EFFECTS OF AR EMBODIMENT ON COGNITIVE LOAD AND LEARNING EXPERIENCE IN ROBOTIC CONSTRUCTION TASKS

Yalan Mei[1,2], Tan Tan[1,2], Steve Rowlinson[1], Mudasir Hussain[1,2], Kunyi Sun[1]

[1] *Department of Real Estate and Construction, The University of Hong Kong, Hong Kong SAR*
[2] *National Center of Technology Innovation for Digital Construction (NCTI-DC) Hong Kong Branch, The University of Hong Kong, Hong Kong SAR*

Abstract

Augmented Reality (AR) has the potential to enhance human–robot construction interaction and operational efficiency in Industry 4.0 framework. Grounded in embodied cognition theory and cognitive load theory, this study initially investigates how varying levels of embodiment in AR interaction design influence learners' performance, cognitive load, and learning satisfaction. A between-subjects pilot experiment was conducted with 12 participants to compare a high-embodiment interface against a low-embodiment interface. The results preliminarily supported three main hypotheses, revealing that the high-embodiment AR condition significantly reduced extraneous cognitive load and physical demand while substantially improving objective assembly time and subjective learning performance. This research provides pilot foundations to further experimentally explore the theories and hypotheses with wider experiment participations for developing human-centric, task-adaptive AR training and operation systems in the construction industry.

Keywords: AR Embodiment, Robotic Construction, Performance Evaluation, Experiment Research, Interaction Design

1. Introduction

Conventional construction methods gradually reach their technical limits, and the industry is experiencing a technological disruption driven by construction automation and the upcoming ubiquity of robotics (Bock, 2015; De Soto et al., 2018). Complex structures and a large amount of components in construction sites may lead to complicated and dangerous assembly processes (Amouzgar & Willebrand, 2025; Guo et al., 2017; Liang et al., 2021). For construction workers, it is challenging to master the latest automation technologies for robotic assembly due to steep learning curves and high knowledge barriers. Thus, providing construction workers (i.e., robotic operators) with effective and human-centric training and intuitive guidance is significant to ensuring both safety and task success (Park & Kim, 2013). Traditionally, operator rely on 2-Dimensional (2D) manuals or conventional computer screens to understand 3-Dimensional (3D) spatial relationships and sequential actions. This constant mental translation between 2D instructions and the 3D physical workspace imposes a significant cognitive load (Council et al., 2005). When the human working memory is overwhelmed by this extraneous cognitive demand, it not only increases the likelihood of critical operational errors but also severely impairs the overall learning experience, often leading to frustration and decreased motivation (Chu & Ko, 2021).

To address these challenges, 3D visualizations are used to reduce operators' cognitive effort (Li et al., 2024; Yang & Wang, 2023), and there has been an increase in Augmented Reality (AR) instructional designs employing different 3D visualizations (Santos et al., 2013). In engineering, AR could bridge the gap between abstract concepts and physical manipulation, and has been widely applied in hazards recognition and identification, safety training and education, safety instruction and inspection, and so on (Li et al., 2022; Li et al., 2018; Whyte, 2007). Additionally, the advantage is particularly beneficial for complex tasks such as robotic assembly (Mitterberger et al., 2020). By allowing operators to explore objects through actions like rotating and moving virtual models, AR interaction enhances spatial understanding and reducing cognitive load (Buchner et al., 2022; Liu et al., 2021; Thees et al., 2020). Besides, AR learning environments that integrate embodied interaction stimulate operators' sensory and cognitive engagement at multiple levels (Lindgren et al., 2016; Skulmowski & Rey, 2018). In particular, the combination of bodily movements (such as walking and gesture-based manipulation) with virtual content "embodies" abstract engineering concepts, thereby reducing the reliance on working memory alone and improving comprehension and retention (Castro-Alonso et al., 2019; Johnson-Glenberg, 2018; Mayer, 2002). These findings provide guidance for designing AR-based human-robot integration modes: they should encourage operators to "learn by doing and do by learning" in AR, deepening knowledge construction through interaction between the body and the AR environment.

