Methodology for Extended Reality–Enabled Experimental Research in Construction Engineering and Management

Nan Li, M.ASCE[1]; Jing Du, M.ASCE[2]; Vicente A. González, M.ASCE[3]; and Jieyu Chen, S.M.ASCE[4]

Abstract: Extended reality (XR) technologies are increasingly being used as a novel research instrument to facilitate scientific inquires in the construction engineering and management (CEM) domain. By allowing humans to interact with immersive environments in controlled and monitored experimental settings, XR technologies have opened new opportunities for researchers to conduct CEM research involving human participants or concerning human behavior. Yet, XR-enabled research, as an independent, rigorous methodology for the CEM domain, is still underexplored. This paper serves as an effort to build an organized knowledge base and workflow for using XR technologies in various CEM research areas and methodological contexts. The paper first investigates the status quo of XR-enabled CEM research, by identifying current research areas in the CEM domain where XR technologies are considered the preferred or recommended methodological solutions. A process model for XR-enabled research is then proposed, with actionable recommendations about how XR-enabled research should be planned, designed, implemented, analyzed, verified, and validated. This process model is demonstrated with two illustrative case studies. Last, the paper discusses the philosophical, methodological, and technological roots of the evolution of XR-enabled CEM research and describes our vision of more enabling, adoptable, and value-adding XR-enabled research in CEM in the near future. **DOI: 10.1061/(ASCE)CO.1943-7862.0002367.** *© 2022 American Society of Civil Engineers.*

Author keywords: Extended reality (XR); Experimental research; Methodology; Process model; Construction engineering and management (CEM).

Introduction

Extended reality (XR) is a collective term for immersive technologies, including virtual reality (VR), augmented reality (AR), and mixed reality (MR), that can create various real and virtual combined environments and immerse users through visuals, audio, and potentially olfactory and haptic touch cues (Alizadehsalehi et al. 2020). XR technologies are continually gaining momentum in the recent decades, offering innovative ways for users to engage and interact with artificial digital environments that either block out and replace or are overlaid on the real world.

While the construction industry is traditionally believed to be slow in adopting new technologies, we have observed that XR technologies are gradually paving their way to the construction engineering and management (CEM) domain. Methodologically speaking, XR technologies have been used in two different ways in the CEM domain. In many cases, XR technologies are explored by academics to help develop new solutions to a variety of practical problems in the construction industry. Examples of such XR-based solutions include interactive design systems (Reffat et al. 2008), construction project monitoring tools (Golparvar-Fard et al. 2011), engineering education tools (Behzadan and Kamat 2013), and construction safety training protocols (Jeelani et al. 2020). Alternatively, XR technologies have also been employed as a novel research instrument to facilitate scientific inquiries in the CEM domain. By allowing humans to interact with immersive environments in controlled and monitored experimental settings, XR technologies have opened new opportunities for researchers to conduct CEM research involving human participants or concerning human behavior. Here we would like to draw a distinction between the XR-based solutions and XR-enabled research. Specifically, the former refers to technological solutions developed for professionals of the construction industry, where XR technologies are integrated as an essential or peripheral component of the practical solutions; whereas the latter refers to CEM research, which usually addresses human behavior–related research questions, that is conducted in real and virtual combined environments, where XR technologies serve as an enabling research instrument that creates immersive environments needed for achieving certain research goals.

Prior studies that developed and tested XR-based solutions for the construction industry have been well documented and discussed in a few recent review papers (Alizadehsalehi et al. 2020; Li et al. 2018; Wang et al. 2018; Zhang et al. 2020). In contrast, there has been a scarcity of effort to highlight and synthesize XR-enabled CEM research, in spite of an increasing amount of literature that involves XR-enabled methodologies. Based on a holistic analysis of relevant literature and drawing on our own years of experience in conducting XR-enabled research, we believe that there are at least three benefits that XR technologies can bring to the CEM research agenda. (1) The nature of CEM involves both engineering (e.g., mathematical models, simulation) and social sciences (e.g., action research, case studies) research approaches. Thus, in traditional CEM research, it is extremely difficult to design and implement controlled experimental settings involving individuals

[1]Associate Professor, Dept. of Construction Management, and Hang Lung Center for Real Estate, Tsinghua Univ., Beijing 100084, China (corresponding author). ORCID: https://orcid.org/0000-0002-7272-4273. Email: nanli@tsinghua.edu.cn

[2]Associate Professor, Dept. of Civil and Coastal Engineering, Univ. of Florida, Gainesville, FL 32611. Email: eric.du@essie.ufl.edu

[3]Professor, Dept. of Civil and Environmental Engineering, Univ. of Alberta, Edmonton, Canada AB T6G 1H9. Email: vagonzal@ualberta.ca

[4]Ph.D. Student, Dept. of Construction Management, Tsinghua Univ., Beijing 100084, China.

Note. This manuscript was submitted on October 25, 2021; approved on May 11, 2022; published online on July 23, 2022. Discussion period open until December 23, 2022; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, © ASCE, ISSN 0733-9364.

and the complexity of projects. One may need to develop, at best, quasi-experimental scenarios in real projects to test hypotheses, as there are many compounding effects to isolate from independent and dependent variables, and it is difficult to test hypotheses as such. To this end, XR technologies reveal an unexplored avenue to experimentally controlled environments. (2) XR technologies provide notable enrichment of the toolbox of academics to obtain various types of research data. This is particularly important for scenarios in which accessing the data is technically, economically, or ethically difficult or prohibitive. (3) XR technologies offer a possible pathway to ensure reproducibility of findings, as the virtual, augmented, and mixed realities created and used in CEM research can be easily shared among or reproduced by fellow academics. Reproducibility is a major principle of scientific methods. It means that a result obtained by an experiment or observational study should be achieved again with a high degree of agreement when the study is replicated with the same methodology by different researchers. Unfortunately, the implementation of traditional CEM methods (e.g., interviews, case studies) and the data they produce are usually highly context dependent and extremely difficult to reproduce. The reproducibility provided by XR technologies can significantly enhance the replicability, validity, and generalizability of CEM research outcomes.

Meanwhile, a number of major challenges have thus far prevented a more prevalent adoption of XR technologies in CEM research. These challenges include the following. (1) The way the XR technologies are used in prior CEM studies is in most cases ad hoc, and XR-enabled research, as an independent, rigorous methodology for the CEM domain, is still underexplored. (2) There is still a lack of an organized knowledge base and workflow for using XR in various CEM research areas and methodological contexts. (3) The ecological validity of the XR-based experiments, namely the extent to which experiment participants' perceptions and responses can be generalized to real-life settings, is debatable and sometimes criticized.

In this paper, we aimed to investigate the status quo of XR-enabled CEM research, by identifying current research areas in the CEM domain where XR technologies are considered the preferred or recommended methodological solutions. More importantly, we aimed to develop a process model with actionable recommendations about how XR-enabled research should be planned, designed, implemented, analyzed, verified, and validated. The process models are then demonstrated with two illustrative examples from recent studies. Last, we also discuss the philosophical, methodological, and technological roots of the aforementioned challenges in XR-enabled CEM research and describe our vision of more enabling, adoptable, and value-adding XR-enabled research in CEM in the near future.

Background

XR technology has witnessed continuous technological advancements as well as commendable infiltration into CEM research over the past three decades. This section briefly reviews the development of XR and peripheral technologies over time and the expansion of XR-enabled research topics in CEM.

Development of XR and Peripheral Technologies

The origin of XR technology dates back to the 1960s, when Ivan Sutherland invented a head-mounted three-dimensional display named "The Sword of Damocles" (Sutherland 1968). According to Milgram and Kishino (1994) and Flavián et al. (2019), XR can be regarded as a continuum spectrum, as shown in Fig. 1. VR, at one end of the spectrum, creates an immersive virtual environment without the involvement of visual sensation from the real world. MR blends the virtual world and real world. AR can be seen as one specific type of MR that augments virtual information, such as objects, text or video, onto a real scene. Meanwhile, a narrower definition of MR, also known as pure mixed reality (PMR), is also adopted in the literature (Flavián et al. 2019). PMR creates an environment in which the real world and virtual contents can be totally merged and the incorporation of reality and virtuality is bidirectional. In PMR, users can interact with both real and virtual contents, and the elements from the real world and virtual world can also interact (Flavián et al. 2019).

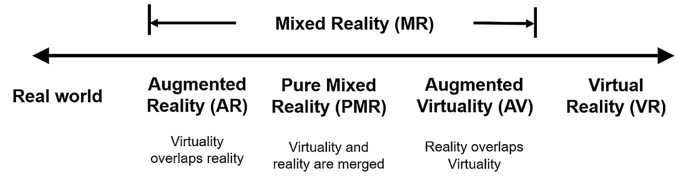

Fig. 1. XR spectrum. (Adapted from Milgram and Kishino 1994; Flavián et al. 2019.)

VR technology ushered in the first commercial boom in the 1990s. But the high cost of VR products and the immaturity of peripheral technologies limited its wide adoption by the public at that time. The second boom occurred in the 2000s, which led to the gradual expansion of VR usage in CEM research. There are different forms of VR technology, including the desktop VR system, Cave Automatic Virtual Environment (CAVE), and head-mounted display (HMD). Among them, HMD provides the highest level of immersion for users and allows users' head orientation and movement to be monitored with advanced tracking technologies. As for AR technology, it is categorized into mobile AR (e.g., smartphones and tablets) and wearable AR (e.g., Google glasses) according to the output device. AR systems commonly use sensor-based or vision-based tracking technologies (Ashwini and Patil 2020). The establishment and release in 1999 of "ARToolKit," a typical computer tracking library developed using the marker-based method (Kato and Billinghurst 1999), is a milestone in the history of AR.

Later on, the tracking methods began to migrate from marker-based to markerless-based (Chi et al. 2013). The incorporation of markerless-based tracking technology provided new opportunities to apply AR in construction fields. Compared with V/AR, the launch of PMR technology was much later, and few devices using PMR are available in the market. Currently, the only technologies that can be truly regarded as PMR are the holographic devices Microsoft Hololens (Microsoft 2021) and Magic Leap (Magic Leap 2021). PMR devices are equipped with sophisticated sensors that are capable of capturing the contour and content of the workspace so that virtual objects can intelligently integrate into users' environments.

In addition to advancements in the functionality, usability, and affordability of XR technology, the rapid expansion of XR usage in CEM research is inspired by the increasing availability of three classes of peripheral technologies that are used for improving XR environment modeling, facilitating human-computer interactions, and assisting experimental data collection. For XR environment modeling, various existing 3D modeling and animation technologies can be used to improve modeling efficiency and quality. Instead of creating XR environments directly in XR engines, researchers can either create as-is 3D models of the environments with fast modeling technologies (e.g., photogrammetry, laser scanning), or leverage models of different formats [e.g., building information model (BIM), Maya, 3ds Max] that are created for other applications during the lifecycle of the construction projects. Meanwhile, to facilitate human-computer interactions in XR environments, various novel multisensory devices are becoming available, which can lead to more immersive, holistic, and realistic user experience. For example, haptic devices such as gloves and exoskeletons can be used to accurately track hand and body postures and reflect force feedback to users; heater, smell, and vibrating devices can be used to create rich sensory stimuli from the XR environment and elicit complex user responses. In addition, rich behavioral data are produced when users interact with the XR environment. The collection of XR experimental data, including behavioral and physiological data, relies on the use of XR-embedded sensors. For user behavioral data collection, XR manufacturers leverage body-carried motion trackers. These motion trackers are fueling research in human factors and ergonomics. XR-embedded physiological sensors do not directly affect users' experience in XR environments; however, these sensors can provide various physiological measurements [e.g., electromyography (EMG), electrocardiography (ECG or EKG), galvanic skin response (GSR), electroencephalography (EEG), and functional near-infrared spectroscopy (fNIRS)]. We observe a clear trend in the recent literature that neurophysiological sensors are integrated into XR systems for a more effective human assessment. For example, eye trackers have been embedded in mainstream VR/AR headsets. Researchers are leveraging the embedded eye trackers in cognitive and attention pattern studies as part of the XR methods (e.g., Sipatchin et al. 2021). Most recently, neuroimaging devices are also being implemented and even embedded in the XR headsets. As a result, we project that the integration between neurophysiological sensors and XR systems will become even more prevailing.

