

# Enhancing Primary Students' Oral Proficiency With AI-Powered Chatbots: Introducing Learn

Suryani Atan<sup>1</sup>, Aw Guat Poh<sup>2</sup>, Kartini Anwar<sup>3</sup>, Dharmaraj Sundararaj<sup>4</sup>, Dharani Kanna Ravi<sup>5</sup>, Lim Kheng Yong<sup>6</sup>, Muhammad Hidir Mohd Ezat<sup>7</sup>

<sup>1,2,3,4,5,6,7</sup> Asian Languages and Cultures (ALC), National Institute of Education (NIE)-Nanyang Technological University (NTU), Singapore,  
suryani.atan@nie.edu.sg<sup>1</sup>, guatpoh.aw@nie.edu.sg<sup>2</sup>, kartini.anwar@nie.edu.sg<sup>3</sup>,  
sundararaj.dharmaraj@nie.edu.sg<sup>4</sup>, dharani.ravi@nie.edu.sg<sup>5</sup>, khengyong.lim@nie.edu.sg<sup>6</sup>,  
hidir.ezat@nie.edu.sg<sup>7</sup>

---

**Abstract**– This study presents LEARN (Language automated Evaluation by generating Answers/Questions from caRtoons), an innovative AI-powered system designed to improve oral proficiency among Primary 1 and 2 students learning their Mother Tongue Languages (MTL) in Singapore. Jointly developed by the National Institute of Education (NIE) – Nanyang Technological University (NTU), and the Singapore Institute of Technology (SIT), LEARN engages students in interactive conversations based on picture-based prompts. The project aims to enhance the home-based learning environment by leveraging conversational AI chatbots that simulate engaging picture-based dialogues. Incorporating advanced technologies such as visual dialogue engines, automated speech recognition, and gamification, LEARN targets a 20% increase in oral proficiency over two years. Data collected from 659 students across ten primary schools were used to train LEARN. Preliminary findings and achievements suggest strong feasibility, linguistic impact, and scalable potential. This paper discusses LEARN as an edtech solution to improve MTL oral proficiency.

**Keywords**– AI in education, automated speech evaluation, chatbot learning, Mother Tongue Language, oral proficiency, primary education

---

## I. INTRODUCTION

Singapore's bilingual education system places significant emphasis on developing oral proficiency in Mother Tongue Languages (MTL) – Chinese Language, Malay Language, Tamil Language – among young learners. However, many Primary 1 and 2 students lack sufficient home-based oral practice, resulting in limited confidence and fluency. With the growing capabilities of AI in natural language processing and speech recognition, this paper explores the development and implementation of LEARN, a chatbot system aimed at bridging this gap through technology-mediated oral practice.

## II. RESEARCH PURPOSE AND DESIGN

This section discusses the research purpose and design of LEARN.

### A. Purpose and Objectives

This research project aims to tackle the challenge of enhancing the Mother Tongue Language (MTL) learning environment at home for P1 and P2 students, striving to increase oral proficiency by 20% within two years. The proposed solution involves the implementation of AI conversational chatbots that engage students in short, simple conversations based on the Picture Description Task. These AI chatbots will facilitate spoken exchanges in the MTL, allowing students to practice their language skills. The goal is to enable realistic, engaging, coherent and meaningful short conversations between AI and student based on the Picture Description Task. Our innovative approach encompasses several key components, including the automatic generation of age-appropriate questions and answers from cartoon images, the development of conversation AI chatbots for data collection and assessment, automated speech recognition for young children speaking a mixture of English and MTL, and the use of automatic picture captioning techniques to extract and match key topics in speech and reference content.

### B. Research Design and Method

A total of 659 Primary 1 and 2 students from ten Singapore primary schools participated in a data collection exercise in July 2023. The collected data, consisting of students' reading and picture conversation recordings, were used to train and fine-tune the AI systems within LEARN.