Well-designed AR instructional environments can significantly boost learning effectiveness while lowering cognitive load (Lyu & Deng, 2024). However, the cognitive mechanisms underlying AR interaction remain unclear. While existing studies generally agree that embodied interaction (e.g., head-mounted AR) enhances immersion and bodily engagement, there is still a lack of empirical evidence on how design variables such as spatial alignment, information presentation, and interaction density can be optimized to balance cognitive load and learning experience (Kruijff et al., 2010). Therefore, this study aims to investigate, from an integrated perspective of embodied cognition theory and Cognitive Load Theory (CLT), how varying levels of embodiment in AR interaction design influence operators' performance, especially regarding cognitive load and learning satisfaction.

2. Hypotheses

Current AR-aided applications can be divided into three categories: spatial AR, handheld AR, and head-mounted display (HMD) AR (Fang et al., 2023). These three have different operating distance from the AR system to the human and different operating mode. Besides, a systematic review identifies three conceptual types based on spatial proximity and system architecture for AR-assisted human-robot interaction modes: 1) remote modular interface, 2) proximal modular interface, and 3) proximal integral interface (Tan et al., 2026). Building on these distinctions in spatial proximity and interaction design, this study compares two of the aforementioned AR categories to represent different levels of embodied interaction: 1) HMD AR with proximal modular interface (high-embodiment) and 2) spatial AR with remote modular interface (low-embodiment), to examine how these AR interaction design shape operators' learning states and proposes the following three hypotheses.

We hypothesize that in tasks with high spatial cognitive demands, using high-embodiment AR leads to a decrease in extraneous processing and thus in extraneous cognitive load (H1) compared to low-embodiment AR, by aligning physical actions with digital responses and reducing the mental effort required for spatial mapping. Furthermore, we hypothesize that the level of embodiment has an influence on task outcomes through specific mechanisms. We expect that the positive effect of high-embodiment AR on task performance (e.g., completion time) is mediated by embodied experience, as intuitive physical interaction facilitates problem-solving (H2). Finally, we propose that high-embodiment AR significantly enhances operators' subjective learning satisfaction (H3). We expect this effect occurs because the direct manipulation capabilities lead to increased perceived control and greater immediacy in task feedback, which partially mediate the relationship between the high-embodiment condition and the resulting satisfaction.

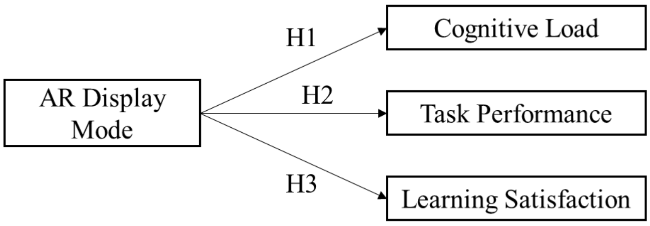

Figure 1. Hypothesis

These three hypotheses form the core theoretical model of this study. Together, they reflect one expectation: that AR interfaces with greater embodiment will not only ease cognitive burdens (H1), but also translate to better task performance (H2) and more positive operators experiences (H3), especially under conditions of high task complexity where such support is most impactful. The model not only systematically examines the instructional effects of different levels of embodiment in AR systems on cognitive and affective outcomes but also attempts to reveal the mediating mechanism (embodied experience) and moderating conditions (task complexity), thus promoting the integrated application of embodied cognition theory and CLT in the context of engineering education. By empirically testing these hypotheses in an engineering education context, we aim to clarify the role of embodiment in AR interaction design and its multifaceted effects on learning.

3. Methodology

3.1 Participants

In a between-subjects design with two conditions, the level of embodiment was manipulated in a robotic construction task using an AR interface. One group received a high-embodiment design, which utilized an HMD AR display to present the target structure from a first-person, egocentric perspective, dynamically aligned with the participant's viewpoint (Group 1). The other group received a low-embodiment design, in which the task was displayed via a screen-based AR setup in a fixed, axonometric view, such that the digital information was spatially separated from the physical workspace (Group 2). Dependent variables are extraneous cognitive load, operators' subjective learning satisfaction, and task performance (measured by completion time).

