Education Tech Research Dev
https://doi.org/10.1007/s11423-024-10360-2

RESEARCH ARTICLE

Effects of robot-based multiple low-stakes assessments on students' oral presentation performance, collective efficacy, and learning attitude

Darmawansah Darmawansah[1,2] · **Gwo-Jen Hwang**[1,3,4]

Accepted: 26 February 2024

Abstract
Low-stakes assessment has gained attention in recent years due to its link to enhancing learning effects and its essential role in learning evaluation. Unlike high-stakes assessments, low-stakes assessments have little or no consequences for learners' academic performance, and are designed to support the feedback-oriented learning process. Providing multiple low-stakes assessments to students yields significantly greater long-term retention of knowledge and skills. However, learners may not give their best efforts when taking low-stakes assessments, which could lead to poorer learning outcomes. Using emerging technologies such as social robots in the learning environment could foster interactive learning, engagement, and motivation for learning assessments. Therefore, integrating low-stakes assessments and robots might encourage students to exert greater effort while performing learning tasks. This study aimed to discover the impacts of robot-based multiple low-stakes assessments on students' oral presentation performance, collective efficacy, and learning attitude. A quasi-experiment was conducted in two sixth-grade classes of elementary students. The Robot-based Multiple Low-Stakes Assessment (Robot-MLSA) was randomly assigned to one class, while the Computer-based Multiple Low-Stakes Assessment (C-MLSA) was assigned to another class. The findings showed that the Robot-MLSA could enhance students' oral presentation performance, support their collective efficacy, and improve their learning attitude toward robots. Furthermore, an in-depth discussion of students' learning perceptions and experience is provided to explore the effectiveness of the Robot-MLSA.

Keywords Social robot · Low-stake assessment · Oral presentation · Collective efficacy · Learning attitude

Introduction

Emerging technologies such as robots for communication are simply defined as social robots. They are mainly programmed to have interactive communication with people (van den Berghe et al., 2019), whether the robot is controlled by a person or without a person in real time. The notion of a social robot offers many behavioral norms typical of

Extended author information available on the last page of the article human interaction due to the existence of a physical body. To some extent, the employment of robots in learning settings may lead to further development of the students' language and communication skills.

Recent studies have highlighted the efficacy of robots in various educational roles, such as leading conversational practice (Engwall et al., 2022), creating digital storytelling (Chen Hsieh & Lee, 2021; Yang et al., 2022), and providing real-time feedback to learners (Al Hakim et al., 2022). In the context of robot-assisted language learning (RALL), there is evidence suggesting that students can learn effectively with either human teachers or robots. Some research, such as a study with a single session format involving a robot (e.g., Engwall et al., 2022), even indicates superior learner engagement and interaction compared to multi-session formats which seem to have a minimal impact on learning engagement (e.g., Al Hakim et al., 2022). This leads to the hypothesis that the novelty of robots may diminish over time. Prior studies on RALL have posited that robots can significantly aid in developing students' verbal and non-verbal communication skills, which are crucial for current language learning challenges (e.g., Chen Hsieh, 2021; Randall, 2019; Wedenborn et al., 2019). However, it is vital to carefully design learning materials for robots, especially when using RALL in language assessments focused on thinking skills.

This brings us to the concept of thinking skill-based assessment in language learning. As we shift from the general efficacy of RALL to its specific application in assessing thinking skills, it is important to recognize the unique challenges this presents. To counter the potential sheer novelty effect of robots in multi-session learning environments, researchers and practitioners must develop robust frameworks for meaningful learning. This involves integrating learning strategies, such as the 6E learning model (see Hsiao et al., 2022), promoting collaborative work (e.g., Rosenberg-Kima et al., 2020), and allowing students to contribute to robot-based learning content (e.g., Chen Hsieh & Lee, 2021). Furthermore, thinking skill-based assessments should encompass a wide range of cognitive abilities, including comparing and contrasting, classification, reasoning, decision making, and problem solving (see Burke & Williams, 2012). In this context, it becomes imperative for teachers to focus their instruction and assessment methods not only on language proficiency but also on enhancing the quality of formal learning by targeting the development of students' thinking skills, alongside their interactions with both peers and robot assistance.

Low-stakes assessments, while providing feedback to students, primarily serve to offer more opportunities for applying and utilizing learned content. They crucially shift the emphasis from assessments that solely inform instructional decisions to those that actively engage students in the learning process. In other words, they allow students to experiment with and understand concepts without the high pressure of significantly impacting their overall grades. This approach aligns seamlessly with the educational objectives emphasized by Hsia and Sung (2020) and Srisawasdi et al. (2018), where the priority is on meaningful learning rather than on using instruction and assessment for entertainment purposes only.

In the context of designing robot content, breaking down information into manageable chunks and delivering it over suitable intervals can profoundly enhance the learning process. As suggested by Sweller (2010), this method can maximize the germane cognitive load, making the learning experience more effective and meaningful. Low-stakes assessments can be strategically utilized in this scenario to facilitate engagement with these information chunks. The reduced consequences for errors in these assessments encourage a deeper exploration and understanding of the content, fostering a more robust and creative learning environment.

The concept of Classroom Assessment Techniques (CATs), as coined by Cross and Angelo (1988), further supports this approach. CATs are designed to assist students in examining their language development and help instructors, or in this case, the robot used in the study, to prioritize learning goals. The robot was employed as a tool to support students during low-stakes assessments. This integration aimed to enhance students' learning through innovative and student-centered methods. Thus, low-stakes assessments, when combined with effective instructional strategies such as CATs and supported by technology such as robots, can significantly enhance the learning experience. They shift the focus from traditional high-stakes testing to a more holistic, engaging, and learner-centric approach.

This study emphasizes the intersection of low-stakes assessment in the form of social robot deployment, an approach that is called Robot-based Multiple Low-Stakes Assessment or Robot-MLSA. The insights and pedagogical implications in this current study are generally applicable to the RALL context using multiple low-stakes assessment tasks, as well as the CATs which offer various assessments related to thinking skills (e.g., assessing prior knowledge and understanding, critical thinking, and students' reactions to instruction and class activities).

Based on the purpose of this study, the following questions were proposed:

(1) To what extent do students in the Robot-based multiple low-stakes assessment (Robot-MLSA) group make better English oral presentations than students in the computer-based multiple low-stakes assessment (C-MLSA) group?
(2) To what extent do students in the Robot-MLSA group have higher collective efficacy than students in the C-MLSA group?
(3) To what extent do students in the Robot-MLSA group have better learning attitudes than students in the C-MLSA group?
(4) What are the students' learning experiences and perceptions of the two learning environments?

Literature review

Social robot employment and learning strategies in the RALL context

Due to the primary function of social robots of interacting and communicating with people, their use in educational settings, particularly in RALL, has gained significant traction. Empirical studies, such as those by Wellsby and Pexman (2014), employed social robots to allow students to naturally interact with the real-life physical environment. This is especially beneficial in language development, where the humanoid appearance of robots plays a critical role. However, it is noteworthy that while previous studies have made strides in exploring morphosyntax skills (i.e., word learning) (van den Berghe et al., 2019), there is a need for more extensive research on speaking skills, particularly in RALL contexts, as some studies were based on small samples and lacked control conditions, resulting in tentative evidence only.

With the mainstream acceptance of social robots in RALL, understanding the best practices for deploying this tool has become crucial. This includes addressing key issues such as instructional design and teaching strategies for language learning. There is a growing body of studies investigating the employment of social robots in RALL (e.g., Liang &

Hwang, 2023; Yang et al., 2022), emphasizing the need for well-chosen and well-designed learning strategies to foster meaningful student interactions.

Many studies have investigated the influence of learning strategies in RALL environments (Balkibekov et al., 2016; Hong et al., 2016; Min et al., 2019; Wu et al., 2015). Some have explored the impact of integrating teaching strategies and the role of social robots in young language learners' learning (Alemi et al., 2015; Han et al., 2008; Lee et al., 2011; Tanaka & Matsuzoe, 2012). While many have focused on the social aspects of robot use, such as maintaining students' interest and emotional connection, there is a gap in the research specifically targeting the enhancement of cognitive abilities related to language skills. For example, the social robot's ability to retrieve information from individual student inquiries or to reciprocate students' utterances may affect students' productive language skills, such as their English speaking. However, such effects may not necessarily enhance their linguistic skills in more formal situations, such as public speaking or speeches.

Therefore, while the integration of social robots and learning strategies in RALL has been well documented in many ways, there remains a need for concrete recommendation to enhance English speaking skills through RALL. A promising approach, as suggested by Engwall and Lopes (2020), is the application of pedagogical theories and teaching strategies tailored to English language learning, especially in elementary schools.

For effective RALL, it is proposed that English language learners engage with social robots in various roles such as peer learners, opponents, and content creators. The use of robots should be grounded in pedagogical theories such as communicative and collaborative language teaching, facilitating interactive practices between students, robot, and teacher. Incorporating multiple low-stakes assessment tasks could further enhance the learning experience, making RALL an innovative and effective tool for language education.

Multiple low-stakes assessments

Low-stakes assessment refers to offering learning assessments or tasks to students without having any consequences for students' assessment results (see Wise & DeMars, 2003). Due to the extensive collection of low-stakes assessments in learning assessments, practitioners and researchers view it as a viable option for evaluating and assessing learning. Unlike high-stakes assessments, low-stakes assessments have minimal impact on students' final grades. This is exemplified in many studies in which a low-stakes assessment was investigated for its effectiveness in terms of guiding the learning process and enhancing learning performance (e.g., Schüttpelz-Brauns et al., 2020; Timpe-Laughlin et al., 2022; van der Lans et al., 2018).

