Automation in Construction 119 (2020) 103367

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

The impact of engineering information formats on learning and execution of construction operations: A virtual reality pipe maintenance experiment

Yangming Shi[a], Jing Du[a,*], Darrell A. Worthy[b]

[a] Department of Civil and Coastal Engineering, Engineering School of Sustainable Infrastructure and Environment (ESSIE), Herbert Wertheim College of Engineering, University of Florida, Gainesville, FL 32611, United States

[b] Department of Psychological & Brain Sciences, College of Liberal Arts, Texas A&M University, College Station, TX 77843-4235, United States

ARTICLE INFO

Keywords:
Working memory
Engineering information
Cognitive load
Virtual reality
Eye-tracking

ABSTRACT

Given the increasing complexity of construction tasks, and the growing number of construction operations within confined workplaces, construction workers rely heavily on working memory. In this context, working memory is defined as the short-term and temporary storage of information related to near future events to ensure the seamless execution of construction tasks. Although a strong relationship between engineering information format and task performance has been observed in the relevant literature, there remains an obvious theoretical disagreement—in particular, about the cognitive mechanisms for explaining why different information formats affect working memory development and retrieval in distinct ways. This study presents a human–subject experiment (n = 120) to examine the impact of information format on the performance of a pipe maintenance task, and the implications of cognitive costs in both working memory development (information encoding) and retrieval (information recalling). Participants were required to review the operational instructions for a pipe maintenance task for a short period, and then perform the task from memory. Participants were divided into four groups depending on the format of information they received: 2D isometric drawing of the plate heat exchanger with bulleted-text operational instructions (2D-simple group); 2D isometric drawing of the plate heat exchanger with rich-text operational instruction (2D-complex group); an interactive 3D model of the plate heat exchanger with bulleted-text operational instructions (3D group); or an immersive Virtual Reality (VR) environment with bulleted-text operational instructions (VR group). The results indicated that 3D and VR groups outperformed 2D-simple and 2D-complex groups in both operation time and maintenance accuracy. A further cognitive load analysis (based on surveys and pupil dilation) suggested that the superior performance of these groups is driven by more efficient usage of working memory, measured by how easily the encoded information can be recalled in the operation phase. Larger pupil dilation during encoding, indicative of successful working memory formation, was associated with better subsequent performance. These findings provide more evidence about cognitive mechanisms engaged by different information formats, help to resolve the current theoretical disagreement within the construction literature, and may inspire designs of cognition-driven information systems that can improve working memory in construction workers.

1. Introduction

The Architecture, Engineering, and Construction (AEC) industry is experiencing a rapid digital revolution. The United Kingdom (UK)'s Construction 2025 report [95] suggests that adopting and advancing digital technologies throughout all phases of AEC projects, from the design phase to the construction and operational phases, should be the first strategic priority for the AEC industry in the next decade. One of the representative phenomena is the emergence of new visualization technologies for intuitive communication of complex engineering information, including Building Information Modeling (BIM), Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR). Many studies in the AEC industry have already demonstrated the potentials of VR and AR technologies for project communications [21,22,59], energy use optimization [36,37], construction safety [76,81,110], training and education [2,75,109], and facility management [38,82,101]. In addition, there is a growing interest in examining the impact of emerging information communication methods, such as 3D models and VR/AR, on construction worker task performance [3,69,77]. However, the existing literature provides conflicting findings. For example, some

* Corresponding author.
E-mail addresses: shiyangming@ufl.edu (Y. Shi), eric.du@essie.ufl.edu (J. Du), worthyda@tamu.edu (D.A. Worthy).

https://doi.org/10.1016/j.autcon.2020.103367
Received 4 November 2019; Received in revised form 22 May 2020; Accepted 16 July 2020

scholars have claimed a positive impact of emerging visualization technologies (3D models and VR) on task performance of construction operations [3,69,77,89], while other scholars have not found beneficial effects [10,27,73]. Clearly, there is a knowledge gap regarding the role of information formats on task performance in construction operations, and further research is needed. Moreover, the underlying cognitive mechanisms whereby different information formats might lead to different levels of task performance remain poorly understood.

Working memory is one cognitive mechanism that is critical for learning, retention and execution of complex construction operations. Another construct related to working memory is cognitive load. Different information presentation formats may lead to different levels of cognitive load that can tax working memory resources. For example, one format may appear very dense, and require enhanced focus and concentration during learning, while another format may allow for easier interpretation. In this paper we examine the cognitive mechanisms underlying successful learning and execution of a pipe-maintenance task in a VR environment by comparing task performance across groups that differ based on the format of information they are given to learn the task. We also collect proxy measures of working memory and cognitive load such as pupil dilation, which has shown to be related to cognitive load [40,54,62,97]. These proxy measures should allow us to provide a better mechanistic account of the effects of information format on learning and task execution of complex plans.

