AMERICAN JOURNALS
Open Access, Peer Reviewed Journals

American Journal of Pedagogical and Educational Research
ISSN (E): 2832-9791| Volume 43, |December - 2025

INTEGRATING MULTIMODAL AI INTO TBLT: A FRAMEWORK FOR ENHANCING STRATEGIC AND INTERACTIONAL COMPETENCE IN HIGHER EDUCATION

Suvankulova Aziza Nuriddin qizi
PhD Student, Gulistan State University
aziza.suvankulova1991@gmail.com

A B S T R A C T

This article presents a new theoretical framework that demonstrates how multimodal Artificial Intelligence (AI), specifically GPT-4o and the Whisper speech-recognition model, can assist in the formation of strategic and interactional communicative competence among B1-B2 non-philological students in higher education. While previous research has focused mostly on AI's function in enhancing language correctness, vocabulary retention, and individualized grammar assistance, the discourse-level components of communicative competence remain insufficiently addressed. These include learners' abilities to negotiate meaning, regulate turn-taking, correct misunderstandings, and maintain coherent interaction in spontaneous communication. Building on new developments in multimodal AI, the paper proposes that tools capable of processing audible, visual, and textual input can better simulate real communicative situations than traditional teaching methods. GPT-4o can simulate adaptive, dynamic conversational contexts, whereas Whisper generates accurate transcriptions of learners' speech, allowing for extensive review of disfluencies, hesitation markers, repair patterns, and interactional timing. When combined with a Task-Based Language Teaching (TBLT) cycle, these technologies provide more possibilities for pre-task preparation, performance-based scaffolding, and post-task reflection. The paper introduces the "AI-TBLT Strategic and Interactional Competence Framework", which describes how learners can enhance discourse-level skills through task design, multimodal input, data-driven feedback, and teacher mediation. The model prioritizes pedagogical and ethical factors, ensuring that AI functions as a supplement rather than a substitute for human interaction. In all, the study emphasizes the potential of incorporating multimodal AI into communicative pedagogy to support non-philological students in developing the strategic and interactional flexibility necessary for academic and professional communication in multilingual, technologically advanced environments.

K E Y W O R D S

Artificial Intelligence, communicative competence, strategic competence, interactional competence, Task-based language teaching, multimodal AI, non-philological students, AI-pedagogy.

INTRODUCTION

Artificial intelligence (AI) is rapidly developing into an influential factor in modern higher education, shaping not just how students access and consume information, but also how they interact and collaborate in academic and professional settings. Although AI-powered solutions have shown great promise in improving linguistic accuracy, grammatical correction, and personalized vocabulary growth (Zhang & Zou, 2022; Son et al., 2023), their use in cultivating discourse-oriented communication skills is limited. Many university students, particularly non-language majors, lack strategic competence (the capacity to overcome communication challenges) and interactional competence (the ability to regulate conversational flow). Students in STEM-oriented faculties frequently achieve intermediate (B1-B2) linguistic proficiency, but face difficulties in spontaneous communication tasks like clarifying uncertainty, responding to unanticipated inquiries, negotiating meaning, and controlling turn-taking (Yilmaz, 2022). Traditional, heavily text-based classroom instruction does not offer the fluid, unpredictable conversational settings in which such skills naturally develop (Walsh, 2020). Therefore, these learners often find it difficult to integrate classroom knowledge into genuine academic or professional interactions. Recent advances in multimodal AI have created novel opportunities for bridging this gap. Tools such as GPT-4o can understand multimodal input (text, audio, and visual cues), simulate authentic conversational partners, and provide instant, context-sensitive feedback on learners' communicative behaviors (Kim & Kim, 2024). Meanwhile, the Whisper speech-recognition system can adequately transcribe spontaneous, accented second language speech, revealing detailed information on pauses, hesitation indicators, repairs, and turn-taking patterns (Li, 2023; Rasheed, 2023). Together, these methods enable the analysis of communication not just at the linguistic level, but also at the interaction and strategic levels. Notwithstanding these possibilities, current AI-in-ELT research has a major emphasis on linguistic competency rather than discourse-level performance (Li & Ni, 2021). Similarly, while Task-Based Language Teaching (TBLT) is generally acknowledged for fostering meaning negotiation and interactional competency (Long, 2015), its systematic integration with multimodal AI remains underexplored (Ho & Wong, 2024). This article addresses these gaps by proposing the AI-TBLT Strategic and Interactional Competence Framework, which explains the way GPT-4o and Whisper can be used within a TBLT cycle to support non-philological B1-B2 students in developing the communicative flexibility needed in multilingual academic and professional contexts.