The participants were recruited from the university student body. Twelve participants (N=12) were recruited and randomly distributed into the two groups (n = 6 per group). All participants possessed basic technical literacy regarding construction and digital fabrication but indicated limited prior experience in robotic arm operation. All subjects gave their informed consent for inclusion before they participated in the study.

3.2 Research Design

All students participated in a modular construction simulation task, involving the assembly of a wall-like structure composed of blocks measuring 5 × 5 × 5 cm and 5 × 5 × 10 cm. The task was executed using a UR5e collaborative robotic arm, which participants controlled remotely via a teach pendant. The task required a sequence of grasping, placement, movement, and adjustment actions to construct the target structure. The AR visualization presented only the final goal state of the modular assembly, without providing intermediate step-by-step instructions. Consequently, students were required to interpret the target configuration and plan the corresponding robotic actions.

In the high-embodiment condition, the visualization supports an embodied interaction mode, allowing participants to directly couple their visual perception with the robot's actions in the physical workspace. This enables participants to observe the task from close range and multiple angles. Conversely, in the low-embodiment condition, participants are required to remain in front of a static display, observing a fixed axonometric view of the digital representation alongside the physical task environment (Wenk et al., 2023). All other aspects of the task, including AR content, block dimensions, task objectives, and the robotic control interface, were held constant across conditions to ensure the task involved the same high spatial cognitive demands.

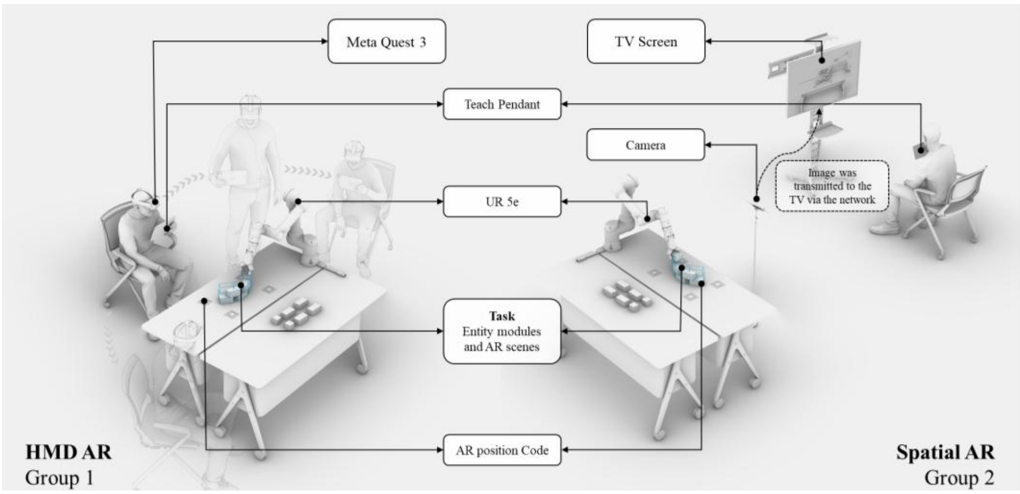

Figure 2. Experimental setup

3.3 Data Analysis

To measure cognitive load and task load, the NASA Task Load Index (NASA-TLX), developed by Hart and Staveland (1988), was used to assess dimensions such as mental demand, physical demand, and effort. To measure subjective learning satisfaction, we adapted the work of Han et al. (2022) and Ibáñez and Delgado-Kloos (2018) and developed a context-specific questionnaire adapted to the AR-supported robotic task (see appendix). Task performance was recorded based on the accuracy of the final physical structure compared to the digital model and the total completion time. Given the exploratory nature of this pilot study, the data were analyzed using descriptive statistics, internal consistency checks, and effect size estimation to evaluate the differences between the high-embodiment and low-embodiment conditions.

