https://doi.org/10.2352/EI.2026.38.2.SDA-342

A Comparative Study on Memory Strategy Adaptation in XR Vocabulary Learning

Nicko R. Caluya, Yang Minxu, and Damon M. Chandler; Ritsumeikan University, Ibaraki City, Osaka, Japan

Abstract

As extended reality (XR) technologies in education—particularly Augmented Reality (AR) and Virtual Reality (VR)—advance rapidly, there is a growing need to understand how these platforms influence cognitive learning processes. Vocabulary acquisition, a core aspect of second-language learning, relies heavily on memory strategies. However, it remains unclear how platform-specific features, such as contextual anchoring in AR or spatial immersion in VR, interact with different strategies to affect learning outcomes. This study examines how two key memory strategies, semantic association and spatial positioning, perform in AR and VR environments during second-language vocabulary learning. Specifically, it investigates whether a strategy–platform compatibility exists, in which certain strategies may be more effective depending on the platform's cognitive affordances.

Situated Vocabulary Learning in XR Environments

XR environments offer capabilities that are difficult to achieve in traditional computer-assisted language learning systems. First, vocabulary items can be presented with rich visual and spatial context. Words can be paired with images, objects, or scenes that represent them, and these representations can help clarify their meanings. Second, XR environments support embodied interaction. The learner is not limited to simply looking at screens but can move, look around, and interact with learning content through natural actions. VR and AR afford these features differently.

VR creates a fully virtual experience in which all visual and spatial information is digitally constructed. This affords control over layout, visual complexity, and interaction design, and may enable deep, immersive learning experiences due to heightened presence [8]. AR, by contrast, overlays visual information onto the physical environment. Vocabulary items may be attached to physical objects in the environment, aiding learners' use of familiar places as memory cues. Due to these differences, VR and AR may afford different ways of attending to, encoding, and retrieving information [7].

Encoding Strategies in Memory

For vocabulary learning using XR technologies, several approaches can be used for memory encoding. Two common methods are called semantic encoding and spatial encoding. Semantic encoding aims to link a vocabulary item to its meaning through visual or conceptual cues, using objects or scenes that represent the target word. Emphasis is placed on direct meaning associations, a common choice in XR vocabulary systems [6, 9-11].

Spatial encoding also associates words, phrases, or other items with locations or paths through space. The student is expected to recall words by "remembering" where they were placed in the environment or by thinking of a spatial structure. A related category is techniques derived from "spatial mnemonics," which use human spatial memory to aid recall of items arranged in a spatial layout or path (sometimes called a "memory palace") [3][5]. XR environments are particularly suitable for this strategy because they can present stable three-dimensional spaces that learners can explore and remember.

Although both methods have theoretical backing, they use very different cognitive methods: one emphasizes meaning and imagery, while the other emphasizes location and spatiality. For this reason, their effectiveness may depend on how the material is presented and how students learn through the XR world.

Comparing AR and VR as Learning Environments

While interest in XR-based vocabulary learning is growing, important questions remain. Most existing XR vocabulary systems attempt to use semantic encoding strategies, showing pictures or 3D objects of the word or concept they are trying to convey to convey meaning. By contrast, spatial encoding strategies are less frequently examined, even though spatial memory appears to be a very powerful mechanism for learning and memory [4]. How semantic and spatial strategies compare in XR remains an open and interesting question.

Furthermore, many prior studies focus on either virtual reality or augmented reality in isolation, with few examining how the same vocabulary-learning strategy performs across both platforms. However, VR and AR differ in immersion, environmental motionlessness (e.g., persistence, stability), and interaction style [1]. These differences may affect how learners attend to semantic cues or leverage spatial information during learning. Without a direct comparison, it is difficult to determine whether a strategy that works effectively in VR will also be beneficial in AR, or vice versa.

Another challenge is understanding the cognitive demands of different strategy–platform combinations. XR environments can increase engagement, but they may also introduce additional cognitive load related to navigation, interaction, or sensory processing [13]. Spatial strategies may place greater demands on working memory, whereas semantic strategies may reduce cognitive effort by presenting meaning more directly [12]. The interplay between these demands and the characteristics of VR and AR remains underexplored in current research.