The reason we selected pipe maintenance as the test case is that pipe maintenance plays an important role in modern industrial construction projects. Construction workers or maintenance professionals often have to enter confined spaces to perform pipe maintenance tasks such as routine maintenance, repairs, and inspections. It is worth noting that workers only have a limited amount of time (approximately 15 min) to perform work in confined spaces to comply with the safety codes [60] and have limited access to engineering information while constrained by the physical environments. Therefore, most of the pipe maintenance and repair tasks require workers to highly rely on their working memory to memorize the operational instructions and perform the tasks. Meanwhile, limited space, time pressure, and unfavorable working environments can increase the risk for severe or fatal injuries which often occur in confined spaces. According to a report from the Bureau of Labor Statistics (BLS), fatal occupational injuries involving confined spaces increased by 15% to 166 in 2017 from 144 in 2016 [93]. Moreover, 52% of these injuries happened when workers entered confined spaces to perform the pipe maintenance task of routine maintenance, repairs, and inspections [45]. Therefore, there is an urgent need to better understand how different information formats can affect a construction worker's ability to learn and successfully execute a pipe maintenance task. The results of this study will help construction professionals design better maintenance and inspection protocols.

This study aims to investigate how different information formats (2D isometric drawings, 3D models, and VR models) affect performance in a pipe maintenance task. We hypothesize that the 3D and VR formats will facilitate the best learning, and that this effect will be driven by reduced cognitive load in the learning phase (i.e., fewer mental difficulties when memorizing information during the engineering review), and when recalling the same information in the operation phase [32]. We developed an interactive VR system with eye-tracking functions to simulate a pipe maintenance task of replacing a plate heat exchanger. Metrics of participants' task performance (including operation time and operation accuracy) were used as the indicators of working memory quality, while pupil dilation data and cognitive load questionnaires were used to evaluate the cognitive load levels in review and operation sessions. The remainder of this paper introduces the point of departure of this study, the research method and the experiment, and the findings and recommendations.

2. Literature review

2.1. Working memory and sensory-processing

Working memory plays a critical functional role in cognitive tasks such as learning, reasoning, and comprehension [4], and it represents a complex cognitive process [31,46,85]. Several different theories related to working memory have been proposed—including working memory theory [5], the multimedia learning theory [55], and Wickens' model of human information processing [107]. According to these theories, there are two main phases when a person processes information: the encoding/learning phase (i.e., digesting and storing information temporarily in the working memory) and the retrieval/execution phase (i.e., recalling temporarily stored information in knowledge-based tasks) [107]. In the learning phase, different types of sensory information such as visual, auditory, or haptic is processed in parallel; this leads to a multi-component structure of working memory [4–8]. Visual and auditory information is further segregated by what type of processing it requires. For example, some visual information may be processed by a 'phonological loop' that processes text, while other visual information may be processed based on its color, shape, or motivational significance. Despite the variety of information formats, much of learning involves two main types of information: phonological information (i.e., auditory-verbal information or text information) and visuospatial information (i.e., visually presented information or spatial related information). Thus, how information is visualized and presented to the construction worker affects the method and quality of learning and subsequent task completion.

Amid the fast development of information technologies, diverse ways of engineering visualization technologies have been introduced to the architecture, engineering, and construction (AEC) industry, such as Building Information Modeling (BIM), photogrammetry, Virtual Reality (VR) [21,22], and Augmented Reality (AR). However, how these emerging visualization technologies affect the learning and execution of construction tasks is still not fully understood, and as a result, the literature tends to present conflicting findings. Some researchers have found that more immersive visualization (using 3D models and VR/AR) can improve individual working memory [24], procedural memory [70], spatial configuration memory [68], and spatial cognition [98]. For instance, Dünser et al. [24] found that VR/AR technologies were effective in improving students' spatial memory. Meanwhile, the addition of visual display features (such as stereoscopic visuals and head-tracked viewing) have been shown to improve test subjects' ability to identify gaps in visual geometries significantly [3,69]. Sampaio et al. [77] compared students' learning outcomes for design and construction projects with different visual content. They found that 3D modeling and VR models delivered better outcomes than traditional 2D methods. Tavanti and Lind [89] conducted two experiments to investigate spatial memory empirically based on 2D and 3D displays. They found that 3D displays could improve performance in designed spatial memory tasks. In contrast, cumulative evidence indicates that in certain situations 3D or VR visual representations are no better than the traditional 2D method, due to the cognitive burden of processing additional information such as texture, colors, and orientations [10,27,73]. For instance, Bliss et al. [15] compared three visual methods for a navigation task in an unfamiliar building using 2D blueprint, VR, and no-training (control group). They found there was no significant difference between the VR and blueprint groups regarding speed and accuracy of the navigation performance. Richards and Taylor [72] conducted a human-subject experiment (n = 129) to compare biology students' learning gained from class content by using a 2D method and 3D virtual models. They found that the 2D visualization led to a better learning outcome than using 3D virtual models. They explained that 3D virtual models might provide more detail and more distractors to users, causing cognitive overload during the learning phase. Cockburn [19] reproduced the experiment designed by Tavanti and Lind with a stricter control to explore the effectiveness of spatial memory by using 2D and 3D displays. They also found that 3D visualizations did not improve spatial memory development in monocular static displays.

