

<https://doi.org/10.5281/zenodo.18449940>

Multimodal Linguistic Technologies for Solving Spoken Interaction Problems in EFL Classrooms

Baneen Asghar

Foreign Language Teaching Assistant, Michigan State University,
asgharb1@msu.edu

Maria Habib

(Corresponding Author)

Riphah International University, Pakistan, mariahabib1997@gmail.com

Abstract: *Spoken interaction is a core component of communicative competence in English as a Foreign Language (EFL) learning; however, many learners experience persistent difficulties in oral communication due to limited exposure, anxiety, pronunciation problems, and insufficient interactional feedback. Recent developments in multimodal linguistic technologies—integrating speech recognition, visual analytics, gesture modeling, and automated feedback—offer innovative solutions to these challenges. This study investigates the effectiveness of multimodal linguistic technologies in solving spoken interaction problems in EFL classrooms through an experimental research design. A mixed-methods approach was employed, involving pre- and post-speaking assessments, classroom observations, system-generated interaction data, and learner interviews. Forty EFL learners were divided into experimental and control groups, with the experimental group receiving multimodal technology-assisted instruction over an eight-week period. The findings reveal that learners exposed to multimodal technologies demonstrated significant improvement in pronunciation accuracy, fluency, turn-taking skills, and pragmatic appropriateness compared to those receiving traditional instruction. Qualitative data further indicate increased learner confidence, reduced speaking anxiety, and higher engagement levels. The study concludes that multimodal linguistic technologies provide effective pedagogical support for enhancing spoken interaction by offering immediate, multimodal feedback and authentic communicative practice. These results highlight the potential of frontier linguistic technologies to address real-world linguistic problems in EFL classrooms and contribute to the advancement of technology-enhanced language education.*

Keywords: Multimodal technologies; Spoken interaction; EFL classrooms; Speech recognition; Applied linguistics

1. Introduction

Spoken interaction plays a central role in second and foreign language learning, as it enables learners to negotiate meaning, express ideas, and participate actively in communicative contexts (Lantolf 2000). In EFL classrooms, however, spoken interaction often remains underdeveloped due to a range of linguistic, psychological, and pedagogical constraints. Learners frequently encounter problems related to pronunciation accuracy, fluency, interactional competence, turn-taking, and pragmatic appropriateness. These issues are particularly pronounced in contexts where English is not widely used outside the classroom, limiting opportunities for authentic communication. Traditional EFL instruction has primarily focused on grammatical accuracy and written skills, often neglecting the dynamic and interactive nature of spoken communication (Abdushukurova 2024). As a result, learners may possess adequate grammatical knowledge but struggle to engage in real-time spoken interaction. Furthermore, factors such as speaking anxiety, fear of making mistakes, and lack of immediate corrective feedback further hinder oral participation. Teachers, constrained by time and large class sizes, may find it difficult to provide individualized feedback on spoken performance (Ehrenberg, Brewer et al. 2001).

In recent years, advancements in frontier linguistic technologies have transformed the landscape of

language education. Multimodal linguistic technologies, which combine auditory, visual, and interactive modalities, have emerged as powerful tools for addressing spoken interaction problems. These technologies include speech recognition systems, pronunciation visualization tools, virtual avatars, gesture and facial expression modeling, and AI-driven feedback mechanisms (Tajik 2025). By engaging multiple sensory channels, multimodal technologies support more effective language processing and learning. The theoretical foundation of multimodal learning suggests that meaning is constructed through the integration of multiple semiotic resources. In spoken interaction, communication extends beyond verbal language to include gestures, facial expressions, gaze, and prosody. Multimodal technologies replicate these features by providing learners with rich, interactive environments that simulate authentic communicative situations (Xie, Yang et al. 2025). Such environments allow learners to practice speaking in low-anxiety settings while receiving immediate and personalized feedback. Despite the growing interest in technology-enhanced language learning, empirical research examining the specific role of multimodal linguistic technologies in solving spoken interaction problems remains limited. Many studies focus on isolated aspects such as pronunciation or fluency, without considering the broader interactional and pragmatic dimensions of speaking. Therefore, there is a need for comprehensive experimental research that investigates how multimodal technologies can holistically improve spoken interaction in EFL classrooms (Yi, Dong et al. 2024). This study aims to address this gap by examining the effectiveness of multimodal linguistic technologies in solving spoken interaction problems among EFL learners. By adopting a mixed-methods experimental design, the study provides both quantitative and qualitative insights into learners' spoken performance and perceptions. The findings are expected to contribute to applied linguistics research and offer practical implications for EFL pedagogy.