Expansion of XR-Enabled CEM Research Topics

Researchers started to use XR technology to conduct CEM research involving human participants or concerning human behavior in the early 2010s. In these studies, the use of XR can be broadly categorized as a visualization tool, a medium of communication, or a stimulus to participants, distinguished by their respective purpose and human-computer interaction characteristics. As a visualization tool, XR makes use of visual elements and visual interaction strategies to present abstract or complex contents that are difficult to understand, with the aim to support users' perceptual inferences instead of arduous cognitive comparisons and computations. When XR acts as a medium of communication, it enables user-user interactions in the XR environment, with the help of text and voice chat tools, visual sharing affordance, and virtual avatars, to facilitate information sharing and collaboration between users. As a stimulus to participants, XR creates multisensory human-computer interactions, with the aim to reproduce hazardous and stressful environments in a relatively safe and controlled environment and evoke users' emotional and behavioral reactions to the virtual stressors. Among all the XR-enabled studies involving humans, we have observed based on a review of relevant literature that five topics are the most studied by CEM scholars, including construction safety, emergency evacuation, human–building interaction, architectural and urban design, and training and education. These classifications are not mutually exclusive; a study may involve multiple topics. For studies on each topic, Table 1 summarizes the main concerned research questions, the type of XR technology used, the role XR technology plays, and commonly used peripheral technologies.

For studies on construction safety and emergency evacuation, XR technology acts as a stimulus to participants and provides new opportunities for researchers to observe participants' responses and behaviors in a virtual dangerous environment, while the participants are not exposed to real risks of physical injuries. Existing XR-enabled studies on construction safety focus on (1) exploring environmental factors [e.g., sounds (Lu and Davis 2016)] and cognitive factors [e.g., cognitive load (Han et al. 2021) and mental fatigue (Tehrani et al. 2021)] that influence workers' safety; and (2) investigating safety intervention methods that address identified ergonomic risks (Hasanzadeh et al. 2020). By allowing humans to interact with immersive environments in controlled and monitored settings, existing XR-enabled studies on emergency evacuation aim

Table 1. Main research topics explored in existing XR-enabled CEM research

Research topic	Main research questions	Type of XR	Role of XR	Examples
Construction safety	• Exploring the factors influencing workers' safety from the perspectives of the physical environment and cognition theory	VR	Stimulus to participants	Han et al. (2021), Lu and Davis (2016), and Tehrani et al. (2021)
	• Investigating safety intervention methods based on the identification of ergonomic risks	AR/PMR	Stimulus to participants	Fu et al. (2019) and Hasanzadeh et al. (2020)
Emergency evacuation	• Exploring how physical environments and human cognition process influence evacuation performance	VR	Stimulus to participants	Cao et al. (2019) and Lin et al. (2020a)
Human–building interaction	• Assessing thermal comfort, lighting and acoustic comfort, visual comfort	VR	Visualization tool	Heydarian et al. (2015), Li et al. (2020), and Yeom et al. (2019)
	• Identifying occupants' spatial usage patterns and preferences	VR	Visualization tool	Khashe et al. (2015)
Architectural and urban design	• Investigating how design features influence occupants' physiological and psychological status	VR	Visualization tool	Fich et al. (2014)
	• Exploring how to facilitate the understanding and communication among various stakeholders thus making design tasks more efficient	VR/AR/PMR	Visualization & communication tool	Wang and Chen (2009)
Training and education	• Investigating how to improve skills and performance of workers and students	VR/AR/PMR	Stimulus to participants; visualization & communication tool	Kandi et al. (2020) and Kwiatek et al. (2019)
	• Exploring how to change the unsafe behaviors of workers	VR/AR/PMR	Stimulus to participants	Shi et al. (2019)

to explore factors that influence the evacuation performance from two aspects including the external environment [e.g., signage (Fu et al. 2019), alarm (Xia et al. 2021), and crowd (Lin et al. 2020b)] and the human cognition process (Lin et al. 2020a). Participants usually need to complete a virtual emergency response task, such as evacuating from an indoor space, in such studies, and their responses and behaviors in different controlled settings are analyzed (Cao et al. 2019).

For studies concerning human–building interaction, XR mainly acts as a visualization tool presenting dynamic or static scenes to users in the XR environment. In these studies, participants need to assess thermal comfort (Yeom et al. 2019), lighting and acoustic comfort (Heydarian et al. 2015), and visual comfort (Li et al. 2020) in virtual built environments. Occupants' spatial usage patterns and preferences (Khashe et al. 2015) are also investigated in XR environments. For XR-enabled studies on architectural and urban design, the main investigated issues include (1) how architectural and urban design features influence occupants' physiological and psychological status (Fich et al. 2014); and (2) how to facilitate the understanding and communication among various stakeholders, making design tasks more efficient (Wang and Chen 2009). For the first issue, XR plays the role of a visualization tool presenting various design features that are difficult to manipulate in reality. For the second issue, XR mainly acts as a communication tool that helps designers and stakeholders share ideas and collaborate with each other in an experimental setting.

For training and education research, XR technology can simulate large-scale operations and provide inexpensive and effective training scenarios. The visualization of learning materials can help workers and students understand knowledge better and faster. Among prior XR-enabled training and education studies, a few focused on measures to improve the skills and performance of workers and students under the settings of operational task training (Kwiatek et al. 2019) and architectural design education (Kandi et al. 2020), where XR acts as a visualization tool and a medium of communication. Safety training is another active research topic, where XR training and education systems are typically designed as serious games to explore how to change the unsafe behaviors of workers (Shi et al. 2019) when they are exposed to various hazards on construction sites and in built environments.

Conducting XR-Enabled Experiments in CEM Research

The goal of a scientific experiment is to establish empirical evidence of a relationship between an independent or experimental variable and a dependent variable that is affected by it (Bernold and Lee 2010). To illustrate the input-output nature of a scientific experiment and the flow of phases that need to be followed to achieve this goal, we developed a process model for XR-enabled experiment in CEM research, as illustrated in Fig. 2. This process model is developed based on (1) references to the scientific tradition of experimental design that considers planning, designing, and performing an experiment, followed by analyzing data, confirming results, and evaluating conclusions (Berger et al. 2002); (2) references to the classic human subject research methodologies in the experimental psychology domain (Bennett-Levy et al. 2004) and the experimental research process model developed specifically for CEM research (Bernold and Lee 2010); (3) considerations of unique challenges in XR-enabled human subject research in CEM and possible approaches to address them; (4) references to best practices and lessons learned from prior XR-enabled CEM studies, examples of which include Heydarian and Becerik-Gerber (2017) and Zhu et al. (2018); and (5) effort to achieve a proper level of detail, so as to make sure sufficient details are provided for scholars to put the process model into practice, while avoiding giving too much detail to limit the applicability of the process model in various research contexts.

The developed process model consists of four sequential phases, including needs identification, XR environment design, experimental design, and data collection and analysis, and an additional phase of verification and validation that applies to the entire process. The activities that need to be undertaken in each phase, and the particular concerns and considerations associated with each activity, are discussed in detail below.

Needs Identification

Before beginning to design an XR-enabled experiment, researchers first should identify the need for building their research on an XR-based methodology and select the appropriate technologies. There are four activities researchers are recommended to perform at this phase.

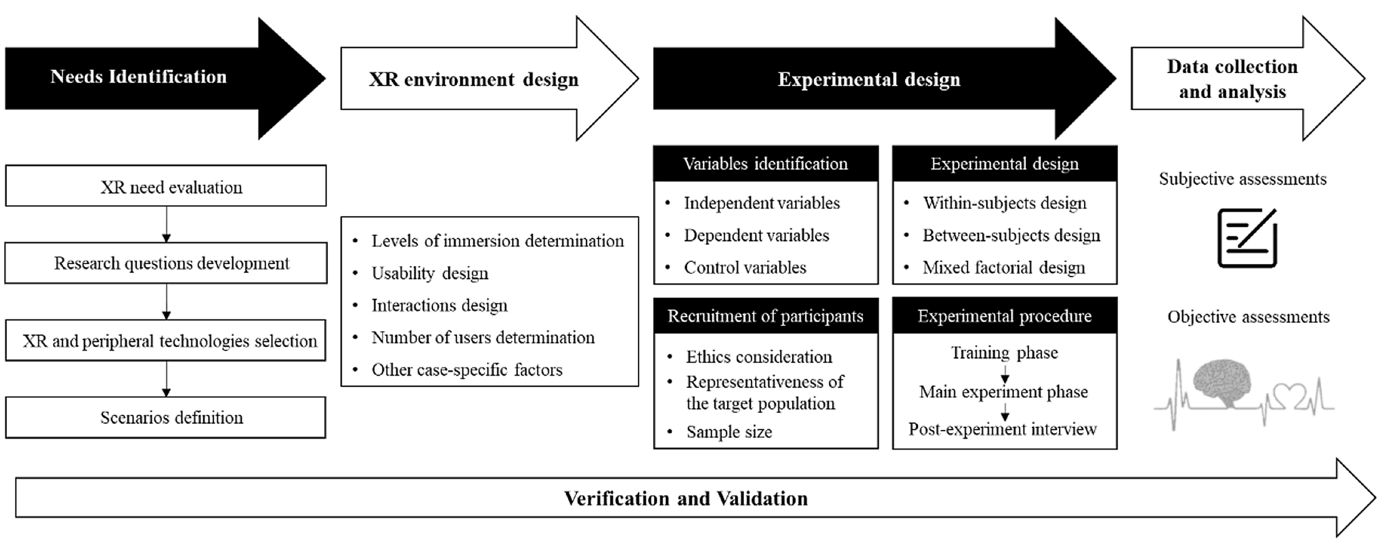

Fig. 2. The process model for XR-enabled experiment.

First, the use of XR in the context of a specific study should be well justified to avoid a temptation to use XR for everything. A key justification for using XR in construction-related research is embodied cognition. Embodied cognition refers to the cognitive process and physical interactions of the body with the world (Barsalou 1999; Wilson 2002). Evidence suggests that embodied cognition is the process by which humans use their sensory neural structures to create multisensory representations of their environment, and humans can reconstruct their cognitive structures when they mentally imagine an object or action (Barsalou 2003, 2008). With that said, the most basic criterion for researchers to deem whether XR is suited for their research is whether embodied cognition is the core of the investigation. For example, if a researcher aims to examine how a worker interacts with the physical processes and the corresponding situational awareness on the job site, XR is more appropriate than other research instruments (e.g., playing a video on a screen), because such situational awareness is based on the embodied cognition about physical interactions. We propose the following metrics for evaluating the need for XR in research: sense of presence, emotion arousal needs, spatial sense, and level of hazards. If the four metrics are evaluated with a higher score, XR methods are more suitable. Note that, for some of the measurement metrics, existing assessment instruments can be used, such as the Slater-Usoh-Steed presence questionnaire (Usoh et al. 2000) for measuring the sense of presence. Nonetheless, there is a general lack of agreement on how these metrics can be measured for each specific research study.