Although promising, low-stakes assessments are not without their pitfalls. One such pitfall is the possibility of judging learners' status based on insufficient information from a one-time assessment. Traditionally, in educational settings, assessment was used as an evaluation tool for students' knowledge and to facilitate learning (see Kulik & Kulik, 1988). However, a one-time assessment might not have much more effect than the repeated assessment before administering the final test. One explanation behind this argument is that repeated assessment promotes active retrieval of information from memory and re-encoding knowledge or information, which is critical for long-term retention (Roediger III & Butler, 2011; Wiklund-Hörnqvist et al., 2014). This is called the testing effect; taking multiple tests usually enhances later performance. In the context of this study, taking multiple low-stakes assessments hypothetically has greater effects on students' oral presentation performance than a single low-stakes assessment.

One aspect to consider regarding low-stakes assessments is whether they differ from formative assessments. In terms of language learning, it is important to distinguish between the two types of assessment. While both share the common goal of supporting learning and providing feedback, their focuses and applications in language learning environments differ significantly (see Liu et al., 2015; Wise & DeMars, 2003). Firstly, formative assessments are primarily designed to enhance learning and offer continual feedback as part of a student's learning journey (Sadler, 1989). They are a key component in instructional decisions, aimed at helping students identify their strengths and areas for improvement (Schüttpelz-Brauns et al., 2020). In the language learning context, formative assessment could take the form of regular quizzes or writing exercises, where the feedback provided directly informs and shapes the ongoing teaching and learning process.

Low-stakes assessments, on the other hand, tend to have minimal or no impact on grading. They are not heavily weighted in determining a student's final grade, thereby creating a less stressful environment for learners (Liu et al., 2015). This approach is particularly beneficial in language learning, as it encourages students to focus on applying their skills and understanding the material without the pressure of high-stakes outcomes. For example, in this current study, several low-stakes assessments involve ungraded or minimally graded activities such as conversation circle activities or vocabulary exercises. The primary aim is to provide students with ample opportunities to practice and apply their language skills.

Furthermore, while formative assessment in language learning is often detailed and structured to provide comprehensive feedback for student improvement, low-stakes assessments may offer feedback, but they are not primarily designed to inform instructional decisions (Schut et al., 2020). Instead, they serve as a means for students to continuously engage with and apply what they have learned in a practical context.

Another key difference lies in the frequency and formality of these assessments. Low-stakes assessments are typically more frequent and informal, focusing on continuous learning and application. This is particularly relevant in language learning, where frequent, informal practice is crucial for skill acquisition and retention. Although low-stakes assessments are often student-centered, the role of the teacher in designing these assessments, providing feedback, and using the outcomes to informally track students' progress remains significant. The lesser formality of low-stakes assessments does not diminish the teacher's involvement; rather, it indicates a different approach to assessment where the emphasis is on the learning process rather than solely on the measurement of learning outcomes.

Another aspect to consider is whether to treat low-stakes assessment as a progress test or as feedback-oriented assessment. Progress tests aim at assessing the knowledge that students recently learned. Although they can identify students' strengths and weaknesses and can consequently make them more focused on their learning activities, this type of test may increase students' serious test-taking behavior as they might become anxious about taking the tests (Schüttpelz-Brauns et al., 2020). Meanwhile, treating low-stakes assessments as feedback-oriented assessments increases the mnemonic benefits of the testing. For instance, feedback that includes the correct answers can increase the learning effects due to enabling the students to correct their errors and maintain their correct responses (Pashler et al., 2005). This is why previous studies suggested feedback-oriented after-retrieval practices, which are commonly used to enhance students' memory retention by allowing them to practice what they have learned. Multiple low-stakes assessments would help to ensure that the learning treatment is successful in the future (e.g., Butler et al., 2013).

Classroom assessment techniques (CATs)

In 1988, Cross and Angelo coined the term CATs, defining them as a learning design approach to aid teachers in finding out what students have learned in the classroom and how well they are learning the given topic (Cross & Angelo, 1988). In school settings, CATs could be adopted in both the "learner-centered" and "teacher-directed" instruction modes for observing and improving students' learning performance based on the teachers' knowledge and experience to respond to the information gained through the CATs. In terms of fostering students' English productive language skills (e.g., speaking), it is generally a challenge to teachers and students, as some of the students need to acquire new English language words to complete the tasks. For instance, some students may feel overwhelmed by the teacher's instruction and responses if there is no division between students with upper- and lower-level skills. Consequently, students could not show their complete language abilities. Scholars have indicated that CATs not only enable teachers to know the learning status of individual students, but also help students examine their productive language development (Carduner, 2002).

Regarding classroom assessment, there is a growing conviction that classroom assessment resources are limited in scope. One problem is that, in conventional instruction and settings, students are typically, and historically, evaluated by a "summative" examination or other measurement methods (e.g., completing a final report). In these settings, students often only receive feedback at the end of a unit, lesson, or course. This implies that they have few opportunities to provide or receive feedback from the teacher and their peers during the process of achieving the learning goals (Mertler, 2016). Another problem is that the teachers often assess students' learning based on their understanding at the lower levels of thinking. For instance, instead of assessing the students' critical thinking or problem-solving abilities, teachers often measure students' ability of memorizing and comprehending (i.e., lower-level thinking) what the teachers have presented in the classroom. Addressing these issues, Cross and Angelo (1988) adopted Bloom's Taxonomy (Bloom, 1971) to classify the assessment skills and techniques in the classroom from simple to complex assessment. This brings us to the concept of multiple low-stakes assessments, which intersects with the theories of thinking skills and CATs. In this current study, we designed the multiple low-stakes tasks that draw on this intersection (see Fig. 1). These tasks are integral to providing a more nuanced understanding of student learning. Multiple low-stakes assessments offer frequent, informal evaluation of students' progress. This approach aligns well with CATs, as both aim to enhance learning through continuous feedback and adjustment.

In order to understand what and how CATs are considered effective for students, researchers have linked the CATs to important outcome variables of subject teaching such as in mathematics (e.g., Jawad et al., 2021) and foreign language teaching (Konstantakis et al., 2022). Because experimental evidence in CATs for foreign language teaching is still rare, the causality of these related studies might not be readily established. To some extent, recent studies mainly discussed how technology-based classroom assessment techniques could enhance students' motivation and interest in learning a foreign language (e.g., Bui & Nguyen, 2022). Whenever possible, based on previous studies' findings, we could indicate the connection between previous studies and this study by examining how CATs and social robots might potentially have a cause-and-effect relationship.

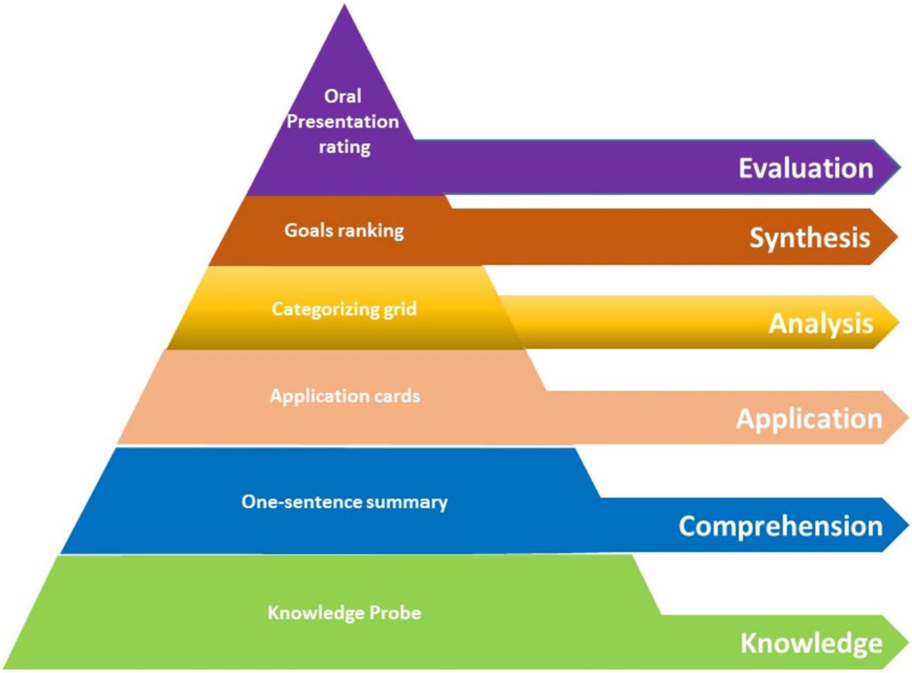

Fig. 1 The proposed multiple low-stakes assessments in this current study

The robot-based multiple low-stakes assessment (Robot-MLSA) learning environment

The Robot-MLSA learning environment involves tripartite elements in the learning process: teachers, learners, and robots. The teacher provided learning content about preserving a lake ecosystem as the learning context, utilized a learning database derived from a learning management system (i.e., Google Classroom), and implemented several CATs for this proposed learning environment. Meanwhile, Kebbi, the AI robot, was programmed to input learning content, accompany students during CATs, and was part of the learning database. On the other hand, the learners were set to provide user responses and to be content creators in some tasks of the Robot-MLSA learning activities. In comparison, other students in the non-Robot-MLSA learning group used Chromebooks for computer-assisted learning throughout the entire learning process. This learning group was named C-MLSA and was taken as the control group (see Fig. 2).