TABLE I DATA COLLECTION

No.	Total Number of Students is 659	
	Breakdown of Students	No. of Students
1.	The total number of Male students.	275
2.	The total number of Female students.	384
3.	The number of P1 Male students.	120
4.	The number of P2 Male students.	155
5.	The number of P1 Female students.	206
6.	The number of P2 Female students.	178
7.	The number of P1 Male Malay students.	46
8.	The number of P1 Male Chinese students.	65
9.	The number of P1 Male Tamil students.	9
10.	The number of P2 Male Malay students.	40
11.	The number of P2 Male Chinese students.	101
12.	The number of P2 Male Tamil students.	14
13.	The number of P1 Female Malay students.	96
14.	The number of P1 Female Chinese students.	94
15.	The number of P1 Female Tamil students.	16
16.	The number of P2 Female Malay students.	84
17.	The number of P2 Female Chinese students.	72
18.	The number of P2 Female Tamil students.	22

The system was designed with the following components, as shown in Fig. 1. The diagram in Fig. 1 presents a structured overview of a multi-component system designed to support dialogue-based learning and automated assessments. It is organised into four main work packages (WPs), each contributing a critical function to the overall system:

- 1) Question and Answer Generation (WP1a) & Dialogue Engine (WP1b):  
 Handles picture-based question generation (VQG), answer understanding (VQA), and dialogue coherence (NLG/NLU). It enables dynamic, image-based interactions to stimulate learner engagement.
- 2) Speech Engine (WP2):  
 Integrates Text-to-Speech (TTS), Automatic Speech Recognition (ASR), and a Translation Manager (TM), enabling multilingual and multimodal communication.
- 3) Automated Evaluation Engine (WP3):

- Assesses student responses using a Speech Evaluation Module (SEM) and a Language Evaluation Module (LEM), allowing for real-time, automated feedback on learners' spoken language performance.
- 4) Gamification (WP4):  
 Includes gamified elements into the learning experience to sustain engagement and motivation.

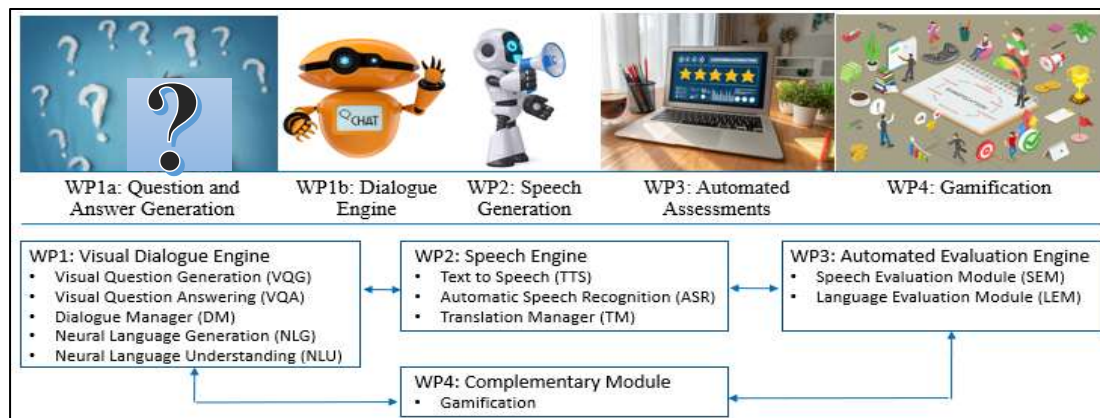


Fig. 1 Components of LEARN Design

Arrows between modules in the diagram indicate dynamic interactions, such as the flow of data from question generation to speech processing, and from speech output to evaluation. The integration of gamification and assessment modules suggests a feedback loop that supports continuous learning and improvement.

