Lee, S.-M., Ahn, T. Y., & Wu, J. G. (2026). Self-regulated speaking with AI-empowered NPCs in VR. *Educational Technology & Society*, *29*(2), 84–99. https://doi.org/10.30191/ETS.202604_29(2).RP05

Self-regulated speaking with AI-empowered NPCs in VR

Sangmin-Michelle Lee[1], Tae youn Ahn[2*] and Junjie Gavin Wu[3]
[1]Kyung Hee University, Korea // [2]Korea National Sport University, Korea // [3]Macao Polytechnic University, Macao SAR // sangminlee@khu.ac.kr // ahntyn@gmail.com // gavinjunjiewu@gmail.com
[*]Corresponding author

(Submitted October 27, 2024; Revised July 8, 2025; Accepted July 10, 2025)

ABSTRACT: In English as a foreign language (EFL) contexts, where authentic speaking opportunities are often limited, virtual reality (VR) platforms have emerged as a potential technology to provide an interactive space for learners to engage in realistic conversations with AI-empowered Non-Player Characters (NPCs). This study investigates the relationship between self-regulation strategies and the improvement of speaking skills among Korean EFL learners, specifically exploring whether the students' use of self-regulation strategies enhances their speaking performance in VR environments, and whether VR-based learning further cultivates the development of these strategies. Sixty Korean university students were divided into two groups: a desktop-based VR (DVR) group and an immersive VR (iVR) group. Data were collected through pre- and post-surveys using the Strategic Self-Regulation for EFL Speaking Scale and pre- and post-speaking tests. Student reflections were also collected for qualitative analysis. The results indicated that while both groups improved in speaking performance, the iVR group exhibited significantly greater gains in self-regulation strategies. Furthermore, students in the iVR group reported higher levels of enjoyment, interest, and reduced speaking anxiety compared to the DVR group. These findings underscore important pedagogical considerations when selecting appropriate VR technologies to enhance language learning. Suggestions are made regarding the need for refined measurement tools to accurately assess self-regulation strategies in VR environments.

Keywords: EFL speaking, Self-regulation strategies, Virtual reality, AI-empowered NPCs

1. Introduction

Virtual Reality (VR) refers to computer-generated, immersive environments where users can interact with digital elements in real time. In developing second language (L2) speaking skills, VR enables lifelike, contextualized speaking practice, reducing anxiety and increasing engagement (Lee et al., 2024b). Moreover, VR facilitates self-regulated learning, which is crucial for L2 development. Self-regulation (SR) involves learners' ability to plan, monitor, and evaluate their progress through goal-setting and adaptive strategy use (Zimmerman & Risemberg, 1997). Research highlights that self-regulated learners actively track their progress, reflect on their learning process, and make strategic adjustments to enhance their proficiency (Chen et al., 2020). This ability to reflect on and modify learning strategies allows learners to refine their speaking skills more effectively over time.

However, traditional EFL classrooms often lack sufficient opportunities for authentic speaking practice and individualized feedback, often confined to role-plays or scripted interactions that fall short of real-life communication (Lee et al., 2024a). With minimal opportunities to converse with native English speakers, improving speaking skills through meaningful interactions is challenging for EFL learners (Lee & Lee, 2020). To address these limitations, advancements in generative AI (GenAI) chatbots have introduced new possibilities for L2 speaking instruction, offering distinct pedagogical advantages such as personalized learning experiences, reduced speaking anxiety, and unlimited access to speaking practice (Zhang et al., 2024). The integration of GenAI chatbots into VR environments presents an opportunity to expand L2 learning possibilities and maximize pedagogical benefits for speaking skill development. Within VR environments, GenAI chatbots are typically implemented as Non-Player Characters (NPCs) with human-like appearances, enabling natural interactions with learners. Unlike traditional NPCs with pre-set dialogues, AI-empowered NPCs can initiate, react to, and sustain conversations with users. These NPCs exhibit sophisticated language understanding capabilities, allowing them to engage in meaningful conversations on various topics in real time (Lee, 2024), enhancing conversational flexibility and effectiveness in L2 speaking practice.

In contrast to traditional teacher-centered environments where instructors provide systematic guidance, VR-based language learning necessitates a heightened degree of SR from students due to its autonomous nature of the learning environment (Wu et al., 2024). As a result, SR strategies are essential for developing speaking skills in VR settings. However, despite the importance of SR in VR-based instruction, empirical studies on students' use of SR strategies in VR environments remain underexplored. Although prior research has explored SR in VR, most studies have primarily focused on its general effects on learning outcomes rather than the specific strategies students employ to regulate their learning in such environments (Liu et al., 2023). Moreover, the integration of AI-powered NPCs into VR-based L2 speaking instruction presents a promising yet underexplored research avenue. While research has investigated SR processes in VR (Li et al., 2024; Sobocinski et al., 2024), studies on AI-driven NPCs are notably scarce, leaving their potential impact on learner autonomy and engagement in VR-based language instruction unexplored.

In light of these research gaps, the present study investigates the relationship between students' use of SR strategies and their L2 performance in VR environments, as well as whether VR-based L2 speaking practice fosters the development of these strategies. Specifically, this research compares two distinct types of VR environments: desktop-based VR (DVR) and immersive VR (iVR). iVR, which utilizes head-mounted displays (HMDs), creates fully immersive experiences that enhance learners' sense of presence and interaction authenticity. In contrast, DVR operates on standard computer screens, offering greater accessibility and lower cost. Although iVR is gaining traction for its ability to enhance immersion and engagement, DVR remains widely used, particularly in educational settings where cost-effectiveness and access matter.

Research suggests that iVR provides a heightened level of immersion, which enhances retention (Di Natale et al., 2020), whereas DVR's lack of full sensory immersion may reduce the authenticity of communicative practice (Sadanala et al., 2024). However, Liu et al. (2024) reported that, unlike iVR, which may introduce higher cognitive load and motion sickness, DVR showed significant benefits in knowledge acquisition and SR skills, particularly in structured learning environments where students can control their pace. These differences highlight the need for a deeper understanding of how varying levels of immersion influence SR strategies and L2 development.

Despite the growing interest in VR and AI integration in language learning, no study has systematically examined the use of SR strategies in VR-based L2 speaking practice. Additionally, empirical research on the impact of AI-powered NPCs in supporting SR across iVR and DVR environments remains scarce. To address this gap, the present study explores how Korean EFL learners regulate their speaking practice in these two VR environments and how the affordances of each setting shape their use and development of SR strategies.

2. Literature review

2.1. Self-regulations skills for L2 speaking

One of the primary goals of foreign language education is to help students become autonomous (Zimmerman & Risemberg, 1997). It is essential that students acquire the ability to regulate their own language learning process, so that they can continuously improve their language skills and take more responsibility for learning (Wang & Sun, 2024). Speaking, as a complex cognitive skill, places heavy demands on EFL learners because it requires them to apply linguistic knowledge in productive tasks. In East Asian countries such as Japan, China, and Korea, speaking is often underemphasized in English education because it is not a central focus in academic assessments (Tan et al., 2024). As a result, many students lack exposure to speaking practice, which hinders their ability to develop effective SR speaking skills (Lee & Lee, 2020; Tan et al., 2024).

For the regulation of language learning, learners employ various strategies, either consciously or unconsciously, to enhance outcomes (Oxford, 2011). Rubin (1981) classified SR strategies into direct and indirect types that learners employ in L2 learning. Oxford (1990) further expanded Rubin's types into six categories: memory, cognitive, compensation (direct strategies), and metacognitive, affective, social (indirect strategies). Later, Oxford (2011, 2017) refined these categories into cognitive, affective, and sociocultural-interactive strategies, all of which play key roles in language learning, especially for the improvement of speaking ability. Uztosun (2020) emphasized the role of motivation within SR, demonstrating its strong influence on the development of speaking. These studies illustrate that variables such as cognitive strategies, emotional regulation, social interaction, and motivation all play critical roles in shaping the effectiveness of SR in speaking.