As a comparison, Table 1 shows the learning process in each CATs in the Robot-MLSA and C-MLSA learning environments. The Robot-MLSA group employed robots in various roles such as robot as a learner's opponent, robot as a teaching assistant, robot as a learning tool, robot as a peer learner, and robot as a peer presenter. Also, Chromebooks were provided for students to access some real-time learning apps, which provided an engaging and interactive learning experience for the students (e.g., Pear deck, Google Jamboard). On the contrary, the C-MLSA group was only offered real-time learning apps which could be accessed through Chromebooks or iPads.

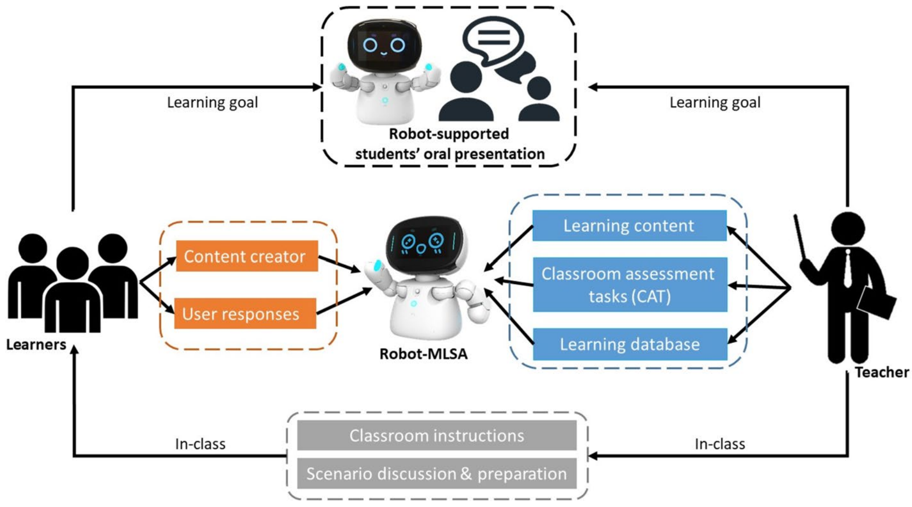

Fig. 2 The structure of the Robot-MLSA learning environment

This study proposes the employment of the Kebbi robot, a NUWA robotics company product (https://www.nuwarobotics.com/), for AI robot-assisted learning to facilitate students' learning activities and oral presentation performance. A few studies have employed the Kebbi robot in students' learning environments (e.g., Chen Hsieh & Lee, 2021; Liang & Hwang, 2023). As shown in Fig. 3, the robot has some interactive features such as an AI voice system and object recognition, customizable dialogue, vibrant body language with emotion, body movement, touch, and facial recognition. For this study, we programmed the robot using "Content editor" and "Quiz editor" features, allowing us to design, to personalize, and to deploy learning content so that all students received and completed the same tasks and assessments. The consistent deployment of uniform learning content across all CATs maintained a standardized pace of instruction for every student. See Fig. 4 for some learning scenes of CATs with robot employment.

Besides the teacher who was able to design robot-based learning content for every task, during the CATs, students were also able to use cloud-based development tools to create learning content for the robot's system. These tools offer text, picture, and video options for students to interact with both their peers and the robot itself. The teacher was able to observe students' content and interactions, correct any misconceptions, provide guidance on different aspects, and help answer more difficult language programming questions that had not yet been solved.

In class, students completed robot-based content tasks and reflected on and analyzed their work. For reflection, each group of students worked collaboratively to create their robot-based learning content. Then, they reflected on the robot's movement and responses by observing the effects of the deployment of the content on the robot.

For analysis, the teacher provided sample robot-based learning content for students to use as examples. Once the students were confident that their robot's program was close to the standard, they could deploy and present it in class. The teacher could review all the robot programs created by students in the back-end system, and provide feedback and guidance to each group via the system. Please see Fig. 5 for details.

Table 1 Learning process in the Robot-MLSA and C-MLSA groups

Classroom assessment techniques (CATs)	Implementation	Robot-MLSA group	C-MLSA group
Knowledge Probe	Skimming through answers to reveal important misconceptions or concepts that need to be reviewed.	*Robot as an opponent*: Each group of students queues to respond to the robot's questions (touch sensing)	*Real-time learning app*: The questions are shown on the screen, and the students answer them anonymously
One sentence summary	Each group responds to and refines the questions within the groups and presents a summary report to the class	*Robot as a teaching assistant*: The robot gives questions to each group. Students discuss some possible answers, then the leader responds to the robot by selecting the text	*Real-time learning app*: The questions are given through slides. Students drag the possible answer
Application cards	Students within the group discuss and design the range of applications for the principle or idea	*Robot as a tool*: Students create simple content in the robot's system for the application of principle or idea (biodiversity)	*Real-time learning app*: Students draw together for the actual application of the principle or idea (biodiversity)
Categorizing grid	Providing a list of techniques and asking the students to categorize them as "helpful" or "not helpful."	*Robot as a peer*: Students answer Yes/No statements as categorized by "helpful" and "not helpful."	*Real-time learning app*: The statements are shown on the slides, and students respond with Yes/No answers
Goals ranking	Students create a composite "level of challenge" rank ordering that reflects the best oral presentation performance	*Robot as a teaching assistant*: Robot provides follow-up language expression that students can use in their oral presentations (short voice response).	*Real-time learning app*: Slides of follow-up language expression that students can use in their oral presentation. (drag and drop activity)
Oral presentation rating	Each group rates other groups' performance and presentation content	*Robot as a peer presenter*: Student + peer + robot make oral presentations. Students rate other groups' performance by online form	Online form + teacher: Student + peer make oral presentations. Students rate other groups' performance by online form

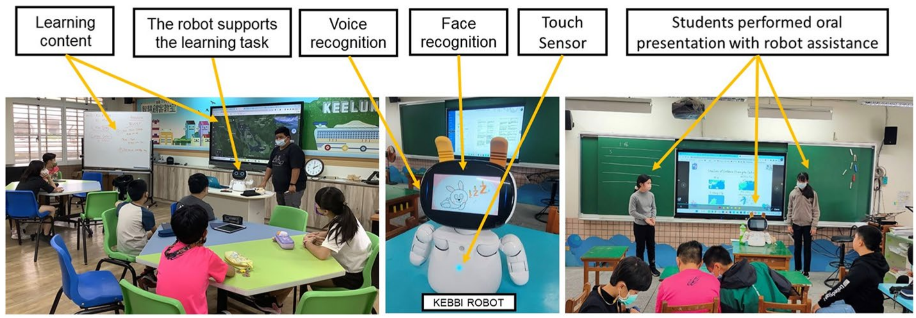

Fig. 3 Some learning scenes in the Robot-MLSA learning environment

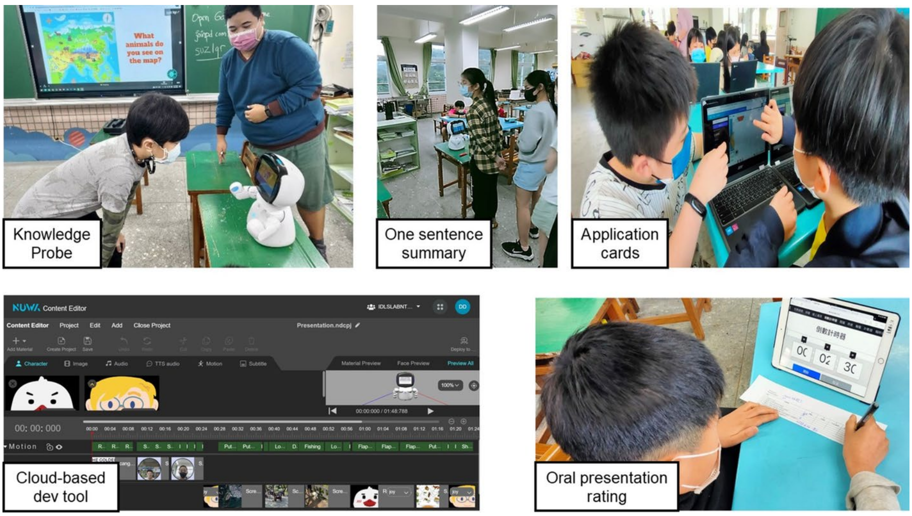

Fig. 4 Learning scenes of the proposed classroom assessment techniques (CATs)

Experimental design

To investigate the effects of the Robot-MLSA learning environment, this study collected quantitative data to analyze students' oral presentation performance, collective efficacy and learning attitude, and qualitative data of the students' reflective writing about their learning experience and perceptions.

Participants

A quasi-experiment was employed for the sixth-grade students ($N = 58$) in an elementary school in northern Taiwan. Students from one class were assigned to the Robot-MLSA group ($n = 28$, Female = 12, Male = 16), while students from another class were assigned to the C-MLSA group ($n = 30$, Female = 14, Male = 16). Based on the previous term's

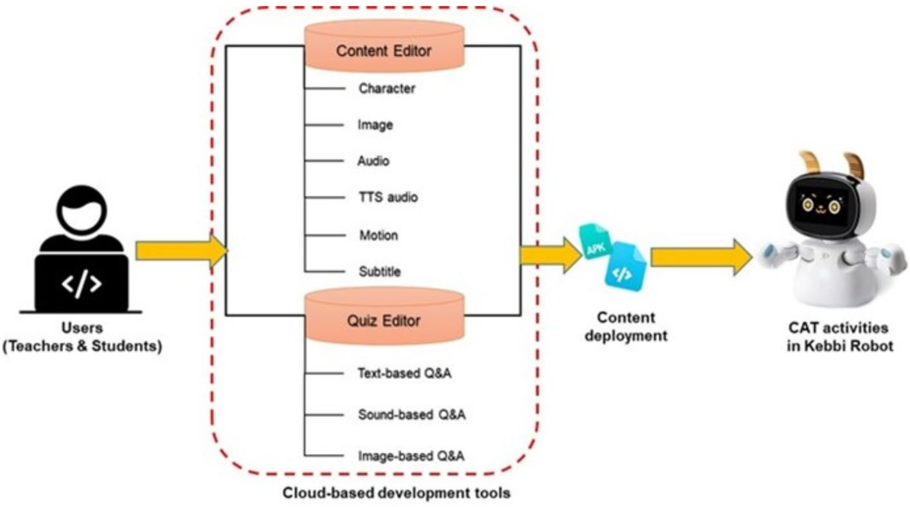

Fig. 5 The system structure deployment of Robot-MLSA

report in both classes, students' English proficiency levels varied from the pre-A1 to the A2 level based on the Common European Framework of Reference for Language (CEFR). Prior to conducting the study, students were familiar with using Chromebook laptops for learning activities. However, none had previous experience of learning with physical robots such as Kebbi Robot. All students were taught by two instructors who had taught natural science for over 15 years and English as a Foreign Language (EFL) for over 10 years, respectively.