These limitations lead to a lack of consensus on how designers and researchers can select effective mnemonic strategies across different XR platforms. A systematic comparison of semantic versus spatial strategies across VR and AR is therefore needed to understand how XR affordances affect vocabulary learning outcomes.

Experiment Task

For this study, we designed and developed an XR-based vocabulary-learning system to enable a controlled comparison of semantic and spatial encoding strategies across VR and AR platforms. The system is designed to deliver vocabulary-learning tasks in immersive environments while maintaining consistency in learning logic and task structure across all experimental conditions.

There are three main components of the system: a learning interface, an interaction mechanism, and a testing interface. The learning interface displays target vocabulary items using semantic or spatial-based encoding strategies, depending on the condition to which participants were allocated. The interaction mechanism allows the learners to explore the learning content in XR, and the testing interface is used to assess vocabulary learning performance post-learning.

A key design principle was consistency across conditions. The same vocabulary set, exposure duration, and task flow were used across all conditions. Only the encoding strategy and the XR platform varied. This was done to ensure that any observed differences were due to experimental factors rather than differences in system design.

Both VR and AR versions of the system were developed using a shared design framework, so that, while the graphics differ and the environments are different, the underlying learning functions and interaction logic are the same across both platforms. This commonality supports side-by-side comparison of encoding strategies, comparatively, across XR platforms under controlled conditions.

During the learning phase of the experiment, participants studied a set of 15 target vocabulary items in a classroom environment presented in XR. All vocabulary items were displayed simultaneously at fixed locations in the environment and remained visible throughout the learning phase. Participants were free to look around the classroom and view different items during that phase.

Spatial Encoding Condition

In the spatial encoding condition, learning vocabulary words is facilitated by having learners form location-based associations, in which each vocabulary item is linked to a specific place in the XR environment. Then, recalling the new word requires remembering its location in the environment rather than its definition.

In the learning phase, words or references are scattered throughout the environment. To review a word, one walks around a structured grid-based space and re-enters the location where the item is. This structured layout does not change through the learning period, allowing stable location-word encoding via a 2D coordinate system-based system (e.g., recall via rows and columns).

Interaction in the spatial condition involves exploring and traversing the XR environment, where learners change their viewpoint and position to access various locations of vocabulary items, learning through repeated spatial exposure to the visual referent of a word without semantic imagery. Learners are shown only a blue halo virtual object to indicate the closest location with a word (Figure 1).

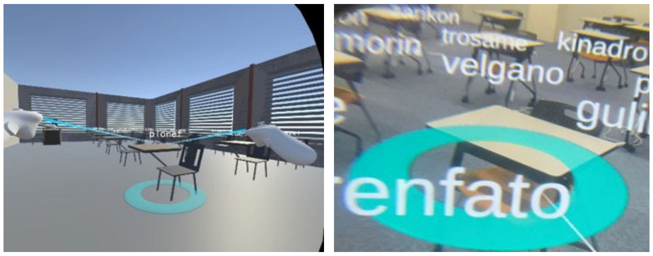

Figure 1. User's perspective for the VR (left) and AR (right) views of the Spatial condition.

Semantic Encoding Condition

In the semantic encoding version of the vocabulary learning conditions, each vocabulary item is paired with graphic representations of its meaning (as images or virtual objects) to optimize learning.

Interaction in the semantic condition emphasizes observation and inspection rather than navigation. In this condition, participants view the pictures from different perspectives and spend varying amounts of time inspecting each vocabulary item. (Figure 2). Again, specific navigation or movement-based memory strategies need not be used. The learning is predicated on forming meaning-based associations between the vocabulary items and the pictures.

By stripping away even more spatial structure but keeping the presentation format constant, this condition avoids the structured layout of the spatial condition and further encourages meaning-based recall.

VR and AR Implementations

The vocabulary learning system was implemented in both virtual reality and augmented reality to enable a direct comparison across XR platforms. Both implementations share the same learning logic, task, and interaction flow, so any differences in learning outcomes can be attributed to the XR platform.

