

Improving English Oral Communication through Multimodal Learning Strategies for Secondary School Students

Aya Ragab,¹ Nur Arifah Drajati², Eka Bhudi Santosa³

¹Faculty of Education, Sebelas Maret University, Surakarta, Indonesia.

²Faculty of English Education, Sebelas Maret University, Surakarta, Indonesia.

³Faculty of Education, Sebelas Maret University, Surakarta, Indonesia.

Abstract

This report presents a conceptual framework for a multimodal learning platform designed to address critical educational needs in the 21st century: cultivating English speaking proficiency and instilling ethical digital citizenship. The proposed framework integrates visual, auditory, reading/writing, and kinesthetic instructional methods to create an engaging and inclusive learning environment. By leveraging the power of multimodal artificial intelligence (AI), the platform provides personalized, real-time feedback that fosters language acquisition and reinforces responsible digital conduct. The framework is specifically adapted to the unique challenges and opportunities of the Indonesian educational context, aligning with national policies like the Kurikulum Merdeka while providing strategies to navigate issues of digital access and teacher professional development. The findings indicate that such an integrated, technology-driven approach can significantly enhance learning outcomes and prepare students to become informed, responsible, and effective global citizens.

ARTICLE HISTORY

Received 25 July 2025

Accepted 25 November

2025

KEYWORDS

Digital Citizenship, English Speaking Skills, Multimodal Learning, Secondary School Students.

INTRODUCTION

The "Cultivating Global Citizen Skills through Integration of Technology in Education" conference theme underscores a fundamental shift in the definition of global competence. In today's interconnected world, being a global citizen is no longer limited to geographical or cultural awareness; it is inextricably linked with digital literacy and ethical conduct (UNESCO, 2023). A global citizen, as defined by leading organizations, is an individual who is "informed and compassionate" and whose "ethical compass drives them to lead lives and make decisions that contribute to a more just, equitable, peaceful and sustainable world" (AFS Framework, 2025).

According to the Organization of Economic Co-operation and Development (OECD, 2022), Indonesian students' performance in literacy and communication remains below the international average, reflecting persistent challenges in preparing youth for global participation. Similarly, a report by the Central Bureau of Statistics (BPS, 2023) noted that although access to technology in schools has increased, the integration of technology into pedagogical practices often remains superficial. Prior studies (Salma et al., 2023; Umroh et al., 2025) have highlighted the role of technology in language learning, yet their focus has largely been limited to improving motivation or general achievement rather than strengthening communication skills for global citizenship.

At the conceptual level, global competence involves not only linguistic proficiency but also intercultural awareness, digital literacy, and ethical online conduct (OECD, 2018). Multimodal learning—defined as the integration of visual, auditory, and interactive resources in instruction—has been recognized as a promising approach to developing these competencies (Aiello &

✉Corresponding author: ayaragab650@gmail.com

Mongibello, 2019). However, while multimodal strategies have been explored in higher education contexts, their application to secondary school students, particularly in Indonesia, remains underexplored.

The context of this research is the teaching of English oral communication for secondary school students in Indonesia. Although English is a compulsory subject, classroom practices often emphasize written grammar exercises and rote learning, leaving limited opportunities for authentic speaking practice. This mismatch between curriculum goals and classroom realities creates a pressing need for innovative approaches that combine language learning with the cultivation of digital responsibility.

Based on the review above, the state of the art shows that existing studies focus either on digital tools or linguistic skills in isolation, without addressing their integration to foster both oral communication and responsible digital citizenship. This reveals a clear research gap: the lack of platforms that use multimodal learning to simultaneously improve speaking skills and cultivate ethical global citizenship among Indonesian secondary students.

Therefore, the objective of this study is to design and conceptualize a multimodal learning platform that enhances English oral communication while instilling digital citizenship values. This manuscript discusses the pedagogical foundations of the platform, its technical design, and its adaptation to the Indonesian educational

This study contributes theoretically by integrating oral communication, multimodal learning, and digital citizenship into a unified conceptual model that reflects 21st-century learning needs. It advances the literature by synthesizing multimodal strategies with global competence, ethics, and responsible digital participation.

Practically, the framework offers actionable guidance for teachers and schools implementing the Kurikulum Merdeka, supporting technology-enhanced speaking practice, ethical online behavior, and improved student engagement in Indonesian secondary schools.