To promote SR in L2 speaking, it is important to integrate strategies that address students' cognitive, emotional, and adaptive skills. One effective approach is providing students with opportunities to practice speaking in diverse contexts, helping them adapt their communication style to different situations (Derakhshan & Fathi, 2024). Another key strategy is fostering motivation, as students often benefit from engaging, motivational environments that encourage active participation (Uztosun, 2021). Additionally, giving students ownership of their learning and opportunities for self-reflection and task repetition enables them to better monitor progress and take responsibility for improving speaking skills (De Vrind et al., 2024).

As digital learning environments, such as VR, become increasingly integrated into language education, the need for effective SR strategies in these contexts grows even more critical. VR provides unique opportunities for learners to develop SR skills across multiple dimensions. From a cognitive and metacognitive perspective, VR environments enable learners to apply organizing, elaborating, goal-setting, and self-monitoring strategies while completing tasks (Parong & Mayer, 2020). However, immersive and complex stimuli in VR can increase cognitive load, making it difficult to apply strategies consistently. As a result, learners' ability to efficiently manage cognitive resources becomes critical to their language learning success (Lan, 2020). Affectively, VR contexts elicit authentic emotional responses due to their interactive nature (Lee et al., 2024b), requiring learners to regulate anxiety, frustration, and excitement that naturally arise during immersive experiences. Behaviorally, the exploratory freedom of VR requires enhanced SR as learners make ongoing decisions about meaningful interactions while avoiding distractions (Parmaxi, 2020). Socially, the effective use of social SR strategies in VR environments correlates with increased engagement, reduced anxiety, and improved communicative competence in language learning tasks (Lee et al., 2024b; Parmaxi, 2020).

While VR provides a flexible yet complex learning space where learners must actively regulate their speaking practice and adapt to different communicative situations, capturing SR within these digital spaces poses methodological challenges (Saint et al., 2022). Alvarez et al. (2022) provide a systematic review of digital tools designed to support SR, highlighting dependencies on specific platforms and inconsistencies in operational definitions of SR constructs. These inconsistencies hinder cohesive understanding of SR in digital contexts. To advance research in this area, there is a need for a more standardized approach to measuring SR in VR environments, one that accounts for its cognitive, affective, and social dimensions. This study addresses this gap by employing a comprehensive framework for measuring self-regulated L2 speaking to ensure consistency and contribute to a more reliable, holistic assessment of SR in VR.

2.2. L2 learning in VR

Given the increasing role of VR in facilitating SR, it is important to examine how VR enhances L2 learning experiences. By offering contextual learning, fostering motivation, and enabling self-directed practice, VR creates powerful learning opportunities for language learners and makes language learning tasks more meaningful and effective (Lee et al., 2023; Wu et al., 2024). VR is particularly valuable for speaking practice, as it recreates real-world contexts where students practice language dynamically. Additionally, VR lowers speaking anxiety, which is prevalent in traditional L2 speaking classrooms particularly in Asian countries and hinders students' development of speaking skills (Tan et al., 2024; Wu et al., 2024).

From an affective and metacognitive perspective, VR fosters motivation and engagement by sustaining students' interest (Makransky & Petersen, 2021), and this increased motivation and engagement, in turn, enhances SR learning and enables them to take a more active and autonomous role in the learning process (Lee & Ahn, 2025; Oxford, 2017). Unlike traditional classrooms, where students often passively receive information, VR empowers students to become active participants (Yeh & Lan, 2018). Specifically, the reduced need for direct teacher intervention further supports their autonomy (Wu et al., 2021), allowing them to explore and learn at their own pace. This active participation extends to cognitive and metacognitive processes. Cho and Lim (2017) observed that VR learning environments promote higher-order thinking skills and metacognitive awareness, both of which are essential components of SR learning. Makransky and Petersen (2021) also found that self-paced learning in VR environments led to improved learning outcomes, increased self-efficacy, and SR learning.

However, different types of VR can affect students' learning experiences. VR is typically categorized into iVR and DVR. iVR provides a fully immersive experience, which blocks external stimuli and places users within a vivid, first-person virtual environment. DVR, on the other hand, offers a comparatively less immersive experience, where users interact with a 3D virtual environment through a monitor using a mouse and keyboard. While iVR provides full immersion, DVR allows users to engage with the environment from an external observer's perspective. Although direct comparisons of iVR and DVR in L2 learning remain limited, examining their broader educational impact provides insights into their potential effectiveness in language learning as their affordances - such as immersion, engagement, and interactivity - are particularly relevant to L2 learning, where authentic communication and contextualized practice are crucial. Existing literature shows that while iVR offers heightened immersion and engagement, its impact on learning compared to DVR remains inconclusive. For instance, Di Natale et al.'s (2020) systematic review found iVR enhances presence, motivation, and experiential learning through interactive, first-person environments. However, research by Alrehaili and Osman (2022) demonstrated no significant difference in learning between iVR and DVR environments. Despite being less immersive, DVR has been shown to contribute positively to academic achievement and foster emotional and cognitive development (Liu et al., 2024). Similarly, Shen et al. (2025) reported that even semi-immersive VR (spherical video-based) environments significantly enhanced EFL learners' SR strategies, especially in metacognitive and motivational/affective regulation, when tasks were contextualized. This suggests that well-designed, task-based DVR settings may support SR development even without full immersion.

Beyond the type of VR used, the integration of AI chatbots enhances VR-based learning by offering effective support functions such as answering inquiries, providing immediate feedback, and facilitating personalized learning (Nong et al., 2025). In L2 speaking, AI chatbots serve as valuable pedagogical tools, providing authentic, context-rich practice (Zhang, 2024). They allow learners to practice independently, anytime and anywhere, thereby enhancing language fluency and confidence. Furthermore, they foster engagement, support self-paced learning (Yin et al., 2021), and help reduce speaking anxiety through independent practice (Jeon, 2022). Recent research confirms the positive impact of AI-powered conversational agents on learners' SR learning. Specifically, Du (2025) found that interactions with intelligent agents significantly improved EFL learners' metacognitive awareness and consistent use of SR strategies, such as planning, monitoring, and emotional regulation, thereby leading to better language retention. These findings underscore the pedagogical value of AI-integrated systems in fostering SR in language learning.

Makransky et al. (2019) noted that chatbots (pedagogical agents) in VR offer instant feedback, addressing challenges in traditional classrooms where a single teacher must manage multiple students. However, most existing studies (e.g., Jeon, 2022; Kim et al., 2022) have focused on text-based or voice-based AI chatbots without anthropomorphic representations, limiting their effectiveness in fostering embodied learning experiences. Embodied cognition (Makransky & Petersen, 2021) posits that learning improves when grounded in sensory-motor experiences. From this perspective, AI-driven NPCs in VR offer a more immersive and interactive learning experience by integrating visual, auditory, and kinesthetic modalities, distinguishing them from traditional AI chatbots.

AI-empowered NPCs in VR expand beyond traditional chatbot and scripted interactions by leveraging advanced Natural Language Processing (Nong et al., 2025). Unlike conventional VR systems that rely on static, pre-scripted dialogues, AI-empowered NPCs dynamically generate responses based on user interactions, creating a more immersive and adaptive experience. While Nong et al. (2025) explored AI-empowered NPCs in immersive game environments, their focus was on enhancing natural language interactions, not education. Research on AI-empowered NPCs in VR for language learning remains scarce, highlighting the need to explore their role and their educational potential.

Therefore, this study aims to examine 1) whether students' SR strategies influence their speaking performances after practicing with AI-empowered NPCs in VR settings, and 2) whether such learning experiences contribute to further development of SR strategies for L2 speaking. Furthermore, this study intends to find out potential differences between DVR and iVR as contextual variables that may influence students' use of SR strategies. In light of these considerations, this study addresses the following research questions:

- RQ1: What is the influence of students' SR strategies in speaking on their speaking performance after practicing with AI-empowered NPCs in VR environments? To what extent does the type of platform (DVR vs. iVR) affect the results?
- RQ2: Does practicing L2 speaking with AI-empowered NPCs in VR environments contribute to the development of students' SR strategies in speaking? Are there any platform-specific differences in the development of SR strategies between DVR and iVR environments?