Fig 1. Multimodal technology in EFL classrooms

2. Literature Review

2.1 Spoken Interaction in EFL Learning

Spoken interaction is widely recognized as a fundamental component of communicative competence (Sun and research 2014). It involves not only linguistic accuracy but also fluency, interactional management, and pragmatic appropriateness. In EFL contexts, learners often face difficulties in

producing spontaneous speech due to limited exposure to authentic spoken input and insufficient opportunities for interaction. Research indicates that learners may struggle with pronunciation, hesitation, limited vocabulary, and inappropriate turn-taking, which negatively affect communicative effectiveness (Osuchukwu, Ogayi et al. 2019).

2.2 Challenges in Traditional Speaking Instruction

Traditional approaches to speaking instruction often rely on teacher-centered activities, such as drills and controlled dialogues, which may not adequately prepare learners for real-life communication (Sattarova 2024). These methods provide limited opportunities for meaningful interaction and individualized feedback. As a result, learners may lack confidence and remain passive during speaking activities. The absence of immediate and detailed feedback further restricts learners' ability to self-correct and improve their spoken performance (García and Martínez 2018).

2.3 Multimodality and Language Learning

The concept of multimodality emphasizes that communication and learning occur through the interaction of multiple modes, including speech, text, images, gestures, and sounds. Multimodal learning theory suggests that integrating multiple modalities enhances comprehension and retention by engaging different cognitive processes (Gilakjani, Ismail et al. 2011). In language learning, multimodal input supports learners in understanding complex linguistic features and improves their ability to produce meaningful spoken output (Cheetham 2019).

2.4 Multimodal Linguistic Technologies

Multimodal linguistic technologies incorporate various technological tools to support spoken interaction. Speech recognition systems provide real-time feedback on pronunciation and fluency, while visual tools display stress patterns and intonation contours. Virtual avatars and interactive platforms simulate conversational scenarios, allowing learners to practice turn-taking and pragmatic strategies (Morton and Jack 2005). Studies have shown that such technologies can improve oral proficiency, increase learner motivation, and reduce speaking anxiety.

2.5 Research Gap

Although existing studies highlight the benefits of technology-enhanced language learning, limited research has focused on the integrated use of multimodal technologies to address spoken interaction problems comprehensively (Xie, Yang et al. 2025). There is a lack of experimental studies that examine how these technologies influence multiple dimensions of spoken interaction, including pronunciation, fluency, and pragmatic competence. This study seeks to fill this gap by providing empirical evidence on the effectiveness of multimodal linguistic technologies in EFL classrooms (Zhang, Akoto et al. 2023).

