

LEARN: Improving Young Learners' Oral Performance Through AI-Guided Picture Conversations

Kartini Anwar, National Institute of Education, Singapore

The European Conference of Education 2025
Official Conference Proceedings

Abstract

Oral language development is a foundational skill in early literacy, yet traditional classroom interactions often provide limited opportunities for structured, interactive, and individualized oral practice. Research has shown that AI-powered language tools can enhance oral performance by providing adaptive, interactive, and real-time feedback, as well as sustained engagement. AI-driven conversation depends on pedagogically sound content design, particularly in how picture-based prompts are selected and structured. **LEARN** (Language automated Evaluation by generating Answers/questions from caRtoonNs) is an AI-powered chatbot designed to improve the oral performance of young children in their mother tongue language (MTL). Jointly developed by the Singapore Institute of Technology (SIT) and the National Institute of Education, Nanyang Technological University (NIE-NTU), **LEARN** facilitates curriculum-aligned picture conversations for young learners, providing visual and verbal stimuli. The key component of its AI-driven conversation is the thematic picture selection based on (i) curriculum alignment, (ii) dynamic and action-based imagery, (iii) diversity and inclusion, and (iv) integration of progressive questioning. This paper presentation discusses the creation of picture-based prompts and pedagogical considerations for leveraging LEARN to facilitate oral performance in mother tongue language learning.

Keywords: AI-chatbot, thematic picture conversation, oral performance, mother tongue language

iafor

The International Academic Forum
www.iafor.org

Introduction

AI-powered chatbots have gained significant attention in education, particularly in language learning. These conversational agents have the potential to facilitate a learning context, provide customized and personalized learning experiences, and enable learners to control their advancement. Chen et al. (2021) provide a comprehensive analysis of the evolution of personalized language learning (PLL), examining trends, methodologies, and emerging research areas. The study identifies that adaptive learning technologies, including conversational agents, chatbots, and intelligent tutoring systems, have played a crucial role in tailoring language instruction to individual learners. The findings suggest that AI-driven tools and adaptive learning models continue to shape the future of PLL.

Xu et al. (2022) show that dialogic reading with a conversational AI agent enhances children's engagement and story comprehension comparable to human-led interactions. Recent studies have demonstrated that chatbot-mediated interaction promotes a learner-friendly environment, reducing anxiety and enhancing communicative confidence (Guo & Yan, 2025). Research has shown that learners exhibit a greater willingness to communicate after interacting with a chatbot (Ayedoun et al., 2019; Tai & Chen, 2023). AI-powered chatbots also serve as virtual teaching assistants, providing instant feedback, corrective suggestions, and conversation simulations that enhance oral language performance (Labadze et al., 2023). Oral language performance has a substantial impact on the overall language proficiency of young learners. AI-powered chatbots offer learners an interactive and engaging environment, providing a personalized learning experience. The integration of AI chatbots into instructional settings is crucial in optimizing language learning outcomes. This supports the growing potential of AI chatbots supplementing home-based educational activities among young learners.

In Singapore, adaptive learning technologies are increasingly integrated into educational frameworks. AI Singapore had launched the "AI in Education Grand Challenge," which aims to encourage the development of innovative approaches that use artificial intelligence to enhance Mother Tongue Language (MTL) learning for primary-level students in Singapore.

Project Objectives

LEARN (Language automated Evaluation by generating Answers/questions from caRtoons) is an AI agent that aims to enhance the MTL learning environment at home to increase proficiency among Primary 1 and 2 students in Singapore. Jointly developed by the Singapore Institute of Technology (SIT) and the National Institute of Education (NIE), **LEARN** facilitates curriculum-aligned picture conversations for young learners, providing visual and verbal stimuli. This project aims to enhance MTL home-based learning by developing an AI platform that engages students in short conversations and offers personalized feedback. **LEARN** incorporates a gamification element to motivate, instill a desire to learn, and consequently improve students' MTL (Chinese, Malay, and Tamil) proficiencies. The project objectives include:

- To enable AI-supported multilingual oral practice in MTL (Chinese, Malay, and Tamil).
- To support task-oriented, picture-based conversational practice.
- To offer real-time, personalized feedback to increase learners' confidence in MTL use.
- To achieve pedagogical equivalence across languages through a unified framework.

