Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

An Embodied Projection Recognition System for Situated Learning to Enhance Learning Effectiveness and Self-Reflection Ability

Hui-Ting LIU[a], Zi-Ting DING[a], Su-Hang YANG[c], Jian-Yu WU[a], Jen-Hang WANG[b*], Po-Yao CHAO[d], Yung-Yu ZHUANG[a] & Gwo-Dong CHEN[a]

[a]*Department of Computer Science and Information Engineering, National Central University, Taiwan*

[b]*Research Center for Science and Technology for Learning, National Central University, Taiwan*

[c]*Department of Hospitality Management, Chien Hsin University of Science and Technology, Taiwan*

[d]*Department of Information Communication, Yuan Ze University, Taiwan*

*harry@cl.ncu.edu.tw

Abstract: In service-oriented disciplines such as Hospitality Management Education and Nursing Education, multimodal elements such as bodily movements and facial expressions play crucial roles in addition to learning through oral instruction and written materials. Current digital reality learning systems provide excellent spaces for situated learning and real-time assessment. However, they often lack an environment that simultaneously considers real-time assessments of physical actions and spatial awareness. This omission can prevent students from effectively applying their learned knowledge in real professional settings, thereby reducing the practicality and effectiveness of their learning. To address this issue, our study has developed a teaching assistance system aimed at enhancing learners' experiences through readily available technologies. We established a new learning environment and model that can be used in the classroom, incorporating the concepts of personas and spatial computation proposed by Apple Vision Pro to facilitate real-time assessment and scaffolded guidance in service-oriented learning. This model, which complements traditional learning methods, supports embodied cognition and situated learning, allowing users to interact simultaneously with the physical environment and digital reality. Students can engage in both language and embodied action exercises within this environment. Through the system's real-time evaluation and feedback, students can immediately reflect and adjust their behavior in a nearly real interaction setting, enhancing their practical experience and thus improving their ability to absorb and apply learned skills. Moreover, students' performances in the digital reality are displayed on a large screen in the classroom, allowing peers in the audience area to simultaneously observe each other's acts, thereby creating a theater-style learning environment. This transformation not only allows students to see their own and others' real-time performances but also shifts their role from mere participants to performers. This shift helps build a sense of responsibility and enhances their motivation to learn. This method was tested in the 'Hospitality Practices and Japanese' course at a university of technology, with results showing significant improvements in learning outcomes in practical courses that require physical and verbal interactions compared to traditional digital realities.

Keywords: Situated Cognition, Situated Learning, Embodied Cognition, Multimodal Learning, Self-Reflection, Immediate Feedback

1. Introduction

Traditional classroom instruction often relies on oral explanations and printed materials, placing learners in a passive role where they receive knowledge without significant practical involvement or application experience. The theory of Situated Cognition emphasizes the importance of situated learning within real-life contexts, where learners acquire knowledge through sensory and bodily interactions (Brown et al., 1989). Furthermore, the theory of Embodied Cognition extends this concept by underscoring that learners create multimodal memories through physical movements, which are crucial for cognitive development (Wilson, 2002; Foglia & Wilson, 2013).

Multimodal Learning emphasizes the integration of diverse sensory data such as language, images, music, and movement to enhance student learning, particularly in aspects of understanding and information processing (Moreno & Mayer, 2007; Ochoa & Worsley, 2016). Presenting multiple modal information synchronously can enhance learners' encoding and memory of knowledge, thereby improving learning outcomes (Sankey et al., 2010).

Studies show that when learners engage their bodily senses and movements during interactions, this not only increases their level of engagement in mixed reality simulations (Lindgren et al., 2016), but also enhances their confidence and sense of achievement through the system's provision of immediate positive feedback (Howes & Ying-Leh, 2019). Providing feedback to learners is crucial for enhancing learning outcomes (Black & Wiliam, 1998). Past research on multimodal learning utilizing virtual reality has demonstrated embodied learning effects, such as controlling digital avatars through handheld devices or using depth cameras for motion capture. However, these methods rarely consider interactions with actual physical spaces, such as those involving objects and other individuals, which are essential for applying the results of multimodal learning to real-world settings. Reducing the burden of transferring and applying knowledge from virtual to real environments, and applying this knowledge more authentically, are pressing issues that current research needs to explore further.