More importantly, the four measurement metrics can still be extended with the new development in embodied cognition science. Therefore, we urge more attention from the construction community to work together on developing a widely accepted measurement method for evaluating the suitability of the XR method in human-centered research. In addition, the use of XR technologies comes at a cost in the form of, e.g., investment in equipment, lab space, and staff's technical skills; extra time spent on XR model development; potential threats to and criticism of the ecological validity of the findings; and so on. The cost should also be considered by researchers in XR need assessment to make informed decisions.

Second, once the need for conducting XR-enabled experimental research is justified, specific research questions should be carefully developed. The questions the researchers aim to explore fundamentally determine which of the aforementioned roles of XR technologies and which specific XR technologies should be used, what experiment hypotheses should be tested, what treatment should be applied in the experiment, and what data should be collected during the experiment.

Third, given the distinct features of VR, AR, and MR technologies and the growing variety of peripheral technologies, researchers have to decide which technologies to use for developing the technical capabilities needed for their studies. While this decision is fundamentally driven by the research needs that vary significantly from case to case, a noninclusive list of factors to consider involves the required levels of immersion, video quality, and interactivity of the XR environment; the type of experimental sites (lab versus construction site); the portability of devices (PC-connected/standalone headset, smartphone-based, or projector-based) and associated control modes and movement tracking methods; needs for multisensory feedback (e.g., haptic and olfactory senses); and needs for wearable physiological sensors (e.g., eye trackers and inertial sensors).

Finally, scenarios that need to be reproduced in XR environments should be clearly defined. The scenarios are defined by a range of variables, such as storyline (e.g., scenes, events), characters, tasks to be undertaken, and so on. These variables in turn determine various technical requirements of XR environments and experimental design decisions, such as whether the XR environments should have multiuser functionality, whether multisensory stimuli should be used to evoke certain emotions in users, how the virtuality and reality should be combined, and so on.

XR Environment Design

After identifying research needs and selecting appropriate XR technologies, factors regarding the development of XR environments should be carefully considered when planning XR-enabled experiments. There are generally five groups of factors worthy of being recommended to researchers.

First, levels of immersion and presence using practical methods need to be determined. The sense of immersion, whether complete (e.g., VR) or partial (e.g., AR), is important to consider when developing XR environments, as it is one of the main factors that can blur the boundary between the virtual and real worlds. An appropriate level of immersion should be determined based on the selected XR technologies and the proposed experimental needs, with efficient capability to support the adopted research design. Perception, situational awareness, and field of regard should be considered to provide high levels of immersion (Lemoine et al. 2003).

Second, usability should be designed and optimized through user experience (UX) design. Although there are numbers of techniques to develop UX for XR projects, the Waterfall method introduced by Royce (1987) and the Agile UX design method the Wheel, proposed by Hartson and Pyla (2019), are two suitable software development approaches when it comes to XR design. The selection of the UX design method also relies on the chosen XR technology and research needs. Evaluation of usability should be carefully conducted to improve the development of XR research. In addition, affordance and side effects (i.e., cybersickness) should be considered. Affordance measures how easy users can follow the application to achieve proposed outcomes or follow the designed storyline with the corresponding stimulus created in XR environments (Gibson 2014). It can be critical in optimizing usability to demonstrate characteristics of interactions (Shin 2017).

Third, interactions need to be designed to fulfill the requirements of the proposed immersion level and UX design specifications. Two categories of interactions should be considered, namely sensory interactions and user interactions. Sensory interactions are simulated through stimuli in the XR environments, which generally include visual, auditory, and tactile senses. Navigation, manipulation, and system control are three main areas of user interactions (Bowman et al. 2001).

Fourth, the number of simultaneous users should be carefully weighed. It should be determined by the intended experimental outcomes. Single-user XR would be practical to observe human psychology and behavior and individual cognition (e.g., decision-making, analysis ability, self-efficacy, and emotional arousal), whereas multiuser XR is useful to study different social mechanisms and dynamics (e.g., collaboration, communication, and trust). Multiuser XR also requires server maintainability, digital representatives' management, performance evaluation, and potential forensic monitoring, among other requirements (Siedler et al. 2021; Taylor et al. 2019).

Fifth, attention needs to be paid to scenario-specific factors. With the UX design being conducted properly and high levels of immersion and presence achieved, CEM scenarios involving tasks, context, situations, and resources can be simulated with great flexibility and effectiveness via XR applications. From this point of view, these scenarios should be designed in accordance with the research needs while maintaining high compliance with the general XR project's modeling requirements.

J. Constr. Eng. Manage., 2022, 148(10): 04022106

Experimental Design

To perform the experimental design, CEM researchers first need to identify the independent variables, dependent variables, and control variables of the experiment, which are critical to guide the following activities in experimental design. An independent variable refers to the variable that is specifically manipulated or observed to occur before the dependent variable to assess its effect or influence, whereas a dependent variable refers to the outcome that is observed to occur or change after the occurrence or variation of the independent variable (APA 2021). In addition, because the causal relationship cannot be inferred from the statistical analysis, there are variables that may have an effect on the response measure but that themselves are not of particular interest to the researchers. These variables, referred to as the control variables, should be controlled in an experimental design. Two methods can be used to reduce or avoid the influence of control variables. The first way is to control the influence experimentally, for instance, keep the possible control variables the same in each treatment. When the influence cannot be eliminated by the experimental method, the control variables can be controlled by the statistical method; these control variables are called covariates. The proper treatment to control variables is critical to the internal validity of the study.

The second activity in the experimental design phase is to determine whether to use the within-subjects design or between-subjects design. The within-subjects design refers to an experimental design in which the effects of treatments are seen through the comparison of scores of the same participant observed under all the treatment conditions (APA 2021). Within-subjects design can control the influence of participants' characteristics on experimental results but is highly affected by treatment orders and participants' learning effects. Between-subjects design refers to a study in which participants are assigned to only one treatment or experimental condition and each participant provides only one score for data analysis (APA 2021). Such a design needs a large number of participants to mitigate possible effects of individual differences, with shorter sessions compared with within-subject design. In addition, the above two design methods can be combined in a single study, dubbed as mixed factorial design. Under this design, one variable is altered between participants and another is altered within participants. This design is suitable for studies that consider two or more variables.

With regard to the recruitment of participants, there are generally three major considerations. The first is the research ethics consideration. Researchers have to make sure their research design minimizes the risks to participants and adheres to the Belmont Report's principles of beneficence, respect, and justice (The Belmont Report 1979). Researchers should provide participants with adequate information about the research, obtain their informed consent, and allow them to choose to quit the experiment at any time. In some countries, the entire research design needs to undergo institutional review and approval before the research can be conducted. Second, the selection of participants should consider the representativeness, in terms of age, gender, ethnicity, occupation, education, and so on, of the target population to whom the research findings are expected to be applied. While ease of access to participants usually plays a major role in participant recruitment, researchers should be aware of the potential bias and consequent limitation in the validity and generalization of their findings. Third, the sample size should be enough to ensure sufficient statistical power. It is recommended that an a priori power analysis (Cohen 1988) be done before the experiment to determine the minimum number of participants required for controlling the Type II or β error probability of falsely retaining an incorrect null hypothesis while avoiding excessive time and resources spent on unnecessarily large sample sizes (Wang et al. 2021).

While the specific experimental procedure varies from case to case depending on the objectives of research and technological settings, an XR-enabled experiment typically involves a training phase and a main experiment phase. The former allows the participants to become familiar with the operation of XR devices and the sense of immersion in the virtual environments. A mockup virtual environment can be used at this phase. During the main experiment phase, participants are instructed to complete a series of tasks by interacting with the virtual scenarios. Typically, they are also required to fill in certain questionnaires before and/or after conducting the virtual tasks. All the independent, dependent, and control variables are monitored and recorded during the above procedure. An additional postexperiment interview with the participants can also be conducted to gain further insights about the experimental outcomes. Last but not least, pilot experiments are usually recommended to be carried out before the full-scale experiment. They help to evaluate the feasibility of the study, calibrate certain model settings, and improve on various aspects of the experimental design.

Data Collection and Analysis

Various sensing and data collection technologies are now available for XR-enabled research that help capture high-fidelity human performance data during and after the experiments, providing multisource evidence for a more accurate evaluation. A comprehensive human assessment data collection should cover both behavioral and functional aspects. Therefore, both subjective assessments (e.g., questionnaires) and objective assessments (e.g., neurophysiological measurements) are needed. The former set of measurements is rooted in the psychometrics discipline (Shrout and Lane 2012), focusing on the self-reported experience in the XR environment. Representative questionnaires/surveys used in the XR literature include NASA Task Load Index (TLX) (Hart 2006), presence questionnaires (Slater 1999), and cybersickness surveys (Martirosov and Kopecek 2017). These measures can be used to build the baseline data for benchmarking XR research findings or solicit perceptions of interesting aspects. Neurophysiological measurement focuses on the simultaneous tracking of cognitive, behavioral, and physiological statuses via neurofunctional monitoring devices, eye trackers, and other body-carried sensors. Eye tracking has been a widely adopted technique for decades in the XR literature, from the efforts of using screen-mounted eye trackers, portable eye trackers, to the recent VR/AR headset-integrated eye trackers. The literature has demonstrated the effectiveness of using eye-tracking data to examine gaze focus areas, pupillary size changes, and blink rates as metrics for attention and cognitive load measures, in situational awareness and information system studies (Hasanzadeh et al. 2017; Xu et al. 2019a). Meanwhile, scholars are interested in body-carried sensors to obtain corresponding physiological status in XR environments, such as inertial measurement units (IMUs) for motion tracking (Yan et al. 2017a, b), EMG for musculoskeletal modeling (Xu et al. 2019b), and ECG for heart rhythm and electrical activity tracking (Kim et al. 2021). These measurements can be used in construction ergonomics, safety, and perceptional studies. Most recently, there is a growing interest in exploring the neuroimaging methods combined with the XR systems, including EEG and fNIRS. EEG tracks the electrical potentials on the surface of the skull, while fNIRS uses infrared light signals to infer the hemodynamic responses in interested brain areas. Although both are proven techniques well integrated with XR platforms for a more explicit cognitive status tracking, evidence shows that EEG can be affected by motor artifacts and thus is less effective in motor-intensive experiments (Shi et al. 2020b).

A challenge that deserves more efforts of the CEM research community is the automated labeling of the collected data. Most XR studies in the CEM literature reproduce realistic work settings, where test subjects are required to complete the tasks in a nonstop manner. In contrast, many types of collected data are event driven. For example, the use of EEG data to infer the awareness of a particular construction hazard would require a clear event marker when the test subject is exposed to the hazard (Hasanzadeh et al. 2017). In an ergonomics study that involves a continuous walking task, researchers need to split the collected IMU data based on the specific events during the task (such as a near-miss fall) (Yang et al. 2016). An automated method for event marker labeling will help improve the efficiency of the research (Zhu et al. 2021).

Verification and Validation

Any XR simulation is a conceptual representation of the reality. It requires a certain degree of confidence on the XR simulation that it serves as a valid representation of the real systems/processes, and in return, decisions and analyses derived from it can be useful and convincing. Model verification and validation (V&V) should be used to examine to what extent an XR-enabled experiment acts as designed purpose and generates results that are similar to observations.

Verification refers to examining whether a model works as designed (North and Macal 2007). To verify any XR-enabled experiment to be a correct realization of the designed purpose, the following steps are recommended. (1) Code debugging, i.e., detecting and correcting the apparent programming mistakes that affect the renderings, functions, and data collection of the XR-enabled experiment. (2) Logic examination, i.e., identifying the hidden logic errors that may not affect the apparent use of the XR-enabled experiment, but indeed affect the reasonableness of the desired purpose. For example, the erroneous collider settings in a VR environment can result in inappropriate human–object interactions in the experiment. (3) Unit test and toy case test. This is to assess the functionality of the XR-enabled experiment in a holistic way, i.e., whether desired results can be obtained from the experiment. Unit test is to ensure each unit of the XR-enabled experiment works as designed (e.g., the light rendering works as designed, or the gaze tracking can generate the desired output files). The toy case test allows the designer to walk through the XR environment and experimental workflow completely, to make sure that the user experience, object interactions, and data tracking functions all function as designed.