LITERATURE REVIEW

Challenges of English Teaching in Indonesia

Traditional educational models, often characterized by a "one talk" or "cramming" approach, face significant challenges in preparing students for the complexities of the digital age (Wang, 2023). This is particularly evident in English language education in Indonesia, where long-standing issues such as a shortage of qualified English teachers, limited learning resources, and insufficient opportunities for students to practice the language hinder proficiency gains (Renandya et al., 2018; Suharjati, 2010). Students often lack exposure to English outside the classroom, which impedes language acquisition and fluency (Renandya et al., 2018). The traditional, teacher-centered approach fails to engage students actively, limiting the development of their creative and autonomous learning abilities (Wang, 2023; Mutlu & Eroz-Tuga, 2013).

Theoretical Foundations of Multimodal Learning

The academic literature consistently demonstrates the effectiveness of a multimodal approach in fostering English speaking proficiency. By integrating visual, auditory, reading/writing, and kinesthetic instructional methods, multimodal learning creates a richer and more inclusive experience that boosts engagement and improves retention (Litonjua, 2025; Absorb LMS, 2025; Workramp, 2025). This approach transforms the student's role from a passive recipient of information to an active, engaged participant (Mutlu & Eroz-Tuga, 2013; Wang, 2023). Studies have shown a positive consensus from educators towards the use of AI and virtual reality (VR) in multimodal language teaching, noting that these technologies enhance both receptive skills like listening and reading, and productive skills like speaking and writing (Achivida, 2025). Specific pedagogical strategies such as role-play activities, debates, and peer feedback are considered powerful tools within this framework for providing practical speaking opportunities and fostering a

supportive learning environment (ICCOMIPE, 2025.; EFL Cafe, 2024).

Multimodal Learning and Oral Communication

The integration of technology and AI is a recurring theme in recent literature. Research confirms that technology can improve learning efficiency, enhance educational accessibility, and strengthen teachers' teaching abilities (Harun, Utami, & Gunawan, 2019; Salma Rabani, Annisaul Khairat, Xie Guilin, & Deng Jiao, 2023). In particular, AI-based systems can process diverse data inputs—including text, voice, and images—to provide personalized, real-time feedback that fosters language acquisition and fills knowledge gaps (GeniusEE, 2025). The use of Large Multimodal Language Models (LMFMs) has been identified as a breakthrough for educational research, with capabilities in advanced data analysis, literature reviews, and creating multimodal materials (Salma Rabani, Annisaul Khairat, Xie Guilin, & Deng Jiao, 2023). A study by Aiello and Mongibello (2019) found that the use of an AI speech-recognition software significantly improved students' attitudes and abilities toward oral and speaking skills. However, the use of these technologies also introduces ethical concerns, such as data privacy and algorithmic bias, which can unfairly assess students from diverse linguistic backgrounds (Dwivedi et al., 2023; GESR Journal, 2025).

Digital Citizenship and Ethical Online Behavior

In parallel, a growing body of work explores the concept of digital citizenship. Ethical digital citizenship extends beyond basic internet safety to a comprehensive practice of responsibility, safety, ethics, and good communication in online spaces (StopBullying.gov, 2018; PALCS, 2025). Scholars argue that digital multimodal composing, such as creating videos or podcasts, can help students develop a "critical mindset" by prompting them to make informed choices and challenge dominant discourses (Salma Rabani, Annisaul Khairat, Xie Guilin, & Deng Jiao, 2023). This reinforces the idea that developing English proficiency and ethical digital behavior are mutually reinforcing skills. Effective digital citizenship education must be integrated into the core curriculum, often through project-based learning (Edutopia, 2025).

AI, Multimodality, and Global Competence

Despite the clear benefits of technology integration, a significant digital divide persists, particularly in remote regions of Indonesia where connectivity is limited and devices are scarce (World Bank, 2020; Featherstone, 2024; ResearchGate, 2023). Research suggests that simply providing access may not be enough, as digital inclusion can sometimes amplify pre-existing social inequalities (Featherstone, 2024). Furthermore, many educators and students lack the necessary digital literacy and training to effectively utilize these new tools, highlighting the need for a framework that not only provides access to technology but also teaches the principles of safe, ethical, and productive digital engagement (Renandya et al., 2018; Suharjati, 2010).

Digital Citizenship and Ethical Online Behavior

The literature indicates that oral communication and multimodal strategies are mutually reinforcing: multimodality increases student engagement while providing authentic speaking opportunities. At the same time, global competence, digital citizenship, and ethical awareness are essential components of modern language education. Yet these domains remain fragmented in previous studies. This gap highlights the need for an integrated conceptual framework that aligns language learning with multimodal input, ethical digital behavior, and global citizenship competencies.