3.4 Experiment Procedure

The study took place in a controlled laboratory setting in the university. After the researcher introduced the research to the participants, they were asked for their consent and briefed on the general use of AR tools and the robotic assembly task. After that, the participants performed the construction task using either the head-mounted or the screen-based AR visualization. The participants were asked to continuously compare the physical construction with the AR-defined goal during the process. Afterwards, the NASA-TLX and the learning satisfaction questionnaire were administered. In the end, the session was completed by the researcher.

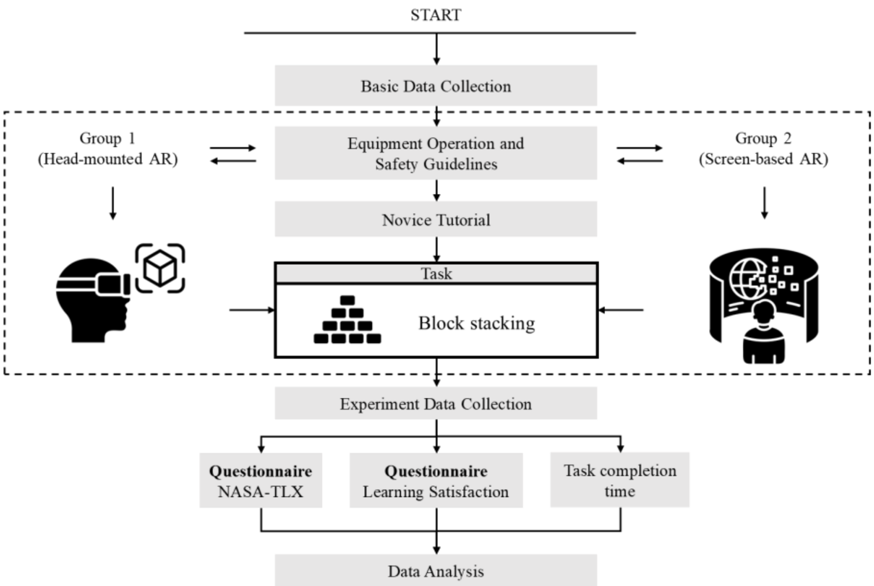

Figure 3. Procedure

Figure 4. Experiment photos

4. Results

To test hypothesis H1, we analyzed the NASA-TLX subscale scores. While the difference in Mental Demand was not statistically significant ($p = 0.153$), the results for Effort ($p = 0.041$) and Frustration ($p = 0.023$) provided support for the hypothesis. Group 1 reported significantly lower levels of Effort and Frustration compared to Group 2. Additionally, Physical Demand was significantly lower for Group 1 ($p = 0.047$). These findings suggest that while the intrinsic mental difficulty of the task remained similar across groups, the high-embodiment condition successfully reduced the extraneous load associated with physical operation and interface interaction, thereby lowering the overall effort and frustration required to map spatial actions.

Table 1. ANOVA results for cognitive load dimensions measured by NASA-TLX

Dimension	F value	P value	Significance
Mental Demand	2.39	0.153	Not Sig.
Physical Demand	5.15	0.047	Sig.
Temporal Demand	0.03	0.870	Not Sig.
Performance	3.76	0.081	Marginal Sig.
Effort	5.52	0.041	Sig.
Frustration	7.20	0.023	Sig.