In light of this, this study proposes an innovative situated learning system that utilizes embodied projection recognition technology to capture learners' facial expressions, body movements, and interactions with objects and individuals in the physical space, projecting these elements into a virtual environment. This enables synchronization of the learners' positions and movements between the real and virtual worlds, providing an immersive learning experience. The system offers immediate positive feedback through motion recognition and spatial computation. By adopting a third-person perspective, learners can observe their own performance in real environment, receive instant feedback, and engage in self-reflection, which is expected to enhance learning outcomes, increase engagement, and foster confidence and a sense of achievement. To validate the research findings, we propose the following research questions:

(1) Can the application of embodied projection recognition technology in situated learning enhance learners' engagement and sense of presence in virtual environments?
(2) Does projecting learners' actual physical movements and spatial interactions into digital reality aid in multimodal learning and memory formation, thereby improving learning outcomes?
(3) What impact does the system's provision of immediate feedback and third-person self-reflection have on enhancing learners' motivation, confidence, and sense of achievement?
(4) Overall, compared to traditional teaching methods, what advantages does the embodied projection recognition situated learning system offer in terms of enhancing learning outcomes?

2. Literature review

2.1 Situated Learning and Self-Reflection

Learning should be placed in real-life contexts, where knowledge is acquired through bodily senses and practical interaction. This perspective is known as Situated Cognition (Brown et al., 1989). Situated learning emphasizes conducting education in relevant real-world contexts, not only to learn professional knowledge but also to develop the ability to solve practical problems (Lave & Wenger, 1991).

2.1.1 Self-Reflection and Immediate Feedback

McLellan (1996) summarized eight key factors constituting situated learning, namely apprenticeship, storytelling, collaboration, reflection, coaching, multiple practices, articulation of learning skills, and technology. Reflection is a critical component of the learning process, enabling the integration of new experiences into existing cognitive frameworks (Schön, 1983). Research indicates that immediate feedback allows learners to recognize their mistakes in real-time, fostering a deeper understanding of the subject matter (Shute, 2008). Learners are more likely to achieve a state of deep engagement and optimal experience when they receive immediate feedback that challenges them just beyond their current capabilities (Csikszentmihalyi, 1990). The system's provision of immediate positive feedback also helps to enhance confidence and a sense of achievement (Howes & Ying-Leh, 2019).

2.1.2 Situated Learning in Virtual Environments

Creating environments that closely mirror the learning content in reality can be challenging; thus, many studies have begun to explore simulating learning situation through virtual reality technology. For example, allowing students to manipulate objects on the screen with a mouse and keyboard, and utilizing 3D design to explore users' spatial abilities (Li et al., 2020). Alternatively, head-mounted display-based immersive virtual realities enhance interactivity and the sense of presence by allowing users to control a virtual avatar (Kwon, 2019).

However, in service-oriented disciplines such as hospitality management and nursing education, in addition to learning knowledge through scenarios created in virtual environments, interpersonal interactions and the practice of operating relevant equipment are also important.

2.2 Embodied Learning and Multimodal Learning

In recent years, Embodied Cognition and Multimodal Learning have gained increasing attention, demonstrating that learning is not merely an abstract cognitive process in the brain but also critically involves bodily actions and multiple sensory inputs.

2.2.1 Embodied Cognition Theory

Embodied Cognition Theory posits that thinking and cognitive processes are closely linked to physical sensations and motor behaviors, suggesting that our cognition is realized through bodily perceptions and actions (Wilson, 2002). Embodied Learning emphasizes the involvement of the body's sensory cognition and sensorimotor systems in the learning process, indicating that cognition is not limited to the brain but is closely linked to bodily cognitive experiences and behaviors (Lindgren & Johnson-Glenberg, 2013). Compared to traditional teaching methods, embodied learning more effectively stimulates learners' various senses, enhancing engagement and interest in learning.

2.2.2 Multimodal Learning

Multimodality refers to the concept involving interactions among multiple modes of communication, such as language, images, music, and movement (Kress, 2010). Multimodal Learning involves capturing and processing multiple sensory data to facilitate student learning, especially in terms of understanding and information processing (Moren & Mayer, 2007; Ochoa & Worsley, 2016). By synchronously presenting images, speech, and movement, multimodal learning can enhance learners' encoding and memory of knowledge, thus improving learning outcomes.