### C. Literature Review

In Singapore, primary school students are required to learn a Mother Tongue Language (MTL) as a second language subject. Typically, students study the language aligned with their ethnic background – Chinese students learn Chinese, Malay students learn Malay, and Indian students learn Tamil [11]. This policy aims to preserve cultural heritage and promote bilingualism, a cornerstone of Singapore's education system. However, despite these efforts, research and national observations have highlighted a growing trend: students are increasingly using English as their dominant language, both at home and in social settings. This shift has led to a decline in MTL proficiency among younger learners [12]. The Ministry of Education has acknowledged this challenge and introduced initiatives such as structured reading programmes and increased exposure to MTL in kindergartens to address the issue [12]. To support these national efforts and bridge the gap in oral language competency, we developed LEARN, a conversational AI chatbot tailored for young learners. LEARN is designed to provide engaging, scaffolded speaking practice in MTLs through interactive dialogues, cartoon-based scenarios, and gamified learning experiences. By leveraging AI technologies and culturally relevant content, LEARN aims to make MTL learning more accessible, enjoyable, and effective for today's digitally native students. Recent studies have demonstrated the efficacy of AI-powered dialogue systems in language education. For example, it is found that AI-driven tutoring systems enhanced students' engagement and performance in language tasks [9]. Conversational agents, such as chatbots, are increasingly integrated into educational platforms, with researchers noting their ability to provide real-time, scalable interaction in second language acquisition [5]. Automated Speech Recognition (ASR) tools have also been shown to assist in evaluating pronunciation and fluency [10]. However, many existing systems do not adequately support young learners, especially in multilingual settings like Singapore. LEARN addresses this gap by incorporating child-friendly interfaces, multilingual support, and contextually relevant content through cartoon imagery. In the context of bilingual education, Cummins' (1979) theory of Common Underlying Proficiency (CUP) suggests that skills learned in one language can support development in another [4]. This is highly relevant to Singapore's education system, where students often develop language proficiency across English and a Mother Tongue Language simultaneously. Bilingual learners benefit from structured language input and interaction, which LEARN facilitates through scaffolded dialogues tailored to their developmental stage. From an e-pedagogical perspective, frameworks such as the TPACK model

(Technological Pedagogical Content Knowledge) emphasize the need for coherent integration of content, pedagogy, and technology [7]. LEARN aligns with this model by combining AI technologies with research-based pedagogy and age-appropriate linguistic content. In addition, Laurillard's Conversational Framework (2002) reinforces the importance of dialogic learning interactions [6] – a principle embedded in LEARN's AI-driven picture conversations. These frameworks support the pedagogical validity of technology-enhanced oral language learning environments. Scaffolding is a foundational concept in educational psychology that refers to the structured support provided to learners as they acquire new skills or knowledge. Originally introduced by Wood, Bruner, and Ross (1976), scaffolding involves breaking down learning tasks into manageable components and gradually reducing assistance as learners gain competence [1]. In the context of language education, scaffolding is particularly effective in supporting oral language development, as it allows learners to build confidence and fluency through guided practice. Research has shown that scaffolded instruction enhances learner autonomy and improves language outcomes [2]. For young learners, especially those acquiring a second language, scaffolding can take the form of visual aids, sentence starters, modelled dialogues, and interactive feedback. These strategies help learners internalize linguistic structures and vocabulary in meaningful contexts. LEARN incorporates scaffolding through its tiered learning design, which progresses from visual comprehension and vocabulary acquisition to conversational practice. Each level is carefully structured to support learners at their developmental stage, ensuring that they are neither overwhelmed nor under-challenged. This approach aligns with Vygotsky's (1978) theory of the Zone of Proximal Development (ZPD), which emphasizes the importance of providing appropriate support to help learners move from what they can do independently to what they can achieve with guidance [3][8]. By embedding scaffolding into its AI-driven interactions, LEARN not only facilitates language acquisition but also fosters learner confidence and engagement – key factors in successful bilingual education.

### III. LEARN DEVELOPMENT

This section discusses the development of LEARN.

#### A. Description of Portal

LEARN leverages cartoon imagery to generate contextual conversations, allowing for spontaneous and guided speech practice. It supports three Mother Tongue languages – Chinese, Malay, and Tamil – offering multilingual flexibility in real-time conversations. Its adaptive scaffolding model supports varying student proficiencies through tiered prompting and automatic feedback. Gamification features enhance user motivation and facilitate sustained oral practice at home. Fig. 2 shows the homepage of LEARN portal.



Fig. 2 Homepage of LEARN Portal

To reflect the multicultural fabric of Singapore, LEARN features three families: a Chinese family, a Malay family, and an Indian family (see Fig. 3). These characters are embedded in cartoon-style scenarios that depict everyday life. This design choice serves two purposes:

1) Cultural Familiarity:

Students see their own and others' cultures represented, fostering inclusivity.

2) Visual Appeal:

The cartoon format is engaging and age-appropriate, increasing motivation and enjoyment in learning.



Fig. 3 Multiracial Characters in LEARN  
B. Pedagogical Framework

LEARN is thoughtfully designed to scaffold student's speaking skills through a structured, progressive approach (see Fig. 4). The system is divided into three levels, each building upon the previous to enhance language acquisition and oral proficiency.

