A Study on the Exploring Strategies for the Construction of a Language Laboratory Based on AI & VR: Focusing on Improving L2 Learners' Speaking Proficiency

Yuan FENG · Gyun HEO[†]

Pukyong National University(student · [†]professor)

인공지능과 가상현실 기반 언어실험실 구축전략 탐색 연구: L2 학습자 말하기 능력 향상을 중심으로

펑위안 · 허 균[†]

국립부경대학교(학생 · [†]교수)

Abstract

This study proposes a framework for a language laboratory that integrates Artificial Intelligence (AI) and Virtual Reality (VR) to improve learners' speaking abilities. Literature analysis identified key challenges in speaking proficiency, including insufficient skills, lack of real-life context, and language anxiety. Based on these findings, this study developed an AI- and VR-integrated speaking output training system framework. A qualitative method was used, with expert interviews to validate the framework's theoretical rationale and technical feasibility. Expert feedback indicated that the framework is innovative and valuable, particularly in emotional regulation, personalized learning paths, and immersive interactions. However, further studies are suggested to focus on empirical research, technological costs, emotion recognition accuracy, and privacy protection.

Key words : Speaking output, Language laboratory, Artificial intelligence, Virtual reality, Emotion recognition

I. 서 론

Language laboratories are specialized learning environments designed to enhance students' proficiency in English and other foreign languages. Unlike traditional classrooms, these labs are equipped with advanced tools and facilities for systematic training in listening, speaking, reading, and writing. Through structured programs and targeted exercises, they aim to improve students' pronunciation, grammar, and vocabulary while also stimulating interest and motivation in foreign language learning. As a result, language laboratories play a pivotal role in second language acquisition (Carvalho, 2023; Park, 2024).

Although language laboratories provide abundant tools for listening and pronunciation practice, they often fail to support learners' speaking output. Speaking proficiency, a core goal of language learning, is vital for academic, professional, and social communication, and its weakness can hinder overall English improvement (Kong, 2017).

[†] Corresponding author : 051-629-5970, gyunheo@pukyong.ac.kr

However, traditional language laboratories limit speaking opportunities due to teacher-centered approaches, insufficient resources, and lack of authentic contexts (Xie and Li, 2021). This issue is intensified in non-native environments, where limited interaction opportunities make it difficult to enhance fluency and expression. Besides, oral anxiety and fear of mistakes further diminish learners' motivation to speak (Islam and Roy, 2024).

In recent years, the rapid development of information technology has opened new pathways for learning (Kim et al., 2022). Artificial Intelligence and Virtual Reality are transforming traditional language education models with their interactivity, intelligence, and immersive experiences. AI, through Natural Language Processing and Learning Analytics, offers personalized feedback, real-time error correction, and adaptive learning paths. As these features improve, AI has become widely used in language learning (Hu-Au & Lee, 2018). Meanwhile, VR creates immersive language scenarios, addressing the lack of contextuality in traditional classrooms (Ling, 2023). The integration of AI and VR not only enhances the learning experience but also turns language labs into multifunctional, multimodal interactive platforms.

Constructivism emphasizes that knowledge is built through cognitive processes and experiences (Simpson, 2002; Jiang and Ding, 2012). people construct different types of knowledge in various contexts (Cobb and Bowers, 1999). However, Traditional language laboratories often fail to provide this contextual depth (Bao, 2014), making AI and VR integration particularly relevant for creating authentic learning environments.

Situated cognition theory thinks that language learning is inseparable from authentic contexts and cultural practices (Liu, 2009; Tang and Shen, 2004),

with knowledge serving as a tool for cultural adaptation (Wang, 2005). VR technology consists with this framework by enhancing motivation and reducing anxiety in language learning (Zheng et al., 2021; Kaplan-Rakowski and Gruber, 2023). Thus, this theory emphasizes that authentic contexts are essential for knowledge acquisition, validating VR's role in creating realistic language learning environments.

The Output Hypothesis argues that language acquisition requires active production (Swain, 1995). While traditional classroom settings often inhibit speaking practice due to emotional barriers (Shan and Liu, 2020; Fei, 2023), technological solutions can provide safe practice environments with immediate feedback (Go'zal, 2024), reducing psychological pressure (Liang, 2006) and improving learning outcomes.