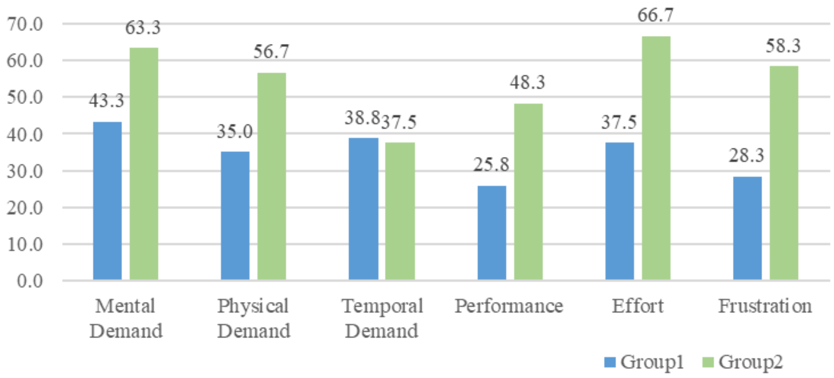

Figure 5. NASA-TLX results comparison

Regarding H2, the data revealed an advantage for the high-embodiment condition. In terms of objective efficiency, Group 1 demonstrated a substantially faster Assembly Time (M = 20.00 min) compared to Group 2 (M = 33.16 min). Subjective measures also corroborated this. Group 1 reported significantly higher ratings for Learning Performance (M = 6.17) and Learning Efficiency (M = 5.83) compared to Group 2 (M = 3.83 for both). Although the NASA-TLX Performance subscale showed only marginal significance ($p = 0.081$), the combination of objective time data and self-reported proficiency strongly supports the main effect proposed in H2, indicating that the intuitive physical interaction facilitated superior problem-solving and execution.

Table 2. Task performance and satisfaction

Dimension	**Questionnaire Item (Abbreviated)**	**G1(Mean)**	**G2(Mean)**
Usability	Confidence in using AR to complete the task	5.83	4.33
Usability	Proficiency in accessing key functions	6.33	4.33
Spatial Guidance	Ability to align blocks using AR overlays	5.67	4.00
Overall Satisfaction	Overall satisfaction with AR mode	6.00	4.83
Interaction Quality	Interaction feels smooth and controllable	5.17	3.67
AR Guidance	AR guidance is clear and meets needs	5.00	4.67
Frustration	Rarely felt frustrated during interaction	5.67	3.33
Learning Support	AR guidance helped reduce assembly errors	5.17	3.83
Cognitive Load	Reduced switching burden during task	5.83	3.83

Behavioral Intention	Intend to continue using this AR mode	6.00	4.17
Behavioral Intention	Intend to use AR for more complex tasks	5.50	3.67
Behavioral Intention	Willing to use AR for future learning	6.00	4.67
System Quality	System ran stably with little disruption	3.83	3.00
AR Accuracy	AR overlays were accurately registered	3.50	3.67
Learning Efficiency	AR improved assembly efficiency	5.83	3.83
Learning Performance	AR improved construction performance	6.17	3.83
Motivation	AR increased learning motivation and engagement	6.17	4.50

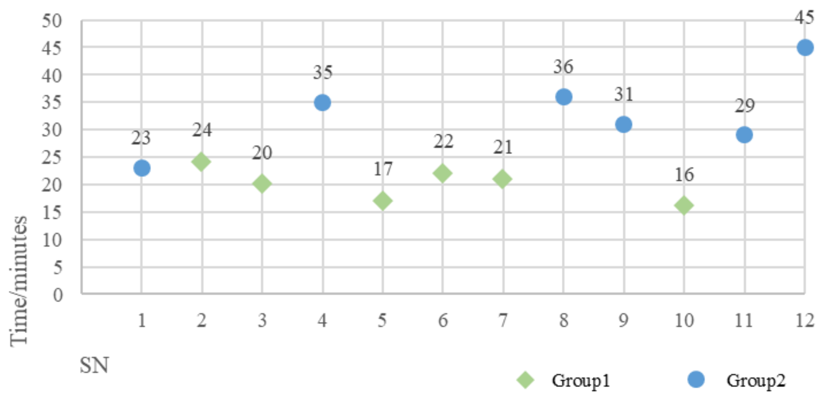

Figure 6. Task completion time comparison

Finally, to test H3, we analyzed the Learning Satisfaction questionnaire dimensions. Group 1 reporting higher Overall Satisfaction (M = 6.00) than Group 2 (M = 4.83). Group 1 rated the smoothness and controllability of the interaction significantly higher (M = 5.17) than Group 2 (M = 3.67). Moreover, Group 1 reported significantly lower frustration during the interaction (M = 5.67 vs. M = 3.33). These findings confirm that the high-embodiment condition provided the necessary immediacy and control to foster a more satisfying learning experience.