Experimental procedure

Figure 6 shows the experimental procedure from the first to the eighth week. All students spent the first week having the learning environment orientation. The Robot-MLSA group was introduced with Kebbi Robot and some real-time learning apps (Canva and Pear Deck) that they used during the classroom activities. Meanwhile, students in the C-MLSA group were introduced to the same online platforms, excluding Kebbi Robot.

In the first and second weeks, both groups were taught about another lake's ecosystem compared to the following lesson about the local lake near their school. Students were given an orientation of the learning environment and learning systems. In the third to sixth weeks, students in the Robot-MLSA group performed CAT activities with Kebbi Robot and Chromebook laptops as learning tools.

In contrast, students in the C-MLSA group performed CAT activities using the Chromebook laptop only. In the seventh week, students in both groups were given the questionnaires of learning attitude and collective efficacy. Moreover, the students were invited to write their learning reflections to elicit their perceptions of the proposed learning environment. However, only a few of them ($n=10$) were willing to write and submit their learning reflections by the end of Week 7. In the eighth week, students from both groups performed their oral presentations. Their presentation materials were collected, and their presentation was recorded as the posttest of the study.

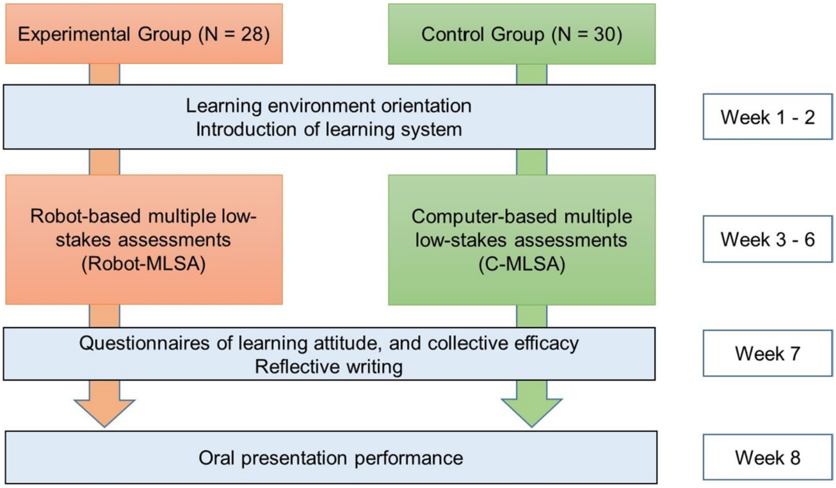

Fig. 6 Experimental procedure

Measuring tools

The measuring tools employed in the present study consisted of the scoring rubrics for oral presentation performance, a questionnaire of collective efficacy, a questionnaire of learning attitude, as well as reflective writing.

The scoring rubrics for oral presentation performance were modified from Mercer et al. (2017). The framework has been used by Stordeur et al. (2022) to assess students' oral presentations in elementary schools. As shown in Table 2, the framework defines the general skills, relevant skills, and some particular tasks for students when performing oral presentations. It consists of five general skills, namely voice, presentation content, interaction with audience, body language and gesture, and linguistic skill. Each relevant skill has particular tasks which need to be fulfilled by students in their oral presentation performance. The relevant skills are rated from 1 to 4 points, with 1 being the lowest and 4 being the highest. Two experienced teachers (an English language and a science teacher) rated the students' oral presentation performance, and the rating result was collected as the posttest data. The Kendall's ω values of the grading from both groups of the five criteria were 0.81, 0.89, 0.83, 0.83, and 0.92 ($p < .001$), respectively. This indicated that the grading of the two teachers was significantly correlated.

The collective efficacy survey was adopted from Wang and Lin (2007) based on the questionnaire validated by Pintrich et al. (1993). It consists of eight items with a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*), for example, "I believe that our group can achieve a superior outcome for this learning task" and "I believe that our team members can learn the basic concepts instructed in the learning task." The Cronbach's alpha was .91.

Learning attitude was based on the development and validation of the attitude scale by Sisman et al. (2019). The questionnaire consists of 17 items divided into four sections, namely engagement (5 items), intention (4 items), enjoyment (4 items), and anxiety (4 items), with a 5-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly*

Table 2 Scoring rubrics of students' oral presentation

Criteria	4	3	2	1
Voice	Student speaks clearly and confidently with appropriate pace, pitch, and volume	Student speaks mostly clearly and confidently, but may have some issues with pace, pitch, or volume	Student speaks with some clarity and confidence, but has noticeable issues with pace, pitch, or volume	Student speaks unclearly or lacks confidence and may have significant issues with pace, pitch, or volume
Presentation content	Student presents material that is well-organized, informative, and engaging	Student presents material that is mostly well-organized, informative, and engaging, but may have some minor issues in one or more areas	Student presents material that is somewhat organized and informative, but lacks engagement or has noticeable issues in one or more areas	Student presents material that is poorly organized, lacks information, and is not engaging
Interaction with audience	Student engages the audience, and maintains consistent eye contact or uses appropriate gestures or questions	Student engages the audience, but may not maintain consistent eye contact or use appropriate gestures or questions	Student attempts to engage the audience, but lacks consistency or appropriateness in eye contact, gestures, or questions	Student does not attempt to engage the audience or lacks appropriateness in eye contact, gestures, or questions
Body language or gesture	Student uses appropriate body language and gestures to enhance the presentation	Student uses mostly appropriate body language and gestures, but may have some minor issues in one or more areas	Student uses some appropriate body language and gestures, but lacks consistency or appropriateness in one or more areas	Student does not use appropriate body language or gestures and may be distracting or inappropriate
Linguistic skill	Student demonstrates a strong command of language, using proper grammar and vocabulary	Student demonstrates some proficiency in language, but may have some errors in grammar or vocabulary	Student demonstrates limited proficiency in language, with frequent errors in grammar and vocabulary	Student demonstrates poor proficiency in language, with frequent errors in grammar and vocabulary.

agree). The Cronbach's alpha values for each questionnaire were 0.850 for engagement, 0.808 for enjoyment, 0.815 for anxiety, and 0.830 for intention.

In terms of the learning experience reflection, students were asked to answer the questions, "What do you think you obtained the most when learning in your class using this approach? What part did you like and learn the most from?". In terms of their final oral presentation, students were asked, "In what aspects have you done well and what still needs to be improved?". The outline of the questions referred to Hwang et al. (2009). The students' reflective writing was examined by the two researchers. Their reflective writing was coded based on a thematic analysis approach to further reveal students' perceptions of the proposed learning environment. Students' writing was translated into English. This study employed ATLAS.ti 9, computer-assisted qualitative data analysis software (CAQDAS), for the coding process. We used grounded theory to identify each theme (Creswell, 2016). We began by coding the transcriptions. We carefully reviewed each code to ensure that we reached a consensus. Next, we grouped the codes into thematic code groups. We performed an analysis of the code-document table and counted the codes in absolute frequencies, which allowed us to establish the groundedness of each code. Finally, we used the coded quotations to demonstrate the number of occurrences in the reflective writing.

Results

Oral presentation performance

Regarding the grading results, we employed the Mann–Whitney *U* test, a non-parametric statistical analysis, to examine the results due to the small sample size of this study. In addition, the Shapiro–Wilk test shows that our sample had a non-normal distribution. Table 3 shows that the results of the Mann–Whitney test indicated that the Robot-MLSA group had better oral presentation performance than the students in every criterion.

In terms of voice, students in the Robot-MLSA group (Mdn = 3) scored higher than those in the C-MLSA group (Mdn = 3) with U = 268.00, z = − 2.671, $p = .008$, and $r = .35$. Regarding presentation content, students in the Robot-MLSA group (Mdn = 4) scored higher than those in the C-MLSA group (Mdn = 3) with U = 313.00, z = − 1.9, $p = .05$,

Table 3 The Mann–Whitney *U* test result for the English oral presentation of the two groups

Criteria	Group	*N*	Mean-Rank	Sum of Ranks	*U*	*z*	*r*
Voice	Robot-MLSA	28	34.93	978.00	268.00	− 2.671**	0.35
	C-MLSA	30	24.43	733.00			
Presentation content	Robot-MSLA	28	33.32	933.00	313.00	− 1.9*	0.25
	C-MLSA	30	25.93	778.00			
Interaction with audience	Robot-MSLA	28	37.21	1042.00	204.00	− 3.901***	0.51
	C-MLSA	30	22.30	669.00			
Gesture	Robot-MLSA	28	37.43	1048.00	198.00	− 4.041***	0.53
	C-MLSA	30	22.10	663.00			
Linguistic skill	Robot-MLSA	28	37.96	979.00	267.00	− 2.745**	0.36
	C-MLSA	30	24.40	732.00			

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

and $r = .25$. In terms of interaction with the audience, students in the Robot-MLSA group (Mdn = 3) scored higher than those in the C-MLSA group (Mdn = 2) with U = 204.00, z = − 3.901, $p = .001$, and $r = .51$. For body language or gesture, students in the Robot-MLSA group (Mdn = 3) scored higher than those in the C-MLSA group (Mdn = 2) with U = 198.00, z = − 4.041, $p = .001$, and $r = .53$. Lastly, in terms of linguistic skill, students in the Robot-MLSA group (Mdn = 3) scored higher than those in the C-MLSA group (Mdn = 2) with U = 267.00, z = − 2.745, $p = .006$, and $r = .36$.