With this foundation, the research will investigate the development of a language laboratory framework that integrates AI and VR technologies to aid second language learners in improving their oral performance. The research focuses on the following key questions: How could AI and VR technologies be used to construct a framework to address issues learners face during speaking practice, such as lack of context, insufficient interaction, and emotional barriers? How could the technologies collaborate within the framework to improve learners' speaking experience? How is the functionality design and application value reflected in expert evaluations?

Ⅱ. Research Methods

1. Target Audience and Scope of Application

The framework of this study is primarily aimed at second language learners, especially those who

face difficulties or anxiety in speaking output. The framework intends to enhance learners' speaking ability and confidence by providing an immersive, real-time feedback training environment using AI and VR technologies. It is applicable in language laboratory settings and can be used both as a self-learning tool and in classroom teaching under the guidance of instructors, to meet the needs of various teaching scenarios.

2. Research Methods

This study adopts a qualitative research approach, employing literature analysis and expert interviews to construct and preliminarily validate a framework for supporting speaking output through the collaborative application of AI and VR technologies.

a. Literature Analysis

Research has found that learners' weakest area in language learning is speaking ability, with factors such as insufficient language proficiency, lack of context, sociocultural barriers, and psychological factors contributing to this. Language anxiety is a common obstacle affecting learners' speaking output, and effective emotional management can enhance learners' confidence in speaking and improve their language output skills (Wei et al., 2021). Traditional methods such as practical activities and language exchange courses are effective but have limitations. AI and VR technologies, through tools like speech recognition, grammar correction, and

simulation of real-life situations, offer learners more opportunities to practice speaking. Additionally, by integrating emotion recognition technology, they can effectively alleviate language anxiety, enhancing learners' confidence and fluency in speaking.

b. Expert Interviews

To verify the feasibility of the framework, six experts were invited for semi-structured interviews, including three PhDs in educational technology (focused on the application potential of AI and VR in education) and three PhDs in applied linguistics (focused on strategies and practical value for overcoming speaking output obstacles). Interviews were conducted online, with each expert's session lasting 30-40 minutes. The key interview points are shown in <Table 1>.

Ⅲ. Research Results

1. System Design Positioning

Language laboratories serve both language teaching and language research functions. Learners' language skills cover listening, speaking, reading, and writing, but this study focuses specifically on enhancing learners' speaking output abilities.

The speaking output training system in this study is a technology-driven language learning platform designed to createa personalized, immersive oral training environment for learners through the deep integration of artificial intelligence and virtual reality

<Table 1> Key Points of Expert Interview Outline

Dimensions	Specific Content
Situation Analysis	Speaking Weakness; Solution Effectiveness; Feasibility
Technical Implementation	AI and VR collaboration; Applicaiton of Emotion recognition; Challenges
Emotion Recognition	Necessity; Function; Implementation difficulties
Practical Application	Equipment requirements; Cost analysis; Accuracy; Suggestions

technologies, with the aim of comprehensively improving their speaking output capabilities. The system is intended to help learners overcome issues such as the lack of context, insufficient personalization, and emotional barriers in traditional language teaching, thus enabling more efficient language acquisition. By utilizing intelligent speech analysis technology, the system can identify pronunciation, grammar, and fluency issues in real-time and provide accurate feedback. In combination with virtual reality technology, the system will construct highly realistic language usage scenarios, allowing learners to engage in immersive language interaction in a virtual environment.

In addition, the system will integrate emotion recognition technology to dynamically capture learners' emotional states, such as tension and anxiety, during the training process. Based on learners' psychological states, the system will adjust task difficulty and feedback methods in real time, creating a more supportive learning atmosphere. Through contextual simulation design, the system will embed language objectives into real-life scenarios, such as information exchange and problem-solving, to enhance learners' motivation for speaking output.

Moreover, the system will utilize data analysis technology to track learners' speech performance, emotional states, and task completion status,

generating detailed learning reports to inform personalized path adjustments and teaching optimizations. This system is positioned to meet the needs of individual language learners, the development of new language laboratories, and school education. It will not only provide learners with an efficient speaking output tool but also offer an intelligent, data-driven solution to address oral output obstacles, helping to improve learners' speaking abilities and promote the overall enhancement of their language proficiency.