5. Discussions and Conclusions

This study focuses on the cognitive challenges of complex construction tasks by examining AR-assisted instructional environments through the integrated perspectives of embodied cognition and CLT. This research compared the effects of high-embodiment (HMD AR with proximal modular interface) and low-embodiment (Spatial AR with remote modular interface) interfaces on operators. The investigation reveals that the level of physical engagement and spatial alignment in AR interaction design significantly shapes operators' cognitive load, task performance, and overall satisfaction. By evaluating these two AR-assisted human-robot interaction modes, the findings establish a foundation for future empirical studies. By testing these theories and hypotheses with larger participant samples, future work can drive the development of human-centric, task-adaptive AR training and operational systems in the construction industry. Given that this is a pilot study with a sample size of 12, the reported p-values should be interpreted as indicative rather than definitive. We intend to validate these preliminary findings through a larger-scale study involving over 50 participants.

Acknowledgement

This research was supported by the Teaching Development Grant (TDG), funded by the University of Hong Kong. Ethical approval for this research was also granted by the Human Research Ethics Committee of the University of Hong Kong (Reference Number: EA250708).

Appendix

Learning Satisfaction Questionnaire

Questions	Score				
	1	2	3	4	5
In this task, I feel confident using the AR system to complete the robotic-arm assembly/construction.					
In this task, I can proficiently use this mode to access key functions (e.g., positioning).					
In this task, I can align the blocks to the correct position and orientation based on the AR overlays.					
Overall, I am satisfied with using this AR mode as a learning support tool in this task.					
In this task, the interaction with this mode feels smooth and controllable.					
In this task, the AR guidance is clear and meets my needs when constructing the robotic arm.					
In this task, I rarely felt frustrated due to the interaction or the system’s responses.					
In this task, the AR guidance helped me reduce assembly errors (e.g., misalignment, missing parts, incorrect orientation, incorrect sequence).					
In this task, this mode reduced the burden of switching between “looking up information—performing the operation—rechecking.”					
Based on this experience, I intend to continue using this AR mode to support my robotic-arm construction learning in the future.					
If the learning task becomes more complex, I would still intend to use this AR mode.					
I am willing to use this AR mode for future learning.					
In this task, the system ran stably (with little lag/interruption) and did not noticeably disrupt the construction process.					
In this task, the AR overlays were accurately registered/aligned (the virtual cues matched the real parts’ positions).					
In this task, using this AR mode improved my efficiency in completing the assembly task.					

In this task, using this AR mode improved my construction performance (accuracy/completeness/quality).	
In this task, using this AR mode increased my learning motivation and engagement.	
*The scores ranged from 1, which means "strongly disagree," to 5 which means "strongly agree."	

Reference

Amouzgar, K., & Willebrand, J. (2025). A novel XR-based real-time machine interaction system for Industry 4.0: Usability evaluation in a learning factory. *Journal of manufacturing systems*, *82*, 254–283.

Bock, T. (2015). The future of construction automation: Technological disruption and the upcoming ubiquity of robotics. *Automation in construction*, *59*, 113–121.

Buchner, J., Buntins, K., & Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, *38*(1), 285–303.

Castro-Alonso, J. C., Ayres, P., & Sweller, J. (2019). Instructional visualizations, cognitive load theory, and visuospatial processing. In *Visuospatial processing for education in health and natural sciences* (pp. 111–143). Springer.

Chu, C.-H., & Ko, C.-H. (2021). An experimental study on augmented reality assisted manual assembly with occluded components. *Journal of manufacturing systems*, *61*, 685–695.

Council, N. R., Earth, D. o., Studies, L., Sciences, B. o. E., Committee, G. S., Spatially, C. o. S. f. T., & Curriculum, T. I. o. G. I. S. A. t. K.-. (2005). *Learning to think spatially*. National Academies Press.

De Soto, B. G., Agustí-Juan, I., Hunhevicz, J., Joss, S., Graser, K., Habert, G., & Adey, B. T. (2018). Productivity of digital fabrication in construction: Cost and time analysis of a robotically built wall. *Automation in construction*, *92*, 297–311.