1) Level 1: Visual Comprehension and Basic Interaction

Students begin by observing a contextual image and responding to multiple-choice questions (MCQs). This stage helps learners engage with visual cues and develop basic comprehension skills in a low-pressure environment.

2) Level 2: Vocabulary and Pronunciation Support

At this level, students are introduced to a broader range of vocabulary relevant to the scene. Pronunciation support is provided through audio prompts, enabling students to practice articulation and build phonological awareness.

3) Level 3: Conversational Practice

Students engage in interactive conversations with the system (see Fig. 5). This simulated dialogue allows learners to apply their vocabulary and pronunciation skills in a dynamic, communicative context, fostering fluency and spontaneity.

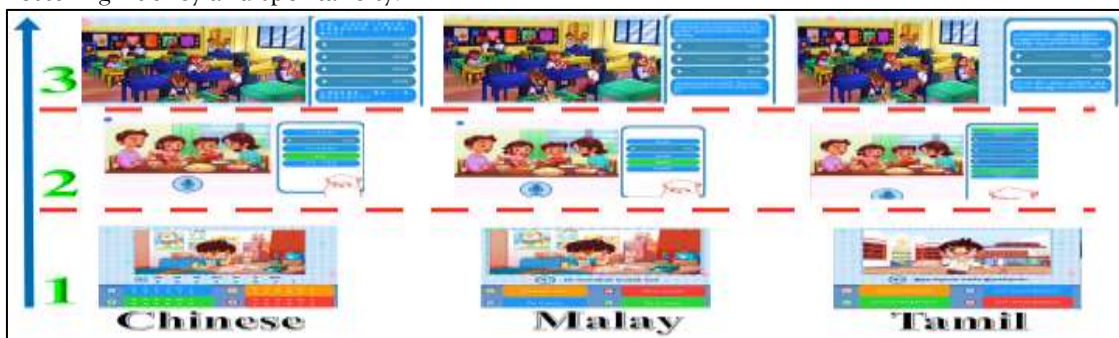


Fig. 4 LEARN Learning Design and Scaffolding



Fig. 5 Malay Language Conversation in LEARN

LEARN exemplifies how digital tools can be effectively designed to scaffold speaking skills while promoting cultural inclusivity. Its structured progression, multimodal inputs, and contextual relevance make it a promising tool for language educators seeking to enhance oral proficiency in diverse classrooms.

#### IV. FINDINGS AND DISCUSSIONS

The early phases of LEARN's trial yielded encouraging results, particularly in its ability to engage young learners and support oral language development. Feedback collected from Primary 1 and Primary 2 students across three language streams – Chinese, Malay, and Tamil – indicates a high level of enjoyment and perceived ease of use (see Table II).

TABLE III STUDENT FEEDBACK

Questions	Primary 1 and Primary 2 Students		
	Chinese Language (%)	Malay Language (%)	Tamil Language (%)
Q1. Did you enjoy the mini game?			
a) I enjoyed it very much!	53.6	66.7	80.0
b) I enjoyed it!	40.2	27.4	20.0
c) I did not enjoy it!	6.2	5.9	0
Q2. Are the questions easy?			
a) Yes, they are very easy.	78.2	82.8	80.0
b) No, they are not easy.	21.8	17.2	20.0

##### A. Student Engagement and Enjoyment

Across all language groups, most students reported enjoying the mini game component of LEARN. Notably, 80% of Tamil language students selected “I enjoyed it very much!”, followed by 66.7% of Malay

and 53.6% of Chinese language students. These responses suggest that the interactive and gamified elements of LEARN are effective in capturing students' interest and sustaining engagement.

#### B. Perceived Ease of Use

When asked whether the questions were easy, over 78% of students in each language group responded affirmatively. This high percentage reflects the effectiveness of LEARN's scaffolding approach, which gradually introduces learners to new vocabulary and conversational structures in a supportive and intuitive manner.