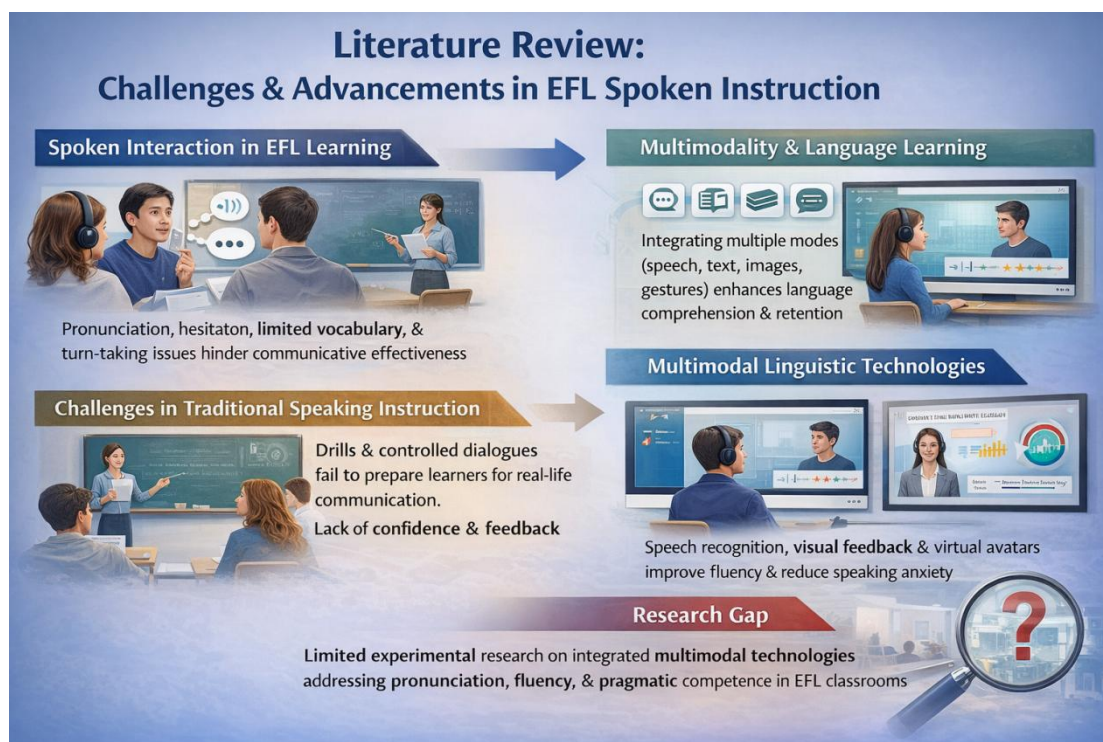


Fig 2: Literature review on EFL spoken instruction

3. Research Method

The methodology for this study is designed to investigate the impact of multimodal linguistic technologies on spoken interaction in English as a Foreign Language (EFL) classrooms (Peng 2019). By adopting a mixed-methods experimental research design, this study combines both quantitative and qualitative approaches to offer a comprehensive analysis of learners' spoken performance, experiences, and perceptions. The integration of multiple methods ensures a holistic understanding of the research problem, capturing both measurable outcomes and contextual insights (Curry, Nembhard et al. 2009).

3.1 Research Design

This study utilized a mixed-methods experimental design to examine how multimodal linguistic technologies impact the development of spoken interaction skills in EFL learners (Fatwassani, Hartono et al. 2025). A mixed-methods approach was deemed appropriate because it allows for a detailed exploration of both quantitative outcomes (such as improvements in pronunciation, fluency, and interactional competence) and qualitative insights (such as learners' perceptions, anxiety levels, and engagement) (Zega and Studies 2025). The experimental design provides a controlled environment where the experimental group engages with multimodal technologies, while the control group participates in traditional classroom activities. This design helps to isolate the effects of multimodal technologies from other variables that might influence learners' performance. The use of both pre- and post-tests in speaking assessments, classroom observations, and semi-structured interviews facilitates the collection of data from multiple angles, contributing to a more robust and reliable conclusion.

3.2 Participants

The participants in this study were 40 undergraduate EFL learners enrolled in an English language program at a university. These learners were selected based on their intermediate-level proficiency in English, which was assessed prior to the study using a standardized language proficiency test. This ensures that all participants had a similar starting point in their language skills.

The 40 learners were randomly assigned to one of two groups:

- **Experimental Group (n = 20):** This group participated in multimodal technology-assisted

instruction. They used speech recognition tools, interactive dialogues, and other multimodal platforms designed to improve spoken interaction.

- **Control Group (n = 20):** This group received traditional, teacher-centered speaking instruction, focusing on grammar, vocabulary, and scripted conversations without the integration of any technological tools.

Random assignment was employed to minimize selection bias and ensure that the groups were comparable at the beginning of the study, both in terms of language proficiency and demographic characteristics. The age range and gender distribution of the participants were not controlled for, but any potential influence of these factors was monitored during data analysis.

3.3 Instruments

A variety of **data collection instruments** were used to gather both quantitative and qualitative data to ensure the reliability and validity of the findings:

Pre- and Post-Speaking Tests: These tests assessed the learners' spoken performance at the beginning and end of the study. The tests included tasks designed to evaluate pronunciation, fluency, accuracy, and turn-taking abilities. The results from these tests provided the primary quantitative data for analyzing the effectiveness of the multimodal technologies.

Speech Recognition-Based Pronunciation Analysis Tools: These tools were used to provide real-time feedback on learners' pronunciation. They measured segmental and suprasegmental features such as stress, intonation, and rhythm. The feedback generated from these tools was analyzed to assess improvement in pronunciation accuracy.

Classroom Observation Checklists: These checklists were used to observe and record behavioral changes in the learners during speaking activities. Teachers and researchers used these checklists to monitor engagement levels, turn-taking behaviors, and interactional competence during the lessons. The checklist data provided qualitative insights into how the multimodal tools impacted learners' participation and interactions.