Validation is to make sure a model works as observed (North and Macal 2007). In other words, what is being evaluated is the "validity" or the degree to which a "test, model, measurement, simulation, or other reproduction provides an accurate representation of its real equivalent" (Harris et al. 2020). Validation is a challenging but critical task to carry out experiments in XR environments that are both robust and reliable. When it is done effectively, meaningful inferences can be obtained given a CEM problem under study. In this paper, three types of validity are considered relevant for XR-based human-centric research in CEM: content validity, face validity, and construct validity. Content validity judges how appropriately a metric, tool, or method incorporates the aspects or "items" that are essential to measure what it is required to be measured (Westen and Rosenthal 2003). For instance, in wayfinding research content, validity refers to all those items necessary to measure "wayfinding pedestrian behavior" (Feng et al. 2022). Face validity is the subjective appreciation held by users in relation to how realistic a simulation can be (Harris et al. 2020). It is highly dependent on visual characteristics of the XR environment, influencing technical and functional features, e.g., how the input of a user is associated to actions. Thus, the simulation design and the system technical capabilities such as immersion are relevant factors of face validity. Construct validity refers to the degree to which a metric, tool, or method appropriately assesses the construct or concept it intends to assess (Westen and Rosenthal 2003). Good construct validity within an XR environment means that it is responsive to performance variation between and within individuals. This can suggest consistency in principles, rules, stimuli, and response aspects between real and XR-based simulated tasks.

To achieve face and construct validity, it is necessary to ensure certain levels of fidelity (Harris et al. 2020). Fidelity helps to assess to what degree an XR environment recreates a real-world system, assessing not only appearance but also the affective, cognitive, and behavioral responses triggered in users (Perfect et al. 2014). According to Harris et al. (2020), there are four types of fidelity to be considered: physical, psychological, affective, and ergonomic and biomechanical. Physical fidelity represents the realism level produced by the physical environment within an XR simulation (e.g., visual information involving field of view, objects' realistic behavior, alignment with laws of physics, level of functionality). Physical fidelity, as for face validity, helps to induce users' feeling of presence and the illusion of plausibility (i.e., illusion that the simulated scenario is really happening) (Slater 2009). Psychological fidelity is the extent to what an XR simulation reproduces the "perceptual-cognitive" demands of the real activity or task performed in the simulated XR environment (Gray 2019). For example, a high-fidelity VR environment to identify construction workers' unsafe behaviors in the workplace should require participants to be exposed to a construction-like working environment (e.g., with other workers, machinery, construction work underway) and demand a similar level of attention and effort to undertake a task (e.g., transport heavy materials) as if they were involved in real-world construction work (Gao et al. 2022). Affective or emotional fidelity needs the XR environment to induce a realistic emotional user's response such as stress or excitement (Harris et al. 2020). This is very important when developing XR experiments or training environments that involve users' tasks that are too dangerous to carry out in the real world. Ergonomic and biomechanical fidelity helps to produce users' realistic movement patterns within XR environments, with immersion being the single most important factor for ergonomic fidelity (Harris et al. 2020). In spite of the developments of XR technology such as haptic gloves and muscle stimulation, the supply of realistic haptic input for XR simulation is still regarded as a major technical challenge (Lopes et al. 2017).

Based on Feng et al. (2022), Harris et al. (2020), and the authors' research experience, Table 2 is proposed to guide CEM researchers in the validation process, suggesting the key question to be tested and the approach to test it in each of the validity and fidelity categories. Researchers should bear in mind that not all of the validity and fidelity categories will be relevant, which will depend on the key research questions posed by the researchers.

XR-Related Challenges

XR is an imperfect replica of the real-world systems. As such, it still faces a few challenges in human-centered experimental research in CEM that should be considered when implementing the process model. First, many CEM problems involve motor-intensive tasks. It is still nontrivial to reproduce physical interactions and motion data in XR environments. Second, CEM applications pertaining to the human–building interactions require a high-fidelity model of the built environment to trigger realistic behavioral responses of the test

Table 2. Validity and fidelity categories, and overall driving questions and testing approaches

Validity or fidelity category	Driving question	Testing approach
Content validity	Do the metric, tool, or method used in the XR simulation include those items necessary to measure what is meant to be measured?	Comparison of collected behavioral data against theoretical behavior observed in the literature (e.g., decision-making, task performance, physical behavior)
Face validity	Does the XR environment look and feel realistic?	User's self-reports regarding plausibility
Construct validity	Does the XR environment provide an accurate characterization of the construct assessed such as real task performance, safe behavior, teamwork, communication, trust?	Capability of the XR environment to differentiate, for instance, between individuals' expertise (real-world expertsversus novices) and track improvements
Physical fidelity	Is there a high level of detail and realism in the physical environment of the XR simulation?	User's self-report and psychophysiology associated to realism and measures of presence
Psychological fidelity	Does the XR simulation accurately represent the perceptual and cognitive features related to the real-world situation, activity, or task being simulated?	Assessment and comparison of various perceptual and cognitive metrics such as mental effort, gaze behavior, neural activity, between real and virtually simulated situation, activity, or task
Affective fidelity	Does the XR environment induce emotional arousal or responses such as stress, anxiety, fear, similarly as the response generated by the real-world situation, activity, or task?	Users' self-reported experiences or monitoring of psychophysiological indices of affect
Ergonomic fidelity	Does the XR environment trigger realistic motor responses or motions?	Realism evaluation of XR motion parameters using, for instance, motion tracking, and comparison analysis of amplitude, speed, interjoint coordination, among other metrics, with real actions

subjects. The reproduction of a photorealistic model is computationally expensive and should be carefully planned to achieve a balance between the expected benefits and development costs. Lastly, concerns still exist related to the potential impacts of XR environments on human subjects, such as the cybersickness, privacy of personal data, and implications of exposure to extreme conditions in XR.

Our proposed workflow aims to mitigate the challenges via a carefully thought-out cross-checking process. First, delicate decisions should be made to determine whether XR is deemed to be the proper research tool, i.e., the need identification. Innovative embodied cognition criteria are proposed to minimize the use of XR when alternatives are available. Second, if XR is determined to be proper, the models will be verified to ensure quality and reproducibility. The proposed verification standards can evaluate to what extent the developed XR environments can trigger the desired realistic behavioral responses. Third, a set of measures can be used to ensure the safety of XR in human subject experiments, such as the sense of presence questionnaires and cybersickness evaluation, to ensure the quality of the collected data as well as protect the experiment participants. These added steps are different from the classic human-centered studies owing to the new features of XR. Finally, the recommended data analysis and validation activities also reflect the advantages and limitations of XR methods. The XR-embedded sensors, such as neurophysiological sensors, provide unprecedented access to rich human functional data. Meanwhile, validation of the collected data is challenging, requiring a new approach for validity assessment. The method proposed by this study sets a cornerstone for the research community to explore more in this critical area.

Illustrative Examples of XR-Enabled Research in CEM

This section presents two illustrative examples of XR-enabled CEM research. We use these examples, each of which addresses a different problem in the CEM domain and uses XR technologies in a different way, to demonstrate how every phase and activity outlined in the last section can be implemented in the context of specific use cases.

Example One: VR for Exploring Cognition Load in Altered and Stressful Construction Tasks

Needs Identification

As the US is facing historic challenges with the renewal of critical civil infrastructure systems, there is a pressing need for a quality workforce adapted to the changing pace of construction works. A significant challenge for future construction workers is the ability to digest and comprehend the complex engineering information in workplaces in a timely manner, while still mastering the motor skills for the operations. A representative example of such challenges is industrial turnaround maintenance. Turnaround maintenance is an event wherein each part of the facility is shut down in a rotation turn for renewal, ensuring that the overall facility still functions during the maintenance (Duffuaa and Ben Daya 2004). To minimize the impact of the turnaround schedule, the work is usually done in a 24/7 manner, and hence workers are always under extreme pressure (Duffuaa and Ben Daya 2004). In addition, owing to the growing complexity of the engineered facilities, workers often need to memorize or digest a large amount of engineering information in a very short period. As a result, turnaround maintenance is one of the most dangerous jobs. Growing evidence has linked the root causes of incidents to human errors mostly due to the miscommunication, misunderstanding, and misuse of information, driven by the excessive work-related stress or neglect (Garrett and Teizer 2009). It remains a pressing need for researchers to examine the cognitive load of construction workers who work under substantial work-related pressure, given the varying task contexts, environment, and requirements.

The problem under investigation, i.e., turnaround maintenance, is a combined task that involves both cognitive and motor skills. To successfully complete the maintenance task, embodied cognition plays a critical role in both activity and motor planning processes. As a result, to examine the performance and cognitive functions of participants in a combined task, a realistic environment is needed to trigger both the proper cognitive processes and motor activities. In contrast, traditional experiment instruments, such as decisional tasks facilitated with imagery or video stimuli, would not be sufficient, as sensorimotor processes can hardly be engaged. In addition, the context of the turnaround maintenance task is a potentially stressful environment, where the workplace could be austere and human perception could be impaired. Therefore, simulation of stressors is also needed to examine the secondary impacts of environmental stress on performance and function. All these suggest that a realistic, interactive and immersive environment is needed for this study that allows experiment participants to interact with the simulated physical systems in a natural and engaging way. Therefore, using the XR-based experiment is deemed to be proper. Among all the existing XR technologies, VR has been deemed to be able to reproduce hazardous and stressful work contexts in a relatively safe and controlled environment. VR is therefore selected in this research and used to trigger measurable cognitive load of the participants when they perform virtual turnaround maintenance of a piping system.

VR Model Design

In a recent research (Shi et al. 2020a), participants were asked to memorize sequences for turning or closing the valves before they replaced the plate heat exchanger for a turnaround maintenance task. The pre-start-up sequences to cut off the hot water and cold water consist of 10 sequential steps, which were developed based on the operation instruction manual of Alfa Laval plate heat exchangers (Alfa Laval 2016).

Owing to the difficulty of reproducing this pipe maintenance task in the real world, an interactive and immersive VR system was developed based on our previously well-validated VR systems. First, we obtained the pipe skid structure in the format of AutoCAD version 2016 and created the 3D model in Blender. To enable the immersive experience, we added metallic texture. Then the Blender model was transferred to Unity 3D (Unity 2021) for VR modeling [Fig. 3(a)]. HTC VIVE handheld controllers were programmed to interact with the valves, where the contract would trigger a prerecorded animation of valve rotation. We also added collisions, and thus the participants had to maneuver through the busy pipelines to avoid collisions with the structure [Fig. 3(b)]. Furthermore, to raise the cognitive load levels of the participants, we added a countdown function. When a certain time point passed, an explosion would happen [Fig. 3(c)].

Experimental Design

A total of 30 participants were recruited via the university mailing list for the experiment. All participants were college students in engineering-related majors (18 males and 12 females). Before the experiment sessions, we collected background information from participants that might influence performance, such as demographic traits, spatial cognition, and gaming experience (related to VR familiarity). Then all participants were asked to familiarize themselves with the VR devices and the virtual environment in the training session. Experiment investigators were able to ensure that participants' eyeball movements were accurately captured by the eye tracker after several calibration trials. Participants were also given instructions about how to use the two controllers to interact with the virtual valves. The review session was used for participants to review and memorize the pipe maintenance sequence. The review time was limited to 5 min, because some participants might feel sickness (nausea, headache, dizziness, and lightheadedness) if using the virtual environment for 10 min or longer. In the retention session (5-min duration) participants were given another shape memory test. The purpose was to intervene in the working memory storage of the participants and to trigger relatively high cognitive load in the following task. After the retention session, participants were asked to perform the pipe maintenance task in the VR environment (with no time limit). After completing the operation session, participants were given a Slater-Usoh-Steed (SUS) questionnaire (Usoh et al. 2000) to evaluate their presence in the virtual environment. The SUS questionnaire has been proven effective in presence evaluation about human–computer interaction (HCI) (Ling et al. 2012). At the end of all experimental stages, participants were asked to fill out a questionnaire on cognitive load measurement proposed by Leppink's cognitive research (Leppink et al. 2013, 2014).