3. Methods

3.1. Study design

The present study involved 60 college students from different majors who were enrolled in the "College English" course, designed to improve students' communicative skills in various real-life situations. All participants were native Korean speakers, and their English proficiency level ranged from low intermediate (A2) to advanced (C1). Most of the students reported that they had had minimal prior opportunities to practice speaking English both inside and outside the classroom.

The study spanned six weeks and utilized the commercial VR program *Immerse*, designed for language learning. The program offered a DVR platform (low fidelity) and an iVR platform (high fidelity). In *Immerse*, students could virtually visit 35 different real-world locations, such as a restaurant, airport, park, and gym, where they can interact with others to practice oral language. Each location featured NPCs that play different roles relevant to the context and scenario, such as a doctor in a doctor's office or a cashier/server in a restaurant.

Two class sections were randomly assigned to different platform conditions: one class used the DVR platform (Group 1) and the other class used the iVR platform (Group 2). In terms of language proficiency, 21 students were intermediate and 10 were advanced in Group 1, and 19 students were intermediate and 9 were advanced based on the pre-test in Group 2. The students participated in the activity in *Immerse* for six weeks (75 minutes per session, two sessions per week).

To ensure instructional consistency across the two platforms, all tasks, scenarios, and instructor-led activities remained identical for both groups. While the iVR platform offered a more immersive experience, the DVR platform provided the same interactive tasks using a traditional PC interface. Scripted instructional prompts and structured communicative tasks were used to maintain a consistent pedagogical approach across both conditions.

3.2. Instruments

The present study employed a quantitative research methodology as primary data analysis, using Sun's (2022) Speaking Self-Regulation Scales, a comprehensive and detailed measurement of students' SR in speaking tasks. The survey instrument comprised 52 items across 13 categories, encompassing cognitive (*Cognitive processing, Remembering, Idea planning, Goal-based monitoring and evaluation,* and *Self-reflection*), affective (*Anxiety control, Interest enhancement,* and *Motivational self-talk*), and social (*Peer learning, Feedback management, Interactional practice,* and *Environmental control*) domains. Pre- and post-surveys were administered before and after the *Immerse* sessions. In the post-survey, four additional items were included that asked about students' perceptions of enjoyment, interest, motivation, and anxiety related to their experience in the VR-based speaking practice, based on previous studies that focused on interest, enjoyment, and motivation (Lee et al., 2024a; Makransky & Petersen, 2021) and anxiety (Sadler & Thrasher, 2023) in VR. The additional questions were designed to measure students' affective factors, which, according to Makransky and Petersen (2021), are related to self-directed learning and influence learning outcomes.

Additionally, pre- and post-speaking tests were conducted during the initial and final sessions, respectively. They spoke for two minutes on the topic "Describe your symptoms," one of the most common everyday topics but the most difficult for EFL students due to the medical terms. They recorded their responses using their mobile phones and submitted the audio files to the teacher. The topic was chosen based on the following criteria: 1) it should be a monologue so that the students would not need a conversation partner for the test, and 2) it should be at an advanced level so that the students' learning outcomes after speaking practice in VR could be compared. The students took a posttest about the same topic after completing the VR sessions. Speaking tests were scored on the 5-point scale using the TOEFL Speaking rubric by two trained English writing instructors with PhDs in English education who have taught English writing courses for more than 10 years (See Appendix A). Before scoring, the raters reviewed sample responses and discussed scoring criteria. To ensure scoring consistency, each student's recording was scored independently by both raters, and any discrepancies greater than one point were resolved through discussion and consensus. They scored each student's recording separately and compared the results. The interrater reliability was .88.

Finally, a more robust research approach was implemented by collecting and analyzing qualitative data from students' reflection papers. In these one-page reflections, students documented their user experiences while practicing speaking in the VR environments (DVR and iVR) after completing assigned tasks. These reflections complemented the quantitative data by providing learner perspectives on the learning environment and the use of AI-NPCs. Together, the quantitative and qualitative data align with the pedagogical goals of the course by assessing not only measurable improvements in speaking proficiency and self-regulation, but also capturing learners' engagement, perceived challenges, and overall responses to the VR-based speaking practice.

3.3. Procedure

During the first half of the semester, the courses focused on spoken language skills, for which the students used *Immerse*. In the first week, the students learned how to use the platform, interact with the objects, and talk to the AI-empowered NPCs. Group 2, in particular, needed time to learn how to use the HMD and hand controllers (Figure 1), and the instructor allowed both groups to freely explore the locations so that they could become familiar with the VR environment.

Several measures were taken to mitigate potential confounding variables such as the VR novelty effect, technical challenge, and differences in prior VR experience. First, all students received an orientation session in Week 1 and the students in Group 2 were provided with additional practice time to reduce the impact of unfamiliarity with the immersive interface. Second, the instructor conducted troubleshooting at the beginning of each class and technical issues (e.g., network instability) were addressed immediately to minimize technical disruptions. Third, students completed a background questionnaire that assessed their prior experience with VR technologies. While none of the students had previously used a HMD, some had limited exposure to VR environments in the context of PC-based video games. However, none had prior experience using VR for language practice.

From the second week on, the students visited a selected location each week and practiced talking to the NPCs. To give the students more opportunities to practice speaking, the instructor also embedded additional communicative tasks in each location (e.g., initiating a conversation to make a friend at the bar). The students could also converse with other students in addition to the AI-empowered NPCs. Prior to the study, the purpose of the study was explained and written consent was obtained from the students. Since using the HMD could cause discomfort, students were told to take a break or stop using the HMD at any time if they felt uncomfortable.

Figure 1. Students with VR headsets

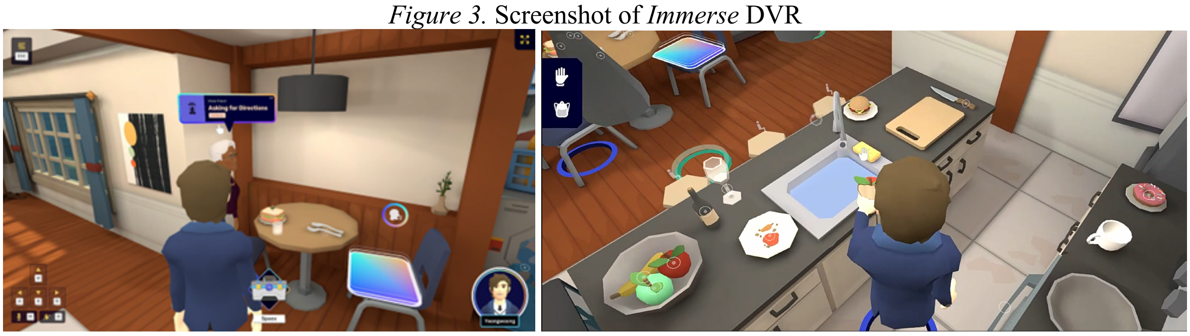
Despite differences in technological affordances, instructional delivery remained uniform. Both groups followed the same communicative tasks. The primary difference was that DVR students interacted using a keyboard and mouse, while iVR students used motion tracking and hand controllers. To account for these differences, the instructor provided iVR participants with additional practice time in Week 1 to avoid technical difficulties disrupting language practice. Throughout the sessions, technical issues such as network instability were addressed immediately by the instructor. The DVR platform had fewer technical disruptions, whereas some iVR participants initially reported difficulty adjusting to the immersive interface. To minimize interference with language learning, the instructor facilitated troubleshooting sessions at the beginning of each class to resolve technical concerns before engaging in communicative tasks.

In *Immerse*, clicking on an AI-empowered NPC prompted students with a specific communicative task. Each NPC offered 2-3 tasks (e.g., describing symptoms to the doctor, ordering food) with different levels of language proficiency. When a topic was chosen, a list of communicative tasks was shown in a checklist format (see Figures 2 and 3). Figure 2 shows that the students were able to view the scene from the 1st person perspective and manipulate objects. As can be seen in the dialogue window (Figure 2), NPCs were multimodal, enabling the students to listen to NPCs and read the text simultaneously, and this feature was available on both platforms. When students spoke, the conversation was transcribed into text for review and then sent to NPCs. If unsatisfied, students could speak again until they had a satisfactory result. Additionally, hints for unknown words were available to support learners in real-time.