Although the findings presented above are somewhat mixed, models of information processing such as Wickens' multiple resource model of human information processing [105,107] suggest that information formats that can be processed by multiple different sensory modalities, like 3D or VR, may enhance learning. This model posits that information is processed in parallel by multiple sensory modalities. For example, phonological information can be processed simultaneously along with visuospatial information, while there is a limited store of information that can be simultaneously processed within the same sensory modality. Extensive support for multiple resource theory comes from dual-task studies where people must perform two tasks simultaneously [41,106]. People are better at performing two tasks that require different sensory modalities, like an auditory and a visual task, than they are at concurrently performing two tasks that require the same sensory modality, like two visual tasks. This theory is relevant to the present work because our 2D information formats may not be processed by fewer sensory modalities than the 3D and VR formats. The latter formats may engage a broader variety of sensory modalities. This could lead to learning of information about the task execution plan in parallel through multiple sensory modalities, which should lead to enhanced learning overall. Although, it's possible that aspects of the 3D and VR environments may create a distraction when trying to process the text-based information, multiple resource theory suggests that this information will not be distracting because it is not processed by the same sensory modality.

2.2. Indicators of information processing

To provide a better mechanistic explanation for why certain information formats lead to better learning we collected data on pupil dilation during the training and test phases of the experiment. Pupillometry has a long history in studies of memory and attention [32,44,46,61]. Some have suggested that pupil dilation reflects a 'summed index' of neural activity during cognitively demanding tasks [32,40,63]. Pupil dilation has been shown to increase under greater cognitive load. As a task becomes progressively more difficult pupil dilation increases. In our experiment enhanced pupil dilation during encoding may be correlated with better performance during the test phase. Based on Wicken's multiple resource information processing model, it's possible that the VR and 3D conditions will lead to greater pupil dilation, indicative of greater cognitive load, because information will be spread more diffusely across different modalities than in the 2D conditions, allowing more information in total to be processed. This would be consistent with studies that have found that larger pupil dilations during training are associated with better subsequent recall of information [35,46]. In addition to pupil dilation we also included a cognitive load questionnaire after the experiment.

3. Research methodology

3.1. Overview

This study focused on the cognitive processes related to information processing and learning and execution of a complex task, as illustrated in Fig. 1. To rule out the impacts of other cognitive processes, we designed a controlled experiment where participants mainly relied on their working memory to perform a sequential operation. In other words, the experiment design did not require higher level cognitive processes, such as reasoning; rather, memorizing information about the operational sequence in the review phase and using this stored information in the operation phase were the main things participants needed to do to successfully complete out experimental task. Our experiment consisted of two phases: a training phase where participants saw the information in different information formats and a test phase where participants completed the task in a VR environment. In addition, we controlled the types of information based on the working memory theories reviewed above. This included phonological information (text information in review phase) and visuospatial information (2D isometric drawing or 3D model in review phase). By carefully designing the experiment and tasks, we minimized the influences of other cognitive processes. An interactive VR system with eye-tracking functions was used in our human-subject experiments to create an immersive virtual environment that realistically simulated a pipe maintenance task in a confined space. The method consists of two main steps: data collection and data analysis.

3.2. Data collection

For data collection, we used an interactive VR system with motion tracking and eye-tracking functions to collect participants' behavioral data. A virtual working scenario of replacing a plate heat exchanger in a confined space was designed as the task. In the experiment, participants were asked to memorize a 10-step pipe maintenance sequence using one of the four operational instructions. According to multimedia learning theory [56] and working memory theory [5], working memory development involves dual-channel information processing driven by two main types of information: phonological information (i.e., auditory-verbal information or text information) and visuospatial information (i.e., visually presented information or spatial related information). Therefore, the operational instructions shown to the participants consist of two parts: operational text narratives and visual representation of the plate heat exchanger. To investigate the influence of visuospatial information on working memory development and retrieval, we first designed three groups: a 2D isometric drawing of the plate heat exchanger with bulleted text operation instruction narratives on the monitor (2D-simple group); an interactive 3D model of the plate heat exchanger with bulleted text operation instruction narratives (3D group); and a headset-enabled VR model of the plate heat exchanger with bulleted text operation instruction narratives (VR group). The operational text instructions were same across these three groups. To explore the influence of phonological information on working memory development and retrieval, we designed the information format for the 2D-complex group as a 2D isometric drawing of the plate heat exchanger with rich text operating instruction narratives compared to 2D-simple group. Thus, there were a total of four groups: 2D-simple, 2D-complex, 3D, and VR. After the review session, participants were asked to perform the task based on their memory of the plan from the review session in the VR environment. Participants' motion, operation time, operation accuracy, gaze data, and pupil diameters were collected during the experiment. In addition, we used a background questionnaire to collect participants' demographic information, and we applied two ability tests (the spatial cube test and the shape memory test) to evaluate participants' spatial cognition and memory abilities. The cognitive load questionnaire was used at the end of the experiment to evaluate participants' cognitive load and validate the results assessed by the pupillary data.