2. Design of Speaking Output Support Framework Based on AI and VR
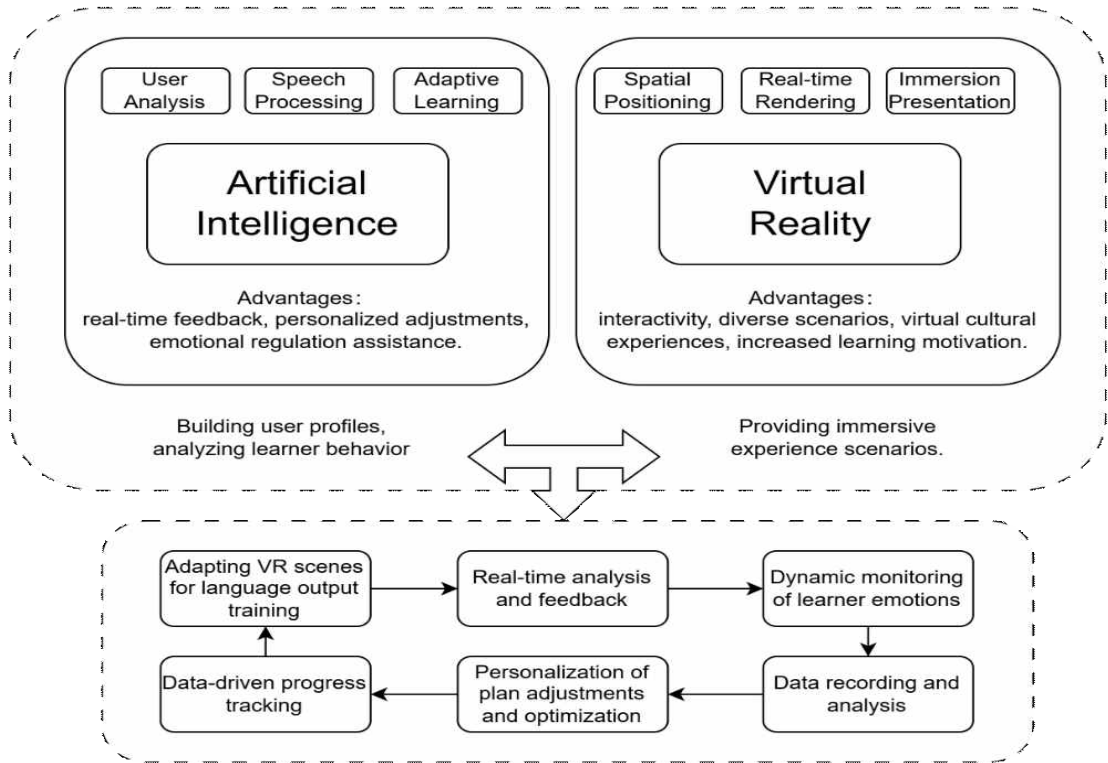
a. Collaborative Application of AI and VR

The collaborative application of AI and VR provides significant advantages for learners' speaking output training, as shown in [Fig 1]. The figure illustrates that these two technologies possess unique functional modules, which, through close interaction and collaboration, jointly create an intelligent and immersive learning environment.

On the AI side, the system is equipped with user analysis, speech processing, and adaptive learning functions. These modules work in coordination to provide real-time feedback, personalized adjustments, and emotional regulation support. By continuously analyzing learners' behavior patterns and learning data, the AI system

<Table 2> System Design Positioning

Elements	Specific Description
Target Users	Learners who need to improve their foreign language speaking skills.
Core Positioning	Personalized, immersive, and interactive speaking training environment.
Usage Scenarios	Autonomous learning in and outside the classroom.
System Features	Real-time emotion monitoring and dynamic adjustments ; AI companionship
Design Purpose	Improve learners' speaking output ability ; reduce language output anxiety
Design Value	Meet the needs for multi-level and multi-scenario language output training

[Fig. 1] Collaborative Application Mechanism of AI and VR in the Speaking Output Training System.

constructs comprehensive user profiles, providing data support for optimizing the learning experience.

