

AI-enabled Multimodal Feedback for Chinese as a Second Language Learners

Xinyu GUO & Yun WEN *

National Institute of Education, Nanyang Technological University, Singapore

[*yun.wen@nie.edu.sg](mailto:yun.wen@nie.edu.sg)

Abstract: Recent research has shown that AI-enabled multimodal feedback is effective and meaningful for language learners. However, there is limited understanding of how to support students, particularly young learners, in interacting with such feedback to reflect on and improve their learning. This poster presents the design of ARCHe 2.0, an AI-enhanced Chinese language learning system tailored for lower primary students in Singapore. The system provides AI-generated multimodal feedback, including automatic star ratings, audio-based suggestions on grammar and content, and visual feedback in the form of AI-generated images, to support both vocabulary learning within and beyond the classroom. Preliminary results from the ongoing intervention suggest that AI-enabled multimodal feedback can effectively promote continuous improvement in students' learning outcomes during collaborative tasks. This study contributes to the growing field of AI in education by offering insights into the design of engaging, feedback-rich learning environments for young learners in language learning. Future research will further explore how lower primary students actively interpret, respond to, and act upon multimodal feedback, and identify pedagogical strategies that support students in self-regulating and processing the feedback more effectively.

Keywords: AI, Multimodal Feedback, Second language, Primary education

1. Introduction

Automatic feedback has shown great promise in enabling adaptive learning experiences and supporting diverse students' needs in language education (e.g., Wen et al., 2025). With advances in AI technology, recent studies have demonstrated that students engage with multimodal feedback for longer periods than with traditional text-based feedback, and they tend to use multimodal feedback in more self-paced and meaningful ways, resulting in improved learning outcomes (Campbell & Feldmann, 2017). Such feedback has also been shown to enhance young learners' enjoyment and engagement in the learning process (Aslan et al., 2025). Despite the growing interest in AI-generated feedback, there remains a lack of empirical studies examining how language learners with diverse backgrounds interact with multimodal feedback in authentic classroom settings (Kumar et al., 2023). To address this gap, the present study introduces ARCHe 2.0, a self-designed AI-generated multimodal feedback system for lower primary students' Chinese language as a second language (L2) in Singapore. This study aims to investigate whether multimodal feedback can effectively support students in completing group tasks, in their Chinese language learning.

2. Literature review

An increasing number of studies have explored innovative feedback formats to examine the potential of AI-enhanced multimodal feedback in L2 learning. Unlike direct feedback, AI-enabled multimodal feedback operates as a form of indirect feedback. It signals the presence of errors in students' vocabulary practice and writing, encouraging learners to identify and address the issues themselves, rather than relying solely on corrective input from teachers (Musk, 2016). Research has shown that L2 learners often prefer indirect feedback during the revision process, as it promotes deeper cognitive engagement and metalinguistic awareness (Ferris, 2010). Studies have shown that AI-powered multimodal feedback, particularly in the form of images or animations, can enhance learning by visualizing narrative content and

highlighting areas for improvement, thereby supporting learners in improving their performance (e.g., Liu et al., 2024).

3. Method

3.1 Participants and Intervention

A total of 76 Primary 2 students from four classes across two local primary schools in Singapore participated in the study. The intervention was designed to span three academic terms over a 10-month period and comprised four ARChE 2.0 sessions. At the time of reporting, two sessions had been completed, and the intervention was at its midway point. In each ARChE 2.0 session, students in small group were encouraged to engage in vocabulary and sentence-making exercises collaboratively.

3.2 ARChE 2.0 System

During the ARChE 2.0 class-based collaborative learning activity, students are required to construct sentences to describe sample pictures provided by the system (Figure 1). Upon submission, they first receive multimodal feedback in the form of AI-generated star ratings and audio-based suggestions on grammar and content (Figure 1a). This feedback enables students to revise and improve their sentences by evaluating their output based on the star ratings and applying the language suggestions to refine grammatical structure and content clarity. In the next step, each group need to generate an AI-generated image based on their submitted sentence (Figure 1b). This visual feedback allowing students to compare the generated image with the original sample picture. The discrepancy between the two encourages deeper reflection on semantic accuracy and descriptive precision, which may in turn prompt further group discussion and collaborative revision.



Figure 1. ARChE2.0 learning system with AI-promoted multimodal feedback

4. Preliminary Findings, Discussion, and Future Directions

To investigate the impact of AI-generated multimodal feedback on student learning, all group-created sentence artifacts ($n = 216$) were evaluated using a four-level scoring rubric (5, 10, 15, 20) automatically generated by the ARChE 2.0 system. This rubric assessed sentence quality based on grammatical accuracy, vocabulary usage, and semantic clarity. To ensure the reliability of the automated scores, a randomly selected 25% subset ($n = 54$) of the artifacts was separately rated by a researcher using the same rubric. Inter-rater reliability between the AI-generated and human-assigned scores was examined using the intraclass correlation coefficient (ICC), which indicated strong agreement ($ICC = .910, 95\% CI [.851, .946], p < .001$), supporting the consistency and reliability of the automated scoring.