METHOD

This paper is a conceptual literature-based study rather than a field research project. Therefore, the methodology does not involve new data collection but relies on a systematic review and synthesis of prior works. The research approach can be categorized as qualitative-descriptive with a conceptual development design, since its main goal is to propose a pedagogical framework for

integrating multimodal learning and digital citizenship into English language education.

Scope and Sources

The scope of this study covers global and Indonesian educational contexts, with particular attention to digital literacy, multimodal approaches, and English as a foreign language. Literature was gathered from peer-reviewed journal articles, books, policy documents (e.g., UNESCO, 2023; AFS, 2025), and case studies relevant to the field of education technology and language learning.

Search Process and Article Selection

A systematic literature search was conducted using Google Scholar, ERIC, ResearchGate, and Scopus between 2018–2025. The initial search produced 112 studies, of which 48 met the inclusion criteria focusing on multimodal learning, English oral communication, digital citizenship, and global competence. After removing duplicates and non-relevant sources, 32 studies were included for final conceptual synthesis. Although this review is not a full PRISMA systematic review, it follows PRISMA principles in terms of transparency of search, screening, and selection.

Selection Criteria

The sources included were selected based on three main criteria:

1. Relevance to multimodal learning approaches,
2. Explicit focus on English language learning and communication skills,
3. Studies addressing digital citizenship, ethics, and global competence.

Data Analysis

The data from these studies were analyzed using a content analysis approach. Quantitative findings, such as reported improvements in English proficiency, engagement levels, or digital literacy, were extracted and described to illustrate potential impacts. Qualitative findings, including teacher and student experiences, policy recommendations, and case study insights, were coded and synthesized thematically to identify recurring pedagogical strategies.

Integration into Framework

The analysis followed an integrative synthesis process where insights from both quantitative and qualitative studies were combined to form a coherent conceptual model. This model highlights the pedagogical strategies, digital tools, and ethical considerations necessary for fostering English speaking skills and global citizenship.

Research Limitation

Because this is a literature-based framework, the methodology does not involve field data collection, experiments, or surveys. The findings are therefore conceptual in nature and should be viewed as a basis for future empirical validation in Indonesian schools and higher education institutions.

RESULT AND DISCUSSION

A. The Multimodal Framework: A Catalyst for Speaking Proficiency

Multimodal learning is an instructional approach that integrates visual, auditory, reading/writing, and kinesthetic channels to support diverse learner needs and increase engagement (Litonjua, 2025; Absorb LMS, 2025; Workramp, 2025). By combining formats such as videos, diagrams, podcasts, discussions, and hands-on tasks, this approach moves beyond passive “cramming” methods (Wang, 2023) and instead encourages active participation, creativity, and critical thinking (Absorb LMS, 2025). The technology-driven nature of multimodal platforms shifts students from passive recipients to active learners, strengthening autonomy and overall language proficiency (Mutlu & Eroz-Tuga, 2013) while reducing monotony through continuous interactive experiences. Empirical evidence shows its effectiveness in improving English speaking skills

through role-plays, debates, conversation circles, language games, and peer feedback (Renandya et al., 2018; ICCOMIPE, 2025). Large-scale data—including a 32.5% increase in English proficiency in Paraná, Brazil—further confirms its impact, with 95% active classroom participation (Efekta Education Group, 2025). AI-based tools reinforce these gains by providing personalized pronunciation and intonation feedback (ICCOMIPE, 2025; IMARC Group, 2025), and students find them useful and easy to use (Aiello & Mongibello, 2019). However, AI also introduces concerns about privacy and algorithmic bias (Dwivedi et al., 2023), making transparency essential so learners understand how data is collected and processed. Integrating AI ethics into language learning modules transforms abstract concepts into practical experiences, allowing students to develop both speaking proficiency and responsible digital awareness simultaneously.

B. Cultivating Ethical Digital Citizenship

Ethical digital citizenship goes beyond basic online safety to encompass responsibility, ethics, protection of personal data, and respectful communication, forming a natural bridge to English speaking skills (PALCS, 2025; StopBullying.gov, 2018). The interpersonal skills required for effective oral communication—clarity, active listening, empathy, and constructive dialogue—translate directly into positive online behavior (EFL Cafe, 2024; ICCOMIPE, 2025). A well-designed platform can integrate both skill sets through features like a virtual debate module that evaluates language fluency alongside respectful digital conduct (Edutopia, 2024). However, as technology and AI tools become more embedded in education, concerns such as data privacy, algorithmic bias, and inequitable assessment for minority dialect speakers must be addressed (Dwivedi et al., 2023). Teaching students algorithmic literacy—how to evaluate AI-generated content, question data sources, and understand privacy implications—is essential for preparing them to engage critically in an AI-driven world. Digital citizenship education is most effective when embedded within project-based learning, where multimodal tasks naturally require ethical practices such as proper citation, accessibility features, and peer feedback grounded in empathy and responsibility (Edutopia, 2025). This integrated approach nurtures learners who can communicate ethically, think critically about technology, and participate responsibly in global digital spaces.