Research has shown that while multiple representational forms do not necessarily directly improve academic performance, students hold positive attitudes towards multimodal learning elements, finding them helpful for understanding and memorizing learning materials (Sankey et al., 2010). Developing students' multimodal literacy in digital learning environments is also emphasized to meet the demands of the 21st century, covering visual, media, information, and multicultural literacy. Using multimedia tools and multimodal types to foster students' creative expression, multimodal teaching methods help develop students' multicultural literacy, critical thinking, and verbal communication skills, among other interdisciplinary competencies (Fedorenko, 2019).

Some studies have experimented with using voice and gestures for interaction in virtual reality to enhance creativity in design tasks (Wolf et al., 2019). Devices like Kinect have been used for motion capture, allowing learners to answer questions through movements, thereby increasing attention (Kosmas et al., 2019).

Recently, the launch of the Apple Vision Pro has allowed digital content to appear alongside the physical world right before the user’s eyes. We have applied this concept to embodied learning; this study has established a digital theatre where multimodal information, such as learners' expressions, dialogues, movements, and interactions with objects or people in the real space, is synchronized and processed bidirectionally. We hope that through virtual-real interaction, the integration of knowledge, operations, and practical application scenarios can be promoted. The system also provides real-time evaluation and feedback, allowing learners to reflect and adjust during the learning process, thus deepening their understanding of and engagement with the learning practice.

3. System Implementation

3.1 Design Philosophy

In this study, we have developed a teaching assistance system aimed at enhancing learners' experiences through readily available technology. In the context of rapid technological advancement, devices like Apple Vision Pro and Meta Quest 3 are now available on the market. These devices allow users to create personas and interact within digital realities. Our design philosophy is inspired by seamlessly integrating students' real behaviors and interactions into the digital realities, similar to a feature of the Apple Vision Pro. Its camera allows users to simultaneously see the external world and digital content and interact with both.

Unlike Apple Vision Pro, we place cameras within the physical environment. Through this approach, students' digital avatars can remain consistent with their actual locations and movements, allowing them to learn in a more intuitive manner within a digital situational theatre. The system also provides immediate assessments of learners' performances in the scenario, enabling students to self-reflect during learning and cognition moments. Furthermore, teachers can design their teaching materials based on the course content, thereby enabling students to learn in a world composed of both real and digital realities in the classroom.

3.2 Environmental Design

In this study, the classroom is divided into two main areas: the exhibition area and the audience area, as shown in Figure 1.

The exhibition area is equipped with computer, ZED 2i stereo camera, and microphone, along with necessary props such as dining tables and utensils. These props allow students to naturally interact in the physical world, while additional elements are generated digitally through Unity to create scenarios for applying knowledge.

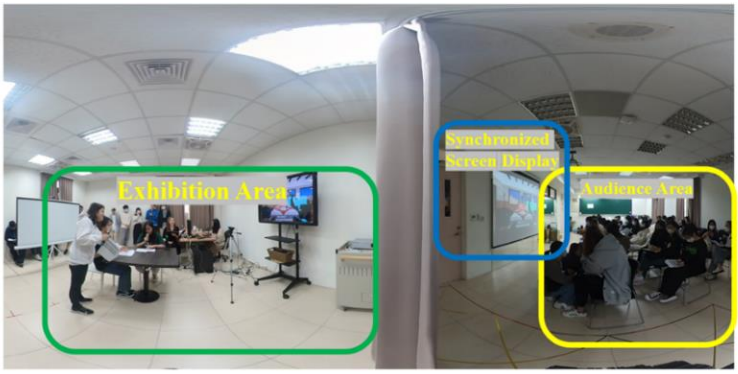

Figure 1. Course Environment Design (360 Degree Camera).

To enhance the learning experience, we have transformed the use of digital reality from a traditional unidirectional interaction, where only the user's perspective was visible, to a bidirectional learning environment. The ZED 2i stereo camera and microphones capture students' dialogue, movements, spatial positioning, and interactions with physical props, synchronously displaying these elements within the digital reality. The system provides immediate assessment results based on the students' performance, combining digital scenery with auditory and visual feedback to offer guidance for behavior correction. Simultaneously, students' performances in the digital reality are projected on a large screen within the classroom, allowing peers in the audience area to synchronously observe each other's performances, thus creating a theater-style learning environment. This transformation not only allows students to see their own and others' real-time performances in the digital theatre, but also shifts their role from mere participants to performers, which helps in building a sense of responsibility, thereby enhancing their motivation to learn.