Semi-Structured Interviews: After the intervention, learners from both groups participated in semi-structured interviews. These interviews allowed for an in-depth exploration of the learners' perceptions, experiences, and emotional responses to the technology-assisted learning environment. Learners were asked about their perceived confidence, anxiety levels, and the usefulness of the multimodal tools in improving their speaking abilities.

System-Generated Interaction Logs: The multimodal platforms used in the experimental group generated logs of learner interactions during speaking activities. These logs included information such as the number of attempts at a task, time spent on each task, and feedback provided. These data points were analyzed to assess the frequency and nature of interaction, providing insights into learners' engagement and progress.

3.4 Procedure

The procedure of this study followed a structured timeline to ensure consistent and systematic implementation. The intervention was carried out over an eight-week period, during which the experimental group engaged in multimodal technology-assisted speaking activities, and the control group participated in traditional instruction.

The experimental group participated in the following activities:

Role-plays: Learners practiced real-world conversational scenarios using multimodal platforms that offered speech recognition, visual feedback, and interactive avatars. These role-plays simulated authentic interactions, allowing learners to practice turn-taking, negotiate meaning, and use pragmatic

strategies in a safe, low-anxiety setting.

Interactive Dialogues: Using the multimodal platforms, learners engaged in real-time, interactive dialogues where they received immediate auditory and visual feedback on their pronunciation, fluency, and conversational strategies.

Simulated Conversations: The platform offered various simulated conversations, providing learners with opportunities to engage in spontaneous speaking tasks, mimicking real-life interactions.

The control group received conventional speaking instruction, which typically involved controlled dialogues, teacher-led discussions, and other traditional methods. The group was provided with regular classroom speaking practice but without the aid of multimodal technologies. Feedback was generally provided by the teacher in the form of verbal correction after the task.

3.5 Data Analysis

The data analysis involved both quantitative and qualitative techniques to assess the impact of multimodal linguistic technologies.

3.5.1 Quantitative Data

Descriptive statistics (mean, standard deviation) were used to summarize the pre- and post-test results for both groups

Inferential statistics (paired t-tests) were employed to determine whether there were significant differences between the two groups in terms of pronunciation accuracy, fluency, interactional competence, and pragmatic appropriateness.

Effect sizes were calculated to measure the magnitude of differences between the experimental and control groups

3.5.2 Qualitative Data

Thematic analysis was used to analyze the semi-structured interview responses and classroom observation checklists. This involved identifying recurring themes and patterns related to learners' experiences, confidence, anxiety levels, and engagement with the multimodal tools. The system-generated interaction logs were analyzed to examine the frequency of interactions, task completion rates, and the nature of learner feedback during the study.

4. Discussion

The findings of the present study demonstrate that multimodal linguistic technologies play a significant role in addressing spoken interaction problems in EFL classrooms. The experimental group's improvement in pronunciation accuracy, fluency, interactional competence, and pragmatic awareness indicates that multimodal technological interventions offer effective solutions to long-standing challenges in EFL speaking instruction. These results provide strong empirical support for the integration of frontier linguistic technologies into language pedagogy. One of the most notable findings is the substantial improvement in learners' pronunciation and fluency. The use of speech recognition systems and visual feedback mechanisms enabled learners to receive immediate and precise feedback on segmental and suprasegmental features, such as stress, intonation, and rhythm. This immediate feedback loop is crucial for spoken language development, as it allows learners to notice discrepancies between their production and target language norms. In line with Schmidt's Noticing Hypothesis, the multimodal feedback heightened learners' awareness of phonological features, thereby facilitating self-correction and sustained improvement. These findings corroborate earlier studies that highlight the effectiveness of technology-mediated pronunciation instruction in EFL contexts. In addition to phonological gains, the study revealed significant improvements in learners' interactional competence. Spoken interaction extends beyond linguistic accuracy to include turn-taking, responsiveness, and the

ability to manage conversational flow. Multimodal platforms simulated authentic communicative scenarios through interactive dialogues, virtual avatars, and real-time prompts, enabling learners to practice conversational strategies in low-anxiety environments. Such simulated interaction aligns with the Interaction Hypothesis, which emphasizes the role of meaningful interaction in second language acquisition. By engaging learners in repeated, scaffolded interaction, multimodal technologies enhanced learners' ability to negotiate meaning and respond appropriately during conversations.