3.3. Data analysis

After the experimental data was collected, we analyzed task performance and cognitive costs in the working memory development and retrieval phases. Participants' operation time (s) and pipe maintenance accuracy (%) were used as indicators of task performance. Pipe maintenance accuracy was defined as the accuracy in performing correct steps, and directly represents how well the participants memorized and performed the pipe maintenance task. Pipe maintenance accuracy was recorded in a range from 0% to 100%. The operation time was defined as the time participants used to complete the task in the virtual environment. This indicator represents how efficiently the participants

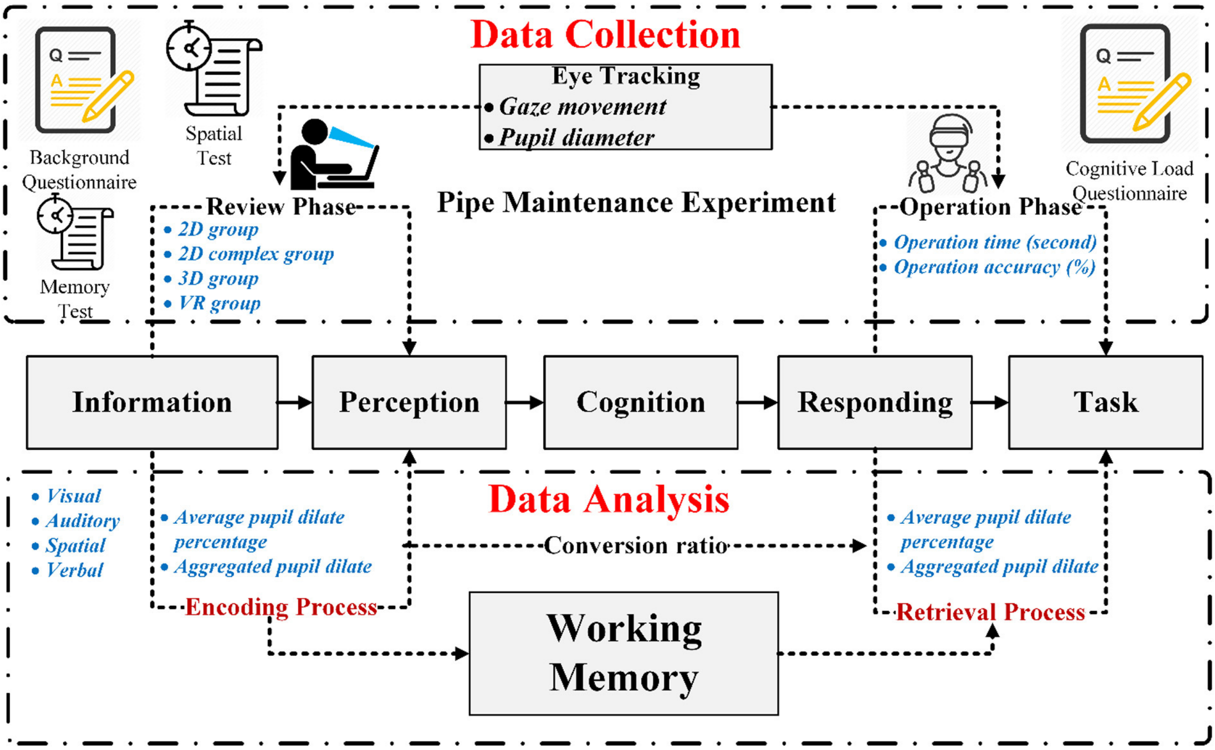

Fig. 1. Research methodology.

finished the task. Furthermore, we used the cognitive load questionnaire and pupil diameter data to evaluate cognitive load levels in the review and operation sessions.

Existing literature in the AEC industry does not provide a readily available approach for pupillary data analysis when light response effects and personal differences are present. We developed a novel pupil diameter analysis approach to evaluate participants' cognitive status based on their pupil diameter data collected by eye trackers, as illustrated in Fig. 2. This approach allowed us to capture participants' cognitive load status related to working memory development and retrieval in real time. Moreover, we also used a well-validated cognitive load questionnaire to validate the cognitive results calculated by the pupillary analysis approach.