In the VR implementation, all vocabulary learning occurs in a fully virtual environment, where everything is digitally constructed, providing more precise control over how vocabulary items are presented and over the space used in the semantic and spatial encoding conditions. In the AR implementation, the digital learning content is rendered above the real environment. The vocabulary items appear overlaid on the “real world,” allowing learners to interact with the learning content while remaining aware of the physical environment. The spatial arrangement of vocabulary items remains fixed relative to the physical environment throughout the learning session. Despite differences in environmental context and visual presentation, the encoding strategies operate the same way in both VR and AR environments. Both semantic and spatial encoding conditions follow the same learning primitives in VR and AR, with differences between platforms limited to the surrounding environment. This design enables a more focused investigation of how XR platform characteristics affect mnemonic encoding.

Learning Materials and Task Design

Target Vocabulary and Visual Materials

To facilitate controlled comparison across experimental conditions while minimizing the influence of prior language knowledge, we created a target vocabulary of novel, artificial words via a pseudolanguage generator Wuggy [2], so that participants would not have any prior concepts or phonological representations to draw on.

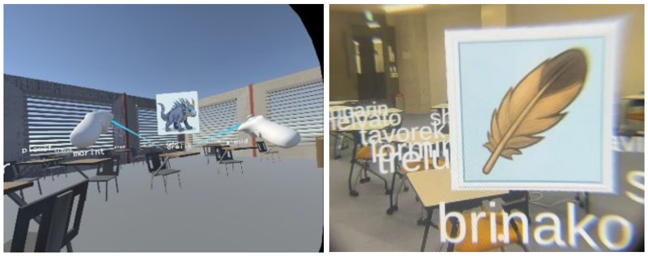

Figure 2. User's perspective for the VR (left) and AR (right) views of the Semantic condition.

Each vocabulary item was paired with a concept that could be mentally represented as a concrete image to promote semantic and spatial encoding. Concepts were selected to be visually distinctive and ideally imageable as a “3D object” (in the sense of a recognizably common object like the sun or a dog). This design decision enabled us to control for similarities in visual stimuli and interpretations.

For the condition in which items were encoded spatially, another set of words was used, but rather than conveying meaning through imagery, each word was presented at a different location within a grid-based XR space, and learners were required to recall the item based on its spatial position and the environment layout. This difference between spatial and semantic layouts is apparent from a top-down perspective (see Figure 3).

In the semantic encoding condition, each target word was represented by a visual cue that denoted its meaning, using static images or basic 3D shapes present within the XR environment. The visual stimuli were not accompanied by additional animation or other extraneous embellishments to reduce unnecessary cognitive load.

Despite these differences appearing to impart different meanings to the words, the vocabulary and visual quality were held constant across conditions to evaluate whether the effect on performance was due to the encoding strategy itself rather than the material used.

Learning Task Design

The learning task was adapted to assess vocabulary acquisition in a controlled XR environment while keeping the encoding strategy and platform constant. Participants completed a learning phase and were instructed to memorize a list of target vocabulary items using the prescribed encoding strategy in VR or AR.

Participants were instructed on how to process the learning task as a function of their condition. In instances where the association may be semantic, participants were asked to help themselves by forming an explicit association between the target word and the picture. In instances where the association may be spatial or route-based between items, participants were asked to pay attention to the location of each word in the room and the route through the environment that connects the two items. No additional mnemonic advice was provided in any form to avoid introducing uncontrolled strategies.

To ensure comparability across XR platforms, we kept item order, task structure, and duration equivalent across VR and AR conditions. The only differences across conditions were the encoding strategy and the platform through which the task was delivered.

Finally, after this learning phase, participants went on to the recall phase. Crucially, there was no feedback on performance in the learning task, nor were participants informed of the results of their recall until the end of the study.

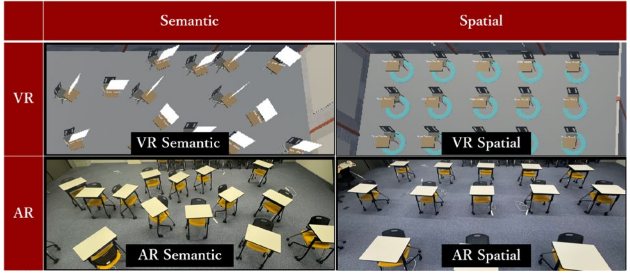

Figure 3. Top-down views of the environment according to the four conditions (Strategy x Platform).