Literature Review

Communicative Competence

Initial concepts of communicative competence (Canale & Swain, 1980) focused on grammatical, sociolinguistic, discourse, and strategic components. Recent updates (Celce-Murcia, 2007) have expanded the role of discourse and interactional skills, emphasizing how learners take turns, signal intentions, negotiate comprehension, and keep conversations coherent. Interactional competence, for example, entails a variety of actions such as initiating engagement, reacting correctly, controlling overlaps, and effectively using discourse markers (Hall, 2019; Ryu, 2021). According to research, these skills rarely emerge automatically. Many higher education contexts place a limited communicative load on learners, limiting opportunities for spontaneous, real-time interaction (Walsh, 2020). As a result, despite having appropriate linguistic skills, many students suffer in unforeseen communicative settings.

Strategic and Interactional Competence

Studies on STEM students indicate that, while students may have an adequate vocabulary for academic work, they frequently lack the strategic flexibility required to maintain discussion when misconceptions form (Yilmaz, 2022; Tang & Taguchi, 2022). Strategic competence, including paraphrase, explanation requests, and self-repair, is rarely explicitly taught (Sucaromana, 2021), limiting learners' ability to handle communication breakdowns.

AI in Language learning

AI's impact on language acquisition has been extensively investigated, primarily in terms of accuracy, individualized learning pathways, and adaptive feedback (Zhang & Zou, 2022; Li & Ni, 2021). While AI chatbots have been utilized for conversational practice, the majority of these tools focus on text-based interactions, which limits their ability to capture and assess discourse-level activities (Son et al., 2023).

Multimodal systems in communicative teaching

Recent developments in multimodal systems have increased the options for oral communication support. GPT-4o responds to spoken input, interprets multimodal cues, and provides pragmatic and discourse-level feedback (Kim & Kim, 2024; Søndergaard, 2023). Whisper provides highly accurate transcriptions of L2 speech, exposing interactional and strategic patterns that are normally overlooked (Li, 2023; Rasheed, 2023). However, incorporating such tools into communicative teaching is still in its early phases.

TBLT and AI integration

TBLT prioritizes real-world communication and provides an ideal setting for incorporating AI-enhanced feedback mechanisms (Long, 2015; Samuda and Bygate, 2008). Tasks inherently facilitate meaning negotiation, strategic decision-making, and interactional management. AI can improve TBLT by generating realistic scenarios, simulating conversations, and providing detailed post-task analysis (Ho & Wong, 2024), though systematic frameworks are still insufficient.

Methodology

This article uses a conceptual-analytical approach, which is appropriate for new fields of inquiry where technology advances faster than empirical evidence. Following Ellis' (2005), Bygate and Samuda's (2008) appeals for theoretical clarity in task-based classroom instruction, this paradigm incorporates three categories of analysis:

- Theoretical foundations include communicative competence models (Celce-Murcia, 2007; Hall, 2019) and TBLT research (Long, 2015).
- Technological possibilities, assessing how GPT-4o and Whisper improve discourse-level performance (Kim & Kim, 2024; Li, 2023).
- Pedagogical alignment involves mapping AI affordances to TBLT phases (Ho & Wong, 2024; Bygate & Samuda, 2008).