These findings imply that students who adopted the Robot-based multiple low-stakes assessment learning environment had better oral presentation performance, measured in every criterion, than those who adopted the computer-based multiple low-stakes assessment learning environment.

Students' collective efficacy

We utilized the Mann–Whitney *U* test, a non-parametric statistical analysis, to examine the results of the survey on collective efficacy due to the small sample size of this study. Additionally, the Shapiro–Wilk test indicated that our sample did not have a normal distribution.

Table 4 reveals that the Robot-MLSA group (Mdn = 4.8) had better collective efficacy than the C-MLSA group (Mdn = 4.1) with U = 262.00, z = − 2.515, $p = .012$, and $r = .33$. These findings suggest that students who used the Robot-based multiple low-stakes assessment learning environment had better collective efficacy than those who used the computer-based multiple low-stakes assessment learning environment.

Students' learning attitude

We conducted the Mann–Whitney *U* test to analyze the learning attitude survey. We chose non-parametric statistical analysis due to our small sample size and non-normal distribution, as determined by the Shapiro–Wilk test.

Table 5 shows that the Robot-MLSA group (Mdn = 5) had higher engagement scores than the C-MLSA group (Mdn = 4.3), with U = 209.00, z = −3.410, $p = .01$, and $r = .45$. In terms of intention, the Robot-MLSA group (Mdn = 5) had higher intention than the C-MLSA group (Mdn = 4.4), with U = 304.50, z = −1.898, $p = .05$, and $r = .25$. Additionally, the Robot-MLSA group (Mdn = 5) had higher enjoyment than the C-MLSA group (Mdn = 4) with U = 175.50, z = −3.925.761, $p = .001$, and $r = .51$. These findings suggest that students who used the robot-based multiple low-stakes assessment learning environment had better engagement, intention, and enjoyment in their learning attitude compared to those who used the computer-based multiple low-stakes assessment learning environment.

Table 4 The Mann–Whitney *U* test result for the students' collective efficacy of the two groups

Group	*N*	Mean-Rank	Sum of Ranks	*U*	*z*	*r*
Robot-MLSA	28	35.14	984.00	262.00	− 2.515*	0.33
C-MLSA	30	24.23	727.00			

*$p < .05$

Table 5 The Mann–Whitney *U* test result for the students' learning attitude of the two groups

Variable	Group	*N*	Mean rank	Sum of ranks	*U*	*z*	*r*
Engagement	Robot-MLSA	28	37.04	1037.00	209.00	– 3.410**	0.45
	C-MLSA	30	33.47	941.50			
Intention	Robot-MLSA	28	33.63	768.50	304.50	– 1.898*	0.25
	C-MLSA	30	25.65	428.50			
Enjoyment	Robot-MLSA	28	38.23	1070.50	175.50	– 3.925***	0.51
	C-MLSA	30	21.35	640.50			
Anxiety	Robot-MLSA	28	23.45	656.50	250.50	– 2.753**	0.36
	C-MLSA	30	35.45	1054.50			

*$p < .05$, **$p < .01$, ***$p < .001$

Furthermore, the Robot-MLSA group (Mdn = 1) had lower anxiety than the C-MLSA group (Mdn = 1.9), with U = 250.50, z = −2.753, $p = .06$, and $r = .36$. This indicates that students who used the Robot-MLSA experienced less anxiety in their learning attitude compared to those who used the computer-MLSA learning environment.

Analysis of students' learning experience

At the end of the learning treatment, five students from the Robot-MSLA group (coded RS021, RS005, RS012, RS017, and RS008) and five students from the C-MLSA group (coded CS015, CS003, CS010, CS019, and CS024) were willing to share their learning experience through reflective writing. Table 6 shows the analysis of the reflective writing results derived from five themes: benefits of technology-supported learning, enhancing oral presentation performance, promoting language learning, increasing learning satisfaction, and supporting collaborative learning. Each theme is discussed in the following sub-sections. The findings are shown in Table 6.

Students' learning experience and perceptions of the Robot-MLSA group

Students in the Robot-MLSA group had positive perceptions of the learning environment. They enjoyed using various technologies in and out of the classroom, making learning more engaging and interactive. For example, they appreciated the opportunity to use and explore the robot's functions as part of their learning process, specifically in the "application cards" where students were given the opportunity to create simple content using the robot system. The robot was seen as a meaningful learning tool that allowed students to experiment with its capabilities and provided them with opportunities for hands-on learning. As RS021 shared, "I like that we used many technologies in and outside the class. It made learning fun and interesting." However, while students enjoyed working with the robots, many were concerned about the limited time to explore their capabilities during the task activities. RS008 wanted more time to work with the robot, saying, "We only had limited time working with the robot. I want more time to interact with the robot."

Similarly, RS005 echoed this concern, saying, "I expected to use the robot frequently because it was only used during the first hour of the learning process in the class. We did not use it anymore until the end of the class." The primary objective of this study was to utilize the robots not just for entertainment purposes. In order to achieve this, the

Table 6 Qualitative analysis of the students' learning experience and perceptions

Theme	Code	The number of occurrences	
		Robot-MLSA	C-MLSA
Benefits of technology-supported learning	Perception of the proposed learning environment	3	0
	Technology acceptance	9	5
	Perception of robots	15	0
	Benefits of technology-supported low-stakes assessments	5	4
	Technology issues	1	1
Enhancing oral presentation performance	Benefits of oral presentation	11	7
	Increasing creative thinking	11	0
	Oral presentation issues	0	2
Promoting language learning	Providing bilingual learning	7	4
	Bilingual learning instruction	7	8
	Teacher-students interaction	2	2
Increasing learning satisfaction	Increasing learning interest	14	4
	Increasing learning engagement	3	3
	Introducing new learning experiences and instruction	19	7
	Perception of the conventional way of learning	3	4
Supporting collaborative learning	Facilitating thinking skills	4	0
	Collaborative tasks in the low-stakes assessment	5	4
	Benefits of collaborative work	0	2
	Collaborative learning issues	1	2

learning strategies in this current study restricted students to only interacting with the robots during the low-stakes assessments. For instance, in the "knowledge probe," students engaged in robot interaction through a game challenge to compete with each other. This activity was not conducted throughout the entire learning session as we intended it to help students recall the subject matter taught in the previous session. Furthermore, using robots to support multiple low-stakes assessments varies and depends on the types of assessment. This indicates the need for a more tailored approach to incorporating robots into learning.

Despite these concerns, students praised the use of technology to support learning in general. For example, RS012 noted, "I think the biggest difference is the use of many technologies in this classroom." RS005 added, "It was fun when we used various tools such as robots, CANVA (an online graphic design platform), laptops, and many more." Using these tools and technologies provided students with a more diverse set of learning experiences and helped keep them engaged in the learning process. However, some students complained about the internet speed, which sometimes disrupted their ability to use the technology effectively. As RS017 noted, "Sometimes the internet was not working at all." This highlights the need for reliable and fast internet access to fully leverage technology's benefits in the learning process.

In terms of the benefits of using R-MLSA to enhance students' oral presentation performance, several advantages were noted by the students. RS005, for example, mentioned that they learned how to design a presentation and speak English and Mandarin in front of the class. Additionally, students expressed excitement about using the robot to assist them in their oral presentations. RS012 stated that creating the robot's speech and making slides for the presentation was so much fun (see Application cards in Table 1). While some students may have initially found performing oral presentations challenging, they were determined to improve their performance for their final task. For example, in the "goals ranking," students had the opportunity to practice their final oral presentation multiple times with the robot. The robot was able to provide immediate feedback on their language expression. It was found that the use of R-MLSA technology has the potential to increase student engagement and motivation, which can ultimately lead to improved learning outcomes.

Regarding language learning, students who participated in the Robot-MLSA program shared their experiences of learning English. A tripartite interaction between students, teachers, and robots allowed them to express their experiences. For example, RS021 described their interaction with the teacher in English and said, "We went to the lake and observed animal activities. Also, we discussed with the teacher in English while taking some pictures for observation." RS005 also mentioned the benefits of interacting with the robot, stating, "The robot can speak with us in the class, and the good thing is we can make the robot speak whatever we want (by creating the text-to-speech in the robot's system)."

Some of the students were highly interested in learning the English language. For instance, RS021 mentioned, "I want to continue having classes like this because I enjoy learning English and want to practice speaking more." The R-MLSA learning environment was designed with the intention of providing bilingual instruction in both English and Mandarin Chinese. This means that not only the teacher, but also the robot, was capable of communicating with the students in either English or Mandarin Chinese. With bilingual instruction, elementary school students may be able to increase their interaction and communication during learning, especially when considering their limited English proficiency and language anxiety when speaking English. By providing more opportunities for students to practice speaking and expressing themselves in English, the Robot-MLSA program may help them develop their language skills and boost their confidence in speaking English.