The host institution, the Singapore Institute of Technology (SIT), does engineering and technical development of **LEARN**. It is based on the Visual Question Generation (VQG)/Visual Question Answering (VQA) and Dialogue Manager (DM) framework. SIT collaborated with 10 primary schools, involving a total of 659 MTL students from Primary 1 and 2 levels, for data collection and test-bedding activities. **LEARN** utilizes an online platform at www.underoneschool.com. Learners can access the online interface from home.

Design and Features

Multimodal Engagement

The key component of **LEARN**'s AI-driven conversation is the picture description tasks. The question-and-answer generation is done directly in each MTL. The Dialogue Manager generates questions in MTL, and students' answers are compared with AI-generated answers and picture descriptions in MTL. If the student's answer is incorrect, two follow-up questions are generated to advance dialogue between the two. If the student is unable to answer the questions after these two attempts, the correct answer will be provided. The integration of visual stimuli (picture prompts) and interactive AI-driven dialogues enhances cognitive processing, reinforcing both comprehension and verbal expression.

Contextualized and Scaffolded Learning

In **LEARN**, picture-based prompts are selected based on (i) thematic relevance and curriculum alignment, (ii) dynamic and action-based imagery, (iii) diversity and inclusion, and (iv) integration of progressive questioning. The selection of pictures for conversation follows a structured framework derived from the primary school curriculum, ensuring alignment with age-appropriate themes and language learning goals.

Through interaction with the chatbot, children engage in scaffolded discussions that promote vocabulary expansion, sentence construction, and expressive communication. The main characters in the cartoons are the general representation of the three major ethnic groups in Singapore. The characters are contextualized within familiar background scenes for the target users. This includes scenarios of activities at school, home/neighborhood, places of interest, and cultural celebrations. The chatbot images reflect varied cultural, linguistic, and social backgrounds, promoting equitable access to language learning among the three MTL. These images help young learners understand the different ethnic groups and their cultural values in Singapore.

Figure 1

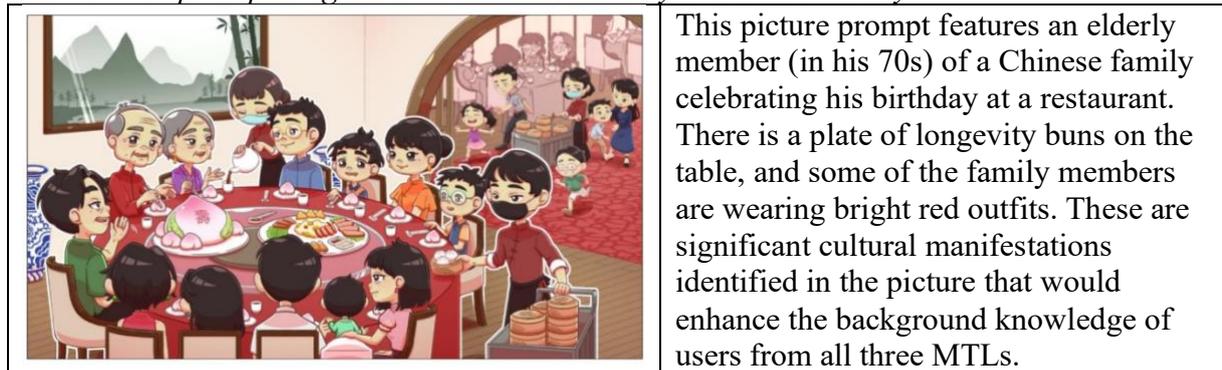
Main Characters From Different Ethnic Groups Used in Picture Prompts



As for thematic relevance, some of the picture prompts feature images of festive celebrations among different ethnic groups, such as preparing for Eid (Malay family), making Kolam (Indian family), and a birthday celebration for an elderly member (Chinese family). While all these festive celebrations hold significance within their respective ethnic groups, AI-driven dialogues are generated in all three mother tongue languages. This exercise expands learners’ language corpus and enhances their understanding of the various cultural contexts in Singapore. This example is shown in Figure 2.