After each performance, different student groups enter the performance area for practice, where the use of a middle curtain effectively separates the performing students from the audience area, ensuring the focus of the performance and the effectiveness of the interaction.

3.3 System Design

The system primarily based on the Unity development platform, to create an embodied projection recognition digital reality space. For multimodal data capture, as illustrated in Figure 2, user information is captured through a ZED 2i stereo camera and microphone, including 3D skeleton data, 3D human coordinates, user background removal images, and audio data. Interaction behaviors and dialogue recognition are processed using the ZED SDK and Microsoft Azure Cognitive Service, which then replicates the users' real-world performances in the digital reality space.

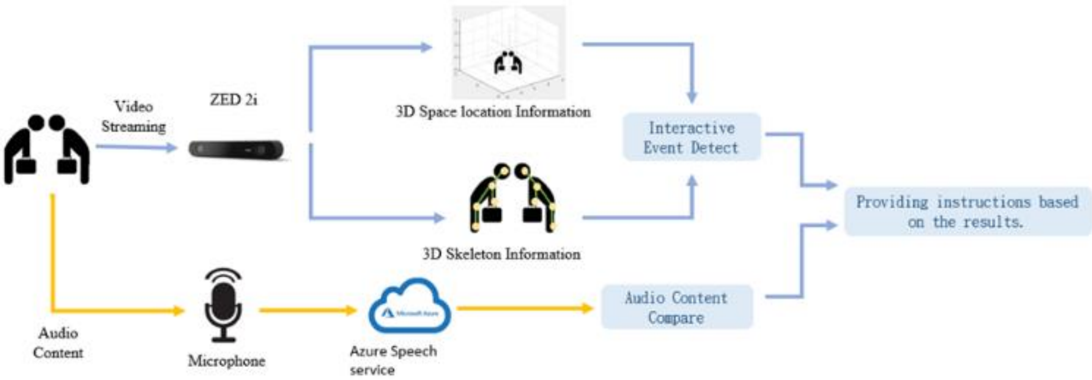

Figure 2. Interactive event Detect Processing Flow.

For the teaching component, as shown in Figure 3, teacher can use the script editing system to develop course content, storing scenario settings relevant to the class in a database.

This includes the design of the script content, objects contained in the scenario, and user action requirements, and designs different responses triggered by various events. The system simulates learning scenario through Unity based on the script content, including architectural elements in the digital reality and object recognition to construct props used in the scene. For example, in this study, the course content was developed in consultation with professor from the Department of Hospitality Management, focusing on hospitality practices. The teaching content covers a complete host process from greeting guests and introducing the menu to finalizing the bill, helping students learn relevant Japanese dialogues and response actions during the hosting process. During the interactive process, the system evaluates whether the user's actions align with the scenario requirements and displays the result on the digital reality. If the actions or the audio content are incorrect, the system will provide visual, auditory, and spatial cues to guide the user to make immediate behavioral adjustments, as show in Figure 4. Finally, the system synchronizes the screen display for other users to watch and record for reflective learning.

Figure 3. Script Editing System and Photo of Teacher Operation.

Figure 4. System Provides Spatial Information and Motion Cues.

4. Research Method

This study aims to explore the impact of a situated learning environment on students' performance within an immersive setting that combines real and digital realities, where personas and multimodal interactions are created using technological products. Our experimental hypothesis is that, compared to a digital situational environment supported only by unidirectional information processing, whether students interacting with peers and receiving real-time multimodal feedback in a digital environment with virtual-real interaction can better understand and apply knowledge content, thereby achieving a better learning outcome.

4.1 Participants

This study was conducted in collaboration with a technological university in Taoyuan City, Taiwan. A total of 82 students from two classes of the 'Hospitality Practices and Japanese' course were selected as subjects. The students were randomly assigned into two groups: one experimental group and one control group. The experimental group included 44 students (11 males and 33 females), and the control group consisted of 38 students (16 males and 22 females). Students were divided into groups of six for the experiment, with each group working independently without interference from the other.

4.2 Test Questions and Questionnaire

To assess the learning effectiveness of the students participating in the experiment, the main content of instruction included basic workplace Japanese dialogues and service etiquette actions used during restaurant service, such as greeting, ordering, serving, and billing. The pre-test and post-test were designed through discussions with the course professor, with both tests focusing on the aforementioned instructional content. The tests included sections on language and physical expressions, with a total of 100 points available (60 points for Japanese and 40 points for actions), and were scored by the same professional teachers.