Regarding learning satisfaction, students in the R-MLSA group enjoyed interacting with the robot. They found the experience stimulating and engaging, and it was evident from their comments that they were enthusiastic about integrating robots and other technologies into their learning tasks. For example, RS017 stated, “I really enjoy learning with the robot. I used the laptop to create the robot’s movements and speech. It is exciting.” RS008 added, “Playing with the robot was my favorite activity. It was also fun to do online quizzes and online drawing.” The students considered interacting with the robot during the CATs, such as “knowledge probe,” “one sentence summary,” or “application cards” activities, which provided them with ample opportunities to engage with the robot. These comments indicate that the students were learning and enjoying themselves. It is also worth noting that the students were eager to extend their learning beyond the classroom. They took the opportunity to conduct a low-stakes assessment task outside of the classroom. For instance, in this current study, students were enthusiastic about visiting the lake and carrying out observations. RS008 even went so far as to say, “We originally learned about biodiversity, but we could also have fun with technology and go outside of the classroom to observe.” This comment highlights the value of integrating technology into learning tasks, as it allows students to take their learning outside of the classroom and into the real world.

Furthermore, the students compared their learning experience with the conventional learning they had previously attended. RS201 explained, “In other classes that adopted conventional learning, we usually just stayed in the classroom and studied from a book. However, we could go out and have fun in this class.” Before the “application cards” activity, students had the chance to visit a nearby lake. The objective was to observe the biodiversity in the lake and include it in their presentation. In the R-MLSA group, students collected their observation findings and input them into the robot system themselves. This suggests that the use of technology and robots in learning tasks has the potential to make learning more engaging and enjoyable for students. These comments demonstrate that integrating robots and other technologies into learning tasks can lead to greater learning satisfaction and a willingness to extend learning beyond the classroom.

The Robot-MLSA learning environment is very effective in terms of encouraging collaborative learning among students. According to RS012, the students did not have to take any tests, but they spent time practicing speaking with teachers and robots and creating slides for oral presentations. This is the advantage of using low-stakes assessment, as students do not experience the same level of anxiety as they do with high-stakes assessment. However, despite the benefits of collaborative learning, this approach can sometimes lead to issues in the learning process. For example, as RS017 reported, some students may not be willing to work with others, which can create a lack of peer support and motivation.

C-MLSA group students’ learning experience and perceptions

The students in the C-MLSA group positively perceived the use of technology in the classroom. They appreciated how technology facilitated their learning by providing interactive and engaging experiences. For example, CS015 expressed excitement about using the online platform CANVA for designing the final presentation, stating, “I know that making the presentation is so hard, but using CANVA is good.” They found the platform user-friendly, which enhanced their creativity and helped them present their ideas in a more organized way. However, the use of technology also has its drawbacks. CS010 mentioned that the teacher instructed them to use Chromebooks for a long period of time. This led to discomfort and eyestrain, negatively impacting their learning and overall experience in the classroom.

Regarding the oral presentation performance, the students shared, “We worked with friends to make presentations and speak in front of the class. It was difficult, but we practiced to make it better.” (CS003). In addition, CS003 mentioned the teacher’s role of supporting their oral presentations: “We learned how to make the presentation, and the teacher was accommodating. They gave examples of making slides and their content, and we also added the slides with some pictures and quotations in our presentation.” CS010 further stated, “I learned to design a presentation and speak in front of the class. I know it was hard, but it was okay since my peers and I encouraged each other.” Unlike the Robot-MLSA group, the C-MLSA group utilized an online slide presentation platform to create their slides and demonstrate practical applications of the subject matter they had learned.

However, despite the students’ positive responses, the oral presentation also had drawbacks for the students. CS015 stated, “Making/designing a presentation was too hard. We had to include many features in our presentation.” Moreover, CS024 added, “I was expecting to learn more about the English language, but we were too focused on making the presentation.” Students’ negative perspectives suggest covering language learning more widely than just making an oral presentation.

Regarding language learning, students in this group had the opportunity to engage in bilingual language transactions with their student-teachers. During their interactions, the students learned many new things about the lake’s ecosystem and how to use English words in class and during oral presentations. The experience of using two languages in class (Mandarin and English) helped them to improve their language skills and perform oral presentations in both languages. CS024 expressed their interest in speaking more English in class, not just writing, which is typically the focus of English language courses. The students appreciated the benefit of having the opportunity to practice and learn in both languages, as they noted that it helped them retain and recall the English vocabulary they had learned previously.

Regarding learning satisfaction, students appreciated the opportunity to conduct a low-stakes assessment task collaboratively outside the classroom. For instance, CS010 commented, “I really like going out of the class, going to the lake, and observing. Maybe we need more similar activities in other classes.” This suggests that students would benefit from more outdoor activities in their learning.

However, the collaborative nature of the tasks raised concerns about peer attitudes. CS019 mentioned, “My friends sometimes did not want to help me or just made jokes about our work, such as deleting the slides. It was annoying.” This highlights a potential issue in collaborative tasks where group members may not be as invested in the work as others. To solve this problem, teachers could focus on group dynamics and offer training to students on how to work collaboratively. SC003 suggested, “I might prefer working with another friend. I was always put in the same group. I felt it would be better to work with my best friends.” This shows that students value the opportunity to work with others they are comfortable with. Teachers should consider students’ preferences when forming groups for collaborative tasks.

Despite the issue of peer attitudes, students appreciated the learning tasks offered by the teachers. For example, CS024 stated, “The teachers gave us many interesting activities such as writing games, online quizzes, and drawing. This is interesting because we did not normally do this in our previous classes.” This shows that students value a variety of learning activities and appreciate teachers who offer creative and engaging tasks.

Conclusion and suggestions

This study introduced the Robot-MLSA learning environment, which was shown to be effective in terms of enhancing students' oral presentation performance, learning attitude, and collective efficacy. The Robot-MLSA learning environment was compared to the C-MLSA learning environment, and showed better results than its counterpart approach.

Moreover, the students who experienced the Robot-MLSA learning environment expressed their agreement that including robots in the learning process brought them several advantages while completing the learning tasks. Compared to the C-MLSA learning environment, where the students were given various technology-supported learning tasks, the use of social robots in RALL environments has been shown to increase learning satisfaction, as students demonstrated a better learning experience in comparison to the conventional way of learning.

In terms of oral presentation performance, there is evidence to suggest that incorporating robots into the learning process can have a positive impact on students' scores in various categories such as voice, presentation content, interaction with the audience, gestures, and linguistic skills. This aligns with the capabilities of RALL. As noted by Iio et al. (2019) and Engwall et al. (2022), robots have been shown to enhance students' oral and speaking skills. They can offer interactive practice, real-time feedback, and a non-threatening audience, which are crucial for fostering public speaking skills. This could potentially be useful in helping students improve their overall performance in oral presentations by integrating the robot into multiple low-stakes assessments of learning tasks.

Furthermore, we argue that given the complexity of oral presentations, it is important for elementary school students in particular to master a series of knowledge, presentation skills, and interpersonal skills. To achieve this, multiple low-stakes assessments can be employed to assist students in preparing, practicing, and rehearsing for oral presentations. For example, students can be trained to speak in front of the class during learning tasks such as knowledge probes and one-sentence summaries, which support the acquisition of complex skills under limited teacher guidance.

Previous studies have highlighted the benefits of using the robot's automated responses to facilitate effective oral practice, particularly through low-stakes assessment (Timpe-Laughlin et al., 2022; van Doremalen et al., 2016). By integrating the robot and MLSA (multiple low-stakes assessments), a socio-cognitive perspective (Bandura, 1986) can be adopted, emphasizing activities such as observation, practice, and feedback to support students' oral presentation skills.

The positive effects of the Robot-MLSA learning environment extend not only to students' academic performance, but also to their overall learning attitudes. In comparison to the more "conventional" C-MLSA learning environment, the Robot-MLSA environment has been found to result in better learning engagement, intention, enjoyment, and reduced learning anxiety among elementary school students.

For example, previous studies have shown that young learners perceive robots as playful and confidence-boosting (e.g., Hsu & Liang, 2021; Liu, 2010; Socratous & Ioannou, 2022). This suggests that incorporating robots into the learning environment can help to create a positive and supportive atmosphere for students. Other studies have also found that using robots in the classroom promotes positive attitudes among students (Serholt et al., 2014; Tung, 2016). By providing students with the opportunity to interact with robots, they may feel more motivated and enthusiastic about learning.

While it is true that some studies suggest a negative attitude towards robots (e.g., Nomura et al., 2006), and have pointed out that over-reliance on RALL could lead to a reduction in human-to-human interaction, we believe that incorporating a moderate level of human features and emotional signals into robot systems can transform students' learning attitudes into higher levels of interaction and engagement. This means that robots can be used as a powerful tool to help students develop learning interest and to foster a positive attitude towards the proposed learning environment.

In this current study, students were expected to engage in many collaborative tasks. The goal was to measure the collective efficacy of students in both groups. The students who learned in the Robot-MLSA group showed higher collective efficacy than those who learned in the C-MLSA group. These findings support previous studies which found that students' collective efficacy in Robot-MLSA groups was better than that of students in control groups. The inclusion of social robots has been shown to positively influence students' learning attitude and collective efficacy. In RALL, this can be attributed to the interactive nature of social robots, which can make learning more engaging and less intimidating, thus encouraging participation and collaborative learning. Moreover, this is due to various reasons such as the higher promotion of creative thinking skills (Yang et al., 2022), higher perceptions of computational thinking skill efficacy (Yang et al., 2023), and the belief that robots are an effective medium to improve collective efficacy (Jaipal-Jamani & Angeli, 2017). Creative thinking skills are particularly important in designing presentation content and creating the robot's movements and sounds. This is because students in the Robot-MLSA group had to think outside the box to come up with innovative ways to make the robot move and sound interesting. They had to be creative in order to capture the attention of their audience.