Interaction Constraints and Controls

Interaction behaviors were intentionally limited and standardized across XR platforms to enable comparisons of encoding strategies. Specifically, all subjects interacted with the system using ray-based pointing and selection in VR and AR conditions.

In the VR condition, participants remained stationary at a fixed virtual location (they did not walk in the environment) and could interact only via head orientation and controller-based pointing. This was done to prevent any impact of locomotion differences on spatial encoding or recall.

By contrast, the AR condition allowed participants to move naturally within the physical space. The core affordance of AR will be that spatial information is grounded in real-world locations and body movement. Although movement was permitted, the spatial layout of vocabulary items and the recall interface remained constant.

Participants

Participants (n=24) were recruited from our university, and all were students at the time of testing. Participation was voluntary, and no participant had previously encountered the artificial vocabulary used in the experiment.

All participants had normal or corrected-to-normal vision; severe visual impairments that would affect their ability to view XR content were screened for. Prior use of VR or AR systems was not grounds for exclusion, though participants were asked to describe their prior experience with HMDs and immersive environments for descriptive purposes. Before the experiment began, they were informed of all procedures, including their purpose, and signed a consent form.

Experimental Conditions

Setup and Instructions

At the beginning of each experimental session, the participants were introduced to the XR setup and the basic interaction techniques that they should use for the study. According to the assigned condition, participants used the same XR device (i.e., Meta Quest 3s), and the experimenter ensured the devices were fitted correctly and the visual content was visible.

The participants received standardized verbal instructions explaining the task's purpose and the general procedure. They were told that the study concerned vocabulary learning and that they would complete a learning phase and a recall phase, but were never told what the hypotheses were or what outcomes to expect.

To familiarize participants with the interaction method, a short practice session was conducted in which they performed ray-based pointing and selection tasks on on-screen objects. The practice task involved no experimental vocabulary items and was designed solely to eliminate any novelty effect associated with XR interaction.

Next, participants were shown how to complete the learning task according to their assigned condition. Participants in the semantic encoding condition were instructed to associate each word with its visual form. Participants in the spatial encoding condition were instructed to remember the words' spatial locations in the environment. They were told not to memorize anything externally.

Once participants confirmed they understood the instructions and felt comfortable with the interaction method, the experiment proceeded to the main experimental tasks.

Session Timeline

Every experimental session followed the same timeline. After a brief introductory discussion, the participant was assigned to their XR devices and completed the interaction practice. After the setup phase, participants completed four learning–testing blocks, one per experimental condition. The order of conditions was counterbalanced across participants.

In each block, participants learned 15 target vocabulary items in an XR classroom for 3 minutes, during which all vocabulary items were presented at fixed locations in the XR environment and remained persistent throughout the learning phase. Participants were free to look around the classroom and attend to different items, but were not allowed to revisit the learning content after the block ended. No performance feedback was provided during the learning phase. Immediately after the learning phase, participants completed a recognition-based multiple-choice test, followed by a writing-based recall test without the device. Both tests were completed before participants proceeded to the next learning block.

After completing all four learning–testing blocks, participants completed a post-experiment questionnaire that collected demographic data and information on perceived learning experience, difficulty, and comfort with computer interaction. After completing the questionnaire, participants were debriefed.

The total duration of each session was kept within a consistent time range for all participants to minimize fatigue effects and ensure comparability across conditions.

Results

A two-way repeated-measures ANOVA was conducted to examine the effects of encoding strategy (semantic vs. spatial) and XR platform (AR vs. VR) on recognition-based vocabulary performance (see Figure 4).