The integration of visual and non-verbal cues also contributed to the development of pragmatic competence. Spoken interaction is inherently multimodal, involving gestures, facial expressions, gaze, and prosody. Traditional EFL classrooms often fail to address these non-verbal dimensions explicitly. However, multimodal technologies provided learners with visual models of appropriate gestures and facial expressions, thereby supporting the acquisition of pragmatic norms. This finding supports multimodal discourse theory, which views communication as a combination of multiple semiotic resources working together to construct meaning. Learners' improved pragmatic performance suggests that multimodal input facilitates a more holistic understanding of spoken interaction. Another important outcome of the study is the reduction in speaking anxiety and the increase in learner confidence. Many EFL learners experience apprehension when speaking due to fear of negative evaluation and limited opportunities for practice. The technology-assisted environment created a supportive space where learners could practice speaking privately or semi-privately, receive non-judgmental feedback, and repeat tasks at their own pace. This aligns with affective filter theory, which posits that reduced anxiety enhances language acquisition. The qualitative findings indicate that learners perceived multimodal technologies as motivating and less intimidating than traditional teacher-centered speaking activities. Furthermore, the findings highlight the pedagogical value of personalized and adaptive feedback. Unlike traditional classroom instruction, where feedback is often delayed or generalized, multimodal systems provided individualized feedback based on learners' performance. This personalized support enabled learners to focus on specific areas of improvement, leading to more efficient learning outcomes. The adaptive nature of these technologies also allowed learners to progress according to their individual needs, thereby addressing learner diversity within EFL classrooms.

The consistency of the present findings with previous research further strengthens their validity. Earlier studies have emphasized the potential of technology-enhanced interaction to improve oral proficiency and learner engagement. However, the current study extends existing literature by demonstrating that an integrated multimodal approach can simultaneously address multiple dimensions of spoken interaction, including phonological, interactional, and affective factors. This holistic perspective contributes to applied linguistics research by offering a comprehensive framework for understanding how technology can solve practical linguistic problems in EFL settings. Despite the positive outcomes, it is important to acknowledge that the effectiveness of multimodal linguistic technologies depends on pedagogical integration. Technology alone does not guarantee improved learning outcomes; rather, it must be aligned with clear instructional objectives and meaningful speaking tasks. Teachers play a crucial role in facilitating technology-enhanced interaction by guiding learners, monitoring progress, and integrating technological tools into communicative activities. Therefore, the findings underscore the need for teacher training and pedagogical awareness when implementing multimodal technologies. In summary, the discussion of the findings suggests that multimodal linguistic technologies offer effective and pedagogically sound solutions to spoken interaction problems in EFL classrooms. By combining auditory, visual, and interactive elements, these technologies support deeper engagement, reduce affective barriers, and enhance communicative competence. The study contributes to both theory and practice by demonstrating how frontier linguistic technologies can bridge the gap between linguistic knowledge and real-world spoken interaction.

4.1 Pedagogical Implications

The findings of this study offer important pedagogical implications for EFL teaching and curriculum

development. First, EFL teachers can effectively integrate multimodal linguistic technologies into speaking instruction to enhance learners' spoken interaction skills. Tools such as speech recognition systems, visual pronunciation feedback, and interactive speaking platforms allow teachers to provide immediate, individualized feedback that is often difficult to achieve in traditional classroom settings. By incorporating these tools into classroom activities, teachers can create more engaging and learner-centered speaking environments that support continuous oral practice. Second, technology-assisted interaction plays a significant role in reducing learner anxiety and increasing classroom participation. Multimodal platforms provide low-pressure, supportive spaces in which learners can practice speaking without fear of negative evaluation. This reduction in affective barriers encourages learners to participate more actively in spoken tasks, thereby improving their confidence and willingness to communicate. As a result, teachers can foster a more inclusive classroom atmosphere where learners at different proficiency levels feel comfortable engaging in oral interaction. Finally, the findings suggest that curriculum designers should systematically incorporate multimodal speaking activities into EFL programs to support effective oral communication development. Integrating technology-enhanced tasks such as simulated conversations, role-plays, and interactive dialogues can help learners develop pronunciation, fluency, and pragmatic competence simultaneously. Aligning multimodal technologies with communicative learning objectives can ensure that EFL curricula address real-world spoken interaction needs and prepare learners for authentic communication contexts.