In terms of findings on the students' reflective writing, both groups agreed that multiple low-stakes assessments were a great and innovative learning experience. Not only did these assessments provide students with a chance to practice their writing skills, they also allowed them to receive feedback on their progress throughout the course. Additionally, students in both groups appreciated the use of various technologies to assist them in dealing with learning tasks that varied in every session. For example, students enjoyed using online tools to collaborate with their peers and complete the assignments. They also found it helpful when teachers used multimedia resources to supplement their lectures. This finding is supported by Schut et al. (2020), who found that students perceived low-stakes assessment as a learning opportunity because of the practical assessment relationship between students and teachers. This relationship is friendly and collaborative. This is when teachers do not show dominance, but rather exhibit friendly behavior. In other words, when teachers support students in a collaborative and non-judgmental way, students are more able to engage with the learning process and feel motivated to improve. In terms of collaborative work among students, it can be particularly difficult for students who struggle or may feel uncomfortable speaking in front of others. As such, teachers need to be aware of these potential issues and be prepared to address them to ensure that all students can benefit from the collaborative learning environment. This might involve providing additional support and guidance to students who are struggling.

Additionally, students in both groups had positive perceptions of using bilingual instruction during class. However, students in the Robot-MLSA group had a higher level of interest in learning than those in the C-MLSA group. This finding may be due to the fact that the use of a robot to support low-stakes assessment can increase learning interest. This argument is supported by Chang et al. (2023), who found that robots in learning environments can enhance students' motivation by increasing their interest in the learning process and raising their curiosity. Furthermore, the use of a bilingual approach may have helped students to feel more engaged with the material and to connect with their classmates who spoke different languages. This is particularly important in a diverse classroom where students may come from different cultural backgrounds.

Moreover, while this study provides valuable insights, it is important to note that there are limitations and challenges that should be taken into account.

Firstly, the limited time students had practicing their English presentation posed challenges in achieving sufficient length and complexity in their spoken English. This might affect the assessment and feedback regarding their language proficiency, subsequently impacting their learning outcomes in the biodiversity topic. For instance, the analysis of students' lexical complexity would have required more than 300 overall word productions (as suggested by McCarthy & Jarvis, 2010), which was not feasible in this current study. However, this limitation highlights an area for further investigation.

Secondly, the robot was only used for low-stakes assessment, which means that students did not always interact with the robot during the learning sessions. As such, it would be beneficial to conduct further studies to explore the extent to which the robot can be used as a peer in whole learning sessions and the impact this has on students' learning opportunities and interactions.

Lastly, while this study had a limited number of participants in each group, there is still scope for further investigation into students' learning behaviors when performing robot-student interactions. This could help to gain a more precise understanding of how social robots can be used to improve students' learning performance in the classroom, and could inform future research in this area.

The integration of social robots into learning tasks, as studied here, presents a wealth of opportunities for educators seeking to maximize student engagement and learning outcomes. Our research has uncovered several key findings and recommendations for future research in this area, which we present below:

(1) In light of the robot's ability to support multiple low-stakes assessments, future studies may wish to consider designing a range of different assessment types to be used in conjunction with social robots, in order to maximize the potential of this technology in the classroom.
(2) However, it is important to note that collecting and managing student assessment results through a robot database may significantly affect teachers' workload. Therefore, it is recommended that future research should focus on developing more flexible and user-friendly interfaces for teachers to interact with the robot's system. While our research has demonstrated promising results for the implementation of social robots in the context of MLSA learning, further investigation is needed to explore the impact of this technology on learners with different backgrounds and learning styles. In particular, future studies should aim to identify the most effective strategies for tailoring the robot's approach to individual learners, so as to ensure that this technology can truly overcome the limitations of conventional learning methods.
(3) Finally, it is worth noting that the integration of social robots into learning environments represents a significant departure from conventional teaching methods, and as such requires careful consideration and planning. We therefore recommend that educators interested in exploring this technology should engage in ongoing professional development and seek out opportunities for collaboration and knowledge-sharing with other practitioners in the field.

Acknowledgements This study is supported in part by the National Science and Technology Council of Taiwan under contract numbers NSTC 112-2410-H-011-012-MY3 and MOST 111-2410-H-011 -007 -MY3. The study is also supported by the "Empower Vocational Education Research Center" of National Taiwan University of Science and Technology (NTUST) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. The authors would also like to thank Ms. Yu Chun Lin for her technical support during the learning treatment of this study.

Declarations

Conflict of interest The authors would like to declare that there is no conflict of interest in this study.

Consent to participate The participants were protected by hiding their personal information during the research process. They knew that the participation was voluntary and they could withdraw from the study at any time.

References

Al Hakim, V. G., Yang, S. H., Liyanawatta, M., Wang, J. H., & Chen, G. D. (2022). Robots in situated learning classrooms with immediate feedback mechanisms to improve students' learning performance. *Computers & Education*, *182*, 104483. https://doi.org/10.1016/j.compedu.2022.104483.

Alemi, M., Meghdari, A., & Ghazisaedy, M. (2015). The impact of social robotics on L2 learners' anxiety and attitude in english vocabulary acquisition. *International Journal of Social Robotics, 7*(4), 523–535. https://doi.org/10.1007/s12369-015-0286-y

Balkibekov, K., Meiirbekov, S., Tazhigaliyeva, N., & Sandygulova, A. (2016). Should robots win or lose? Robot's losing playing strategy positively affects child learning. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 706–711). IEEE. https://doi.org/10.1109/ROMAN.2016.7745196.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bloom, B. S. (1971). *Handbook on formative and summative evaluation of student learning*. McGraw-Hill.

Bui, H. P., & Nguyen, T. T. T. (2022). Classroom assessment and learning motivation: insights from secondary school EFL classrooms. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2022-0020

Burke, L. A., & Williams, J. M. (2012). Two thinking skills assessment approaches: Assessment of pupils' thinking skills and individual thinking skills assessments. *Thinking Skills and Creativity, 7*(1), 62–68. https://doi.org/10.1016/j.tsc.2011.11.002

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, *105*(2), 290. https://doi.org/10.1037/a0031026.

Carduner, J. (2002). Using classroom assessment techniques to improve foreign language composition courses. *Foreign Language Annals*, *35*(5), 543–553. https://doi.org/10.1111/j.1944-9720.2002.tb02722.x.

Chang, C. C., Hwang, G. J., & Chen, K. F. (2023). Fostering professional trainers with robot-based digital storytelling: A brainstorming, selection, forming and evaluation model for training guidance. *Computers & Education*. https://doi.org/10.1016/j.compedu.2023.104834

Chen Hsieh, J., & Lee, J. S. (2021). Digital storytelling outcomes, emotions, grit, and perceptions among EFL middle school learners: Robot-assisted versus powerpoint-assisted presentations. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2021.1969410

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage.

Cross, K. P., & Angelo, T. A. (1988). *Classroom Assessment Techniques: A Handbook for Faculty*. The National Center for Research to Improve Postsecondary Teaching and Learning, The University of Michigan.

Engwall, O., & Lopes, J. (2020). Interaction and collaboration in robot-assisted language learning for adults. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2020.1799821

Engwall, O., Lopes, J., Cumbal, R., Berndtson, G., Lindström, R., Ekman, P., & Mekonnen, M. (2022). Learner and teacher perspectives on robot-led L2 conversation practice. *ReCALL*. https://doi.org/10.1017/S0958344022000027

Han, J. H., Jo, M. H., Jones, V., & Jo, J. H. (2008). Comparative study on the educational use of home robots for children. *Journal of Information Processing Systems*, *4*(4), 159–168. https://doi.org/10.3745/JIPS.2008.4.4.159.

Hong, Z. W., Huang, Y. M., Hsu, M., & Shen, W. W. (2016). Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. *Journal of Educational Technology & Society*, *19*(1), 337–349. https://doi.org/10.2307/jeductechsoci.19.1.337. https://www.jstor.org/stable/.

Hsia, L. H., & Sung, H. Y. (2020). Effects of a mobile technology-supported peer assessment approach on students' learning motivation and perceptions in a college flipped dance class. *International Journal of Mobile Learning and Organisation*, *14*(1), 99–113.

Hsiao, H. S., Lin, Y. W., Lin, K. Y., Lin, C. Y., Chen, J. H., & Chen, J. C. (2022). Using robot-based practices to develop an activity that incorporated the 6E model to improve elementary school students' learning performances. *Interactive Learning Environments*, *30*(1), 85–99. https://doi.org/10.1080/10494820.2019.1636090.

Hsu, T. C., & Liang, Y. S. (2021). Simultaneously improving computational thinking and foreign language learning: Interdisciplinary media with plugged and unplugged approaches. *Journal of Educational Computing Research*, *59*(6), 1184–1207. https://doi.org/10.1177/0735633121992480.

Hwang, G. J., Yang, T. C., Tsai, C. C., & Yang, S. J. (2009). A context-aware ubiquitous learning environment for conducting complex science experiments. *Computers & Education*, *53*(2), 402–413. https://doi.org/10.1016/j.compedu.2009.02.016.