The analysis revealed a significant main effect of encoding strategy, $F(1, 23) = 27.63$, $p < .0001$, indicating that participants achieved higher accuracy under semantic encoding than under spatial encoding. The main effect of the XR platform was not significant, $F(1, 23) = 1.05$, $p \approx 0.316$, suggesting no overall difference in recognition performance between AR and VR platforms. The interaction between encoding strategy and platform was also not significant, $F(1, 23) = 0.04$, $p \approx 0.837$, indicating that the effect of encoding strategy was consistent across platforms. Overall, the results indicated a significant main effect of encoding strategy, with semantic encoding outperforming spatial encoding. No significant main effect of the XR platform was observed.

For the results of the written recall test, which served as the transfer task, normality assumptions were assessed prior to inferential analysis and were found to be violated for the writing test data. Therefore, nonparametric statistical methods were applied. These results were also categorized by strategy and platform type (see Figure 5).

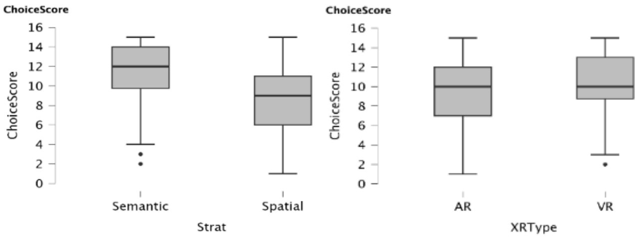

*Figure 4. Average scores from the Immediate Multiple-Choice Recall Test according to **Strategy** and **XR Type***

A Friedman test was conducted to assess overall differences in writing-based recall accuracy across the four experimental conditions (Semantic-AR, Semantic-VR, Spatial-AR, Spatial-VR). The results indicated a significant overall effect of condition, $\chi^2(3) = 10.58$, $p \approx 0.014$, Kendall's $W = 0.147$\$, suggesting that writing performance differed across conditions.

To examine pairwise differences in more detail, we conducted Wilcoxon signed-rank tests. Within each platform, we found that semantic encoding yielded significantly higher writing accuracy than spatial encoding in VR ($Z = 2.74$, $p \approx 0.006$) but not in AR ($Z = 0.95$, $p \approx 0.348$). Across platforms, writing performance under VR semantic encoding was higher than under AR spatial encoding at the uncorrected level ($Z = -2.17$, $p \approx 0.030$), though this difference did not remain statistically significant after correction for multiple comparisons. Other pairwise comparisons did not reach statistical significance. Together, these results indicate that writing-based vocabulary recall was influenced by specific combinations of encoding strategy and XR platform, with semantic encoding in VR showing the most robust performance advantage.

Finally, in both pre- and post-experiment questionnaires, we asked participants to indicate which features were important for memorizing objects alongside words. In the pre-experiment survey, the appearance of objects was deemed most important, with relationships among objects following. After the experiment, however, more participants considered the appearance of the objects the most important feature, rather than the relationships among nearby objects. Notably, participants considered both features more important than the surrounding space or environment (see Figures 6 and 7).

Limitations and Future Work

This study has several limitations that point to directions for future research. First, the relatively small sample size may have reduced statistical power to detect subtle interaction effects between encoding strategy and platform, particularly for the writing-based recall task, where patterns did not always reach conventional significance levels. Increasing the sample size would enable more robust tests of these interaction effects.

Second, the sessions were deliberately short and time-constrained to reflect realistic, brief learning episodes. These constraints may have limited learners' ability to fully benefit from spatial encoding, which often relies on repeated navigation and the consolidation of a stable spatial representation. Longer-term or training-based studies could demonstrate such a benefit.

Third, cognitive load and user experience were assessed via informal self-report rather than standardized instruments. While these responses were informative, future work should incorporate validated cognitive load scales and, potentially, physiological measures to obtain more precise and comparable indices of mental effort.