At the same time, VR technology collaborates spatial positioning, real-time rendering, and immersive presentation functions to create highly realistic virtual environments for learners. The strong sense of immersion offered by VR facilitates the reflection of learners' real emotions (Llanes Jurado, 2024), enabling the AI to monitor emotions in real time and dynamically adjust scenarios to enhance learners' situational engagement. Moreover, the VR environment is not only highly interactive but also offers a wide variety of scenarios and virtual cultural experiences, effectively boosting learners' motivation and engagement in the learning process.

Therefore, the collaborative application of AI and VR technologies in the speaking output training system integrates scenario simulation, real-time feedback, emotional regulation, and personalized learning pathways. This combination significantly improves the efficiency of language learning and enhances learners' motivation. By collaborating these technologies, the system offers a comprehensive and interactive learning environment. It drives innovation in learning methods and deepens the application of advanced technology. Ultimately, the system fosters a more inclusive and engaging learning atmosphere for learners.

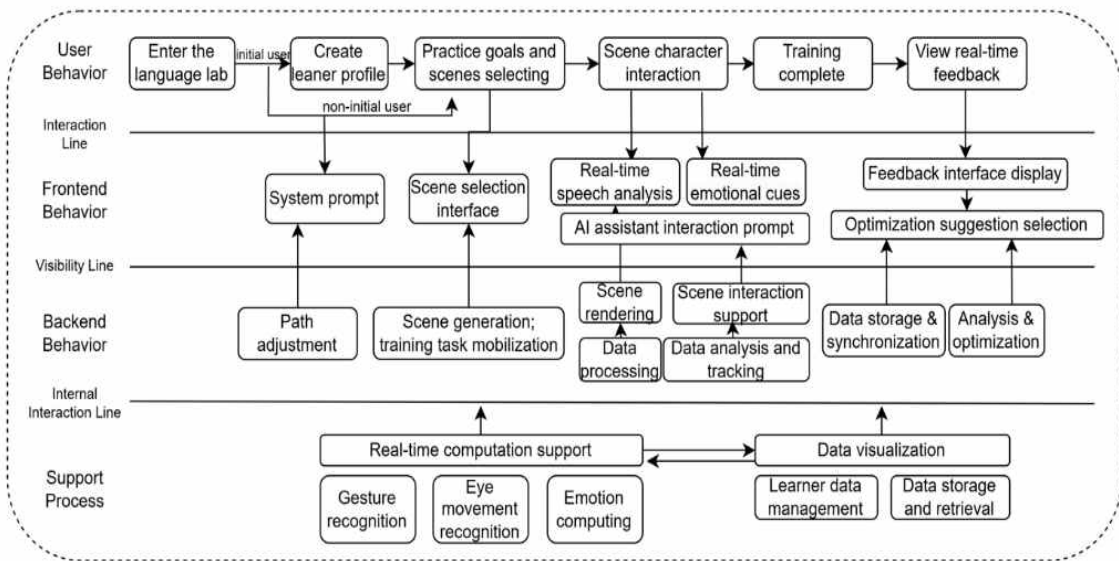b. AI and VR Speaking Output Training System Framework

The AI and VR-powered language laboratory

speaking training system is designed to build a complete, collaborative, and efficient language output support platform for learners. Centered on learners, the system forms a dynamic, closed-loop operating mechanism, including user behavior, front-end interactions, back-end operations, and support processes, as shown in [Fig. 2].

At the user behavior level, learners enter the language laboratory, log into the system, and select their target language, practice tasks, and virtual scenes to begin immersive speaking output training. In these scenes, learners can engage in language output practice, such as situational conversations with virtual characters, simulating communication scenarios in real life, or conducting contextual interactive activities. Each scene is closely tied to a specific language learning objectives, providing immersive contextual support to motivate learners to express themselves in the target language. During thepractice, learners can receive real-time voice analysis and improvement feedback from the system, and after the practice session, they can view personalized learning reports to comprehensively understand their language performance and directions for improvement.