C. Conceptual Framework for an Integrated Learning Platform

The proposed platform integrates English speaking instruction with ethical digital citizenship through a dual-purpose, hybrid design. Core features such as a “Virtual Debate Club” allow students to develop speaking skills through interactive role-play while also learning respectful online communication and media literacy through guided moderation. A “Multimodal Storytelling Project” further promotes creative expression by combining text, images, and audio, supported by a rubric that reinforces proper citation and ethical digital practice. Multimodal AI provides real-time feedback on pronunciation, grammar, and ethical use of content, while an accessible, user-centered interface ensures usability for learners with diverse needs. Together, these elements create a cohesive environment where language development and digital ethics are fostered simultaneously.

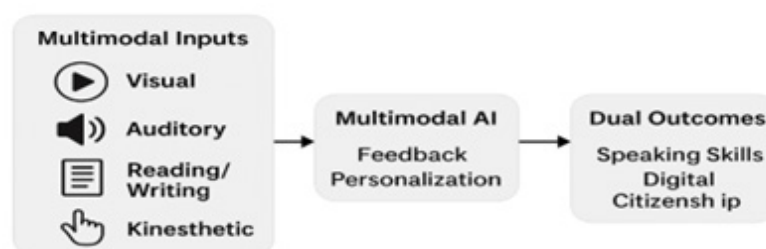


Figure 1. Conceptual Framework Integrating Multimodal Inputs, Multimodal AI, and Dual Learning Outcomes.

The effectiveness of multimodal and technology-integrated learning is well supported by existing evidence. A study from the University of Northern Colorado reported a 20% increase in student retention and comprehension through multimodal presentations, with both students and instructors highlighting improved engagement and enjoyment. In Indonesia, a qualitative study by Umroh, Juhana, and Ruminda similarly showed that digital multimodal projects enhanced students' speaking proficiency and significantly boosted their confidence. Together, these cases reflect a broader educational shift toward dynamic, interactive, and tech-supported learning. The proposed framework builds on these findings, and the accompanying photos of students and teachers using classroom technology help illustrate the type of collaborative, engaging environment it aims to promote.

These images represent the collaborative and technology-rich environment that is crucial for a successful integrated learning platform. They underscore the importance of teachers as facilitators and mentors in a digitally-empowered classroom.

D. Adapting the Framework to the Indonesian Context

Adapting this framework to the Indonesian context requires aligning it with national educational priorities while addressing on-the-ground challenges. The model fits well with the Kurikulum Merdeka, which emphasizes flexibility, student-centered learning, and technology integration, supported by major government investments in connectivity, AI, and digital skills. However, gaps remain between policy ambition and reality, particularly in digital literacy, affordability, and uneven access to infrastructure. To be effective, the platform must be low-cost, mobile-friendly, and functional in low-bandwidth environments, adopting a periphery-centric approach that prioritizes marginalized communities. Because teachers play a decisive role in technology adoption, the framework must include ongoing professional development that supports both pedagogical and technical competence, enabling educators to foster inclusive, multimodal, and ethical digital learning.

CONCLUSION

This report presents a multimodal learning framework aimed at improving English speaking skills and promoting ethical digital citizenship. Multimodal strategies—combining visual, auditory, reading/writing, and kinesthetic elements—have proven effective, with large-scale evidence showing a 32.5% increase in English proficiency and a 95% participation rate (Litonjua, 2025; Absorb LMS, 2025; Efekta Education Group, 2025). The use of multimodal AI strengthens personalized learning while also serving as a tool for teaching digital ethics and critical AI literacy (GeniusEE, 2025; Dwivedi et al., 2023). The framework aligns with Indonesia's Kurikulum Merdeka and digital transformation goals (UNESCO, 2023; Solih & Julianto, 2025), though challenges such as the digital divide, low teacher digital literacy, and limited EdTech investment persist (Renandya et al., 2018; World Bank, 2020; Featherstone, 2024). Addressing these issues requires a localized, periphery-centric approach supported by strong teacher professional development (Featherstone, 2024; ANTARA News, 2025). Policymakers should expand digital infrastructure and enforce data privacy and AI ethics regulations, while practitioners adopt blended learning and ongoing training (World Bank, 2020; Salma Rabani et al., 2023; Mutlu & Eroz-Tuga, 2013; Umroh et al., 2025). Future research should examine long-term impacts on language and ethics and explore scalable strategies for tech-enhanced learning across diverse Indonesian communities (Solih & Julianto, 2025; Featherstone, 2024).