Figure 2

Picture Prompt Depicting a Scenario Celebrated by a Chinese Family

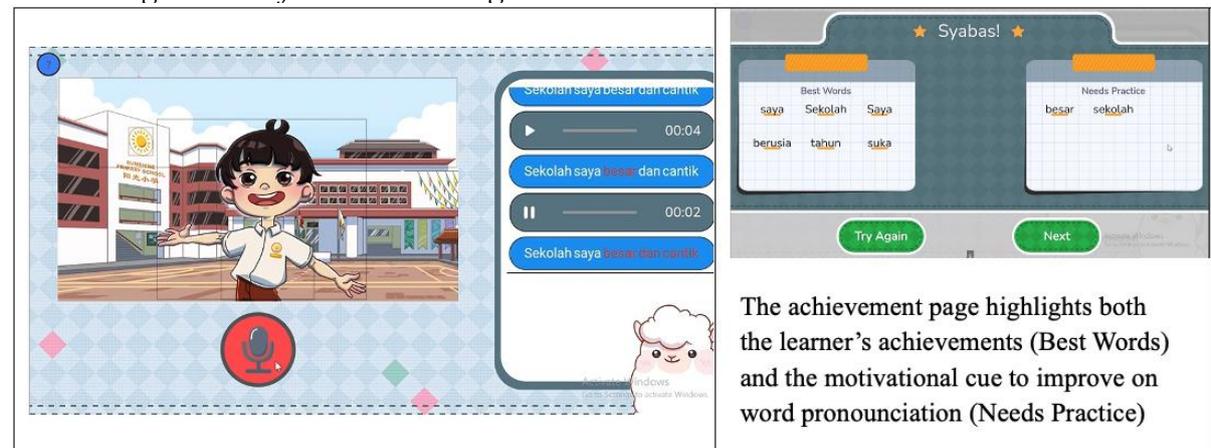


Word Bank – Vocabulary Expansion and Early Syntax Mastery

The AI-assisted picture in **LEARN** exposes young learners to new words, hence reinforcing semantic understanding. At the “word” level, it helps young learners to refine their pronunciation. At the “sentence” level, it helps learners to construct sentences by modelling correct speech patterns. The system is designed to detect errors in word pronunciation or sentence construction by utilizing its automatic speech recognition (ASR). Learners are allowed to record their voice multiple times until they are satisfied before submitting their answers. They are allowed to repeat a maximum of two attempts at answering. If both attempts are not successful, the correct answer will be generated. Each attempt will be followed by a motivational cue that acts as a reward to stimulate interest in the learning process. The reward system helps maintain learner engagement in the task at hand.

Figure 3

Facilitating Oral Performance Through Pronunciation

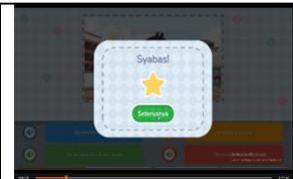


Multiple-Choice Questions and Task Selection

The integration of multiple-choice questions (MCQs) within AI-driven picture-based chatbot interactions serves as a pivotal mechanism for reinforcing cognitive and linguistic development in language learners. In **LEARN**, learners listen to an audio chatbot prompt and select the correct answer from four randomized options. Here, learners are engaged in structured decision-making processes. MCQs facilitate semantic precision and conceptual clarity, prompting the learners to critically analyse visual stimuli and select the most contextually appropriate linguistic expressions. This pedagogical strategy augments lexical acquisition by embedding vocabulary recognition within an interactive framework.

Through gamification, **LEARN** incorporates a real-time adaptive feedback mechanism within the AI chatbot system. This ensures that learners receive immediate corrective reinforcement, making the learning experience engaging and rewarding. There is an achievement page where learners can track the number of exercises completed, the time spent, and exercises with a perfect score.