To understand the students' perceptions, a questionnaire using a five-point Likert scale was designed, ranging from "strongly disagree" to "strongly agree," and the reliability of the questionnaire was assessed using Cronbach's alpha. The questionnaire was based on the Motivated Strategies for Learning Questionnaire (MSLQ) model (Pintrich et al., 1991), which assesses the relationships between learners' motivation, cognitive presence, and learning outcomes. In this study, we explored four dimensions: "Intrinsic Goal Orientation," "Task Value," "Performance and Self-efficacy," and "Peer Learning."

4.3 Experimental Procedure

The experiment was conducted over a period of eight weeks, with the process illustrated in Figure 5. Throughout the experiment, both the experimental group and the control group were taught by teachers specialized in the course. The only difference between the experimental and control groups in the experimental manipulation was in the use of the situational system by the learners; all other aspects, including the learning materials and the learning process, remained the same.

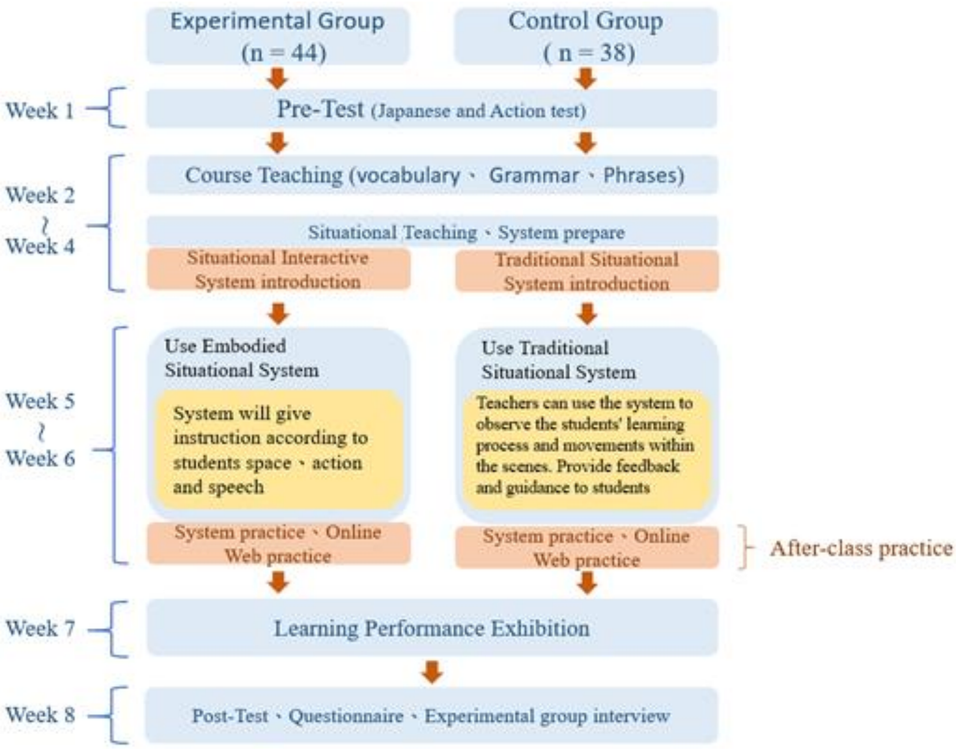

Figure 5. Experimental Procedure.

5. Experimental Results and Analysis

5.1 Pre-test and Post-test Performance Analysis

A total of 82 valid test samples were collected, divided into 44 for the experimental group and 38 for the control group. Data were subjected to tests for randomness and independent sampling, and prerequisites for covariance analysis including tests for normal distribution and homogeneity were conducted, followed by a one-way analysis of covariance (ANCOVA). The covariate was the pre-test score, the dependent variable was the post-test score, and the independent variable was the group.

Independent sample t-tests conducted on the pre-test scores showed no statistically significant difference ($p = 0.250 > 0.05$), indicating equivalent baseline abilities between the experimental and control groups. Normality was tested using the Shapiro-Wilk test, with the significance p-values of the pre-test and post-test results for both groups exceeding 0.05, confirming that all results complied with a normal distribution. Homogeneity of regression was tested, and the interaction of group and pre-test scores showed no significant difference ($F = 0.552$, $p = 0.460 > 0.05$). Levene's Test of Equality of Variance indicated no significant difference in the error variance between the post-test scores of the control and experimental groups ($p = 0.519 > 0.05$).