ACKNOWLEDGMENT

This project has been funded with support of the Kemitraan Negara Berkembang (KNB) Scholarship by the Ministry of Higher Education, Sciences, and Technology of Republic of Indonesia on behalf of the Government of the Republic of Indonesia. This publication/communi-

cation reflects the view only of the author, and the Ministry of Higher Education, Sciences, and Technology of Republic of Indonesia cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

- Absorb LMS (2025). What is Multimodal Learning? Examples, Strategies, & Benefits. <https://www.absorblms.com/blog/what-is-multimodal-learning/>
- ANTARA News (2025). Indonesia aims to provide internet access to 300,000 schools by 2025. <https://en.antaranews.com/news/374517/indonesia-aims-to-provide-internet-access-to-300000-schools-by-2025>
- Common Sense Media (2023). *5 Essential Digital Citizenship Skills for Today's Students*. <https://www.edutopia.org/article/teaching-digital-citizenship-skills/>
- Dwivedi et al. (2023). *Ethical AI for Teaching and Learning*. <https://teaching.cornell.edu/generative-artificial-intelligence/ethical-ai-teaching-and-learning>
- EFL Cafe (2024). *Enhancing Speaking Skills in the ESL Classroom*. <https://eflcafe.net/forums/discussion/enhancing-speaking-skills-in-the-esl-classroom/>
- Efeka Education Group (2025). *Digital learning responsible for 32.5% improvement in English language proficiency*. <https://www.prnewswire.com/news-releases/digital-learning-responsible-for-32-5-improvement-in-english-language-proficiency-302524487.html>
- Featherstone, D. (2024). *Addressing the Digital Divide in Indonesian Higher Education: Insights, Implications, and Potential Solutions*.(((https://www.researchgate.net/publication/377971562_Addressing_the_Digital_Divide_in_Indonesian_Higher_Education_Insights_Implications_and_Potential_Solutions)))
- GeniusEE (2025). *Multimodal AI in EdTech*. <https://geniusee.com/single-blog/multimodal-ai-in-edtech>
- ISTE (2025). *ISTE Standards for Educators*. Retrieved from <https://iste.org/standards/educators>
- Litonjua, E. (2025). *What Is Multimodal Learning?* <https://elearningindustry.com/what-is-multimodal-learning>
- Mutlu, A., & Eroz-Tuga, B. (2013). *Developments on Technology Integration in Language Teaching and Learning*. <https://www.gesrjournal.com/article/developments-on-technology-integration-in-language-teaching-and-learning>
- Ramotion (2024). *Multimodal Interfaces: Importance, Effects & Examples*. <https://www.ramotion.com/blog/multimodal-interfaces/>
- Renandya, W., Hidayat, A., Haryanto, A., & Widodo, H.P. (2018). *English Instruction Challenges and Opportunities in Indonesian Primary Schools: A Systematic Review*. https://www.researchgate.net/publication/381198221_English_instruction_challenges_and_opportunities_in_Indonesian_primary_schools_A_systematic_review
- Solih, A., & Julianto, T. (2025). *Proceedings of the 1st International Conference on Media and Communication Sciences*. <https://conferenceproceedings.ump.ac.id/pssh/article/download/1635/1683>
- Umroh, U., Juhana, J., & Ruminda, R. (2025). *The Use of Digital Multimodal Approach in Teaching English Speaking Skill*. <https://journal.ikipsiliwangi.ac.id/index.php/project/article/view/26397>
- UNESCO (2023). *Technology in education: a case study on Indonesia*. <https://unesdoc.unesco.org/ark:/48223/pf0000387827>
- UNESCO (2023). *What is global citizenship education about?*. <https://unesdoc.unesco.org/ark:/48223/pf0000387827>
- University of Northern Colorado (2024). *Multimodal Presentations Case Studies in Classroom*. <https://www.98thpercentile.com/blog/multimodal-presentations-case-studies-in-classroom/>
- World Bank (2020). *Indonesia's Education Technology During COVID-19 and Beyond*. <https://blogs.worldbank.org/en/eastasiapacific/indonesias-education-technology-during-covid-19-and-beyond>
- Workramp (2025). *What is Multimodal Learning? Examples, Strategies, & Benefits*. <https://www.workramp.com/blog/what-is-multimodal-learning/>