Figure 4
Motivational Cues at Every Stage of Attempts

			
<p><i>Syabas! Seterusnya.</i> Well done! Next. This cue appears when the learner provides a correct answer.</p>	<p><i>Syabas! Ok</i> Well done! Ok. This cue appears when the learner attains all correct answers for a given item.</p>	<p><i>Jawapan kamu salah. Cuba lagi!</i> This cue appears when the learner answers incorrectly and is given another attempt.</p>	<p><i>Jawapan kamu salah. Jawapan yang betul ialah.</i> This cue appears when the learner answers incorrectly on the second attempt. The correct answer will be generated.</p>

LEARN enables learners to select simpler image descriptions that emphasize fundamental vocabulary and sentence structure. As they demonstrate competence, they can attempt other picture-based prompts that offer more complex scenarios, requiring advanced oral expression, to ensure incremental learning progression. The collection of picture prompts enables the system to introduce multi-layered scenarios that require more advanced linguistic skills, such as narrative cohesion, inference, and contextual adaptation.

For instance, initially, learners may engage in simple image prompts that feature a single subject matter requiring the use of nouns and basic descriptive phrases. As learners advance, the system integrates multi-action-based or multi-character visual stimuli, prompting responses that necessitate more sophisticated grammatical structures and discourse skills. The Question/Answer features gradually progress to ensure cognitive engagement, as learners refine sentence fluency and spontaneous conversational ability. Samples of picture prompts are shown in Figure 5.

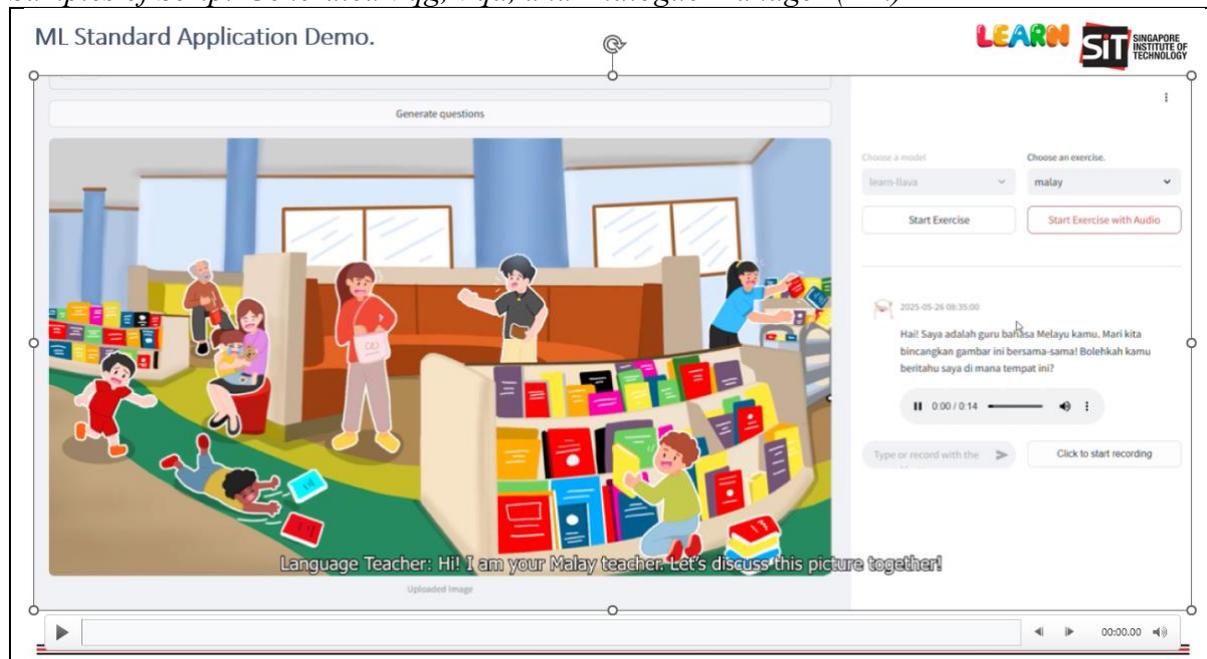
Figure 5*Samples of Picture Prompts, Ranging From Simple to Complex Scenarios***Questions & Answers Features**