This approach enables the creation of a pedagogically sound, ethically responsible framework that can be implemented or experimentally studied in higher education settings.

Proposed Model

Strategic and Interactional Competence Framework

Pillar 1: AI-Enhanced Task Design. With the help of GPT-4o, teachers create realistic, domain-specific communication scenarios that include negotiation, clarification, and collaborative problem-solving. Dynamic scenario variation offers controlled improvements in cognitive and interactional complexity.

Pillar 2: Multimodal Scaffolding during Tasks. GPT-4o's multimodal processing capabilities enable real-time conversational modifications, such as suggesting clarification, providing discourse markers, or modeling acceptable responses, thereby maximizing interactional development.

Pillar 3: Whisper-Based Interactional Analysis. Whisper transcriptions identify specific speech patterns such as pauses, repairs, hesitation signs, overlaps, and timing irregularities. These transcripts serve as the foundation for student reflection and teacher-guided feedback.

Pillar 4: AI-Supported Reflective Feedback. GPT-4o generates individualized suggestions based on transcripts, assisting learners in understanding how strategic and interactional decisions influence communication results. This increases metacognitive awareness.

Pillar 5: Ethical Accountability and Teacher Mediation. Teachers help students make strategic decisions, contextualize insights produced by AI, and make sure AI is utilized responsibly and transparently.

Discussion

The suggested framework demonstrates how multimodal AI might address persistent issues in establishing discourse-level communication skills. While Whisper offers a level of clarity that was previously unattainable in regular classrooms, GPT-4o's ability to simulate interaction allows students to experience real communicative strain (Li, 2023; Rasheed, 2023). Integrating these techniques inside TBLT widens the learning cycle by enabling broader pre-task exposure, greater flexibility in task-phase scaffolding, and deeper post-task reflection (Ho & Wong, 2024). Thus, AI enhances analytical depth and individualization while maintaining pedagogical monitoring, complementing rather than replacing the work of the instructor. However, there are still difficulties. It is necessary to address ethical concerns, the dangers of relying too much on AI, potential errors in automated feedback, and disparities in digital readiness (Hockly, 2023). In order to preserve harmony and guarantee critical engagement with AI-generated feedback, teacher mediation is crucial.

Pedagogical implications

The proposed model has various pedagogical implications for higher education contexts where non-philological B1-B2 students need communicative skills to participate effectively in academic group projects, interdisciplinary collaboration, and workplace scenarios. First, multimodal AI tools offer a scalable approach to enhancing real communication exposure. GPT-4o allows students to practice discourse methods in flexible, semi-structured environments, which is especially effective in settings when teacher time is limited or student-teacher ratios are high. This ensures equal access to communicative practice, regardless of classroom size.

Second, whisper-assisted transcription and analysis provide teachers with a data-driven prism through which to monitor their students' communication behavior, something that is not always possible in real time. This enables teachers to provide more specific feedback, highlight repeating patterns

(prolonged hesitation, ineffective repair moves), and make interactional competence a stated learning goal rather than an unintended consequence. This is consistent with findings demonstrating the value of systematic noticing and reflection in SLA.

Third, the framework promotes the transition from teacher-centered instruction to a more learner-autonomous paradigm. GPT-4o's adaptive prompts motivate students to take the initiative during communicative practice, while AI-generated reflective feedback redirects learners' attention from perfection to communicative effectiveness. This promotes self-regulation, which is an important ability for higher education students navigating multilingual academic environments.

At last, the methodology promotes a rethinking of curriculum design by including AI-supported communication tasks into the existing syllabus. Rather than creating new "AI lessons", educators can incorporate GPT-4o-mediated role plays, problem-solving interactions, and reflective post-task analysis into current course modules. This ensures that meaningful AI deployment is both sustainable and pedagogically linked with the communicative goals of higher education programs.