5. Conclusion

This study demonstrates the significant potential of multimodal linguistic technologies in addressing spoken interaction problems in EFL classrooms. By integrating speech recognition, visual feedback, virtual avatars, and interactive platforms, these technologies effectively enhanced learners' pronunciation, fluency, interactional competence, and pragmatic awareness. The results highlight the advantages of immediate, multimodal feedback in fostering a deeper understanding of spoken language and facilitating more effective self-correction. Learners who engaged with multimodal tools exhibited improved confidence and reduced speaking anxiety, aligning with affective filter theory and offering a supportive environment for practice. Furthermore, the use of visual cues and gesture modeling helped learners develop a more holistic understanding of spoken communication, including non-verbal elements essential for pragmatic competence. The study contributes to applied linguistics by providing empirical evidence that multimodal technologies can solve real-world linguistic challenges, enhancing both language learning outcomes and classroom dynamics. While these technologies proved effective, the research also underscores the importance of pedagogical integration. The role of teachers in guiding the use of these tools and aligning them with communicative objectives is essential for maximizing their potential. Consequently, this research calls for teacher training and curriculum development that incorporates multimodal technologies to create engaging, learner-centered environments that support continuous spoken interaction practice. Overall, this study supports the integration of frontier linguistic technologies into EFL education, offering a valuable framework for future language teaching practices aimed at improving spoken communication skills in authentic contexts.

References

- Abdushukurova, U. (2024). The grammar of fluency: bridging traditional grammar teaching with communicative language practice. INTERDISCIPLINE INNOVATION AND SCIENTIFIC RESEARCH CONFERENCE.
- Cheetham, D. J. A. L. R. (2019). "Multi-modal language input: A learned superadditive effect." **10**(2): 179-200.
- Curry, L. A., et al. (2009). "Qualitative and mixed methods provide unique contributions to outcomes research." **119**(10): 1442-1452.
- Ehrenberg, R. G., et al. (2001). "Class size and student achievement." **2**(1): 1-30.

- Fatwassani, R., et al. (2025). "How Multimodal Communication Enhances EFL Students' Presentation Skills: A Mixed-Methods Study." **10**(2): 200-224.
- García, E. M. and V. G. M. J. M. J. Martínez (2018). "Students' reactions to teacher corrective feedback to oral production: A study on self-correction and autonomy in compulsory EFL university courses." **42**(1): 1-24.
- Gilakjani, A. P., et al. (2011). "The effect of multimodal learning models on language teaching and learning." **1**(10): 1321-1327.
- Lantolf, J. P. J. L. t. (2000). "Second language learning as a mediated process." **33**(2): 79-96.
- Morton, H. and M. A. J. C. A. L. L. Jack (2005). "Scenario-based spoken interaction with virtual agents." **18**(3): 171-191.
- Osuchukwu, C. N., et al. (2019). "Teaching Speaking Skills and the Rule of Turn Taking For Effective Communication: A Pragmatic Approach." **24**(12): 78-83.
- Peng, J.-E. J. S. (2019). "The roles of multimodal pedagogic effects and classroom environment in willingness to communicate in English." **82**: 161-173.
- Sattarova, M. J. M. E. S.-M. J. (2024). "A comparative analysis of traditional and modern approaches in language teaching." **5**(09): 214-223.
- Sun, D. J. J. o. l. t. and research (2014). "From Communicative Competence to Interactional Competence: A New Outlook to the Teaching of Spoken English." **5**(5).
- Tajik, A. (2025). "Integrating AI-Driven Emotional Intelligence in Language Learning Platforms to Improve English Speaking Skills through Real-Time Adaptive Feedback."
- Xie, Y., et al. (2025). "A Review of Multimodal Interaction in Remote Education: Technologies, Applications, and Challenges." **15**(7): 3937.
- Yi, Q., et al. (2024). "Enhancing EFL listening and speaking skills: Strategies and practice for implementing multimedia and multi-modal approaches." **8**(9).
- Zega, Y. S. J. J. o. E., Social and C. Studies (2025). "Effective strategies for enhancing English speaking competence among learners in English education study programs." **2**(2): 88-100.
- Zhang, M., et al. (2023). "Digital multimodal composing in post-secondary L2 settings: A review of the empirical landscape." **36**(4): 694-721.