The results of the one-way ANCOVA revealed a significant difference in post-test scores between the experimental and control groups ($F = 5.309$, $p = 0.024 < 0.05$). Additionally, according to Cohen's effect size indicator (Cohen, 1992), the effect size ($\eta_p^2 = 0.063 > 0.06$) was considered moderate. Thus, these results confirm that the embodied projection recognition situational learning system with virtual-real interaction has a positive impact on learning outcomes compared to digital realities with unidirectional information processing.

Further analysis was conducted on two aspects of the subject matter. In the aspect of Japanese language, there was no significant difference between the groups ($p = 0.098 > 0.05$). In the aspect of physical actions, there was a significant difference ($p = 0.046 < 0.05$). One-way ANOVA results indicated that the experimental group showed significant improvement in the physical action component compared to the control group.

5.2 Questionnaire Results and Interview Analysis

This study's questionnaire was designed based on the Motivated Strategies for Learning Questionnaire (MSLQ) model, with a high reliability coefficient (Cronbach's Alpha) of 0.994. The questionnaire primarily investigated the impact of the system on learners' motivation, divided into four dimensions: "Intrinsic Goal Orientation," "Task Value," "Performance and Self-efficacy," and "Peer Learning." Independent t-tests were conducted to validate the results across these dimensions.

The "Intrinsic Goal Orientation" dimension explored learners' motivation beliefs regarding self-growth and self-efficacy during the learning process, assessing whether the system made participants feel satisfied and strengthened their willingness to learn. The "Task Value" dimension focused on users' understanding and cognitive presence regarding the learning content and scenarios of the course. The "Performance and Self-efficacy" dimension evaluated students' confidence in their abilities and their sense of achievement using the system. The "Peer Learning" dimension addressed whether the system facilitated interaction among peers, as students had to perform together in the system learning activities. Independent t-tests showed that the experimental group scored higher than the control group across all dimensions.

In addition to tests and questionnaires, interviews were conducted with students in the experimental group. In the questionnaire interviews, participants expressed that compared to traditional paper-based learning, experiencing immersive scenarios helped them better understand and remember the content. For example, their comments included: "This is fun, I feel like I'm in the scenes shown to us in the textbook," "I can see myself, and the system gives me prompts to immediately correct my mistakes," "The knowledge from the textbook is no longer an abstract concept," and "I love this teaching method, I can really apply what I learn in class to real situations."

Regarding system usage opinions, participants expressed high interest and curiosity about integrating their movements and expressions into the virtual environment, finding this method novel and fun. They wished the system could more finely represent individual characteristics to enhance identity recognition. Concerning dramatic performance and language learning, some students initially felt nervous but became more natural as they became familiar with the system. Comments included: "Seeing my virtual avatar on the screen mimicking every move I make is an amazing feeling," "It would be even better if I could decorate my avatar," "I was indeed a bit nervous at first use. After all, everyone can see your performance, which is somewhat embarrassing. But as I got used to the system, I began to express myself more comfortably and became more familiar with the content," and "Gradually correcting myself in practice and becoming more skilled was very satisfying; I am pleased with the outcome of my final performance."

Overall, users held a positive attitude towards the use of the embodied projection recognition situational learning system and acknowledged the benefits of engaging in multimodal learning of voice and actions during performance.

6. Conclusion

This study aimed to create an immersive learning environment that integrates real and digital realities to enhance learners' performance. By employing commercially available products such as the Zed 2i stereo camera and generative AI technologies, we successfully projected students' movements and dialogues into the digital reality in real-time, creating personas that are synchronized with their actual actions, while also providing immediate assessment and feedback. The use of this technology not only deepened students' engagement and presence but also enhanced memory formation through multimodal learning, fostering their cognitive connections.

Experimental results demonstrated that this embodied cognition approach significantly improved students' learning outcomes in physical actions compared to digital realities with unidirectional information processing. Particularly, the features providing immediate corrective guidance, feedback, and third-person self-review not only helped students self-adjust during performances but also allowed them to examine their own actions from multiple perspectives, thus boosting their confidence and sense of achievement. Questionnaire and interview results further confirmed the system's positive impact on enhancing learning motivation and effectiveness. Students expressed great satisfaction with the ability to 'see themselves' in the virtual environment and receive immediate feedback, stating that this learning method helped them better understand and remember the instructional content. The digital theatre, where students' actions in the digital reality were displayed on a large screen in the classroom, enhanced learning motivation based on a sense of responsibility.