Each NPC was context-sensitive, assigned a specific role relevant to its location. For example, the AI-NPC at the hospital reception was programmed to handle patient check-ins, leading students through a logical sequence of questions about symptoms and appointment confirmations. Similarly, an NPC in a restaurant setting would take orders, offer menu recommendations, and respond to modifications, mimicking natural service interactions. Unlike traditional chatbot-based language learning tools, the NPCs empowered by GenAI utilized context-aware adaptability. While each NPC was designed to maintain a task-relevant conversation flow, students could diverge from scripted responses and engage in personalized, open-ended interactions. The AI-empowered NPC also responded to non-task-related talk, providing flexibility for more personalized interactions beyond the pre-established scenarios.

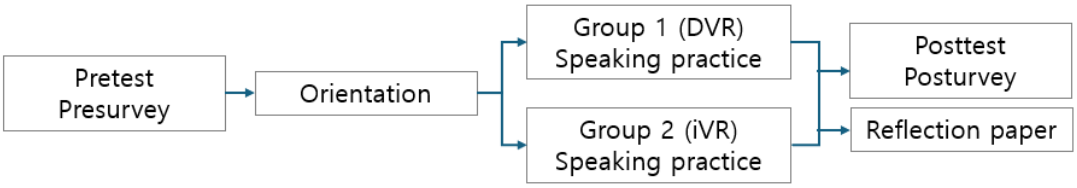
Figure 2. Screenshot of *Immerse* iVR

Figure 3. Screenshot of *Immerse* DVR



The study followed a structured procedure as shown in Figure 4. First, all participants completed a pretest and a presurvey to assess their initial speaking proficiency and SR. After completing the speaking practice sessions followed by an orientation, all participants took a posttest and a postsurvey to measure changes in their speaking performance and the use of SR strategies for L2 speaking.

Figure 4. Procedure of the project and data collection



3.4. Data analysis

Descriptive statistics were utilized to analyze the overall patterns of students' SR strategies. The Cronbach's alpha of the survey was .945, indicating that the survey was reliable. Regarding the learning outcomes, first, the Levene's Test for Equality of Variances was conducted to assess the assumption of homogeneity of variances between the groups. The Levene's Test yielded a *p*-value of .416 and .438, thus, for both variables, the assumption of homogeneity of variances was satisfied. However, since the normality was not confirmed based on the Kolmogorov-Smirnov test and the Shapiro-Wilk test ($p < .05$), a non-parametric test, the Wilcoxon Signed Rank Test, was performed to assess the statistical significance of the differences between the pre and post speaking test scores. Regarding the surveys, there was no significant difference between the groups based on Levene's test ($p > .05$). However, the normality of the surveys was not confirmed ($p < .05$), thus, Wilcoxon Signed Rank Test was performed to examine the differences between the pre- and post-survey scores, and Mann-Whitney U test was used to compare between the groups. Finally, to investigate the potential impact of students'

SR strategies on their speaking outcomes, Generalized Additive Model (GAM), a non-parametric regression, was performed using the categorical variables derived from the SR survey.

The qualitative data underwent thematic analysis through open coding, following an inductive approach. Reflections from the iVR group (Group 2) were organized into three recurrent themes: enhanced immersion and engagement, physical and technical challenges, and individual differences in user experience. Codes under these themes included phrases such as "felt like being inside the situation" (immersion), "got dizzy after wearing the headset" (physical discomfort), and "I had trouble figuring out how to control things at first" (prior experience). In contrast, reflections from the DVR group (Group 1) were more limited and centered on usability issues, with codes such as "mouse dragging was awkward" or "teleporting felt unnatural." Two trained coders, who are also the researchers of the current study, independently coded the data, reaching an initial agreement rate of 87%, with discrepancies resolved through discussion.

4. Results

4.1. Quantitative data analysis

The present study investigated the relationships between the students' SR strategies and their posttest performances. The GAM analysis revealed that the effect of SR strategies on speaking performance in VR showed a non-linear pattern, and all categories of SR strategies had a positive effect on speaking performance in the posttest. Specific categories such as self-reflection, peer learning, and idea planning contribute more significantly, while categories such as interest enhancement and motivated self-talk show weaker effects. More specifically, the variables with a strong effect ($r \geq .60$) include *Cognitive process, Remembering, Idea planning, Goal-based monitoring and evaluation, Peer learning, Interpersonal practice, Assistance seeking* and *Environment control* showed the stronger effects, while the variables with a weak effect include *Interest enhancement, Motivational self-talk*, and *Anxiety control* had weaker effects ($r < .60$). The effect of SR on speaking performance was significant in both contexts, with slight differences between groups, where *Idea planning* had a higher effect in Group 1, while *Goal-based monitoring and evaluation* had a higher effect in Group 2. These findings are summarized in Table 1.

Table 1. Influence of SR strategies on students' speaking performances in VR (GAM)

Category	Coefficients	*r*	GAM Predicted	*p*
CP	.100	.809	.134	< .001
RE	.100	.782	.065	< .001
ID	.141	.754	.108	< .001
GME	.127	.748	.131	< .001
PL	.128	.704	.139	< .001
FM	.113	.664	.136	< .001
IP	.126	.623	.127	< .001
AS	.085	.667	.114	< .001
AC	.117	.566	.103	< .001
IE	.110	.487	.096	< .001
MS	.114	.477	.099	< .001
EC	.126	.654	.126	< .001
SR	.130	.399	.149	< .001

Note. P = Cognitive process, RE = Remembering, ID = Idea planning, GME = Goal-based monitoring and evaluation, PL = Peer learning, FM = Feedback management, IP = Interactional practice, AS = Assistance seeking, AC = Anxiety control, IE = Interest enhancement, MS = Motivational self-talk, EC = Environment control, SR = Self-reflection.

The results of the presurvey showed that the participants' SR appeared moderately high overall ($M = 3.41$). Regarding the categories, both groups exhibited similar patterns; *Assistance seeking* and *Feedback management* marked the highest, while *Peer learning* and *Goal-based monitoring and evaluation* marked the lowest in both groups. The study compared the students' pre- and postsurvey results and found significant differences between the pre- and post-surveys in six categories; *Cognitive processing, Remembering, Idea planning, Goal-based monitoring and evaluation, Peer learning*, and *Environment control* significantly increased after the activities in the VR environments ($p < .05$). There were also significant differences found between the groups; while Group 1 (DVR) showed statistically meaningful increases only in two categories (*Cognitive Processing* and *Idea Planning*), Group 2 (iVR) showed statistically meaningful increases in all the categories except for two categories (*Feedback management* and *Motivational self-talk*). Wilcoxon Signed-Rank tests confirmed that these differences between the presurvey and postsurvey are statistically significant (Table 2). The effect sizes (Wilcoxon's r) of SR in DVR were medium, while the effect sizes of SR in iVR ranged from medium to large. In addition, the four items regarding students' perceptions of the activities included in the post-survey scored high; the students responded that the speaking activities with AI-empowered NPCs in VR were enjoyable (M = 4.4, SD = .767), interesting (M = 4.3, SD = .778), and motivating (M = 4.1, SD = .768) and helped reduce speaking anxiety (M = 4.0, SD = .833). Group 2 scored higher on all four items, and the differences between the groups were statistically significant ($p < .05$), but each effect size was not large (Table 3).