VR-Based Data Collection and Analysis

To obtain cognitive load measures, we used eye tracker and fNIRS devices integrated with the VR system. The Tobii Pro VR eye tracker was integrated into the HTC VIVE HMD. The eye tracker had an accuracy of 0.5°, and the gaze data output frequency was 120 Hz (Tobbi 2019). To achieve the eye-tracking and visualization functions in the virtual environment, several C# scripts were developed based on the Tobii Pro SDK (Tobbi 2019) and the application programming interface (API) in Unity.

Besides the VR-based eye tracking, participants were instrumented with a continuous-wave fNIRS device with a probe map focused on cortical locations defined following the 10-10 international systems using a 16-probe design (Fig. 4). There are 21 channels across a network of brain regions responsible for motor learning and working memory function. The dorsolateral prefrontal cortex (DLPFC) and the premotor regions were chosen because the DLPFC works closely with the premotor and supplementary motor areas for complex motor tasks such as sequence learning (Gerloff 1997), and for stress-related activities (Qin et al. 2009).

We examined the cognitive load changes of participants ($n = 30$). The results indicated that the cognitive load for the 3D and VR groups for memorizing information in the review session was higher, but once the information was encoded, it was much easier for these two groups to retrieve or recall information, showing as a lower cognitive load level. We further examined if the high-fidelity fNIRS data could be used to train a machine learning model for a more precise

(a) (b) (c)

Fig. 3. The VR environment: (a) the VR model of the pipe skid; (b) participant performing the turnaround maintenance task; and (c) hazards pass the given time point for stress triggering. (Images by Jing Du.)

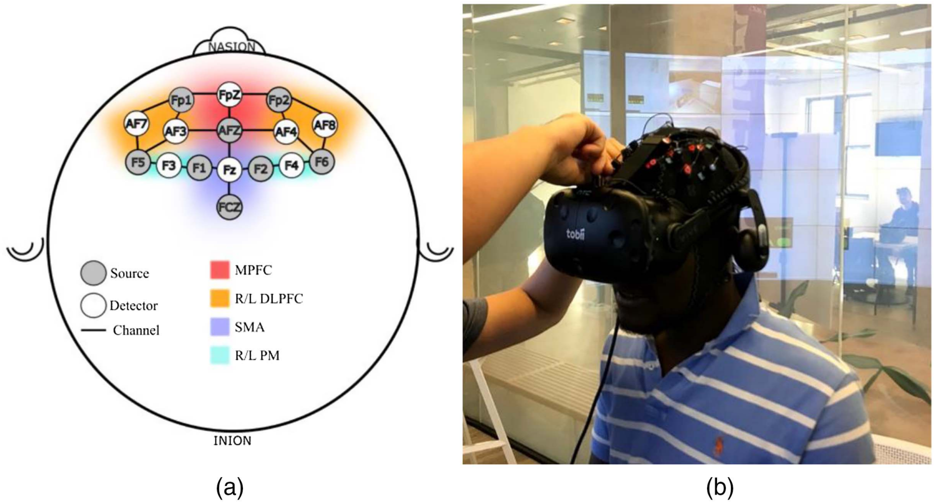

Fig. 4. Cognitive load measurement with VR: (a) probe layout; and (b) participant setting up fNIRS and VR (reprinted from Du et al. 2020, © ASCE).

prediction of the cognitive load changes. We found that our prediction model was successful in distinguishing between encoding and retrieval cognitive states, with an F1 score of 0.844 and accuracy of 79.10% when trained and tested on data collected in both stressful and normal conditions. These results indicated the efficacy of using the VR platform for data collection and model training for predicting cognitive load status in the turnaround maintenance task we studied, suggesting a possible early-warning system in the future.

Verification and Validation

A set of verification and validation activities were performed to ensure the model would trigger the proper and realistic reactions of the experiment participants. First, the pipe skid model and the environment were based on a real heat exchanger model. The scales were also carefully calibrated to reflect the correct ergonomics. The operation steps were based on the manual of Alfa Laval plate heat exchangers. Subject matter experts, i.e., mechanical engineers, were consulted at the end of model development to make sure that the model details, scales, and operation steps were accurate. Second, we programmed a room-scale motion tracking system that allowed the experiment participants to move their bodies in a natural way to finish the operation task. The relative scale of the model over the avatar was calibrated to capture the realistic spatial configuration. Collision detection was added to simulate the constraints in real operations. Finally, the sense of presence questionnaires and NASA TLX surveys were used to validate whether the model triggered sufficient subjective presence sense and high-enough cognitive load for the task. These subjective measures were also compared with the objective sensor data of fNIRS to ensure the validity of the measures.

Example Two: VR Serious Game for Investigating Earthquake Behavioral Responses and Preparedness in Buildings

Needs Identification

Earthquakes are ever-occurring disasters, with an estimation of 100 significant events a year hitting different regions in the world with a range of impacts and damage on structural and nonstructural components of buildings (Cobum et al. 1992; USGS 2021). The structural integrity of a building can be ensured (Ye et al. 2008), but what causes further injuries and casualties in building occupants is largely nonstructural damage (Feng et al. 2020a). Thus, suitable and quick behavioral responses of building occupants during an earthquake and their associated postearthquake evacuation effectiveness are instrumental to increasing chances of survival (Alexander 2012). During an earthquake, "drop, cover, and hold" are the preferred behaviors, and follow-on behaviors should be looked into as the recommended best practice during postearthquake evacuation (NZ-MCDEM 2015). The behavior is the output of cognitive and decision-making process (Aarts et al. 1998). In that regard, investigating cognitive and decision-making processes can provide insights into evacuation behavior.

First, XR technologies, specifically VR coupled with serious games (SGs), provide the possibility to deliver realistic virtual evacuation drills for observing occupants' behaviors and responses in catastrophes with high ecological validity. SG is a kind of video game whose primary goal is education other than entertainment (Wouters et al. 2009). SG can be combined with VR to create highly effective educational and training applications (Huber et al. 2017). VR SG use for earthquake emergencies is still rare. In this case study, VR SG is used to observe and analyze how occupants in built environments make decisions during earthquakes and postearthquake evacuation (Feng et al. 2020a). VR is appropriated from an embodied cognition perspective because when individuals interact with the dynamic changes caused to the indoor environment by the impact of an earthquake in a VR environment, behaviors are triggered that are almost impossible to observe via traditional drills or playing a 2D video game. Second, the fundamental questions this study aims to answer relate to whether and how an individual can be trained in a robust fashion and achieve appropriate learning outcomes and behavioral shift toward best response practices, so that the individual can generate a suitable response during an earthquake and postearthquake evacuation. VR enables a controlled experimental environment in which different training and pedagogical approaches can be trained, and various hypotheses can be reliably tested. Third, the selection of VR software, hardware, and peripherals is guided by the fact that earthquakes and postearthquake evacuation are some of the most challenging and dynamic emergency scenarios. By the time this research was undertaken, state-of-the-art hardware, software, and peripherals were used. Fourth, scenarios and storyline should be designed based on the state-of-the-art research on resilience, preparedness, and response against earthquakes. Practical inputs from experts such as emergency management practitioners as well as end users also play a vital role in the VR SG environment design.

VR Model Design

The fifth floor of the Auckland City Hospital (ACH) was chosen to develop the VR SG environment, which met the requirements for the intended experimental outcomes. A BIM-based workflow was used to represent the dynamics changes in the building components due to earthquake (Feng et al. 2018). Thus, the ACH's building section was rendered using Autodesk Revit (Autodesk 2021), with structural and nonstructural components included in the BIM model. The BIM model was imported into Unity 3D (Unity 2021) to develop the VR and game mechanisms [more details on the BIM to IVR in Lovreglio et al. (2018)].

A qualitative strategy to simulate earthquake-driven damage at the ACH was adopted, which was based on videos and images datasets of building earthquake damage (Lovreglio et al. 2018). The New Zealand Modified Mercalli Intensity Scale was used to mimic indoor building damage and to generate a realistic experience to participants (e.g., in a severe earthquake, "furniture and appliances are shifted. Substantial damage to fragile or unsecured objects" in this scale) (GeoNet 2019). Earthquake dynamics and nonstructural damage such as glass panels breaking, falling ceiling tiles, and toppling partition walls were simulated in Unity3D. To represent building damage, the specific position and orientation of damaged elements (BIM objects) such as ceiling and walls, as well as extra features such as glass that is broken, were modeled in Unity3D to provide visual cues.

A waypoint system for participants' navigation was adopted, which was a set of coordinates identifying a stopping point or where a route could be changed (Veera Ragavan et al. 2011). In each "waypoint," participants faced decision choices, which in turn reflected recommended behaviors to deal with an actual earthquake and the postearthquake evacuation. Merged recommendations from the New Zealand Civil Defense guidelines (NZ-MCDEM 2015) and the Auckland District Health Board Evacuation Plans (ADHB 2009) were used to establish the training framework to be investigated using VR and SG. Fig. 5 shows the gazes in the VR environment (waypoint) to appropriate and inappropriate behavior that a participant could choose from to continue the VR SG experience.

Experimental Design

Ethics approval to carry out this research was secured from the University of Auckland Human Participants Ethics Committee. By using emails and newsletters of the Auckland District Health Board, leaflets and posters distributed through ACH and the University of Auckland, 93 participants (43 males and 50 females) were recruited. Eighty-seven of 93 participants completed the experiment, with 25 ACH staff members (medical and administrative) and 62 visitors. An action-driven narrative method where participants' actions drive the storyline was implemented, consisting of the following main milestones: (1) Participants start the game outside the ACH; (2) Participants meet a doctor in a meeting room; (3) An earthquake strikes; (4) Shaking triggers a series of events, engaging participants in decision-making (e.g., stay under cover, take first aid kit); (5) Participants can get out of the room starting postearthquake evacuation. Other events are triggered (e.g., assist people in need, search for a suitable exit pathway, use stairs); (6) Participants exit the building to reach assembly point; and (7) Post-game assessment is provided. Immediate feedback (such as flashing lights) and postgame assessment were implemented as instructional methods to reinforce in-game best practices, knowledge acquisition and behavioral shift (Feng et al. 2020a).

A meeting room at ACH was used to carry out experiments, which had ethics clearance in place. Before participants went through the VR SG experience, a questionnaire for data collection was handled, assessing (1) demographic information; (2) fire drills and earthquake drills frequency; (3) video games playing frequency; (4) VR experience; and (5) earthquake emergency perceived self-efficacy. On top of that, participants orally answered a five-question knowledge test to assess preparedness or training knowledge. Next, participants were familiarized with VR headsets and controls and were requested to sit in a swivel chair (for health and safety reasons, the whole VR experience was in that chair). In addition, a VR tutorial session was set to assist participants in familiarizing with the VR navigation, interaction, controls, and overall nature of the virtual environment. The experimental session was started by exposing to each of the participants to the same storyline, assessing their decision-making and responses at each waypoint. The VR SG session lasted approximately 20 min, and participants were instructed to perform different tasks responding to the different hazards encountered since they entered the meeting room in the VR environment. Once the session was completed, participants were requested to answer again the five-question training knowledge test. They were also required to complete questionnaires on self-efficacy, training efficacy, and engagement.