Preliminary results from the ongoing intervention indicate that AI-generated multimodal feedback positively impacts students' learning progress. As shown in Figure 2, all four

experimental classes improved their group artifact scores from Lesson 1 to Lesson 2, suggesting the potential of such feedback to foster continuous improvement in collaborative tasks. This progress may be attributed to the iterative nature of multimodal feedback, which encourages ongoing revision and peer reflection. Prior research suggests that multimodal interactions enhance cognitive processing and promote deeper comprehension (Zhang et al., 2024). Teacher observations further revealed that students were motivated to refine their sentences in pursuit of generating more accurate or vivid images and expressed enjoyment when engaging with the system. As students became more accustomed to the feedback environment, they appeared better able to interpret and act on feedback to enhance their work. However, to further determine the impact of multimodal feedback on the quality of students' artifacts, there is a need for longitudinal intervention. We will report this in our subsequent studies and place greater emphasis on investigate how young learners engage with multimodal feedback throughout the learning process in future research.

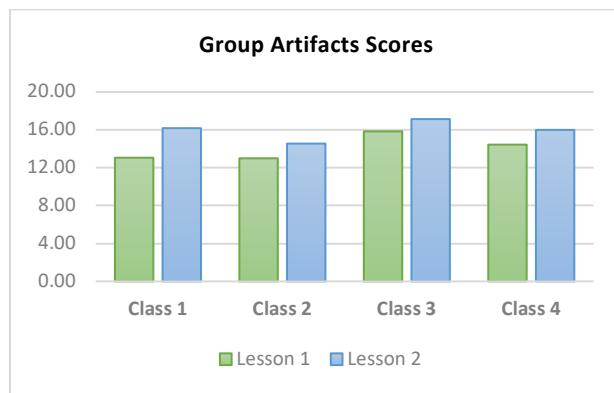


Figure 2. Student Performance Across the First Two ARChE 2.0 Learning Activities

Acknowledgements

This study is funded by Education Research Funding Programme, Ministry of Education (MOE), Singapore, with project nos. ERFP 05/23 WY.

References

Aslan, S., Alyuz, N., Li, B., Durham, L. M., Shi, M., Sharma, S., & Nachman, L. (2025). An early investigation of collaborative problem solving in conversational AI-mediated learning environments. *Computers and Education: Artificial Intelligence*, 8, 100393. <https://doi.org/10.1016/j.caai.2025.100393>

Campbell, B. S., & Feldmann, A. (2017). The Power of Multimodal Feedback. *Journal of Curriculum, Teaching, Learning and Leadership in Education*, 2(2), 1. <https://digitalcommons.unomaha.edu/cgi/viewcontent.cgi?article=1028&context=ctlle>

Ferris, D. R. (2010). L2 Writing Research and Written Corrective Feedback in SLA. *Studies in L2 Acquisition*, 32(2), 181–201. <https://doi.org/10.1017/s0272263109990490>

Kumar, P., Bhatt, G., Ingle, O., Goyal, D., & Raman, B. (2023). Affective Feedback Synthesis Towards Multimodal Text and Image Data. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 19(6), 1–23. <https://doi.org/10.1145/3589186>

Liu, Z., Yin, S. X., Lee, C., & Chen, N. F. (2024). *Scaffolding language learning via multi-modal tutoring systems with pedagogical instructions*. In Proceedings of the 2024 IEEE Conference on Artificial Intelligence (CAI) (pp. 1258–1265). IEEE. <https://doi.org/10.1109/CAI59869.2024.00223>

Musk, N. (2016). Correcting spellings in L2 learners' computer-assisted collaborative writing. *Classroom Discourse*, 7(1), 36–57. <https://doi.org/10.1080/19463014.2015.1095106>

Wen, Y., Chiu, M., Guo, X., & Wang, Z. (2025). AI-powered vocabulary learning for lower primary school students. *British Journal of Educational Technology*, 56(3), 734–754. <https://doi.org/10.1111/bjet.13537>

Zhang, L., Yu, J., Zhang, S., Li, L., Zhong, Y., Liang, G., Yan, Y., Ma, Q., Weng, F., Pan, F., Li, J., Xu, R., & Lan, Z. (2024). *Unveiling the impact of multi-modal interactions on user engagement: A comprehensive evaluation in AI-driven conversations*. arXiv. <https://doi.org/10.48550/arXiv.2406.15000>