Table 2. Wilcoxon Signed-Rank Tests of the pre- and post-survey on self-regulation for each group

Category	Group1 (DVR)					Group 2 (iVR)				
	Pre	Post	*z*	*p*	*r*	Pre	Post	*z*	*p*	*r*
CP	3.6210	3.8065	-1.948	**.050**	.350	3.5370	4.0185	-3.909	**< .001**	.726
RE	3.3011	3.3387	-.419	.675	-	3.3765	3.9012	-3.721	**< .001**	.691
IP	3.0430	2.9355	-.340	.734	-	3.0741	3.7160	-2.801	**.005**	.376
GME	3.1694	3.3468	-1.533	.125	-	2.7593	3.2500	-2.581	**.01**	.476
PL	2.9570	3.1290	-.729	.466	-	2.4815	3.1605	-3.200	**.001**	.594
FM	4.3145	4.2661	-.258	.796	-	4.2500	4.2407	-.259	.796	-
IP	3.7500	3.4113	-1.987	**.047**	.357	3.6389	4.0463	-1.94	**.043**	.376
AS	4.1855	4.1855	-.230	.818	-	4.1944	4.4444	-1.85	**.042**	.375
AC	3.2581	3.1935	-.199b	.843	-	3.0833	3.4722	-2.324	**.02**	.432
IE	3.4194	3.3548	-.133	.894	-	3.0617	3.5802	-2.235	**0.025**	.415
MS	3.9770	3.8571	-.038	.97	-	3.8995	3.9537	-.0330	.974	-
EC	3.6882	3.6022	.000	1.00	-	2.7160	3.4762	-3.848	**< .001**	.715
SR	3.5484	3.4624	-.217	.828	-	3.0247	3.4762	-1.994	**.046**	.370

Note. CP = Cognitive process, RE = Remembering, IP = Idea planning, GME = Goal-based monitoring and evaluation, PL = Peer learning, FM = Feedback management, IP = Interactional practice, AS = Assistance seeking, AC = Anxiety control, IE = Interest enhancement, MS = Motivational self-talk, EC = Environment control, SR = Self-reflection.

Table 3. Students' perceptions after the activity (Mann-Whitney U Test)

Item	Group 1 (DVR)		Group 2 (iVR)		*z*	*p*	*r*
	M	SD	M	SD			
Enjoyment	4.2	.883	4.7	.672	1.955	**.049**	0.314
Interest	4.0	.795	4.5	.669	2.432	**.015**	0.252
Motivation	3.8	.668	4.3	.768	2.884	**.004**	0.372
Anxiety	3.6	.871	4.4	.692	3.364	**<.001**	0.434

Note. $^{*}p < .05$.

The post speaking test scores significantly increased, and the paired-test results showed that the difference between the pretest and posttest scores was statistically significant (t = -6.939, $p < .001$). Wilcoxon Signed-Rank Tests for each group also showed significant increases in the posttest scores after practicing speaking in the VR environment with both groups (Table 4). The effect sizes for each group were large (r = .714 for Group1, r = .703 for Group 2). However, the difference in learning outcomes between the groups was not significant ($p < .05$).

Table 4. Students' pre- and posttest scores (Wilcoxon Signed-Rank Test)

Group	Pre		Post		*z*	*p*
	M	*SD*	*M*	*SD*		
1 (DVR)	3.06	0.934	3.34	0.844	-3.976	**< .001**
2 (iVR)	3.25	1.145	3.66	0.993	-3.787	**< .001**

Note. $^{*}p < .05$.

4.2. Qualitative data analysis

Student reflections revealed distinct patterns of experience between the iVR and DVR groups. Many students in Group 2 reported that the use of HMDs substantially intensified their sense of presence and realism within the iVR setting. The movement-based interactivity, such as physically rotating one's head or virtually navigating through digital spaces, was identified as a critical factor contributing to increased concentration and motivation. The physical and technical challenges associated with HMD use were also frequently mentioned by students in Group 2. While the immersive qualities of iVR were widely acknowledged by the students, many experienced adverse physical symptoms including eye strain, dizziness, headaches, and motion sickness, particularly during extended sessions. Technical difficulties such as unstable connectivity and some instances of hardware malfunction further impeded the learning process and undermined the overall learning experience. Additionally, student reflections from Group 2 showed that individual differences significantly shaped their experiences with iVR environments. Students with a higher tolerance for motion-induced discomfort generally adapted more easily to the HMD environment, whereas students with lower technical proficiency or a higher sensitivity to motion reported greater difficulty adjusting to the iVR environment. These heterogeneous responses underscore the necessity of accounting for user-specific factors, particularly susceptibility to physical discomfort and technical competence, when integrating iVR into SR learning environments for its effective, inclusive, and sustainable use for EFL learning.

In contrast, reflections from Group 1 revealed fewer physical or technical difficulties, likely due to the familiarity of traditional PC interfaces and the absence of HMD-related strain. Most students did not report notable challenges; however, several mentioned issues related to user experience, particularly in terms of movement control and perspective adjustment. For example, some found the need to simultaneously click and drag the mouse to adjust their field of view inconvenient, while others noted that teleport-based movement reduced the sense of realism.

5. Discussion

5.1. Relationships between self-regulation strategies and speaking performance

Concerning the first research question, the results indicated that both groups improved speaking outcomes, suggesting that practicing with AI-empowered NPCs in both DVR and iVR facilitated the development of students' speaking skills. The GAM analysis also confirmed that students' SR strategies positively influenced their speaking performance in VR in all 13 categories examined. However, a closer look at the results showed that the categories had different degrees of influence on speaking performance. While certain SR categories—such as *Cognitive processing, Remembering, Idea planning, Interpersonal practice, Assistance seeking*, and *Environment control*—had a strong impact on students' speaking performance, some other categories, such as *Interest enhancement, Motivational self-talk,* and *Anxiety control,* had weaker effects.

Several factors may explain the categories that strongly influenced students' speaking performance. SR strategies include three dimensions: cognitive, sociocultural-interactive, and affective (Oxford, 2011, 2017). Among the SR categories that had a positive impact on students' speaking performances, four belonged to the cognitive domain (*Cognitive processing, Remembering, Idea planning*, and *Assistance seeking*) and two belonged to the sociocultural-interactive domain (*Interpersonal practice* and *Environment control*). In the cognitive domain, the design of the VR environments likely enabled the students' active control of learning. The open-world design allowed them to actively locate the tasks, initiate conversations with AI-empowered NPCs, and monitor their performance. In addition, when they did not understand the NPCs' remarks, they could use the hint or other help functions embedded in the program. These cognitive processes required their awareness (of communicative tasks), attention (to the context and NPCs), intention (about the goals of the tasks), and effort (to complete the tasks), which are four essential elements of consciousness to activate SR learning (Schmidt, 2010). Furthermore, as the VR environment emotionally engages learners in the learning situation, learning in a VR setting can lead to deeper cognitive processes regarding the content itself (Chen & Hsu, 2020; Vesisenaho et al., 2019), which may suggest that the students' conscious cognitive processes during the activities played a significant role in enhancing their speaking performance.

In the sociocultural-interactive domain, the use of anthropomorphic NPCs in realistic, communicative tasks may have activated the students' SR strategies and positively affected speaking performance. First, according to Lee and Jeon (2024) and Chen et al. (2024), the design of chatbots with human-like features influences how students perceive and interact with them. In their studies, learners tended to perceive chatbots with human-like visuals or voices as social agents/partners, which led to an enhanced learning experience. In a similar way, interacting with AI-empowered NPCs in the VR environments in this study may have helped the students effectively apply their SR strategies for English speaking, such as *Interpersonal practice*, during their speaking practice. Second, the presence of lifelike contexts and responsive NPCs could have facilitated student engagement and allowed them to better utilize their SR strategies such as *Environment control*. With a heightened awareness of how their environmental choices, including the selection of NPCs and the context of their conversations, affected their communication, they may have chosen speaking tasks that made them feel more comfortable to engage in or that they found more useful for learning.

However, strategies for SR within the affective domain, such as *Anxiety control* and *Interest enhancement*, had a weaker impact on students' speaking performance. Oxford (2017) argued that learners select different SR strategies based on their conditions, settings, situations, and needs, suggesting that affective strategies may become unnecessary when learners feel motivated and confident. Additionally, the learning environments were perceived to reduce speaking anxiety. In line with Oxford's (2017) perspective, the findings of the present study indicate that students did not actively employ specific strategies for *Interest enhancement* or *Anxiety control* during the activities, as the VR environment itself may have inherently addressed these aspects.