In summary, this study confirmed the feasibility and benefits of creating an embodied immersive situational learning environment using existing technological products, especially in learning situation that require rich physical and verbal interactions. Future research could further explore how to optimize technology integration and curriculum design to meet a broader range of educational needs. While currently applied to hospitality practice learning, this framework and methodology could be extended to various fields, utilizing different embodied behaviors and cognitive learning styles across different disciplines to explore their extensive impact on learning outcomes.

Acknowledgements

This work was supported by the National Science and Technology Council, Taiwan, under Contract Numbers MOST 110-2511-H-008-004-MY3, MOST 111-2410-H-008-012-MY3, and NSTC 113-2811-H-008-010.

References

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Brown, J. S., Collins, A., & Duguid, P. (1989). *Situated Cognition and the Culture of Learning. Educational Researcher, 18*(1), 32-42. https://doi.org/10.3102/0013189X018001032

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science, 1*(3), 98-101. https://doi.org/10.1111/1467-8721.ep10768783

Csikszentmihalyi, M., & Csikzentmihaly, M. (1990). *Flow: The psychology of optimal experience* (Vol. 1990, p. 1). Harper & Row.

Fedorenko, S. V. (2019). Experience of developing students' multimodal literacy in the digital learning environment of higher education institutions. *Information Technologies and Learning Tools, 69*(1), 12. https://doi.org/10.33407/itlt.v69i1.2405

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(3), 319–325. https://doi.org/10.1002/wcs.1226

Howes, S., & Ying-Leh, L. (2019). The importance of immediate constructive feedback on students' instrumental motivation in speaking in English. *Britain International of Linguistics, Arts and Education, 1*(2), 1-7. https://doi.org/10.33258/biolae.v1i2.58

Kosmas, P., Ioannou, A., & Zaphiris, P. (2019). Implementing embodied learning in the classroom: effects on children's memory and language skills. *Educational Media International, 56*(1), 59–74. https://doi.org/10.1080/09523987.2018.1547948

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kwon, C. (2019). Verification of the possibility and effectiveness of experiential learning using HMD-based immersive VR technologies. *Virtual Reality, 23*(1), 101-118. https://doi.org/10.1007/s10055-018-0364-1

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Li, Y., Yang, Y., Yao, Z., & Xu, G. (2020). Virtual 3D environment for exploring the spatial ability of students. *Virtual Reality & Intelligent Hardware, 2*(6), 556-568. https://doi.org/10.1016/j.vrih.2020.08.001

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by Embodiment: Six Precepts for Research on Embodied Learning and Mixed Reality. *Educational Researcher, 42*(8), 445-452. https://doi.org/10.3102/0013189X13511661

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. Computers & Education, 95, 174-187. https://doi.org/10.1016/j.compedu.2016.01.001

McLellan, H. (1996). *Situated learning perspectives*. Educational Technology.

Moreno, R., & Mayer, R. E. (2007). Interactive Multimodal Learning Environments. *Educational Psychology Review, 19*(3), 309-326. https://doi.org/10.1007/s10648-007-9047-2

Ochoa, X., Worsley, M. (2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics, 3*(2), 213-219. https://doi.org/10.18608/jla.2016.32.10

Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. National Center for Research to Improve Postsecondary Teaching and Learning.

Sankey, M., Birch, D., & Gardiner, M. W. (2010). Engaging students through multimodal learning environments: The journey continues. *Proceedings of the 27th Australasian Society for Computers in Learning in Tertiary Education*, 852-863. http://hdl.handle.net/10072/381625

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Shapiro, L., & Stolz, S. A. (2019). Embodied cognition and its significance for education. *Theory and Research in Education, 17*(1), 19-39. https://doi.org/10.1177/1477878518822149

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research, 78*(1), 153-189. https://doi.org/10.3102/0034654307313795

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625-636. https://doi.org/10.3758/BF03196322

Wolf, E., Klüber, S., Zimmerer, C., Lugrin, J. L., & Latoschik, M. E. (2019, October 14-18). *"Paint that object yellow": Multimodal Interaction to Enhance Creativity During Design Tasks in VR*. In 2019 International Conference on Multimodal Interaction, Suzhou, China.