#### C. Visual Representation of Feedback

To better illustrate these findings, the following chart presents a comparative view of student responses across the three language groups (see Fig. 6):

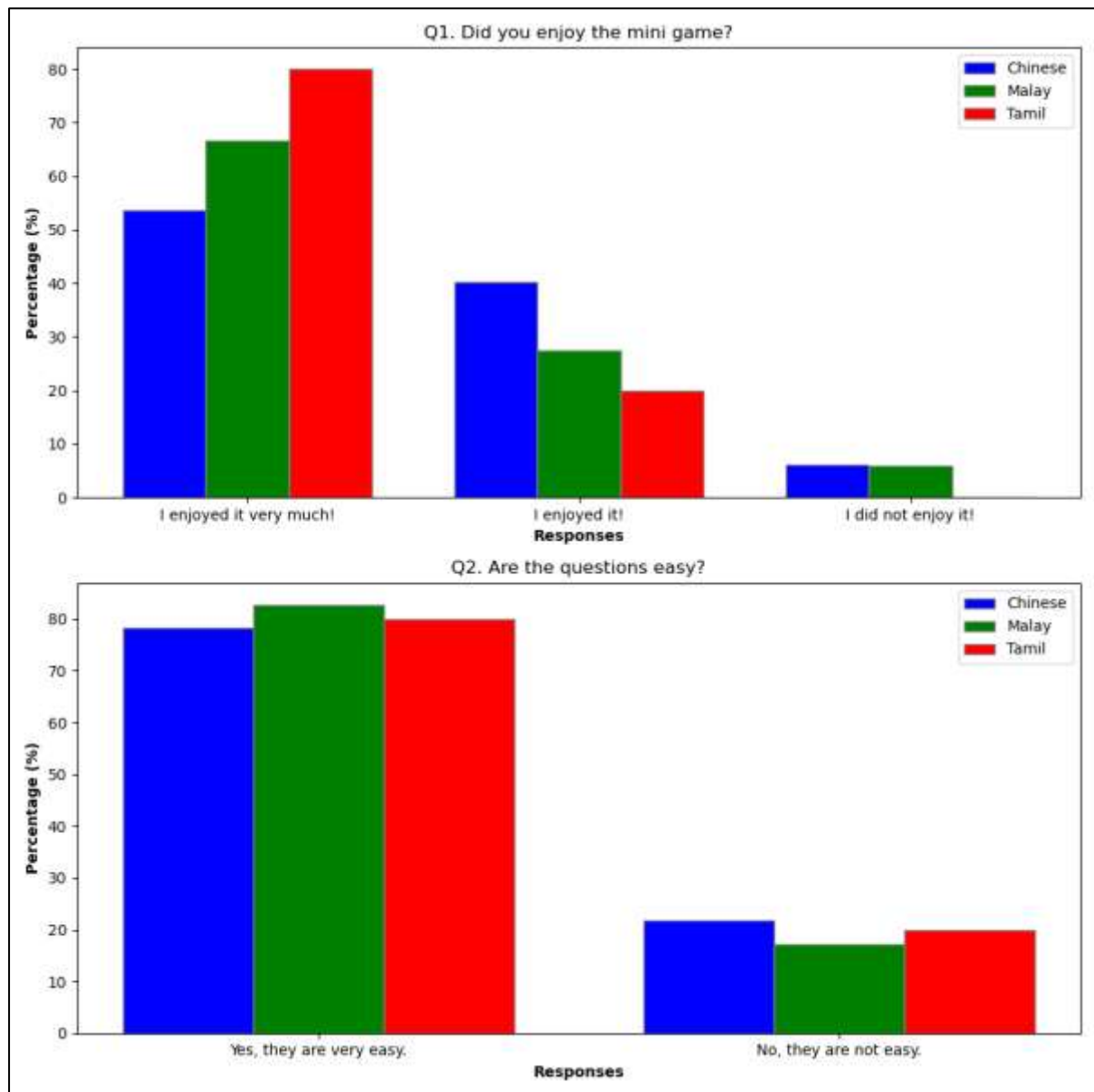


Fig. 6 Feedback Charts

#### D. Interpretation of Results

The positive feedback can be attributed to several key design features of LEARN:

- 1) Scaffolded Learning Design:

The tiered structure of LEARN ensures that students are not overwhelmed. Each level builds upon the previous, allowing learners to progress at a comfortable pace.

2) Multimodal Interaction:

By integrating speech, image, and text, LEARN creates a rich, immersive learning environment that mirrors real-life communication scenarios.

3) Cultural Relevance:

The inclusion of cartoon families representing Singapore's major ethnic groups adds familiarity and relatability, which likely contributes to students' enjoyment and motivation.