VR-Based Data Collection and Analysis

Two instruments were deployed to evaluate preparedness of participants. (1) Knowledge acquisition (five-question knowledge questionnaires completed by participants before and after the experiment), targeting knowledge on three behavioral responses (during an earthquake-indoor space, after an earthquake-indoor

Fig. 5. An example of two alternative actions for participants to choose how to exit the building: (a) crouch below shelf (an inappropriate action); and (b) crouch below table (an appropriate action).

space, and after an earthquake-outdoor space). A knowledge scale based on recommended behaviors was developed, with scores ranging from 1 (no knowledge) to 4 (strong knowledge) (Feng et al. 2020a; Lovreglio et al. 2018). (2) Self-efficacy to cope with earthquake events. Self-efficacy is "a person's belief in his or her ability to successfully accomplish difficult tasks" (Chittaro and Sioni 2015). The self-efficacy questionnaire was completed by participants before and after the experiment and was based on the General Self-Efficacy Scale (Schwarzer and Jerusalem 1995), where participants were asked to rate six self-assessed efficacy statements using a 7-point Likert scale.

The decision-making process of participants during the virtual earthquake evacuation was investigated by assessing the video-recorded reactions during the VR SG experience (Feng et al. 2020b). The results highlighted that people's decision-making process tended to be driven by what most people around them were doing, especially those in authority positions [the doctor nonplayer character (NPC) in this case], thus accompanying other people while evacuating. Participants were also found to be inclined to have wait-or-flight responses in postearthquake evacuation. Those who had flight tendencies were more willing to exit the building immediately and did not consider advice from people in positions of authority (e.g., doctor NPC's advice of sheltering in place). The findings could contribute to improving the current earthquake evacuation guidelines and practice by encouraging people to follow suitable response guidelines and minimizing the flight responses in earthquakes.

Verification and Validation

Various verification, validity, and fidelity levels within the VG SG structure were examined as follows. First, the VR building indoor environment was based on detailed as-built ACH's 3D drawings. The VR earthquake script (code) was designed to replicate VII–VIII intensity damage (modified Mercalli scale), and the dynamic indoor damage was based on the qualitative assessment of damage provided by that scale and video footage of real earthquakes (verification, face validity, physical fidelity). Second, navigation during evacuation within the VR environment was calibrated to make sure movements were as smooth and realistic as possible without motion sickness. The position of users relative to the VR space was calibrated as well, to ensure a credible positioning experience to users (verification, face validity, and physical and ergonomic validity). Finally, subject matter experts (e.g., emergency managers, lifeline engineers) were exposed to the VR SG environment in preliminary piloting sessions during the design of the VR SG prototype to validate the credibility and consistency of the emergency recreated (content and construct validity, psychological and affective validity).

Discussion and Conclusions

As mentioned earlier, there is an increasing trend of adopting XR technologies in CEM research as an enabling tool to conduct lab or field experiments that involve human participants or concern human behaviors. This trend, according to our observation, can be largely attributed to its philosophical, methodological, and technological roots. From a philosophical perspective, many problems concerned in CEM research emerge at the interface between engineering and the social sciences. Given its interdisciplinary nature, the CEM domain has always faced a pressing need to investigate human and organizational behaviors within complex engineering contexts, and to understand engineering processes under the influence of human and organizational influences. However, it is not uncommon that CEM researchers find it highly difficult, and even prohibitive, to develop the necessary engineering contexts or reproduce the accurate human and organizational influences required for their scientific inquiries, for a variety of logistic, technical, and ethnic reasons. Fortunately, the introduction of XR-enabled methodology provides at least a partial solution to the above challenge. Based on a combinatory use of quantitative (deductive) and qualitative (inductive) methods, XR-enabled methodology offers a uniquely enabling approach to flexible, controllable, reproducible, and explainable experiments in CEM research. As illustrated by the two case studies described earlier, this emerging methodology sheds light on a range of problems in CEM that cannot be easily approached with traditional research methodologies.

We also found that recent trends in XR technology have significantly lowered the barrier of XR adoption in CEM research and improved the access to high-resolution human assessment data. First, XR devices are becoming more readily available for a broader population, making scalable participation in human-subject experiments possible. For example, more than 10 million units of Oculus Quest 2 had been sold worldwide by 2021 (The Verge 2021), making citizen participation in XR research more possible. As such, scholars may consider nontraditional approaches to access the broadening pool of human subjects, such as via crowdsourcing. Second, with the fast development of computer vision technologies, the quality of XR rendering has been substantially improved. The introduction of the new PhysX physics engine has also enabled a more realistic rendering of physical processes in the XR environment. All of these advances mean that XR is capable of reproducing real-world physics in an unprecedented manner. It is possible to facilitate studies traditionally thought impossible with XR, such as human-robot collaboration in construction. Third, we also found that recent XR literature is proposing and testing a multisensory integration approach. For example, haptics as new sensory feedback is being widely tested to enable sensory augmentation in XR. As a result, motor-intensive tasks can be easily tested with XR methods. All these trends would expand the horizon of XR-based human subject experiments in various construction tasks. Researchers are empowered by more advanced tools to produce complex, immersive, and interactive scenarios that can induce more realistic individual or collective behaviors by their human participants, and to monitor these behaviors in a more seamless, in-depth, real-time, and non-intrusive manner. This is a main reason behind the rapid growth of the volume of published XR-enabled research in recent years.

In this paper, we summarized concisely the status quo of XR-enabled CEM research. We also presented a process model of how this type of research should be conducted, which aimed to serve as an initial effort to build an organized knowledge base and workflow for using XR to meet the methodological needs in diverse scientific inquiry contexts in CEM. It is also important to note that, drawing on the diffusion of innovations theory, we believe the XR-enabled methodology has been used mainly by early adopters thus far, and it is likely to soon enter a diffusion stage where its penetration is going to grow steadily over time. There are a few factors that may drive the further adoption of XR-enabled methodology in CEM research, including the methodological innovations pertaining to access to participants (e.g., through crowdsourcing XR games) and research verification and validation (e.g., through improved validity and fidelity), as well as technological innovations that extend the functionality boundaries of XR (e.g., through integration of novel multisensory stimuli) and improve the accessibility to XR technologies (e.g., through lower expenses). These factors can be explored collaboratively by engineers, academics, and entrepreneurs in and beyond the CEM domain in the future. Despite the new opportunities brought by the new XR technologies, the fundamental principles of following a scientific process for human-subject studies would not change. For any research, the researcher would still rely on the reasoning process (i.e., deduction or induction) to frame the research questions, define the hypotheses, select the proper research method, and research tools, develop the XR environment if needed, collect measurable data, and interpret and validate the results. That being said, the process we described earlier would not be significantly changed by the new development of XR technologies. However, it is also acknowledged that not all the process model phases and steps should be used in every case. There could be research situations in which the fidelity and interactions within the XR environment are more important (particularly in observational studies) than usability (which can be more relevant in training studies). The level of control on variables and even the levels of validity and fidelity tested may vary from case to case. An important recommendation is that researchers should apply and implement the phases and activities of the process model cautiously and bear in mind the main research questions and the scope of the CEM study underway.

Lastly, this paper bears three limitations that are noteworthy. The two illustrative examples are both based on VR, which reflects its prevailing usage in the CEM domain. However, we acknowledge that AR and PMR technologies are also quickly paving their way to the CEM domain and hence deserve more attention. In addition, because of length limits, we presented only two illustrative cases, but by no means were they supposed to represent all possible scenarios in CEM research that may benefit from XR application. A third limitation is that, as mentioned earlier, numerous studies focus on developing XR-based solutions. In these studies, XR sometimes may also play a limited role in the methodological designs, which is not covered in this paper but may be worth investigation in future research.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

References

Aarts, H., B. Verplanken, and A. Van Knippenberg. 1998. "Predicting behavior from actions in the past: Repeated decision making or a matter of habit?" *J. Appl. Soc. Psychol.* 28 (15): 1355–1374. https://doi.org/10.1111/j.1559-1816.1998.tb01681.x.

ADHB (Auckland District Health Board). 2009. *Auckland district health board evacuation plan 2009*. Auckland, NZ: ADHB.

Alexander, D. 2012. "What can we do about earthquakes?" *Bull. N. Z. Soc. Earthquake Eng.* 45 (2): 49–60. https://doi.org/10.5459/bnzsee.45.2.49-60.

Alfa Laval. 2016. *Instruction manual—Plate heat exchange*. Lund, Sweden: Alfa Laval.

Alizadehsalehi, S., A. Hadavi, and J. C. Huang. 2020. "From BIM to extended reality in AEC industry." *Autom. Constr.* 116 (Apr): 103254. https://doi.org/10.1016/j.autcon.2020.103254.

APA (American Psychological Association). 2021. "APA dictionary of psychology." Accessed October 19, 2021. https://dictionary.apa.org/.

Ashwini, K. B., P. N. Patil, and R. Savitha. 2020. "Tracking methods in augmented reality—Explore the usage of marker-based tracking." In *Proc., 2nd Int. Conf. on IoT, Social, Mobile, Analytics & Cloud in Computational Vision & Bio-Engineering (ISMAC-CVB 2020)*. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.3734851.

Autodesk. 2021. "Autodesk." Accessed July 3, 2022. https://www.autodesk.com/.

Barsalou, L. 2003. "Situated simulation in the human conceptual system." *Lang. Cognit. Processes* 18 (5–6): 513–562. https://doi.org/10.1080/01690960344000026.

Barsalou, L. W. 1999. "Perceptual symbol systems." *Behav. Brain Sci.* 22 (4): 577–660. https://doi.org/10.1017/S0140525X99002149.

Barsalou, L. W. 2008. "Grounded cognition." *Annu. Rev. Psychol.* 59 (1): 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

Behzadan, A. H., and V. R. Kamat. 2013. "Enabling discovery-based learning in construction using telepresent augmented reality." *Autom. Constr.* 33 (Aug): 3–10. https://doi.org/10.1016/j.autcon.2012.09.003.

Bennett-Levy, J. E., G. E. Butler, M. E. Fennell, A. E. Hackman, M. E. Mueller, and D. E. Westbrook. 2004. *Oxford guide to behavioural experiments in cognitive therapy*. New York: Oxford University Press.

Berger, P. D., R. E. Maurer, and G. B. Celli. 2002. *Experimental design*. Belmont, CA: Wadsworth Group.

Bernold, L. E., and T. S. Lee. 2010. "Experimental research in construction." *J. Constr. Eng. Manage.* 136 (1): 26–35. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000085.

Bowman, D. A., E. Kruijff, J. J. LaViola, and I. Poupyrev. 2001. "An introduction to 3D user interface design." *Presence: Teleoperators Virtual Environ.* 10 (1): 96–108. https://doi.org/10.1162/105474601750182342.

Cao, L., J. Lin, and N. Li. 2019. "A virtual reality based study of indoor fire evacuation after active or passive spatial exploration." *Comput. Hum. Behav.* 90 (Mar): 37–45. https://doi.org/10.1016/j.chb.2018.08.041.

Chi, H. L., S. C. Kang, and X. Wang. 2013. "Research trends and opportunities of augmented reality applications in architecture, engineering, and construction." *Autom. Constr.* 33 (Aug): 116–122. https://doi.org/10.1016/j.autcon.2012.12.017.

Chittaro, L., and R. Sioni. 2015. "Serious games for emergency preparedness: Evaluation of an interactive vs. a non-interactive simulation of a terror attack." *Comput. Hum. Behav.* 50 (Sep): 508–519. https://doi.org/10.1016/j.chb.2015.03.074.