In the speaking output system, learners' frontend behaviors span the entire language practice process, with multiple modules providing interactive experiences. First, system prompts serve as the starting point for learning. Then, learners access the scene selection interface to choose specific contexts based on their individual needs. During practice, the system provides real-time speech analysis to offer immediate feedback on pronunciation, grammar, and fluency. Additionally, it integrates emotion recognition technology to detect learners' psychological states, such as anxiety or tension, and provides timely encouragement to maintain stable performance. Meanwhile, the AI assistant interacts with learners in a human-like manner, dynamically adjusting dialogue content to enhance engagement and practice effectiveness. After completing the

[Fig. 2] The Flow Architecture Diagram of Speaking Training System Based on AI and VR.

task, the system presents performance data through a feedback interface, including speech analysis results, completion scores, and phased recommendations. Learners can also select further practice plans based on optimization suggestions.

The backend behavior serves as the technical foundation of the entire system, integrating the collaborative functions of AI and VR technologies. The AI module analyzes learners, dynamically plans personalized learning paths, and recommends suitable scene tasks for future practice. The VR module, through real-time rendering and dynamic adaptation, embeds the AI-generated analysis results into the scenario. For example, when the AI detects that a learner is speaking too quickly or showing signs of anxiety during a task, the VR module can adjust the tone, speed, and content of the virtual character's interactions, creating a more relaxed learning environment and enhancing the learners' sense of engagement and immersion.

The support process is the foundational guarantee of the entire system framework, integrating data visualization records and real-time computation support. By integrating multimodal data, a comprehensive learner profile is created. These profiles not only provide a basis for personalized scene recommendations and task design but also help the system better adapt to learners' needs in subsequent interactions. At the same time, emotion management runs throughout the support process. Through AI algorithms that dynamically recognize changes in learners' emotional state, such as anxiety, distraction, or lack of confidence, the system can adjust the interaction method in real-time in combination with VR scene design. This helps learners relieve stress, boost their confidence in speaking, and enhance the sustainability of their language output.

The notable advantage of this framework lies in its high level of synergy and dynamic adaptability. The combination of AI and VR technologies enables the system to simultaneously meet the immediate needs and long-term development goals of learners, overcoming the limitations of traditional language learning environments that are often single-faceted and static. Through real-time feedback, emotional regulation, and personalized learning path optimization, the system significantly enhances learners' fluency, accuracy, and adaptability in language output tasks. Ultimately, this framework transforms the language laboratory from a platform that solely relies on technological support into an intelligent interactive environment centered on the learner's experience, creating new possibilities and application prospects for language learning.

3. Results and Feedback from Expert Interviews

To validate the theoretical rationale and practical feasibility of the AI and VR-based speaking output training framework, semi-structured interviews were conducted with six domain experts, including three PhDs in educational technology and three PhDs in applied linguistics. The interviews focused on the framework's theoretical foundation, technical feasibility, and the need for emotion recognition. Experts also highlighted potential challenges in practical implementation and provided valuable feedback.

The experts unanimously acknowledged the framework's innovative integration of technologies and its functional design, particularly the application of emotion recognition technology. They highlighted that this feature effectively alleviates learners' language anxiety and enhances engagement, making it a distinctive aspect of the framework. Several

experts emphasized that the synergy between AI and VR technologies addresses issues such as feedback delays and limited contextual diversity in traditional language laboratories, significantly improving learners' speaking output experiences.

However, the experts also noted that the adaptability and accuracy of AI speech analysis are critical for the real-time feedback function. They recommended further optimization of training data to accommodate learners with varying language proficiency levels and accent characteristics. Additionally, they identified the precision and cross-cultural adaptability of multimodal emotion recognition as key areas for improvement. Integrating data from multiple modalities—such as voice, facial expressions, and physiological signals—was suggested to enhance the performance of emotional analysis and provide more targeted emotional support for learners.

In practical application, experts pointed out that the framework may face challenges such as high hardware costs and data privacy and security concerns. The high cost of VR devices could limit the framework's adoption in resource-poor educational areas. It is recommended to develop lighter or low-cost solutions to enhance system accessibility. At the same time, the emotion recognition and personalized recommendation features involve extensive user data, requiring stronger privacy protection and security measures to ensure legal storage and use of the data. These suggestions are crucial for the design of the system's practical application scenarios and promotion strategies.