E. Broader Implications

LEARN demonstrates how AI can be meaningfully integrated into early education environments, especially for oral language development. The use of multimodal data (speech, image, and text) allows for realistic and engaging conversations. Its design also supports scalable deployment across diverse linguistic and socio-cultural backgrounds, making it adaptable beyond the Singaporean context.

## V. CONCLUSIONS

LEARN has shown potential in addressing the challenges of oral language acquisition among young learners of Mother Tongue Languages (MTL) by integrating conversational AI, culturally relevant cartoon imagery, and gamified learning experiences. Its design is grounded in pedagogical principles such as scaffolding, multimodal engagement, and contextual learning, which collectively support the development of speaking skills in a structured yet enjoyable manner. The system's positive feedback across three major language streams – Chinese, Malay, and Tamil – demonstrates its versatility and inclusivity. Feedback from early trial indicates high levels of student engagement, enjoyment, and perceived ease of use, validating the system's design and instructional approach. Beyond its immediate educational impact, LEARN holds significant promise for broader application. Its modular and adaptable framework allows for easy customisation to suit different linguistic and cultural contexts, making it a scalable solution for multilingual education globally. Furthermore, the integration of AI-driven dialogue and real-time feedback positions LEARN at the forefront of innovative educational technology, with strong potential for commercial deployment in both formal and informal learning environments. In conclusion, LEARN is more than a digital tool. It is a transformative platform that reimagines how young learners engage with language. By combining technology with culturally responsive pedagogy, it paves the way for more inclusive, effective, and engaging language education in the 21st century.

## ACKNOWLEDGEMENT

Thank you to the funding agency, AI Singapore (AISG), National Research Foundation, and to our research partner and host institution, Singapore Institute of Technology (SIT).

## REFERENCES

- [1] D. Wood, J. S. Bruner, And G. Ross, "The Role Of Tutoring In Problem Solving," *Journal Of Child Psychology And Psychiatry*, Vol. 17, No. 2, Pp. 89-100, 1976.
- [2] J. Hammond, Ed., *Scaffolding: Teaching And Learning In Language And Literacy Education*. Newtown, NSW: Primary English Teaching Association, 2001.
- [3] L. S. Vygotsky, *Mind In Society: The Development Of Higher Psychological Processes*. Cambridge, MA: Harvard University Press, 1978.
- [4] J. Cummins, "Linguistic Interdependence And The Educational Development Of Bilingual Children," *Review Of Educational Research*, Vol. 49, No. 2, Pp. 222-251, 1979.
- [5] A. Kerlyl, P. Hall, And S. Bull, "Bringing Chatbots Into Education: Towards Natural Language Negotiation Of Open Learner Models," *Knowledge-Based Systems*, Vol. 20, No. 2, Pp. 177-185, 2007.
- [6] D. Laurillard, *Rethinking University Teaching: A Conversational Framework For The Effective Use Of Learning Technologies*. London, UK: Routledge, 2002.
- [7] P. Mishra And M. J. Koehler, "Technological Pedagogical Content Knowledge: A Framework For Teacher Knowledge," *Teachers College Record*, Vol. 108, No. 6, Pp. 1017-1054, 2006.
- [8] K. Shabani, M. Khatib, And S. Ebadi, "Vygotsky's Zone Of Proximal Development: Instructional Implications And Teachers' Professional Development," *Canadian Center Of Science And Education*, Vol. 3, No. 4, Pp. 237-248, 2010.
- [9] S. Wollny, J. Schneider, And H. U. Hoppe, "Conversational Agents For Language Learning: A Systematic Review," *Journal Of Educational Technology & Society*, Vol. 24, No. 1, Pp. 1-17, 2021.
- [10] Y. Zhang, L. Zhao, And Z. Wang, "Speech Recognition And Feedback Tools For Second Language Oral Practice: A Meta-Analysis," *Computer Assisted Language Learning*, Vol. 35, No. 4, Pp. 567-590, 2022.

[11] <https://www.moe.gov.sg/primary/curriculum/mother-tongue-languages/learning-in-school>

[12] <https://www.moe.gov.sg/news/press-releases/20240914-strengthening-efforts-for-students-to-learn-and-use-their-mother-tongue-language-from-young>