5.2. Increased self-regulation speaking skills

Concerning the second research question, the current study aimed to explore whether participating in speaking activities within the VR environments influenced the students' SR. The comparison of pre- and postsurvey results yielded mixed findings. While some categories (e.g., *Cognitive processing, Remembering, Idea planning, Goal-based monitoring and evaluation, Peer learning*, and *Environmental control*) showed significant changes, others did not. The characteristics of the VR environments seem to have contributed to the increase in these categories. The nonlinear, open-world type of these environments provided immersive, interactive experiences, allowing learners to actively engage in communicative tasks that promote self-exploration and a sense of agency (Lee & Ahn, 2025), which, in turn, fostered the development of SR (Tseng et al., 2006; Zimmerman, 2008). The students in the VR environments had to navigate tasks independently, locate NPCs, understand the goals of each communicative task, and perform the tasks mostly on their own. In addition, although each NPC had a pre-defined role with structured conversational tasks, students exercised greater control over their learning by choosing conversation topics aligned with their interests and adjusting the dialogue flow based on their individual preferences. By engaging in these cognitive processes of planning, monitoring, and controlling in the VR environments, the students seemed to have developed SR strategies in certain categories, both directly and indirectly from performing the tasks (Oxford, 2017; Rubin, 1981).

A key implication of these findings is that enhanced SR strategies developed in immersive VR environments may extend beyond the VR setting and influence long-term language learning outcomes. Students who actively practice cognitive and metacognitive strategies in VR may be better equipped to apply these strategies in real-world speaking situations. Additionally, students who develop strong SR strategies in immersive digital environments may become more independent learners, capable of setting goals, seeking out conversational opportunities, and refining their speaking skills through self-initiated practice. On the other hand, a closer analysis revealed clear differences between the two groups regarding improvements in SR strategies. Group 1 (DVR) showed improvements in only two categories (*Cognitive processing* and *Idea planning*), whereas Group 2 (iVR) demonstrated an increase across all categories except for *Feedback management* and *Motivational self-talk*. A key factor that may explain the lack of improvement in *Motivational self-talk* for the iVR group is the immersive nature of VR itself. In traditional learning environments, learners often rely on motivational self-talk to maintain focus, regulate persistence, and stay engaged with the task. However, in iVR, the high level of presence and interactivity naturally promotes engagement (Makransky & Petersen, 2021), which seems to reduce the need for explicit self-motivation strategies. In other words, VR's ability to sustain learner engagement through presence and interactivity may naturally compensate for strategies that are more explicitly needed in less immersive environments. This highlights how SR strategies may function differently depending on the learning platforms.

The characteristics of each platform likely influenced the study's results for three primary reasons. First, the iVR platform, known for its high fidelity, provided more realistic visual representations with a first-person perspective, resulting in a learning experience that was more immersive and closer to real-world experience compared to the DVR platform. As shown in Chen and Hsu (2020) and Vesisenaho et al. (2019), such enhanced feelings of immersion enables students to "feel like they are actually experiencing the context" (Chen & Hsu, 2020, p. 12). Among the categories that Group 2 showed improvements but Group 1 did not, *Environment control* may be particularly relevant to the fact that, unlike DVR, iVR effectively isolates students from their physical surroundings, minimizing distractions from the external environment (Makransky & Petersen, 2021; Lee et al., 2024a). As argued by Makransky and Petersen (2021) and Mulders et al. (2020), through a greater sense of presence (the feeling of "being there") and enhanced engagement with the virtual learning space, this isolation ("physical immersion") may have contributed to the increase of strategies for controlling the environment ("mental immersion") in ways that supported their learning, which was not possible in the DVR environment.

Second, iVR enabled physical embodiment, while DVR did not. Since language is inseparable from its context and is rooted in our sensory and motor experiences of the world, learning in an embodied environment like iVR can facilitate more effective language processing (Chen & Sevilla-Pavon, 2023; Lee et al., 2024; Makransky & Petersen, 2021). The activation of embodied cognition in iVR plays a critical role in language learning (Makransky & Petersen, 2021). In this context, the idiosyncratic features of iVR likely allowed Group 2 to develop their SR strategies and speaking skills more efficiently through physical embodiment, enhancing their cognitive engagement (Chen & Sevilla-Pavon, 2023). In contrast, students in DVR interacted with NPCs and objects from a third-person perspective, which inherently limits spatial embodiment. This lack of direct bodily engagement may have diminished their sense of immersion and presence, making it harder for them to effectively apply their SR strategies to speaking tasks. Consequently, the lack of physical immersion and embodiment in DVR likely hindered the development of SR strategies compared to the experiences in iVR.

Last, the present study demonstrated that Group 2 experienced higher levels of enjoyment, interest, and motivation, along with lower speaking anxiety. Numerous studies confirm the positive correlation between motivation and SR (e.g., Li, 2017; Oxford, 2017; Zimmerman, 2008). Intrinsic motivation, such as enjoyment, interest, and satisfaction, plays a significant role in enhancing self-regulation (Artino, 2008). Furthermore, Cai and Lombaerts (2024) highlight the dynamic interactions among the learning environment, learner motivation, and self-regulation. According to their findings, a conducive learning environment significantly predicts learner motivation, which, in turn, positively affects SR. In the case of iVR, presence and embodiment appear to play a crucial role in strengthening these interactions. While presence enhances engagement and emotional investment, embodiment heightens learners' sense of agency (Klingenberg et al., 2024). From this perspective, the representational fidelity and immediacy of control of iVR greatly affects students' non-cognitive outcomes, such as motivation and interest (Lee & Ko, 2023; Makransky & Lileholt, 2018). Conversely, DVR's limitations in embodiment may have constrained students' abilities to engage in embodied cognition, thereby limiting their self-regulation in L2 speaking tasks.

Based on the results, this study suggests several pedagogical implications for learning in VR environments. Some students in the iVR group reported discomfort due to physical symptoms such as motion sickness, dizziness, and eye strain, while others faced technical barriers such as difficulties navigating unfamiliar interfaces. These negative experiences may have interfered with the effective use of SR strategies. Although student reflections highlighted a generally positive trend in using iVR, the findings underscore the need for careful design considerations when implementing VR for language learning. To ensure inclusive implementation in self-regulated learning environments where teacher presence is minimal or absent, VR-based instruction may need to offer gradual onboarding procedures and adaptive features that accommodate diverse learner profiles (Wu & Lee, 2025). For instance, providing options to adjust navigation speed, simplify interfaces, or receive real-time system guidance during initial practice sessions can help students in varying levels of comfort and technical proficiency build confidence and autonomy before engaging in full iVR tasks.

Variations in user interfaces and interaction methods across VR platforms may also allocate students' cognitive resources differently, influencing their capacity for SR during learning. Although iVR emerged as a more promising environment in this study, the richness of stimuli and complex environmental details can increase extraneous cognitive load. Such cognitive overload, resulting from overly intricate interfaces and intensive interactive features, may leave fewer cognitive resources available for applying SR strategies effectively (Makransky & Lileholt, 2018). To mitigate this issue, clearly defining learning objectives before engaging in iVR activities can help students focus on key tasks, reducing cognitive overload and enhancing their self-regulatory capacity. Therefore, when selecting VR platforms for language learning, educators should carefully balance immersive qualities with interactive features that actively sustain learner engagement without overwhelming learners cognitively.

Moreover, SR strategies may not develop naturally within specific learning environments; instead, they may require teacher intervention and intentional learning design (Ilishkina et al., 2022). Providing explicit modeling of SR strategies, such as demonstrating how to plan, monitor, and adjust learning behaviors in VR, can help students effectively engage with self-regulation in these environments. Lastly, current SR learning scales may not adequately capture the SR strategies necessary for VR-based learning. Consequently, there is a need to develop a new scale specifically designed for VR environments to better reflect the learning processes involved in these settings.

6. Conclusion

Advanced technologies such as AI and VR have demonstrated significant potential in enhancing L2 learning, leading to their increased adoption in L2 classrooms in recent years. While a substantial body of research has explored the effectiveness of AI chatbots and VR for L2 learning separately, investigations into the synergistic use of AI chatbots within VR environments for L2 speaking remain scarce. The current study is significant in exploring these two cutting-edge technologies - perhaps the most important pedagogical tools for L2 speaking - in a single study. Moreover, this study incorporated SR strategies, another under-researched topic in L2 speaking, and investigated how students' SR strategies affected their speaking performance while and after practicing speaking with AI-empowered NPCs in two different VR environments, DVR and iVR, and vice versa. By doing so, this study provides insights into effective L2 instruction in the new era of technology.