Based on expert feedback, the research proposes further optimization directions for the framework. Firstly, on the technical side, it is suggested to improve the adaptability of AI speech analysis and

the accuracy of multimodal emotion recognition to meet diverse user needs. Secondly, in terms of application, exploring more cost-effective device solutions can enhance the feasibility of promoting the system in resource-scarce educational regions. Additionally, strengthening data privacy protection design ensures the system's legality and security. Furthermore, it is recommended to integrate emotion recognition with task dynamic adjustment mechanisms to make feedback and emotional support more personalized, better meeting learners' speaking output needs. These improvements not only enhance the system's practical application effect but also provide clear directions for subsequent framework optimization.

Ⅳ. Conclusion

This study proposes a framework for a language laboratory based on AI and VR technology, focusing on improving learners' speaking output abilities, offering an innovative path for the construction of traditional language labs. Previous studies have demonstrated that AI technologies, such as speech recognition and intelligent assessment, have been widely applied in language learning, significantly improving learning accuracy and effectiveness (Yi, 2024). Meanwhile, immersive scenarios and simulated real-life contexts in VR technology have also been proven to significantly enhance learners' motivation (Li et al., 2023). Building on the standard functions, and unlike previous language learning tools that rely solely on either AI or VR technology, this study attempts to integrate AI and VR technologies more deeply. The framework not only collaborates AI technology for real-time speech analysis and feedback but also enhances contextualized learning experiences through VR technology, addressing issues such as the lack of real-life scenarios and insufficient

interaction in traditional language learning environments. In addition, the system's uniqueness lies in integrating emotion recognition technology, dynamically monitoring learners' emotional states (e.g., nervousness or anxiety), and adjusting task difficulty or providing emotional support to alleviate language anxiety. This emotional regulation enhances learners' confidence and participation, making language output more fluent and natural.

Nevertheless, this study has certain limitations. First, although the results of expert interviews validated the rationality and feasibility of the system framework, the lack of specific user testing and practical application data means that the framework's actual effectiveness and operability require further verification. Then, this study focuses solely on speaking output ability without covering the comprehensive development of listening, speaking, reading, and writing skills. In addition, the reliance on high-performance hardware and data privacy for technical implementation may pose challenges to the framework's widespread adoption.

Accordingly, future research is recommended to focus on the following aspects. First, empirical studies will be critical invalidating the framework's effectiveness and optimizing its design. Through user experiments and feedback data, the framework's practicality and scalability can be further enhanced. Second, future studies could aim to expand the framework's functions to include the integrated development of listening, speaking, reading, and writing skills while optimizing the accuracy and real-time performance of emotion recognition technology. Third, future research could explore the development of low-cost, hardware-light versions of the system while improving data privacy protection mechanisms to ensure sustainability and compliance.

Overall, the proposed framework provides an effective solution for speaking output in language learning and offers a new perspective and practical approach for technology in education. Through continuous improvement and expansion, this framework has the potential for broader application in education, contributing to the modernization, personalization, and equity of language learning.

References

Bao SB(2013). Exploration of Situational Interactive Language Laboratory Construction. Laboratory Research and Exploration, 32(6), 195~198. https://doi.org/10.3969/j.issn.1006-7167.2013.06.053

Carvalho BB(2023). The Use of the Language Laboratory in a Federal University: Reflections on the Relevance of Using the Language Lab for English Learners. Master dissertation, Federal University of Pará.

Cobb P and Bowers J(1999). Cognitive and Situated Learning Perspectives in Theory and Practice. Educational Researcher, 28(2), 4~15. https://doi.org/10.3102/0013189X028002004

Fei W(2023). An Empirical Study on the Impact of Foreign Language Learning Anxiety on College Students' Comprehensive English Ability in a Networked Multimodal Environment. Foreign Language and Educational Technology, (3), 89~105.