To analyze the pupillary data, we first performed pupillary blink response correction to the raw pupil diameter data. The blink responses in the data are characterized as rapid declines towards 0 at blink onset and rapid rises from 0 back to a regular value at blink offset within 400–600 ms [51]. We used linear interpolate method to filter out all eye blinks from the raw data [44,52]. Second, according to previous pupillary studies [29,57], increasing motor task complexity can produce increased pupil diameter. In contrast, increasing motor task precision appears to cause a decrease in pupil diameter in the period between response planning and response execution [29]. Therefore, we designed the experiment in the following ways to control the influences of motor task complexity and motor task precision in this study. In the review phase of the experiment, all participants were instructed to sit or stand still to review the pipe maintenance instructions across different groups. In this phase, no motor tasks were involved. In the operational phase, we designed the same level of task complexity across different groups (10-step motor task). We also developed an innovative approach

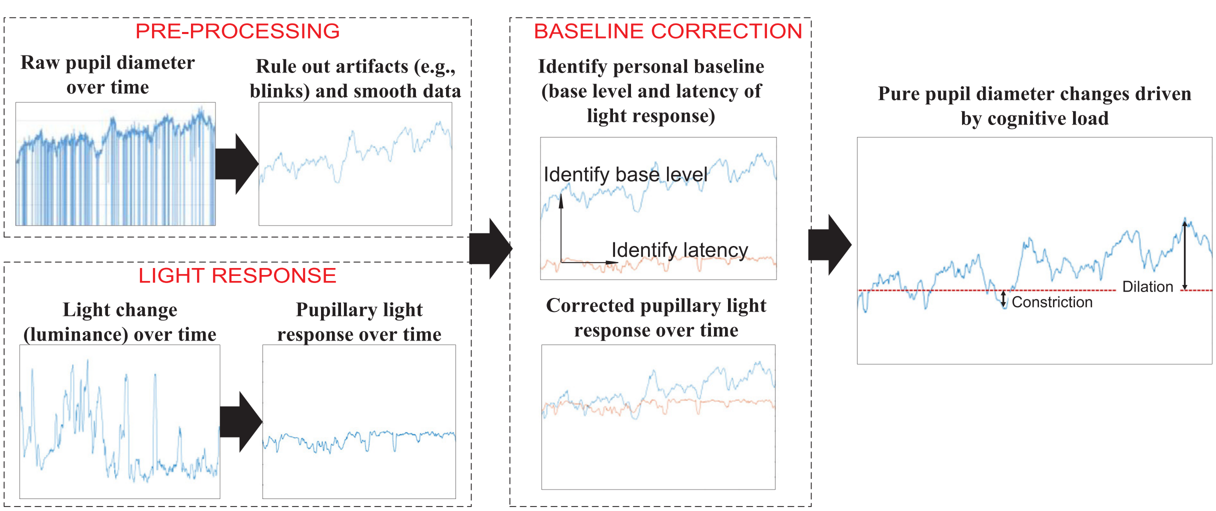

Fig. 2. Workflow of pupil diameter analysis.

to ensure consistent motor precision requirements, called the invisible collider-box method. This method assigns an invisible collider-box to each of valves in the virtual environment. The valve rotation can only be triggered when a participant's "virtual hands" reach the same radius of the collider box. In this way, participants were "forced" to apply the same level of precision in their tasks. As a result, both motor task complexity and precision level were controlled. Finally, in the pupillary data pre-processing phase, we utilized the Hampel filter [65] to remove artifacts and smooth the data [30]. This method has been widely used in different pupillary studies to reject artifacts [67,74]. Third, we also ruled out the *pupil light reflex* [42] in the pupil size data. According to the study by [44], the pupillary light reflex is much larger in magnitude than cognition-induced pupil changes. Thus, we followed standard practice [44,57,99] to maintain constant environmental luminance during the experiment. As a result, luminance changes from the displays (monitor and lens in the headset) were the only lighting factors that affected participants' pupillary light response [44]. Because the experiments were all performed at the same location, the environment lighting was consistent across different groups (identical lighting). As a result, the luminance change from the displays (monitor used in 2D and 3D groups and headset lens used in the VR group) was the only factor that affected participants' pupillary light response. We do want to note that the luminance from the display was different than that from the headset lens. Therefore, we developed an algorithm in the system to calculate the luminance received by both eyeballs based on the RGB values of all the pixels on the display (monitor and lens). Literature [100] shows that no matter what the sources are, the luminance received by human eyes can be computed based on relative strength of the three primary colors – red, green and red. We calculated the real-time luminance following:

$$\text{Luminance} = \text{sqrt}\,(0.299\ R^2 + 0.587\ G^2 + 0.114\ B^2)$$

Then we applied a pupil diameter light response formula developed by [13,17] to quantify the pupil diameter changes caused by the luminance changes from the displays. Using the above approach, we filtered out light effects and obtained the pupillary changes purely driven by cognitive load differences that resulted from our experimental manipulation. Then we selected the mean value of the first 90 samples (1 s) as the baseline of pupil size to perform the subtractive baseline correction [53]. The purpose of using subtractive baseline correction is to adjust the general pupil light response model and capture the personal baseline. Finally, we applied a symbolic approximation (SAX) approach [43,49,102] to correct pupil light response latency, as each individual is different in terms of how fast their pupil size changes according to light changes [9,12,108]. After the baseline correction, the variations of pure pupil diameter can directly represent the participants' real-time cognitive load status. Because well-documented evidence indicates that pupil dilation often relates to an increase in cognitive load [18,32,34,61,62], we focused on assessing pupil dilation frequency and magnitude.

Two features were selected to capture pupillary changes, including average pupil dilation percentage (%) and aggregated pupil dilation (mm) as listed in Table 1. The average pupil dilation percentage is the mean percentage value of pupil dilation over the baseline over time. According to [32,64], the pupil dilation percentage compared to the baseline can represent mental demand levels. Aggregated pupil dilation is the total value of pupil dilation above the personal baseline over time. Aggregated pupil dilation represents total cognitive demand during a task [32,44,46,61].

3.4. Cognitive load validation

To validate the cognitive load measured by pupil dilation, we used a cognitive load measurement questionnaire to evaluate participants' cognitive load levels after they had finished the task. This method is widely used in existing pupil diameter and cognitive load studies [71,80]. Instead of using a traditional cognitive load measurement such as the NASA Task Load Index (TLX), we employed a more advanced cognitive load measurement proposed by Leppink [47,48] that can directly measure the three types of cognitive load previously identified (intrinsic cognitive load, extraneous cognitive load, and germane cognitive load). According to Sweller's Cognitive Load Theory [87], intrinsic cognitive load is affected by the complexity of tasks, extraneous cognitive load is affected by how information is presented, and germane cognitive load is affected by the previous knowledge of the tasks. The results of Leppink's cognitive load measurement serves as a validation of the pupil analysis results.

4. Human-subject experiment

4.1. VR system with eye tracking feature

Owing to the difficulty of simulating complex pipe maintenance tasks in confined spaces in the real world, an interactive VR system was developed based on our previously well-validated VR systems [22,23,81,82]. To obtain precise and high-resolution pupil data, two eye trackers were used in this study (the Tobii Pro X3-120 and the Tobii Pro VR) as shown in Fig. 3. Both eye trackers were manufactured by Tobii and both use advanced Pupil Centre Corneal Reflection (PCCR) remote eye-tracking technique to capture eyeball movement and pupil size [90]. The Near-infrared illuminators in the eye tracker are used to create the reflection patterns on the cornea and pupil of the eye. The cameras in the eye tracker are used to capture high-resolution images. Finally, the advanced image-processing algorithms and a physiological 3D model of the eye are implemented to estimate the position of the eye in the space and pupil size [90]. Since both eye trackers use the same eye-tracking technique and the same Software Development Kit (SDK) provided by Tobii, participants' eye movement and pupil size was consistently captured across different groups. For the 2D and 3D groups, the Tobii Pro X3-120 was mounted to the monitor to record participants' gaze positions and pupil diameters at a frequency of 120 Hz when they stared at the monitor screen. The Tobii Pro X3-120 has an accuracy of 0.4° on the monitor with an operating distance of 50–90 cm [92]. At the end of each experiment trial, the system automatically generated a TSV file with all raw data. For the VR group, the Tobii Pro VR eye tracker was integrated into the HTC VIVE Head Mounted Display (HMD). The Tobii Pro VR integration eye tracker has an accuracy of 0.5° and the gaze data output frequency is 120 Hz [91]. To achieve the eye-tracking and visualization functions in the virtual environment, several C# scripts were developed based on the Tobii Pro SDK [91] and the application programming interface (API) in Unity. In

Table 1
Pupil dilation features.

Feature	Unit	Equation
Average pupil dilation percentage	%	$\text{DP} = \text{mean}\left(\sum_{i=1}^{n} \frac{d_i}{b_i}\right)$ d_i is the pupil dilation of a frame. b_i is the pupil diameter baseline of a frame. n is the number of frames that pupil dilated. DP is the average pupil dilation percentage
Aggregated pupil dilation	mm	$\text{D} = \sum_{i=1}^{n} d_i$ d_i is the pupil dilation of a frame. n is the number of frames that pupil dilated. D is the aggregated pupil dilation

