided within the DA framework.

Discussion

The significant Group $\times$ Time interaction supports the effectiveness of embedding ChatGPT within a DA framework informed by SCT. The experimental group’s gains align with Vygotsky’s ZPD: mediation supplied by structured AI prompts enabled performance beyond independent capability (Vygotsky, 1978; Lantolf & Poehner, 2004). The findings extend earlier DA speaking studies (Estaji & Farahanynia, 2019) by showing that AI can scale interventionist mediation while preserving graduated prompting and contingency.

Qualitative data illuminate how the intervention operated. Learners reported heightened awareness of recurring accuracy errors, reduced hesitation, expanded lexical choices, and improved pronunciation—directly addressing the four rubric criteria. These perceptions resonate with prior AI-in-language-learning research (Crompton & Burke, 2023; Song & Song, 2023) and with Barjesteh’s (2023) demonstration that technology can enhance core language skills when thoughtfully applied. At the same time, challenges of cognitive overload and incomplete contextual understanding reaffirm that AI mediation is most effective when complemented by human teacher mediation (Barjesteh, 2019). The study therefore contributes to the growing consensus that blended DA–AI models can promote both linguistic and socio-cognitive development when carefully designed.

Implications

- **Pedagogical:** Teachers can employ structured ChatGPT prompts as an out-of-class mediational tool, freeing classroom time for higher-order interaction while remaining responsible for interpreting nuanced performance and managing cognitive load.
- **Curriculum design:** Speaking tasks can be sequenced so that AI feedback targets specific criteria (accuracy → fluency → lexical resource/pronunciation) in successive cycles.
- **Teacher education:** Professional development should include prompt engineering for DA principles and strategies for integrating AI feedback with human scaffolding.
- **Research:** Future studies should examine subscale trajectories separately, employ larger samples, and investigate long-term retention and transfer.

Limitations

The modest sample size ($n = 15$ per group) limits generalizability. The study was conducted in a single private institute with adult intermediate learners; results may differ for other proficiency levels or contexts. Although audio submissions were accompanied by transcripts, ChatGPT interaction remained primarily text-based; fully multimodal speech analysis was beyond the scope of the tool configuration used. Qualitative data relied on written questionnaires rather than in-depth interviews. Finally, the six-week intervention, while sufficient to detect short-term gains, does not address longer-term durability of improvement.

Conclusion

Integrating ChatGPT into a theoretically grounded Dynamic Assessment framework can produce statistically significant and pedagogically meaningful improvements in adult intermediate EFL learners' speaking performance across accuracy, fluency, lexical resource, and pronunciation. Learner and instructor perceptions confirm increased motivation, engagement, and awareness of specific linguistic features, while also highlighting the continuing necessity of human mediation. When implemented with attention to cognitive load, prompt design, and teacher professional development, ChatGPT-mediated DA offers a scalable, personalized pathway for enhancing oral proficiency.

References

- Ahmadi Safa, M., & Beheshti, S. (2018). Interactionist and interventionist group dynamic assessment (GDA) and EFL learners' listening comprehension development. *Iranian Journal of Language Teaching Research*, 6(3), 37–56.
- Ahmadi, M. R. (2018). The use of technology in English language learning: A literature review. *International Journal of Research in English Education*, 3(2), 115–125.
- Alemi, M. (2015). The impact of dynamic assessment on Iranian EFL students' writing self-assessment. *TELL*, 9(1), 145–169.
- Annamalai, N., Eltahir, M. E., Zyoud, S. H., Soundrarajan, D., Zakarneh, B., & Al Salhi, N. R. (2023). Exploring English language learning via chatbot: A case study from a self-determination theory perspective. *Computers and Education: Artificial Intelligence*, 5, 100148.

- Barjesteh, H. (2013). A paradigm shift toward a new philosophy of assessment: Dynamic assessment from a critical perspective. *Indian Journal of Fundamental and Applied Life Sciences*, 3(3), 526–535.
- Barjesteh, H. (2019). Dynamicity of transformative L2 materials preparation model in EFL classroom: Place of critical language pedagogy in teacher education. *Journal of Teaching Language Skills*, 38(2), 47–79.
- Barjesteh, H., & Isace, H. (2023). Technology is an asset: Enhancing EFL learners' vocabulary knowledge and listening comprehension through CALL. *International Journal of Research in English Education*, 9(1), 50–69.
- Barzegar, N., & Azarizad, A. (2014). The role of assessment in language teaching. *Procedia – Social and Behavioral Sciences*, 98, 916–920.
- Bozkurt, A., et al. (2022). Chatbot-human interaction and its effects on EFL pupils' L2 speaking performance and anxiety. *Novitas-ROYAL*, 16(2), 113–131.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Chen, C.-H., Koong, C.-S., & Liao, C. (2022). Influences of integrating dynamic assessment into a speech recognition learning design to support pupils' English speaking skills, learning anxiety, and cognitive load. *Educational Technology & Society*, 25(1), 1–14.
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22.
- Dizon, G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning: An investigation of Alexa. *JALT CALL Journal*, 16(2), 107–120.
- Estaji, M., & Farahanynia, M. (2019). The immediate and delayed effect of dynamic assessment approaches on EFL learners' oral narrative performance and anxiety. *Educational Assessment*, 24(2), 135–154.
- Kazu, I. Y., & Kuvvetli, M. (2023). The influence of pronunciation education via artificial intelligence technology on vocabulary acquisition in learning English. *International Journal of Psychology and Educational Studies*, 10(2), 480–493.
- Kozulin, A., Gindis, B., Ageyev, V. S., & Miller, S. M. (Eds.). (2003). *Vygotsky's educational theory in cultural context*. Cambridge University Press.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics*, 1(3), 49–74.
- Lee, D., Kim, H.-H., & Sung, S.-H. (2023). Development research on an AI English learning support system to facilitate learner-generated context-based learning. *Educational Technology Research and Development*, 71, 629–666.
- Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, 14, 1255594.
- Shan, Z., Song, Z., Jiang, X., Chen, W., & Chen, L. (2025). Complementing but not replacing: Comparing the impacts of GPT-4 and native-speaker interaction on Chinese L2 writing outcomes. *Behavioral Sciences*, 15(4), 540.
- Sindermann, C., et al. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI – Künstliche Intelligenz*, 35(1), 109–118.
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.