Cobum, A. W., R. J. S. Spence, and A. Pomonis. 1992. "Factors determining human casualty levels in earthquakes: Mortality prediction in building collapse." In *Proc., 10th World Conf. on Earthquake Engineering*, 5989–5994. Oxfordshire: Routledge.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Du, J., Q. Zhu, Y. Shi, Q. Wang, Y. Lin, and D. Zhao. 2020. "Cognition digital twins for personalized information systems of smart cities: Proof of concept." *J. Manage. Eng.* 36 (2): 04019052. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000740.

Duffuaa, S. O., and M. A. Ben Daya. 2004. "Turnaround maintenance in petrochemical industry: Practices and suggested improvements." *J. Qual. Maint. Eng.* 10 (3): 184–190. https://doi.org/10.1108/13552510410553235.

Feng, Y., D. C. Duives, and S. P. Hoogendoorn. 2022. "Development and evaluation of a VR research tool to study wayfinding behaviour in a multi-story building." In Vol. 147 of *Safety science*, 105573. Amsterdam, Netherlands: Elsevier.

Feng, Z., V. A. González, R. Amor, M. Spearpoint, J. Thomas, R. Sacks, R. Lovreglio, and G. Cabrera-Guerrero. 2020a. "An immersive virtual reality serious game to enhance earthquake behavioral responses and post-earthquake evacuation preparedness in buildings." *Adv. Eng. Inf.* 45 (Aug): 101118. https://doi.org/10.1016/j.aei.2020.101118.

Feng, Z., V. A. González, L. Ma, M. M. A. Al-Adhami, and C. Mourgues. 2018. "Rapid 3D reconstruction of indoor environments to generate virtual reality serious games scenarios." In *Proc., 18th Int. Conf. on Construction Applications of Virtual Reality*, 185–195. Auckland, New Zealand: Univ. of Auckland.

Feng, Z., V. A. González, M. Trotter, M. Spearpoint, J. Thomas, D. Ellis, and R. Lovreglio. 2020b. "How people make decisions during earthquakes and post-earthquake evacuation: Using verbal protocol analysis in immersive virtual reality." *Saf. Sci.* 129 (Sep): 104837. https://doi.org/10.1016/j.ssci.2020.104837.

Fich, L. B., P. Jönsson, P. H. Kirkegaard, M. Wallergård, A. H. Garde, and Å. Hansen. 2014. "Can architectural design alter the physiological reaction to psychosocial stress? A virtual TSST experiment." *Physiol. Behav.* 135 (Aug): 91–97. https://doi.org/10.1016/j.physbeh.2014.05.034.

Flavián, C., S. Ibáñez-Sánchez, and C. Orús. 2019. "The impact of virtual, augmented and mixed reality technologies on the customer experience." *J. Bus. Res.* 100 (Jul): 547–560. https://doi.org/10.1016/j.jbusres.2018.10.050.

Fu, L., S. Cao, W. Song, and J. Fang. 2019. "The influence of emergency signage on building evacuation behavior: An experimental study." *Fire Mater.* 43 (1): 33. https://doi.org/10.1002/fam.2665.

Gao, Y., V. A. González, T. W. Yiu, G. Cabrera-Guerrero, N. Li, A. Baghouz, and A. Rahouti. 2022. "Immersive virtual reality as an empirical research tool: Exploring the capability of a machine learning model for predicting construction workers' safety behavior." *Virtual Reality* 26 (1): 361–383. https://doi.org/10.1007/s10055-021-00572-9.

Garrett, J. W., and J. Teizer. 2009. "Human factors analysis classification system relating to human error awareness taxonomy in construction safety." *J. Constr. Eng. Manage.* 135 (8): 754–763. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000034.

GeoNet. 2019. "Earthquake indensity." Accessed June 1, 2020. https://www.geonet.org.nz/.

Gerloff, C. 1997. "Stimulation over the human supplementary motor area interferes with the organization of future elements in complex motor sequences." *Brain* 120 (9): 1587–1602. https://doi.org/10.1093/brain/120.9.1587.

Gibson, J. J. 2014. *The ecological approach to visual perception: Classic edition*. London: Psychology Press.

Golparvar-Fard, M., F. Peña-Mora, and S. Savarese. 2011. "Integrated sequential as-built and as-planned representation with D^4 AR tools in support of decision-making tasks in the AEC/FM industry." *J. Constr. Eng. Manage.* 137 (12): 1099–1116. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000371.

Gray, R. 2019. "Virtual environments and their role in developing perceptual-cognitive skills in sports." In *Anticipation and decision making in sport*, 342–358. New York: Routledge.

Han, Y., Y. Diao, Z. Yin, R. Jin, J. Kangwa, and O. J. Ebohon. 2021. "Immersive technology-driven investigations on influence factors of cognitive load incurred in construction site hazard recognition, analysis and decision making." *Adv. Eng. Inf.* 48 (Apr): 101298. https://doi.org/10.1016/j.aei.2021.101298.

Harris, D. J., J. M. Bird, P. A. Smart, M. R. Wilson, and S. J. Vine. 2020. "A framework for the testing and validation of simulated environments in experimentation and training." *Front. Psychol.* 11 (Mar): 11605. https://doi.org/10.3389/fpsyg.2020.00605.

Hart, S. G. 2006. "Nasa-task load index (NASA-TLX); 20 years later." In *Proc., Human Factors and Ergonomics Society Annual Meeting*, 904–908. Los Angeles: SAGE.

Hartson, R., and P. Pyla. 2019. "Agile lifecycle processes and the funnel model of agile UX." *The UX book*, 63–80. Burlington, MA: Morgan Kaufmann Publishers (an imprint of Elsevier).

Hasanzadeh, S., J. M. de la Garza, and E. S. Geller. 2020. "Latent effect of safety interventions." *J. Constr. Eng. Manage.* 146 (5): 04020033. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001812.

Hasanzadeh, S., B. Esmaeili, and M. D. Dodd. 2017. "Measuring the impacts of safety knowledge on construction workers' attentional allocation and hazard detection using remote eye-tracking technology." *J. Manage. Eng.* 33 (5): 04017024. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000526.

Heydarian, A., and B. Becerik-Gerber. 2017. "Use of immersive virtual environments for occupant behaviour monitoring and data collection." *J. Build. Perform. Simul.* 10 (5–6): 484–498. https://doi.org/10.1080/19401493.2016.1267801.

Heydarian, A., J. P. Carneiro, D. Gerber, and B. Becerik-Gerber. 2015. "Immersive virtual environments, understanding the impact of design features and occupant choice upon lighting for building performance." *Build. Environ.* 89 (Jul): 217–228. https://doi.org/10.1016/j.buildenv.2015.02.038.

Huber, T., M. Paschold, C. Hansen, T. Wunderling, H. Lang, and W. Kneist. 2017. "New dimensions in surgical training: Immersive virtual reality laparoscopic simulation exhilarates surgical staff." *Surg. Endoscopy* 31 (11): 4472–4477. https://doi.org/10.1007/s00464-017-5500-6.

Jeelani, I., K. Han, and A. Albert. 2020. "Development of virtual reality and stereo-panoramic environments for construction safety training." *Eng. Constr. Archit. Manage.* 27 (8): 1853–1876. https://doi.org/10.1108/ECAM-07-2019-0391.

Kandi, V. R., F. Castronovo, P. Brittle, S. Mastrolembo Ventura, and D. Nikolic. 2020. "Assessing the impact of a construction virtual reality game on design review skills of construction students." *J. Archit. Eng.* 26 (4): 04020035. https://doi.org/10.1061/(ASCE)AE.1943-5568.0000434.

Kato, H., and M. Billinghurst. 1999. "Marker tracking and HMD calibration for a video-based augmented reality conferencing system." In *Proc., 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, 85–94. New York: IEEE.

Khashe, S., A. Heydarian, D. Gerber, B. Becerik-Gerber, T. Hayes, and W. Wood. 2015. "Influence of LEED branding on building occupants' pro-environmental behavior." *Build. Environ.* 94 (Dec): 477–488. https://doi.org/10.1016/j.buildenv.2015.10.005.

Kim, N., J. Kim, and C. R. Ahn. 2021. "Predicting workers' inattentiveness to struck-by hazards by monitoring biosignals during a construction task: A virtual reality experiment." *Adv. Eng. Inf.* 49 (July): 101359. https://doi.org/10.1016/j.aei.2021.101359.

Kwiatek, C., M. Sharif, S. Li, C. Haas, and S. Walbridge. 2019. "Impact of augmented reality and spatial cognition on assembly in construction." *Autom. Constr.* 108 (Nov): 102935. https://doi.org/10.1016/j.autcon.2019.102935.

Lemoine, P., F. Vexo, and D. Thalmann. 2003. "Interaction techniques: 3D menus-based paradigm." In *The first Research Workshop on Augmented Virtual Reality.* AVIR 2003, (VRLAB-CONF-2007–030). Geneva: Augmented Virtual Reality.

Leppink, J., F. Paas, C. P. M. Van der Vleuten, T. Van Gog, and J. J. G. Van Merriënboer. 2013. "Development of an instrument for measuring different types of cognitive load." *Behav. Res. Methods* 45 (4): 1058–1072. https://doi.org/10.3758/s13428-013-0334-1.

Leppink, J., F. Paas, T. van Gog, C. P. M. van der Vleuten, and J. J. G. van Merriënboer. 2014. "Effects of pairs of problems and examples on task performance and different types of cognitive load." *Learn. Instr.* 30 (Apr): 32–42. https://doi.org/10.1016/j.learninstruc.2013.12.001.

Li, J., Y. Jin, S. Lu, W. Wu, and P. Wang. 2020. "Building environment information and human perceptual feedback collected through a combined virtual reality (VR) and electroencephalogram (EEG) method." *Energy Build.* 224 (Oct): 110259. https://doi.org/10.1016/j.enbuild.2020.110259.

Li, X., W. Yi, H. L. Chi, X. Wang, and A. P. C. Chan. 2018. "A critical review of virtual and augmented reality (VR/AR) applications in construction safety." *Autom. Constr.* 86 (Jul): 150–162. https://doi.org/10.1016/j.autcon.2017.11.003.

Lin, J., L. Cao, and N. Li. 2020a. "How the completeness of spatial knowledge influences the evacuation behavior of passengers in metro stations: A VR-based experimental study." *Autom. Constr.* 113 (Feb): 103136. https://doi.org/10.1016/j.autcon.2020.103136.

Lin, J., R. Zhu, N. Li, and B. Becerik-Gerber. 2020b. "Do people follow the crowd in building emergency evacuation? A cross-cultural immersive virtual reality-based study." *Adv. Eng. Inf.* 43 (Jan): 101040. https://doi.org/10.1016/j.aei.2020.101040.

Ling, Y., W.-P. Brinkman, H. T. Nefs, C. Qu, and I. Heynderickx. 2012. "Effects of stereoscopic viewing on presence, anxiety, and cybersickness in a virtual reality environment for public speaking." *Presence: Teleoperators Virtual Environ.* 21 (3): 254–267. https://doi.org/10.1162/PRES_a_00111.

Lopes, P., S. You, L.-P. Cheng, S. Marwecki, and P. Baudisch. 2017. "Providing haptics to walls & heavy objects in virtual reality by means of electrical muscle stimulation." In *Proc., 2017 CHI Conf. on Human Factors in Computing Systems*, 1471–1482. New York: ACM.

Lovreglio, R., V. Gonzalez, Z. Feng, R. Amor, M. Spearpoint, J. Thomas, M. Trotter, and R. Sacks. 2018. "Prototyping virtual reality serious games for building earthquake preparedness: The Auckland City Hospital case study." *Adv. Eng. Inf.* 38 (Oct): 670–682. https://doi.org/10.1016/j.aei.2018.08.018.