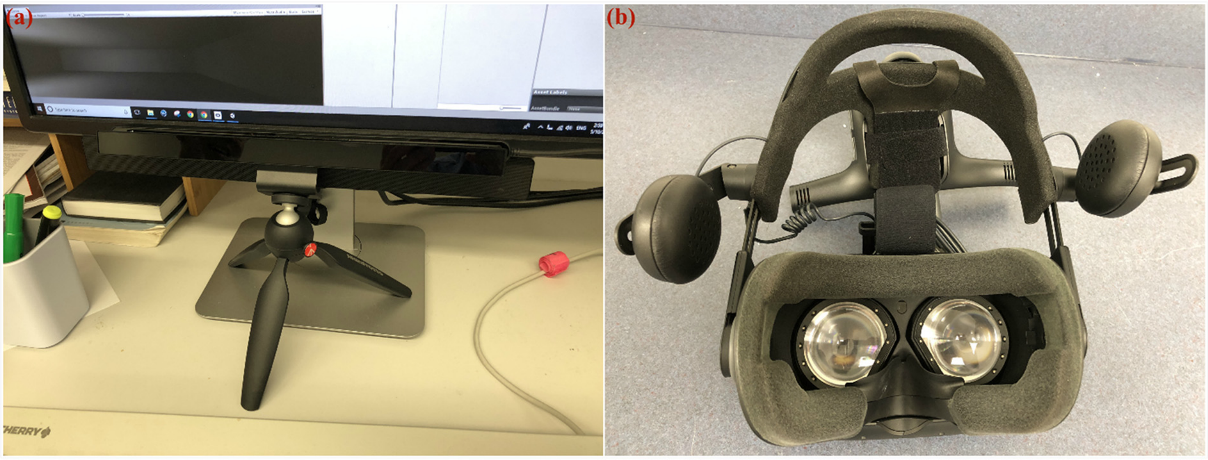

Fig. 3. Eye trackers used in this study: (a) Tobii Pro X3-120 eye tracker; (b) Tobii Pro VR Integration eye tracker.

the virtual environment, data relating to participant gaze movement, head movement and rotation, body movement, hand movement, and pupil diameter were collected by the system with a frequency of 90 Hz. After each VR experimental trial, the developed VR system automatically generated a CSV file with all the raw data. The VR HMD used in the study was the HTC VIVE [39], with a field of view (FOV) of 110° and a resolution of 1400×1600 pixels per eye for the dual displays [39]. The VR system used in the experiment was developed with the Unity 3D-5.6.3f1 version and the pipe model was developed based on the plate heat exchanger model developed in SketchUp. The VR system ran on a workstation with an Intel Xeon CPU at 2.60 GHz with 64 GB of RAM. The workstation graphic card was the NVIDIA GTX 1080. Fig. 4 shows that participants from different groups utilized the eye-tracking system to review the pipe model and memorize the pipe maintenance sequence and performed the task in the immersive virtual environment.

4.2. Experiment task and virtual environment

Participants were asked to memorize sequences for turning or closing the valves before they replaced the plate heat exchanger. The pre-start-up sequences to cut off the hot water and cold water consist of 10 steps, which were developed based on the operation instruction manual of Alfa Laval plate heat exchangers, as listed in Table 2 [1]. To examine the impact of information formats, four types of operation instructions were designed for each group (2D-simple, 2D-complex, 3D, and VR). Operation instructions for the 2D-simple group were designed as a 2D isometric drawing of the plate heat exchanger with bulleted text operating instruction narratives on the monitor as shown in Fig. 5(a). Similarly, the 2D-complex operating instructions were designed as a 2D isometric drawing of the plate heat exchanger with rich text operating instruction narratives displayed on the monitor as shown in Fig. 5(b). The 3D operating instructions were designed as an interactive 3D model of the plate heat exchanger with bulleted text operating instruction narratives as shown in Fig. 5(c). The participants in the 3D group could use the keyboard and mouse to view the operating instruction texts and the 3D model. The VR operating instructions were designed as an HMD headset to review the operating instructions as well as a virtual plate heat exchanger model in an immersive virtual environment. Participants in the VR group could also interact with the plate heat exchanger model while reviewing the operating instructions. To simulate the confined space, participants could see the limited space boundary and they were told not to go beyond the boundary when they performed the task.

4.3. Experiment procedure

Before the experiment sessions, we collected background information from participants that might have influenced performance, such as demographic traits, spatial cognition, and gaming experience (related to VR familiarity). We randomly assigned participants to the four groups based on the background information. The pre-experiment analysis did not find any significant difference among the groups. In addition, all experiments were done at the same location (Francis Hall Rm 101 BIM CAVE at Texas A&M University), with the same devices. The environmental effects can be ruled out as well. We also made sure that the experimental stimuli were clear to participants, without any possible vague interpretations. Specifically, the independent variable used in the experiment was objective information stimuli, which were 2D isometric drawing, 3D models, and the VR model in a pipe maintenance task. The experiment conditions the participants were exposed to were objective. The participants could clearly tell the difference between different types of information formats. Participants were instructed to memorize the correct sequence and spatial configuration of a 10-step pipe maintenance task within a 5 m period, and then to perform the task in the VR environment. Participants were told that their performance would be compared with others and the amount of experimental compensation would be decided by the task performance. The purpose was to motivate participants to memorize the pipe maintenance sequence and perform the task as accurately as possible.