LEARN's chatbot enables learners to engage in a conversation through a Visual Question Generation (VQG)/Visual Question Answering (VQA) and Dialogue Manager (DM) framework. VQG enables the chatbot to generate contextually relevant questions based on an image, prompting learners to engage in structured conversations. Instead of simply describing an image, learners are encouraged to respond to the questions, fostering deeper cognitive engagement. This can be done by generating open-ended and inferential questions. This encourages learners to expand their responses, improving fluency and critical thinking. On the other hand, VQA allows the chatbot to interpret learners' responses and provide adaptive feedback based on the image content. If a learner describes an image inaccurately or incompletely, the system can guide them toward a more precise answer.

The Dialog Manager (DM) serves as the central control system, managing chatbot interactions to ensure coherent and context-based conversations. It manages turn-taking mechanics, personalized learning pathways, and error correction and reinforcement, providing instant feedback to refine pronunciation, grammar, and sentence structure. By integrating VQG, VQA, and DM, **LEARN** creates an immersive, adaptive, and engaging language learning experience. The incremental learning approach ensures that learners progress naturally, building confidence in oral performance.

The picture below serves as a demo for the application. It begins with an introduction provided by an AI-generated teacher's voice. The learner answers by recording his/her voice and submits it to the chatbot. Both visual and oral questions and answers are generated.

Figure 6
Samples of Script-Generated Vqg, Vqa, and Dialogue Manager (Dm)



Teacher	Hai! Saya adalah guru Bahasa Melayu kamu. Mari kita bincangkan gambar ini bersama-sama. Boleh kamu beritahu saya di manakah tempat ini? <i>Hi, I am your Malay Language teacher. Let's discuss this picture together. Can you tell me where this place is?</i>
Learner	Gambar ini di perpustakaan. <i>This picture is at a library.</i>
Teacher	Jawapan yang betul ialah, "Gambar ini menunjukkan sebuah perpustakaan." <i>The correct answer is "This picture shows a library."</i>
Teacher	Jika kamu perhatikan, kamu boleh melihat beberapa orang melakukan pelbagai aktiviti seperti membaca, berbincang dan bermain. Sekarang, lihat ke bahagian kiri gambar. Apakah yang dilakukan oleh budak lelaki yang berseluar merah di atas karpet? <i>If you observe, you can see a few people engaging in various activities, such as reading, discussing, and playing. Now, look at the left side of the picture. What is the boy wearing red shorts doing on the carpet?</i>
Learner	Dia sedang berlari. <i>He is running.</i>
Teacher	Jawapan yang baik. <i>That is a good answer.</i>
Teacher	Apakah yang akan kamu lakukan di perpustakaan? <i>What would you like to do in the library?</i>

Learner	Saya akan membaca buku. <i>I would read a book.</i>
Teacher	Bagus. Hari ini kamu telah belajar perkataan-perkataan baharu seperti; “perpustakaan, berlari, memandang, meminjam buku” dan yang lain-lain. Teruskan berlatih. Jumpa lagi! <i>Great job! Today you have learnt the following new words, for example, library, running, and borrowing books. Remember these words. Continue practicing and see you next week.</i>

Project Contributions

Among the key innovations of **LEARN** is the leveraging of machine learning and natural language processing (NLP), which allows learners to practice oral language performance through interactive conversations. The adaptive learning tailors educational content based on individual learner progress. This ensures personalized instruction that adapts to student needs. It promotes learners’ autonomy and self-directed learning. Learners are allowed to repeat tasks and select tasks according to their learning progression.

The picture prompts are developed based on their relevance in the Singapore context and align with the primary school curriculum content. This includes scenarios with diverse age-appropriate themes and the infusion of values. Most of all, it meets the language learning goals.

The multimodal learning features enable image recognition, voice interaction, text, oral, and audio-based tutoring. The incorporation of VQG, VQA, and DM provides progressive questions that further ensure interactive learning progressions and oral performance. **LEARN** incorporates gamification and engagement strategies that use a reward system, progress tracking, and interactive challenges to enhance motivation and retention. This promotes confidence in oral performance, reduces anxiety, and makes MTL learning immersive for young learners.