To translate these findings into pedagogical practice, language educators can consider integrating AI-NPCs within VR environments as part of communicative speaking curricula to foster learner autonomy and engagement. In particular, tasks that simulate real-world interactions can encourage students to make choices aligned with their personal learning goals and comfort levels. Moreover, offering both DVR and iVR modalities can provide greater flexibility, allowing educators to accommodate learners' preferences, technological constraints, and physical sensitivities. Such flexibility supports more inclusive approaches to immersive L2 instruction.

Despite its contributions, the study has several limitations. First, the small sample size restricts the generalizability of the findings. Given the complex interplay of various factors in different learning contexts, the results should not be overgeneralized without further validation through larger-scale studies. Second, other potential factors such as novelty effects, cognitive load, and technical difficulties were not accounted for in this study. These variables may have significantly influenced or interfered with students' use of SR strategies. Future research should, in particular, explore how variations in cognitive load across VR platforms affect students' ability to self-regulate. Additionally, while tasks in this study were presented at different proficiency levels, allowing students to select tasks according to their ability, it remains unclear whether the AI-NPCs adjusted their language complexity based on students' proficiency levels. Investigating whether the NPCs adapted their responses dynamically would require a detailed analysis of student-NPC interactions, highlighting the need for further research examining dialogues patterns to gain deeper insights into how AI-driven interactions align with learners' linguistic capabilities. Moreover, this study did not focus on proficiency-based differences in speaking performance and SR strategies, as its primary aim was to examine the impact of VR environments (DVR vs. iVR) on SR strategies. However, exploring how learners of different proficiency levels engage with self-regulated learning strategies in VR-based speaking practice could provide valuable insights into how individual differences influence learning experiences in iVR settings. Lastly, as this study was conducted over a short term, the long-term effects of utilizing these technologies for L2 speaking remain unexplored. Future longitudinal studies are necessary to elucidate the sustained impact of AI-empowered VR environments on L2 speaking proficiency and SR strategies. As these technologies continue to evolve, continuous research in this area is crucial in shaping effective, technology-enhanced L2 learning environments.

Acknowledgement

This work was supported in part through Institute of Information & Communications Technology Planning & Evaluation (IITP) under the metaverse support program to nurture the best talents (IITP-2024-RS-2024-00425383) grant funded by the Korea government (MSIT) and in part through NRF-2023S1A5A2A21083590.

References

Alrehaili, E. A., & Osman, H. (2022). A virtual reality role-playing serious game for experiential learning. *Interactive Learning Environments, 30*(5), 922-935. https://doi.org/10.1080/10494820.2019.1703008

Alvarez, R. P., Jivet, I., Perez-Sanagustin, M., Scheffel, M., & Verbert, K. (2022). Tools designed to support self-regulated learning in online learning environments: A systematic review. *IEEE Transactions on Learning Technologies, 15*(4), 508–522. https://doi.org/10.1109/TLT.2022.3193271

Artino, A.R., Jr. (2008). Cognitive load theory and the role of learner experience: An abbreviated review for educational practitioners. *AACE Journal, 16*(4), 425-439. https://www.learntechlib.org/primary/p/25229/

Cai, J., & Lombaerts, K. (2024). Self-regulation matters: Examining the relationship between classroom learning environments and student motivation through structural equation modeling. *Social Psychology of Education, 27*, 411–434. https://doi.org/10.1007/s11218-023-09827-6

Chen, Y-L., & Hsu, C-C. (2020). Self-regulated mobile game-based English learning in a virtual reality environment. *Computers & Education, 154*, 103910. https://doi.org/10.1016/j.compedu.2020.103910

Chen, Y.-T., Li, M., Cukurova, M., & Jong, M. S.-Y. (2024). Incorporation of peer-feedback into the pedagogical use of spherical video-based virtual reality in writing education. *British Journal of Educational Technology, 55*(2), 519–540. https://doi.org/10.1111/bjet.13376

Chen, H. I., & Sevilla-Pavon, A. (2023). Negotiation of meaning via virtual exchange in immersive virtual reality environments. *Language Learning & Technology, 27*(2), 118–154. https://doi.org/10125/73506

Chen, X., Wang, C., & Kim, D. H. (2020). Self-regulated learning strategy profiles among English as a foreign language learners. *TESOL Quarterly*, *54*(1), 234-251.

Cho, Y. H., & Lim, K. Y. (2017). Effectiveness of collaborative learning with 3D virtual worlds. *British Journal of Educational Technology, 48*(1), 202-211. https://doi.org/10.1111/bjet.12356

Derakhshan, A., & Fathi, J. (2024). Growth mindset, self-efficacy, and self-regulation: A symphony of success in L2 speaking. *System, 123*, 103320. http://dx.doi.org/10.1016/j.system.2024.103320

De Vrind, E., Janssen, F. J., Van Driel, J. H., & De Jong, N. H. (2024). Improving self-regulated learning of speaking skills in foreign languages. *The Modern Language Journa*l, *108*(3), 601-624. https://doi.org/10.1111/modl.12953

Di Natale, A. F., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology*, *51*(6), 2006-2033. https://doi.org/10.1111/bjet.13030

Du, Q. (2025). How artificially intelligent conversational agents influence EFL learners' self-regulated learning and retention. *Education and Information Technologies*. Advance online publication. https://doi.org/10.1007/s10639-025-13602-9

Ilishkina, D., de Bruin, A., Podolskiy, A. Volk, M., & van Merriënboer, J. (2022). Understanding self-regulated learning through the lens of motivation: Motivational regulation strategies vary with students' motives. *International Journal of Educational Research, 113*, 101956. https://doi.org/10.1016/j.ijer.2022.101956

Kim, H., Yang, H., Shin, D., & Lee, J. (2022). Design principles and architecture of second language learning chatbot. *Language Learning & Technology, 26*(1), 1-18. https://doi.org/10125/73463

Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning, 37*(1-2), 1-26. http://dx.doi.org/10.1080/09588221.2021.2021241

Jeon, J., & Lee, S. (2024). The impact of a chatbot-assisted flipped approach on EFL learner interaction. *Educational Technology & Society, 27*(4), 218-234. https://doi.org/10.30191/ETS.202410_27(4).RP12

Klingenberg, S., Bosse, R., Mayer, R. E., & Makransky, G. (2024). Does embodiment in virtual reality boost learning transfer? Testing an Immersion-Interactivity Framework. *Educational Psychology Review*, *36*(4), 116. https://doi.org/10.1007/s10648-024-09956-0

Lan, Y. J. (2020). Immersion, interaction and experience-oriented learning: Bringing virtual reality into FL learning. *Language Learning & Technology, 24*(1), 1–15. https://doi.org/10125/44704

Lee, J., & Ko, Y. (2023). Effects of self-regulation, goal orientation, and anxiety on EFL speaking in metaverse and face-to-face contexts. *English Teaching, 78*(4), 219-248. http://dx.doi.org/10.15858/engtea.78.4.202312.219

Lee, J. S., & Lee, K. (2020). Role of L2 motivational self-system on willingness to communicate of Korean EFL university and secondary students. *Journal of psycholinguistic research, 49*(1), 147-161. https://doi.org/10.1007/s10936-019-09675-6

Lee, S. M., & Ahn, T. (2025). L2 learner experiences in a playful constructivist metaverse space. *ReCALL*, *37*(1), 129-145. https://doi.org/10.1017/S0958344024000235

Lee, S., & Jeon, J. (2024). Visualizing a disembodied agent: Young EFL learners' perceptions of voice-controlled conversational agents as language partners. *Computer Assisted Language Learning, 37*(5-6), 1048-1073. https://doi.org/10.1080/09588221.2022.2067182

Lee, S. M., Wang, X., Park, I., & Lestiono, R. (2024a). It feels so real! Situated authentic language learning in immersive virtual reality. *Education and Information Technologies*, 1–23. https://doi.org/10.1007/s10639-024-12807-8

Lee, S. M., Yang, Z., & Wu, J. G. (2024b). Live, play, and learn: Language learner engagement in the immersive FR environment. *Education and Information Technologies, 29*(9), 1052910550. https://doi.org/10.1007/s10639-023-12215-4

Li, K. (2017). *Motivational regulation and L2 Learning*. Palgrave Macmillan.