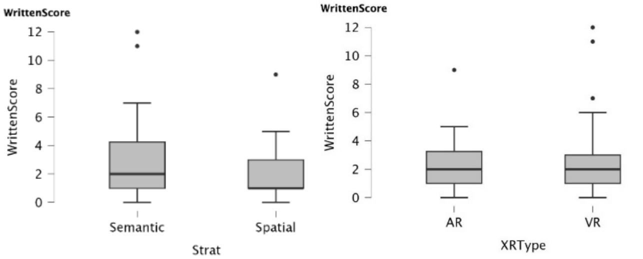

*Figure 5. Average scores from the Transfer Written Recall Test according to **Strategy** and **XR Type***

Fourth, the experiment focused on a controlled classroom-like XR environment and an artificial vocabulary set. This design supports internal validity but limits generalizability to more complex settings, authentic language content, and collaborative or task-based learning contexts. Subsequent studies should test a broader range of content types, spatial scales, and instructional scenarios.

Finally, the present work treated semantic and spatial encoding as distinct strategies. Future research could explore hybrid or adaptive approaches that integrate semantic cues with spatial structure or dynamically adjust mnemonic support based on learner progress and platform characteristics. Such designs may better leverage the strengths of XR to support sustained, flexible vocabulary learning. Overall, these limitations highlight promising avenues for extending the current findings rather than undermining them. By broadening the scope of participants, tasks, measures, and instructional designs, future research can further clarify how XR platforms and mnemonic strategies jointly shape language learning outcomes.

Conclusions

This study investigated how mnemonic encoding strategies and XR platforms jointly influence second-language vocabulary learning. Specifically, it compared semantic and spatial encoding strategies across virtual reality (VR) and augmented reality (AR) environments using a controlled 2x2 experimental design. Vocabulary learning performance was assessed through both immediate choice-based and delayed writing-based recall tasks.

The results showed that the encoding strategy was the principal determinant of the learning outcome: items encoded semantically were remembered better than those encoded spatially, both when recalling items and when recalling word meanings. In the present experiment, this finding suggests that meaningful, but not spatial, visual associations provided better assistance to vocabulary learning. More broadly, this suggests that, typically, semantic cues are useful learning mechanisms for XR-based vocabulary learning, particularly in short bursts, such as during short learning periods.

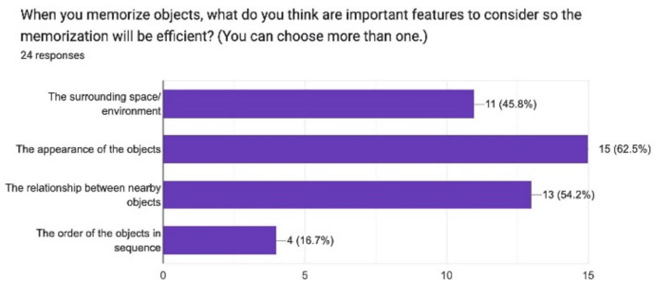

Figure 6. Pre-experiment survey responses for important features considered when memorizing.

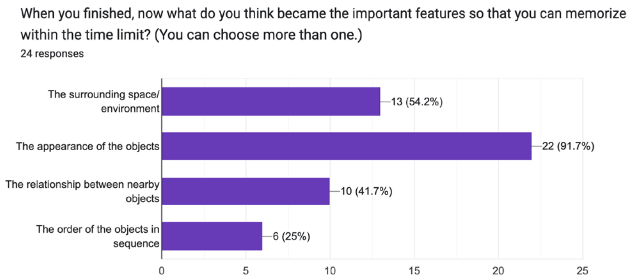

Figure 7. Post-experiment survey responses for important features considered when memorizing.

In contrast, there were no main effects of the XR platform. When controlling task structure, interaction methods, and learning duration, AR and VR supported similar levels of vocabulary learning performance. This suggests that platform modality does not guarantee learning benefits and that instructional design choices may be more important than the intended technology differences.

Although no reliable interaction between encoding strategy and platform was detected in the recognition task, a trend toward an interaction was observed in the writing-based recall task. This finding suggests that platform characteristics may be secondary, task-dependent factors during more demanding types of recall. This caveat should be considered and represents an area for further exploration in future work.

Together, these findings offer insight into how mnemonic strategies operate in XR. Beyond immersion or spatial novelty, the broader implications center on strategy-driven design and on how learners' cognitive capacity and task demands can yield the right cue for deeper encoding. By comparing semantic and spatial strategies across AR and VR formats, this work aims to provide a lens for designing more effective XR vocabulary-learning systems and to suggest areas for further exploration of adaptive, hybrid, and long-term learning in extended reality.