Limitations of the research

Although the suggested structure provides intriguing teaching prospects, numerous constraints must be recognized. First, this research is conceptual rather than empirical. While the model is founded on well-established theories of communicative competence, interactional competence, and TBLT, it has yet to be tested in classroom experiments. Future research should look into how the model performs in real-world instructional contexts, as well as how learner characteristics like competence level, motivation, and technology readiness influence its success.

Second, the model assumes consistent access to multimodal AI tools, which may not be available in all higher education contexts. Implementation may be limited by hardware limits, unreliable internet access, or institutional AI technology prohibitions. Furthermore, while Whisper's speech-recognition accuracy for varied L2 accents is high, there may be occasional errors that affect feedback accuracy.

Third, the rising use of AI in educational settings creates significant ethical concerns. The issues of privacy, data storage, openness, and learner reliance must be addressed. Without appropriate teacher mediation, learners may rely too heavily on AI-generated cues or accept proposed responses uncritically. This underscores the importance of clear norms and ethical frameworks in ensuring appropriate AI-assisted education.

Conclusion

This study presented a comprehensive model that demonstrates how multimodal AI tools may be systematically integrated into TBLT to build strategic and interactional competence among B1-B2 non-philological students. By integrating GPT-4o's dynamic conversational capabilities with Whisper's precise speech analytics, the suggested framework broadens the pedagogical potential of task-based classroom instruction and indicates the way forward for AI-enhanced communicative language teaching. The model not only addresses the gap in existing research on AI in ELT, but it also serves as a framework for future empirical study. With proper instructor mediation and ethical safety measures, multimodal AI can dramatically improve learners' communicative flexibility in both academic and professional settings.

References

1. Bygate, M., & Samuda, V. (2008). Tasks in Second Language Learning. Palgrave Macmillan.
2. Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics, 1(1), 1–47.
3. Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In E. Alcón Soler & M. P. Safont Jordà (Eds.), Intercultural Language Use and Language Learning (pp. 41–57). Springer.
4. Ellis, R. (2005). Principles of instructed language learning. System, 33(2), 209–224.
5. Hall, J. K. (2019). Interactional competence and development. Language Learning, 69(S1), 4–40.
6. Hockly, N. (2023). AI for language teachers: Opportunities and limitations. ELT Journal, 77(4), 421–431.
7. Ho, C., & Wong, M. (2024). AI-supported TBLT tasks for university students. Innovation in Language Learning and Teaching, 18(2), 134–150.
8. Kim, Y., & Kim, J. (2024). Multimodal feedback and conversational AI in speaking development. ReCALL, 36(1), 45–63.
9. Li, M., & Ni, X. (2021). Intelligent tutoring systems in EFL learning. Educational Technology & Society, 24(3), 1–15.
10. Li, Z. (2023). Using speech recognition systems to analyze L2 performance. System, 112, 103012.
11. Long, M. H. (2015). Second language acquisition and task-based language teaching. Wiley Blackwell.
12. Rasheed, A. (2023). Speech-to-text technologies in higher education. Computers & Education, 197, 104676.
13. Ryu, J. (2021). Discourse markers and L2 interactional competence. System, 102, 102639.
14. Son, J. B., Ružić, N. K., & Philpott, A. (2023). Artificial intelligence technologies for language teaching. Journal of China CALL, 1–18.
15. Søndergaard, H. (2023). LLMs as conversational partners in academic English. Journal of Academic Language and Learning, 17(1), 22–40.
16. Sucaromana, U. (2021). Strategic competence in L2 communication. RELC Journal, 52(2), 267–286.
17. Tang, X., & Taguchi, N. (2022). Interactional competence development in EFL learners. Applied Linguistics Review, 13(1), 79–101.
18. Walsh, S. (2020). Classroom discourse and interaction. Edinburgh University Press.
19. Yilmaz, Y. (2022). Learners' interactional competence during task-based collaboration. System, 110, 102912.
20. Zhang, H., & Zou, B. (2022). AI chatbots for speaking practice: Opportunities & challenges. Computer Assisted Language Learning, 35(1–2), 323–347.