Implications and Improvements

LEARN has launched its platform at www.underoneschool.com. The SIT team did a trial at a primary school involving 206 MTL students from Primary 1 and Primary 2. Students were given the audio-based multiple-choice question (MCQ) task and the oral-based question-and-answer task. Feedback gathered from the trial session showed that 6% of the total participants did not enjoy it. When asked about the difficulty level of the tasks, most respondents believe that the questions generated by the system are straightforward, whereas 20% find the questions generated by the system to be challenging. This implies that while most respondents enjoy using **LEARN** to improve their oral performance, some of them are challenged by the questions prompted by the chatbot.

Despite the innovation and contributions of **LEARN**, the AI-powered chatbot system, several limitations remain that could be addressed for improvement. One of the challenges is that AI-generated voices may struggle with ambiguous, nuanced, or multi-turn conversations, resulting in inaccurate or generic prompts or responses. This makes it less effective for learning nuanced languages. Additionally, natural language processing (NLP) limitations affect the ability to interpret idiomatic expressions, sarcasm, and cultural nuances, which are essential for human-like interactions.

The speech recognition technologies embedded in chatbots often fail to accurately assess pronunciation, especially for young learners with developing articulation skills. Young learners may provide fragmented or unconventional responses that are accepted as correct; however, AI chatbots may misinterpret them, leading to incorrect or generic feedback. This may lead to discouraging learners rather than guiding them constructively.

While chatbots can stimulate positive reinforcement, they lack proper emotional awareness, making it challenging to adjust tone dynamically based on the young learner's engagement level. Motivation in young learners is highly dependent on emotional connection. AI-driven interactions may sometimes feel mechanical rather than intuitive.

LEARN could integrate advanced NLP models that recognize child-specific speech patterns in the mother tongue language. By refining the multi-turn dialogue processing, LEARN can provide more adaptive and meaningful conversations tailored to young MTL learners. The responses from the AI chatbot could be adjusted based on tone, engagement levels, and learner frustration cues to improve emotional security in MTL learning.

LEARN could incorporate advanced speech recognition to support pronunciation, enhance AI-driven phonetic analysis, and provide accurate pronunciation feedback without discouraging young learners. This could also be achieved by implementing gradual speech correction models, rather than direct correction, which would help children progressively refine their spoken language.

Conclusion

The LEARN project showcases the design and development of an innovative AI platform for learning mother tongue languages. It addresses the current linguistic and educational landscape, supporting the language literacy of young learners.

The system holds significant potential for improving the oral performance of young learners, particularly through the integration of picture-based prompts, adaptive questioning, and multi-modal interactions. While LEARN fosters engagement, motivation, and oral proficiency, existing limitations necessitate further technological refinement. As technology continues to evolve, LEARN would hold significant potential in the educational landscape, particularly in mother-tongue language learning.

Acknowledgements

This project is funded by the AI Singapore (AISG) project LEARN: Language Automated Evaluation by Generating Answers/Questions from CaRtoonNs, supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-GC-2022-004). Special thanks to the research team from the Singapore Institute of Technology (SIT) and the National Institute of Education (NIE).

References

- Ayedoun, E., Hayashi, Y., & Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education*, 29(1), 29-57. <https://doi.org/10.1007/s40593-018-0171-6>
- Chen, X., Xie, H., Zou, D., & Cheng, G. (2021). Twenty years of personalized language learning. *Educational Technology & Society*, 24(1), 205-222.
- Guo, Z., & Yan, H. (2025). The impact of Artificial Intelligence chatbots on young EFL learners' willingness to communicate, *Curriculum and Teaching Methodology*, 8(06), 112-119. <https://doi.org/10.23977/curtm.2025.080116>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in Education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20, 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Tai, T.-Y., & Chen, H. H.-J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, 31(3): 1485–1502. <https://doi.org/10.1080/10494820.2020.1841801>
- Xu, Y., Aubele, J., Vigil, V., Bustamante, A., Kim, Y. S., & Warchauer, M. (2022). Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. *Child Development*, Mar, 93(2), 149-167. <https://doi.org/10.1111/cdev.13708>