Fang, W., Chen, L., Zhang, T., Chen, C., Teng, Z., & Wang, L. (2023). Head-mounted display augmented reality in manufacturing: A systematic review. *Robotics and Computer-Integrated Manufacturing*, *83*, 102567.

Guo, H., Yu, Y., & Skitmore, M. (2017). Visualization technology-based construction safety management: A review. *Automation in construction*, *73*, 135–144.

Han, P.-F., Zhao, F.-K., & Zhao, G. (2022). Using augmented reality to improve learning efficacy in a mechanical assembly course. *IEEE Transactions on Learning Technologies*, *15*(2), 279–289.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139–183). Elsevier.

Ibáñez, M.-B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education*, *123*, 109–123.

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, *5*, 375272.

Kruijff, E., Swan, J. E., & Feiner, S. (2010). Perceptual issues in augmented reality revisited. 2010 IEEE international symposium on mixed and augmented reality,

Li, N., Du, J., González, V. A., & Chen, J. (2022). Methodology for extended reality–enabled experimental research in construction engineering and management. *Journal of Construction Engineering and Management*, *148*(10), 04022106.

Li, X., Yi, W., Chi, H.-L., Wang, X., & Chan, A. P. (2018). A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Automation in construction*, *86*, 150–162.

Li, X., Zheng, C., Pan, Z., Huang, Z., Niu, Y., Wang, P., & Geng, W. (2024). Comparative study on 2D and 3D user interface for eliminating cognitive loads in augmented reality repetitive tasks. *International Journal of Human–Computer Interaction*, *40*(23), 8008–8024.

Liang, C.-J., Wang, X., Kamat, V. R., & Menassa, C. C. (2021). Human–robot collaboration in construction: Classification and research trends. *Journal of Construction Engineering and Management*, *147*(10), 03121006.

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.

Liu, Q., Yu, S., Chen, W., Wang, Q., & Xu, S. (2021). The effects of an augmented reality based magnetic experimental tool on students' knowledge improvement and cognitive load. *Journal of Computer Assisted Learning*, *37*(3), 645–656.

Lyu, C., & Deng, S. (2024). Effectiveness of embodied learning on learning performance: A meta-analysis based on the cognitive load theory perspective. *Learning and Individual Differences*, *116*, 102564.

Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. 41, pp. 85–139). Elsevier.

Mitterberger, D., Dörfler, K., Sandy, T., Salveridou, F., Hutter, M., Gramazio, F., & Kohler, M. (2020). Augmented bricklaying: Human–machine interaction for in situ assembly of complex brickwork using object-aware augmented reality. *Construction Robotics*, *4*(3), 151–161.

Park, C.-S., & Kim, H.-J. (2013). A framework for construction safety management and visualization system. *Automation in construction*, *33*, 95–103.

Santos, M. E. C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., & Kato, H. (2013). Augmented reality learning experiences: Survey of prototype design and evaluation. *IEEE Transactions on Learning Technologies*, *7*(1), 38–56.

Skulmowski, A., & Rey, G. D. (2018). Embodied learning: introducing a taxonomy based on bodily engagement and task integration. *Cognitive research: principles and implications*, *3*(1), 6.

Tan, T., Liu, X., Walzer, A. N., Ng, M. S., & Hall, D. (2026). Human–robot interaction design with augmented reality in construction: A systematic review of frameworks, applications, and future directions. *Engineering Management*.

Thees, M., Kapp, S., Strzys, M. P., Beil, F., Lukowicz, P., & Kuhn, J. (2020). Effects of augmented reality on learning and cognitive load in university physics laboratory courses. *Computers in Human Behavior*, *108*, 106316.

Wenk, N., Penalver-Andres, J., Buetler, K. A., Nef, T., Müri, R. M., & Marchal-Crespo, L. (2023). Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment. *Virtual Reality*, *27*(1), 307–331.

Whyte, J. (2007). *Virtual reality and the built environment*. Routledge.

Yang, F.-Y., & Wang, H.-Y. (2023). Tracking visual attention during learning of complex science concepts with augmented 3D visualizations. *Computers & Education*, *193*, 104659.