Go'zal R(2024). Language Acquisition Through Artificial Intelligence. News of the NUUz, 1(1.3.1), 304~307. https://doi.org/10.69617/uzmu.v1i1.3.1.1673

Hu-Au E and Lee JJ(2018). Virtual Reality in Education: A Tool for Learning in the Experience Age. International Journal of Innovation in Education, 3(1), 215~226. https://doi.org/10.1504/IJIIE.2017.091481

Islam M and Roy S(2024). Challenges in Developing Learners' English-Speaking Skills at the Tertiary Level in the EFL Context: Teachers' and Learners' Perceptions. Preprint available at Research Square. https://doi.org/10.21203/rs.3.rs-4491944/v1

Jiang XQ and Ding Y(2012). An Initial Exploration of a New College English Teaching Model under Modern Educational Technology. Foreign Language and Educational Technology, (148), 42~46. https://doi.org/10.3969/j.issn.1001-5795.2012.06.007

Kaplan-Rakowski R and Gruber A (2023). The impact

of high-immersion virtual reality on foreign language anxiety. Smart Learning Environments, 10(1), 46. https://doi.org/10.1186/s40561-023-00263-9

Kim YY, Kim DG, Han SJ, Yun HJ, Park MH and Heo G(2022). A Study on the Development of Usability Evaluation Tools of AI Educational Contents for Students through Delphi Method. The Journal of Fisheries and Marine Sciences Education, 34(2), 256~265. https://doi.org/10.13000/JFMSE.2022.4.34.2.256

Kong LC(2017). A Survey of the Current Situation and Correlation Analysis of College Students' English Speaking Learning. University Journal - Research Edition, (6), 37~43.

Li WH, Qian L, Feng QN and Huang J(2023). Does Increased ImmersionImprove Learning Outcomes? －The Influence of Immersion on Learning Outcomes andIts Mechanism. Journal of Educational Technology & Electronics, (12), 55~63.

Liang SX(2006). Positive Effects of Online Chat on Overcoming Psychological Obstacles in Oral Communication. Journal of South-Central University for Nationalities (Humanities and Social Sciences Edition). 26(04), 178~180

Ling W(2023). Artificial Intelligence inLanguage Instruction: Impact on English Learning Achievement, L2 Motivation, and Self-Regulated Learning. Frontiers in Psychology, (14), 1~14. http://doi.org/10.3389/fpsyg.2023.1261955

Liu YL (2009). Research on College English Multimedia Network Teaching Based on Situational Cognition Theory. Educational Technology Research, (7), 113~120.

Llanes Jurado J(2024). Affective computing framework for social emotion elicitation and recognition using artificial intelligence. Doctoral dissertation, Universitat Politècnica de València. https://doi.org/10.4995/Thesis/10251/207112

Park SM(2024) Grounded Theoretical Analysis of the Use of Generative Artificial Intelligence. The Journal of Fisheries and Marine Sciences Education, 36(5), 992~1003. https://doi.org/10.13000/JFMSE.2024.10.36.5.992

Shan H and Liu M(2020). A Strategy Study on College English Listening and Speaking Teaching from the Perspective of Input and Output Hypothesis Theories. Hubei Normal University Journal (Philosophy and Social Sciences Edition), 40(4), 119~123. https://doi.org/10.3969/j.issn.2096-3130.2020.04.022

Simpson T(2002). Dare I Oppose Constructivist Theory? The Educational Forum, (4), 347~354. https://doi.org/10.1080/00131720208984854

Swain M(1995). Three Functions of Output in Second Language Learning. In Cook C and Seidlhofer B (eds), Principles and Practice in Applied Linguistics. Oxford: Oxford University Press.

Tang FL and Shen JL(2004). Theoretical Basis and Teaching Conditions of Situational Cognition. Global Education Outlook, (4), 53~57.

Wang WJ(2005). Situational Cognition and Learning Theory in the Development of Constructivism. Global Education Outlook, (4), 56~60.

Wei XB, Wu LL and Chen X(2021). A Study on the Relationship Between Emotional Intelligence, Language Thinking Patterns, and Willingness to Communicate in a Second Language. Foreign Language World (06), 80~89.

Xie T and Li JN(2021). Practical Experimentation on the College English Speaking Teaching Model. Experimental Science and Technology, 19(5), 108~114.

Yi CY(2024). Implementation of an Intelligent Language Learning System Based on Crawler Technology and Intelligent Voice Q & A Algorithm. Automation and Instrumentation, (112 182~185.

Zheng CP, Lu ZH, Liu HY, Wang LL and Han XH(2021). Exploring Chinese EFL Learners' Conceptions of and Engagement in a Self-Developed 3D Virtual Environment. Computer-Assisted Foreign Language Education, (2), 85~92.