Lu, X., and S. Davis. 2016. "How sounds influence user safety decisions in a virtual construction simulator." *Saf. Sci.* 86 (Jul): 184–194. https://doi.org/10.1016/j.ssci.2016.02.018.

Magic Leap. 2021. “Magic Leap 1 is a wearable computer for enterprise productivity.” Accessed October 19, 2021. https://www.magicleap.com/en-us/magic-leap-1.

Martirosov, S., and P. Kopecek. 2017. “Cyber sickness in virtual reality—Literature review.” In *Proc., 28th DAAAM Int. Symp.*, 0718–0726. Vienna, Austria: DAAAM International. https://doi.org/10.2507/28th.daaam.proceedings.101.

Microsoft. 2021. “Microsoft HoloLens 2.” Accessed July 3, 2022. https://www.microsoft.com/es-es/hololens.

Milgram, P., and F. Kishino. 1994. “A taxonomy of mixed reality visual displays.” *IEICE Trans. Inf. Syst.* E77-D (12): 1321–1329.

North, M. J., and C. M. Macal. 2007. *Managing business complexity: Discovering strategic solutions with agent-based modeling and simulation*. New York: Oxford University Press.

NZ-MCDEM (New Zealand Ministry of Civil Defence & Emergency Management). 2015. *Working from the same page consistent messages for CDEM—Part B: Hazard specific information—Earthquake*. Wellington, New Zealand: Ministry of Civil Defence & Emergency Management.

Perfect, P., E. Timson, M. D. White, G. D. Padfield, R. Erdos, and A. W. Gubbels. 2014. “A rating scale for the subjective assessment of simulation fidelity.” *Aeronaut. J.* 118 (1206): 953–974. https://doi.org/10.1017/S0001924000009635.

Qin, S., E. J. Hermans, H. J. F. van Marle, J. Luo, and G. Fernández. 2009. “Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex.” *Biol. Psychiatry* 66 (1): 25–32. https://doi.org/10.1016/j.biopsych.2009.03.006.

Reffat, R. M., Y. Khaeruzzaman, E. El-Sebakhy, and I. P. Raharja. 2008. “Augmentation of real-time 3D virtual environments for architectural design at the conceptual stage.” *Electr. J. Inf. Technol. Constr.* 13 (Jul): 553–563.

Royce, W. W. 1987. “Managing the development of large software systems: Concepts and techniques.” In *Proc., 9th Int. Conf. on Software Engineering*, 328–338. Washington, DC: IEEE Computer Society Press.

Schwarzer, R., and M. Jerusalem. 1995. “Generalized self-efficacy scale.” In *Measures in health psychology: A user’s portfolio. Causal and control beliefs*, edited by J. Weinman, S. Wright, and M. Johnston, 35–37. Windsor, ON: NFER-NELSON. https://doi.org/10.1037/t00393-00012: 329–345.

Shi, Y., J. Du, C. R. Ahn, and E. Ragan. 2019. “Impact assessment of reinforced learning methods on construction workers’ fall risk behavior using virtual reality.” *Autom. Constr.* 104 (Apr): 197–214. https://doi.org/10.1016/j.autcon.2019.04.015.

Shi, Y., J. Du, and D. A. Worthy. 2020a. “The impact of engineering information formats on learning and execution of construction operations: A virtual reality pipe maintenance experiment.” *Autom. Constr.* 119 (Nov): 103367. https://doi.org/10.1016/j.autcon.2020.103367.

Shi, Y., Y. Zhu, R. K. Mehta, and J. Du. 2020b. “A neurophysiological approach to assess training outcome under stress: A virtual reality experiment of industrial shutdown maintenance using Functional Near-Infrared Spectroscopy (fNIRS).” *Adv. Eng. Inf.* 46 (Oct): 101153. https://doi.org/10.1016/j.aei.2020.101153.

Shin, D.-H. 2017. “The role of affordance in the experience of virtual reality learning: Technological and affective affordances in virtual reality.” *Telematics Inf.* 34 (8): 1826–1836. https://doi.org/10.1016/j.tele.2017.05.013.

Shrout, P. E., and S. P. Lane. 2012. “Psychometrics.” In *Handbook of research methods for studying daily life*, 302–320. New York: The Guilford Press.

Siedler, C., M. Glatt, P. Weber, A. Ebert, and J. C. Aurich. 2021. “Engineering changes in manufacturing systems supported by AR/VR collaboration.” *Procedia CIRP* 96 (Jan): 307–312. https://doi.org/10.1016/j.procir.2021.01.092.

Sipatchin, A., S. Wahl, and K. Rifai. 2021. “Eye-tracking for clinical ophthalmology with virtual reality (VR): A case study of the HTC vive pro eye’s usability.” *Healthcare* 9 (2): 180. https://doi.org/10.3390/healthcare9020180.

Slater, M. 1999. “Measuring presence: A response to the witmer and singer presence questionnaire.” *Presence: Teleoperators Virtual Environ.* 8 (5): 560–565. https://doi.org/10.1162/105474699566477.

Slater, M. 2009. “Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments.” *Philos. Trans. R. Soc. London, Ser. B* 364 (1535): 3549–3557. https://doi.org/10.1098/rstb.2009.0138.

Sutherland, I. E. 1968. “A head-mounted three dimensional display.” In *Proc., Fall Joint Computer Conf., Part I on - AFIPS ’68 (Fall, Part I)*, 757. New York: ACM Press.

Taylor, D. C. P. J., H. Mwiki, A. Dehghantanha, A. Akibini, K. K. R. Choo, M. Hammoudeh, and R. Parizi. 2019. “Forensic investigation of cross platform massively multiplayer online games: Minecraft as a case study.” *Sci. Justice* 59 (3): 337–348. https://doi.org/10.1016/j.scijus.2019.01.005.

Tehrani, B. M., J. Wang, and D. Truax. 2021. “Assessment of mental fatigue using electroencephalography (EEG) and virtual reality (VR) for construction fall hazard prevention.” *Eng. Constr. Archit. Manage.* https://doi.org/10.1108/ecam-01-2021-0017.

The Belmont Report. 1979. “Ethical principles and guidelines for the protection of human subjects of research.” Accessed April 18, 1979. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html.

The Verge. 2021. “Meta’s Oculus Quest 2 has shipped 10 million units, according to Qualcomm.” Accessed March 30, 2022. https://www.theverge.com/2021/11/16/22785469/meta-oculus-quest-2-10-million-units-sold-qualcomm-xr2.

Tobbi. 2019. “Tobii Pro SDK.” Accessed October 20, 2021. https://www.tobiipro.com/product-listing/tobii-pro-sdk/.

Unity. 2021. “Unity.” Accessed October 20, 2021. https://unity.com/.

USGS. 2021. “Earthquake facts.” Accessed October 14, 2021. https://www.usgs.gov/natural-hazards/earthquake-hazards/science/cool-earthquake-facts?qt-science_center_objects=0#qt-science_center_objects.

Usoh, M., E. Catena, S. Arman, and M. Slater. 2000. “Using presence questionnaires in reality.” *Presence: Teleoperators Virtual Environ.* 9 (5): 497–503. https://doi.org/10.1162/105474600566989.

Veera Ragavan, S., S. G. Ponnambalam, and C. Sumero. 2011. “Waypoint-based path planner for Mobile Robot Navigation using PSO and GA-AIS.” In *Proc., IEEE Recent Advances in Intelligent Computational Systems*, 756–760. New York: IEEE.

Wang, P., P. Wu, J. Wang, H. L. Chi, and X. Wang. 2018. “A critical review of the use of virtual reality in construction engineering education and training.” *Int. J. Environ. Res. Public Health* 15 (6): 1204. https://doi.org/10.3390/ijerph15061204.

Wang, X., and R. Chen. 2009. “An experimental study on collaborative effectiveness of augmented reality potentials in urban design.” *CoDesign* 5 (4): 229–244. https://doi.org/10.1080/15710880903320020.

Wang, X., A. Eiselmayer, W. E. Mackay, K. Hornbaek, and C. Wacharamanotham. 2021. “Argus: Interactive a priori power analysis.” *IEEE Trans. Visual Comput. Graphics* 27 (2): 432–442. https://doi.org/10.1109/TVCG.2020.3028894.

Westen, D., and R. Rosenthal. 2003. “Quantifying construct validity: Two simple measures.” *J. Personality Soc. Psychol.* 84 (3): 608–618. https://doi.org/10.1037/0022-3514.84.3.608.

Wilson, M. 2002. “Six views of embodied cognition.” *Psychonomic Bull. Rev.* 9 (4): 625–636. https://doi.org/10.3758/BF03196322.

Wouters, P., E. D. van der Spek, and H. van Oostendorp. 2009. “Current practices in serious game research.” In *Games-based learning advancements for multi-sensory human computer interfaces*, 232–250. Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-60566-360-9.ch014.

Xia, X., N. Li, and V. A. González. 2021. “Exploring the influence of emergency broadcasts on human evacuation behavior during building emergencies using virtual reality technology.” *J. Comput. Civ. Eng.* 35 (2): 04020065. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000953.

Xu, Q., H.-Y. Chong, and P. Liao. 2019a. “Exploring eye-tracking searching strategies for construction hazard recognition in a laboratory scene.” *Saf. Sci.* 120 (Dec): 824–832. https://doi.org/10.1016/j.ssci.2019.08.012.

Xu, S., Q. Ni, and Q. Du. 2019b. “The effectiveness of virtual reality in safety training: Measurement of emotional arousal with electromyography.” In *Proc., 36th ISARC*. Banff, Canada.

Yan, X., H. Li, A. R. Li, and H. Zhang. 2017a. “Wearable IMU-based real-time motion warning system for construction workers’ musculoskeletal disorders prevention." *Autom. Constr.* 74 (Feb): 2–11. https://doi.org/10.1016/j.autcon.2016.11.007.

Yan, X., H. Li, C. Wang, J. Seo, H. Zhang, and H. Wang. 2017b. "Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion." *Adv. Eng. Inf.* 34 (Oct): 152–163. https://doi.org/10.1016/j.aei.2017.11.001.

Yang, K., C. R. Ahn, M. C. Vuran, and S. S. Aria. 2016. "Semi-supervised near-miss fall detection for ironworkers with a wearable inertial measurement unit." *Autom. Constr.* 68 (Aug): 194–202. https://doi.org/10.1016/j.autcon.2016.04.007.

Ye, L., Z. Qu, X. Lu, and P. Feng. 2008. "Collapse prevention of building structures: A lesson from the Wenchuan earthquake." *J. Build. Struct.* 4 (7): 1–9. https://doi.org/10.14006/j.jzjgxb.2008.04.010.

Yeom, D., J.-H. Choi, and S.-H. Kang. 2019. "Investigation of the physiological differences in the immersive virtual reality environment and real indoor environment: Focused on skin temperature and thermal sensation." *Build. Environ.* 154 (May): 44–54. https://doi.org/10.1016/j.buildenv.2019.03.013.

Zhang, Y., H. Liu, S.-C. Kang, and M. Al-Hussein. 2020. "Virtual reality applications for the built environment: Research trends and opportunities." *Autom. Constr.* 118 (Oct): 103311. https://doi.org/10.1016/j.autcon.2020.103311.

Zhu, Q., J. Du, Y. Shi, and P. Wei. 2021. "Neurobehavioral assessment of force feedback simulation in industrial robotic teleoperation." *Autom. Constr.* 126 (Jun): 103674. https://doi.org/10.1016/j.autcon.2021.103674.

Zhu, Y., S. Saeidi, T. Rizzuto, A. Roetzel, and R. Kooima. 2018. "Potential and challenges of immersive virtual environments for occupant energy behavior modeling and validation: A literature review." *J. Build. Eng.* 19 (Sep): 302–319. https://doi.org/10.1016/j.jobe.2018.05.017.