Iio, T., Maeda, R., Ogawa, K., Yoshikawa, Y., Ishiguro, H., Suzuki, K., & Hama, M. (2019). Improvement of Japanese adults' English speaking skills via experiences speaking to a robot. *Journal of Computer Assisted Learning*, *35*(2), 228–245. https://doi.org/10.1111/jcal.12325.

Jaipal-Jamani, K., & Angeli, C. (2017). Effect of robotics on elementary preservice teachers' self-efficacy, science learning, and computational thinking. *Journal of Science Education and Technology*, *26*, 175–192. https://doi.org/10.1007/s10956-016-9663-z.

Jawad, L. F., Majeed, B. H., & ALRikabi, H. T. S. (2021). The impact of CATs on mathematical thinking and logical thinking among fourth-class scientific students. *International Journal of Emerging Technologies in Learning (Online)*, *16*(10), 194. https://doi.org/10.3991/ijet.v16i10.22515.

Konstantakis, M., Lykiardopoulou, A., Lykiardopoulou, E., Tasiouli, G., & Heliades, G. (2022). An exploratory study of Mobile-based scenarios for Foreign Language Teaching in Early Childhood. *Education Sciences*, *12*(5), 306. https://doi.org/10.3390/educsci12050306.

Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*(1), 79–97. https://doi.org/10.3102/00346543058001079.

Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. G., Sagong, S., & Kim, M. (2011). On the effectiveness of robot-assisted language learning. *ReCALL*, *23*(1), 25–58. https://doi.org/10.1017/S0958344010000273.

Liang, J. C., & Hwang, G. J. (2023). A robot-based digital storytelling approach to enhancing EFL learners' multimodal storytelling ability and narrative engagement. *Computers & Education*, *201*, 104827. https://doi.org/10.1016/j.compedu.2023.104827.

Liu, E. Z. F. (2010). Early adolescents' perceptions of educational robots and learning of robotics. *British Journal of Educational Technology*, *41*(3), E44–E47. https://doi.org/10.1111/j.1467-8535.2009.00944.x.

Liu, O. L., Rios, J. A., & Borden, V. (2015). The effects of motivational instruction on college students' performance on low-stakes assessment. *Educational Assessment*, *20*(2), 79–94. https://doi.org/10.1080/10627197.2015.1028618.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381.

Mercer, N., Warwick, P., & Ahmed, A. (2017). An oracy assessment toolkit: Linking research and development in the assessment of students' spoken language skills at age 11–12. *Learning and Instruction*, *48*, 51–60. https://doi.org/10.1016/j.learninstruc.2016.10.005.

Mertler, C. (2016). *Classroom assessment: A practical guide for educators*. Routledge.

Min, Q., Wang, Z., & Liu, N. (2019). Integrating a cloud learning environment into English-medium instruction to enhance non-native English-speaking students' learning. *Innovations in Education and Teaching International*, *56*(4), 493–504. https://doi.org/10.1080/14703297.2018.1483838.

Nomura, T., Kanda, T., & Suzuki, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *Ai & Society*, *20*(2), 138–150. https://doi.org/10.1007/s00146-005-0012-7.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(1), 3. https://doi.org/10.1037/0278-7393.31.1.3.

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813.

Randall, N. (2019). A survey of robot-assisted language learning (RALL). *ACM Transactions on Human-Robot Interaction (THRI)*, *9*(1), 1–36.

Roediger, I. I. I., H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003.

Rosenberg-Kima, R. B., Koren, Y., & Gordon, G. (2020). Robot-supported collaborative learning (RSCL): Social robots as teaching assistants for higher education small group facilitation. *Frontiers in Robotics and AI*, *6*, 148. https://doi.org/10.3389/frobt.2019.00148.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119–144. https://doi.org/10.1007/BF00117714.

Schut, S., van Tartwijk, J., Driessen, E., van der Vleuten, C., & Heeneman, S. (2020). Understanding the influence of teacher–learner relationships on learners' assessment perception. *Advances in Health Sciences Education*, *25*(2), 441–456. https://doi.org/10.1007/s10459-019-09935-z.

Schüttpelz-Brauns, K., Hecht, M., Hardt, K., Karay, Y., Zupanic, M., & Kämmer, J. E. (2020). Institutional strategies related to test-taking behavior in low stakes assessment. *Advances in Health Sciences Education*, *25*(2), 321–335. https://doi.org/10.1007/s10459-019-09928-y.

Serholt, S., Barendregt, W., Leite, I., Hastie, H., Jones, A., Paiva, A., & Castellano, G. (2014, August). Teachers' views on the use of empathic robotic tutors in the classroom. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (pp. 955–960). IEEE.

Sisman, B., Gunay, D., & Kucuk, S. (2019). Development and validation of an educational robot attitude scale (ERAS) for secondary school students. *Interactive Learning Environments*, *27*(3), 377–388. https://doi.org/10.1080/10494820.2018.1474234.

Socratous, C., & Ioannou, A. (2022). Evaluating the impact of the curriculum structure on group metacognition during collaborative problem-solving using educational robotics. *TechTrends*, *66*(5), 771–783. https://doi.org/10.1007/s11528-022-00738-5.

Srisawasdi, N., Pondee, P., & Bunterm, T. (2018). Preparing pre-service teachers to integrate mobile technology into science laboratory learning: An evaluation of technology-integrated pedagogy module. *International Journal of Mobile Learning and Organisation*, *12*(1), 1–17.

Stordeur, M. F., Nils, F., & Colognesi, S. (2022). No, an oral presentation is not just something you prepare at home! elementary teachers' practices supporting preparation of oral presentations. *L1-Educational Studies in Language and Literature*. https://doi.org/10.21248/l1esll.2022.22.1.417

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29–47). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.004

Tanaka, F., & Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, *1*(1), 78–95. https://doi.org/10.5898/JHRI.1.1.Tanaka.

Timpe-Laughlin, V., Sydorenko, T., & Dombi, J. (2022). Human versus machine: Investigating L2 learner output in face-to-face versus fully automated role-plays. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2022.2032184

Tung, F. W. (2016). Child perception of humanoid robot appearance and behavior. *International Journal of Human-Computer Interaction*, *32*(6), 493–502. https://doi.org/10.1080/10447318.2016.1172808.

van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., Van der Ven, S., & Leseman, P. (2019). Social robots for language learning: A review. *Review of Educational Research*, *89*(2), 259–295.

van der Lans, R. M., & Maulana, R. (2018). The use of secondary school student ratings of their teacher's skillfulness for low-stake assessment and high-stake evaluation. *Studies in Educational Evaluation*, *58*, 112–121. https://doi.org/10.3102/0034654318821286.

van Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C., & Strik, H. (2016). Evaluating automatic speech recognition-based language learning systems: A case study. *Computer Assisted Language Learning*, *29*(4), 833–851. https://doi.org/10.1080/09588221.2016.1167090.

Wang, S. L., & Lin, S. S. (2007). The effects of group composition of self-efficacy and collective efficacy on computer-supported collaborative learning. *Computers in Human Behavior*, *23*(5), 2256–2268. https://doi.org/10.1016/j.chb.2006.03.005.

Wedenborn, A., Wik, P., Engwall, O., & Beskow, J. (2019). The effect of a physical robot on vocabulary learning. arXiv preprint arXiv:1901.10461.

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, *5*, 506. https://doi.org/10.3389/fpsyg.2014.00506.

Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, *55*(1), 10–16. https://doi.org/10.1111/sjop.12093.

Wise, S. L., & DeMars, C. E. (2003). Examinee motivation in low-stakes assessment: problems and potential solutions: *Paper Presented At the Annual Meeting of the American Association of Higher Education Assessment Conference*. Seattle, Canada.

Wu, W. C. V., Wang, R. J., & Chen, N. S. (2015). Instructional design using an in-house built teaching assistant robot to enhance elementary school English-as-a-foreign-language learning. *Interactive Learning Environments*, *23*(6), 696–714. https://doi.org/10.1080/10494820.2013.792844.

Yang, F. C. O., Lai, H. M., & Wang, Y. W. (2023). Effect of augmented reality-based virtual educational robotics on programming students' enjoyment of learning, computational thinking skills, and academic achievement. *Computers & Education*, *195*, 104721.

Yang, Y. T. C., Chen, Y. C., & Hung, H. T. (2022). Digital storytelling as an interdisciplinary project to improve students' English speaking and creative thinking. *Computer Assisted Language Learning*, *35*(4), 840–862. https://doi.org/10.1080/09588221.2020.1750431.

Darmawansah Darmawansah Dr. Darmawansah Darmawansah is a postdoctoral research fellow at the Empower Vocational Education Research Center, National Taiwan University of Science and Technology, Taiwan. His research interests include emerging technologies-enhanced learning in language and STEM learning, and argumentation learning.

Gwo-Jen Hwang Dr. Gwo-Jen Hwang is a chair professor at the Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology as well as a chair professor of the Graduate Institute of Educational Information and Measurement, National Taichung University of Education, Taiwan. His research interests include mobile learning, digital game-based learning, flipped classrooms, and artificial intelligence in education.

Authors and Affiliations

Darmawansah Darmawansah[1,2] · Gwo-Jen Hwang[1,3,4]

✉ Gwo-Jen Hwang
gjhwang.academic@gmail.com

Darmawansah Darmawansah
darmawansahlavoe@gmail.com

[1] Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei City, Taiwan

[2] Empower Vocational Education Research Center, National Taiwan University of Science and Technology, Taipei City, Taiwan

[3] Graduate Institute of Educational information and Measurement, National Taichung University of Education, Taichung City, Taiwan

[4] Yuan Ze University, Taoyuan City, Taiwan