Li, Y. F., Guan, J. Q., Wang, X. F., Chen, Q., & Hwang, G. J. (2024). Examining students' self-regulated learning processes and performance in an immersive virtual environment. *Journal of Computer Assisted Learning*, *40*(6), 2948-2963. https://doi.org/10.1111/jcal.13047

Liu, Z., Yu, P., Liu, J., Pi, Z., & Cui, W. (2023). How do students' self-regulation skills affect learning satisfaction and continuous intention within desktop-based virtual reality? A structural equation modelling approach. *British Journal of Educational Technology*, *54*(3), 667-685. https://doi.org/10.1111/bjet.13278

Liu, Z., Zhang, W., & Hu, L. (2024). Can desktop virtual reality effectively enhance academic achievement?—A meta-analysis. *Innovations in Education and Teaching International*, *62*(3), 795-810. https://doi.org/10.1080/14703297.2024.2354753

Makransky, G., & Lileholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development, 66*(5), 1141-1164. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*(3), 937-958. https://doi.org/10.1007/s10648-020-09586-2

Makransky, G., Wismer, P., & Mayer, R. E. (2019). A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. *Journal of Computer Assisted Learning, 35*(3), 349-358. https://doi.org/10.1111/jcal.12335

Mulders, M., Buchner, J., & Kerres, M. (2020). A framework for the use of immersive virtual reality in learning environments. *International Journal of Emerging Technologies in Learning, 15*(24), 208–224. https://doi.org/10.3991/ijet.v15i24.16615

Nong, Y., Zhang, H. T., & Sun, J. Q. (2025). Natural Language Processing for Immersive Game Interactions: Improving NLP Models for More Natural Conversations with AI-driven NPCs in AR/VR Games. *Authorea Preprints*.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Newbury House Publishers.

Oxford, R. L. (2011). *Teaching and researching: Language learning strategies* (1st ed.). Routledge. https://doi.org/10.4324/9781315838816

Oxford, R. L. (2017*). Teaching and researching language learning strategies: Self-regulation in context.* Routledge. https://doi.org/10.4324/9781315719146

Parmaxi, A. (2020). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments, 31*(5), 1448–1469. https://doi.org/10.1080/10494820.2020.1765392

Parong, J., & Mayer, R. E. (2020). Cognitive consequences of playing brain-training games in immersive virtual reality. *Applied Cognitive Psychology*, *34*(1), 29-38. https://doi.org/10.1002/acp.3582

Rubin, J. (1981). Study of cognitive processes in second language learning. *Applied Linguistics, 2*(2), 117–131. https://doi.org/10.1093/applin/II.2.117

Sadanala, G., Xu, X., He, H., Bueno-Vesga, J., & Li, S. (2024). Enhancing distance learning with virtual reality: investigating learners' engagement and outcomes. *Distance Education*, *46*(2), 225-252. https://doi.org/10.1080/01587919.2024.2351929

Sadler, R., & Thrasher, T. (2023). XR: Crossing Reality to Enhance Language Learning. *CALICO Journal, 40*(1), i-xi. https://doi.org/10.1558/cj.25517

Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review of literature. *Computers and Education: Artificial Intelligence, 3*, Article 100060. https://doi.org/10.1016/j.caeai.2022.100060

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker, *Proceedings of CLaSIC 2010*, Singapore, December 2-4 (pp. 721–737). Singapore: National University of Singapore, Centre for Language Studies.

Shen, M., Jiang, M. Y., Jiang, X., & Hung, H.-T. (2025). Harnessing spherical video-based virtual reality to enhance EFL learners' writing performance and self-regulated learning strategy use. *Computer Assisted Language Learning*, 1-46. https://doi.org/10.1080/09588221.2025.2482148

Sobocinski, M., Dever, D., Wiedbusch, M., Mubarak, F., Azevedo, R., & Järvelä, S. (2024). Capturing self-regulated learning processes in virtual reality: Causal sequencing of multimodal data. *British Journal of Educational Technology*, *55*(4), 1486-1506. https://doi.org/10.1111/bjet.13393

Sun, P. (2022). Strategic self-regulation for speaking English as a foreign language: Scale development and validation. *TESOL Quarterly, 56*(4), 1369–1383. https://doi.org/10.1002/tesq.3132

Tan, R., Djonov, E., & Chik, A. (2024). Time matters: A critical multimodal study of an English learning app for children in China. *Media and Technology,* 1–17. https://doi.org/10.1080/17439884.2024.2384857

Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics, 27*(1), 78–102. https://doi.org/10.1093/applin/ami046

Uztosun, M. S. (2020). The development of a scale for measuring the self-regulated motivation for improving speaking English as a foreign language. *The Language Learning Journal, 48(*2), 213–225. https://doi.org/10.1080/09571736.2017.1335766

Uztosun, M. S. (2021). Foreign language speaking competence and self-regulated speaking motivation. *Foreign Language Annals, 54*(2), 410–428. https://doi.org/10.1111/flan.12559

Vesisenaho, M., Juntunen, M., Häkkinen, P., Pöysä-Tarhonen, J., Fagerlund, J., Miakush, L., Miakush, I., & arviainen, T. (2019). Virtual reality in education: Focus on the role of emotions and physiological reactivity. *Journal of Virtual Worlds Research, 12*(1), 1–15.

Wang, Y., & Sun, P. (2024). Development and validation of speaking strategies for self-regulated learning questionnaire (S3RLQ): A multidimensional approach. *The Asia-Pacific Education Researcher*, *33*(6), 1339-1350. http://dx.doi.org/10.1007/s40299-023-00807-0

Wu, J. G., & Lee, S. M. (2025). Immersive literacy. In M. Pegrum., & A. Alm (Eds.), *The palgrave encyclopedia of computer-assisted language learning*. Palgrave.

Wu, J. G., Miller, L., Huang, Q., & Wang, M. (2021). Learning with immersive virtual reality: An exploratory study of Chinese college nursing students. *RELC Journal, 54*(3), 697–713. https://doi.org/10.1177/00336882211044860

Wu, J. G., Zhang, D., & Lee, S. M. (2024). Into the brave new metaverse: Envisaging future language teaching and learning. *IEEE Transactions on Learning Technologies*, *17*, 44–53. https://doi.org/10.1109/TLT.2023.3259470

Yeh, Y. L., & Lan, Y. J. (2018). Fostering student autonomy in English learning through creations in a 3D virtual world. *Educational Technology Research and Development*, *66*, 693–708. http://dx.doi.org/10.1007/s11423-017-9566-6

Yin, J., Goh, T. T., Yang, B., & Xiaobin, Y. (2021). Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance. *Journal of Educational Computing Research*, *59(*1), 154–177. https://doi.org/10.1177/0735633120952067

Zhang, D., Wu, J. G., & Fu, Z. (2024). From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. *System*, *127*, 103501.

Zhang, M. (2024). Enhancing self-regulation and learner engagement in L2 speaking: Exploring the potential of intelligent personal assistants within a learning-oriented feedback framework. *BMC Psychology*, *12*(1), 421. https://doi.org/10.1186/s40359-024-01917-0

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166-183. http://dx.doi.org/10.3102/0002831207312909

Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, *22,* 73–101. https://doi.org/10.1006/ceps.1997.0919

Appendix A. Scoring sample

Student's speaking A (Low-intermediate)	Evaluation Categories	Score
I got headache for a week. It feels like sharpened stepping. Is there any steps to give alleviate this symptoms? And I feel tired. So I think I have to go I have to hit the gym.	General Description Delivery Language Use Topic Development Total (average)	3/5 2/5 2/5 2/5 2.3/5
Student's speaking B (Advanced)	**Evaluation Categories**	**Score**
I'm having a sharp tangling pain in my knee. My knee feels numb and tingly. I've been dealing in with a week long episode of tingling and discomfort in my knee. Is there any remedy or treatment to ease the tingling and discomforting my knee? I've been experiencing increased fatigue recently and I was wondering if you had any suggestion on how to address it. I've been considering going to the gym and incorporating exercise into my routine. Do you have any recommendations?	General Description Delivery Language Use Topic Development Total (average)	4/5 4/5 4/5 4/5 4/5