References

[1] Nicko R. Caluya, Xiaoyang Tian, Damon M. Chandler, "Comparison of AR and VR memory palace quality in second-language vocabulary acquisition (Invited)". In: Electronic Imaging. 2023, pp. 220-1-220-5.

[2] Emmanuel Keuleers and Marc Brysbaert, "Wuggy: A multilingual pseudoword generator". In: Behavior Research Methods 42(2010) pp. 627–633.

[3] Sauvik Das et al. "The memory palace: Exploring visual spatial paths for strong, memorable, infrequent authentication". In: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology. 2019, pp. 1109–1121.

[4] Alexia Larchen Costuchen, Stephen Darling, and Clare Uytman. "Augmented reality and visuospatial bootstrapping for second-language vocabulary recall". In: Innovation in Language Learning and Teaching 15.4 (2021), pp. 352–363.

[5] Eric L.G. Legge et al. "Building a Memory Palace in Minutes: Equivalent Memory Performance Using Virtual Versus Conventional Environments with the Method of Loci". In: Acta Psychologica 141.3 (2012), pp. 380–390.

[6] Oscar Rosello, Marc Exposito, and Pattie Maes. "Nevermind: Using augmented reality for memorization". In: Adjunct Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology. 2016, pp. 215– 216.

[7] Marc Ericson C. Santos et al. "Augmented reality as multimedia: the case for situated vocabulary learning". In: Research and Practice in Technology Enhanced Learning 11.4 (2016), pp. 1–23.

[8] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. “The Experience of Presence: Factor Analytic Insights”. In: Presence: Teleoperators and Virtual Environments 10.3 (June 2001), pp. 266–281.

[9] Maheshya Weerasinghe et al. “Arigatō: Effects of Adaptive Guidance on Engagement and Performance in Augmented Reality Learning Environments”. In: IEEE Transactions on Visualization and Computer Graphics 28.11 (2022), pp. 3737–3747.

[10] Fumeng Yang et al. “A virtual reality memory palace variant aids knowledge retrieval from scholarly articles”. In: IEEE Transactions on Visualization and Computer Graphics 27.12 (2020), pp. 4359–4373.

[11] Adam Ibrahim et al. “Arbis pictus: A study of vocabulary learning with augmented reality". In: IEEE Transactions on Visualization and Computer Graphics 24.11 (2018), pp. 2867–2874.

[12] Fred G. W. C. Paas and Jeroen J. G. van Merrienboer, “The efficiency of instructional conditions: An approach to combine mental effort and performance measures". In: Human Factors, 35.4 (1993), pp. 737–743.

[13] Paul A. Kirschner, John Sweller, and Richard E. Clark, “Why minimal guidance during instruction does not work,” In: Educational Psychologist 41.2 (2006), pp. 75–86.

Author Biography

Nicko R. Caluya *received his B.S. in Computer Science from Ateneo de Manila University (2013) and his M.Eng. and D.Eng. in Information Science at Nara Institute of Science and Technology (2018, 2022). He is currently an Assistant Professor in the Visual Information Engineering Laboratory at Ritsumeikan University. His research interests are augmented/virtual reality, human-computer interaction, and educational technology.*

Yang Minxu *finished his Bachelor of Engineering degree and an incoming Master Student at Ritsumeikan University. His research interests are augmented and virtual reality.*

Damon M. Chandler *received the B.S. in Biomedical Engineering from The Johns Hopkins University (1998); and the M.Eng., M.S., and Ph.D. in Electrical Engineering from Cornell University (2000, 2004, 2005). From 2005-2006, he was a postdoc in the Department of Psychology at Cornell. From 2006-2015, he was on the faculty at Oklahoma State University. From 2017-2021, he was an Associate Professor at Shizuoka University. He is now a Professor at Ritsumeikan University, leading the Visual Information Engineering Laboratory. His research focuses on modeling properties of human vision. He is an Associate Editor for the IEEE Journal of Perceptual Imaging and Journal of Electronic Imaging.*