The experiment consisted of seven sessions: (1) pre-questionnaire, (2) spatial and memory tests, (3) training, (4) review session, (5) retention session, (6) operation session, and (7) post-questionnaire and interviews. The pre-questionnaire session (5–10 m duration) was designed to collect participants' demographic information including age, gender, major, degree level, previous game and VR experience, and knowledge level of the HVAC system. The spatial and memory tests (10–20 m duration) were used to evaluate participants' spatial cognition and spatial memory abilities and to set the baseline for their task performance. We used the cube comparison and shape memory tests developed by the Educational Testing Service (ETS) in our study [20,86]. The training session (5 m duration) was designed for participants to familiarize themselves with the eye-tracking system and interactions in the virtual environment. All participants were asked to familiarize themselves with the VR devices and the virtual environment in the training session. Experiment investigators were also able to ensure 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 (5 m duration) was used for participants to review and memorize the pipe maintenance sequence. The

(caption on next page)

Fig. 4. Participants in review and operation sessions. (a) 2D-simple group; (b) 2D-complex group; (c) 3D group; (d) VR group.

review time was limited to 5 m because some participants might feel sickness (nausea, headache, dizziness, and light-headed) if using the virtual environment for 10 m or longer (based on our previous studies [22,83,84]. In the retention session (5 m 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 [96] to evaluate their presence in the virtual environment. The SUS questionnaire has been proven effective in presence evaluation with regard to human-computer interaction (HCI) [11,50,66]. At the end of all experimental stages, participants were asked to fill out a post-questionnaire to provide comments and feedback. The post-questionnaire was developed based on cognitive load measurement proposed by Leppink's cognitive research [47,48]. Compared to the traditional cognitive load measurement approach (NASA TLX survey), this measurement can evaluate three sources of cognitive load [87]. All of the sessions were conducted in room 101 (BIM CAVE) of Francis Hall at Texas A&M University. The experimental procedure took approximately 60–90 m for each participant. To thank participants, a $10 gift card was gifted to each person after they finished the experiment.

5. Data analysis and results

5.1. Overview

In total, 120 participants (68 males, 52 females) took part in the study, including 53 undergraduate students and 67 graduate students. We performed a power analysis for the one-way ANOVA with group as a factor and operation accuracy as the dependent variable using GPower. Using a large effects size estimate ($\eta^2 = 0.14$), an a priori power analysis with 30 participants in each of our four groups yielded 0.96 power to detect a significant ANOVA for group at $\alpha = 0.05$; using a moderate effect size ($\eta^2 = 0.10$) yielded 0.86 power to detect an effect. A retrospective power analysis using the k sample means method, the observed power is calculated to be 0.99998 with 120 total sample. All participants were recruited via the university emailing list. Participants' ages ranged from 18 to 45 years, with a median age of 23. Participants were from a variety of disciplines, including civil engineering, construction management, and other engineering majors. Their previous game and VR experience was surveyed, as it could affect participants' VR task performance [26]. The participants reported their previous game and VR experiences on a 10-point Likert scale (where 1 = no experience, and 10 = a lot of experience). The average game experience was 5.81 and the average VR experience was 3.18. The results indicate that most participants claimed little VR experience and thus their performance could be compared fairly. Participants were also asked to report their previous knowledge of the HVAC system, their ability to understand text instructions, their ability to read 2D drawings, and their ability to understand 3D models, all measured on a 10-point Likert scale (where 1 = no experience, and 10 = a lot of experience). Participants' average knowledge about the HVAC system was 2.3, indicating that participants have very limited previous knowledge. The average abilities in understanding text instructions, 2D drawings, and 3D models were all greater than 7.5, suggesting that most participants had no issues in comprehending the task. An ANOVA test (normality assumption was supported by the Shapiro-Wilk test) found that there was no significant difference ($p = .2159$) in the cube test score across four groups, showing similar spatial ability levels among the participants in each group. Based on the results of the SUS and post-experiment questionnaires, most participants felt that virtual pipe maintenance scenario was realistic and immersive (Mean = 5.38 and SD = 1.1, on the scale from 1 to 7) and most participants felt mild and bearable sickness while performing the pipe maintenance task in the virtual environment (Mean = 5.38 and SD = 2.66, on the scale from 1 to 10). None of the participants requested to abort the experiment. Table 3 summarizes demographic information for the participants.

5.2. Task performance

First, we evaluated whether task performance was sufficiently different across the four groups. Two task performance indicators were used including pipe maintenance accuracy (%) and operation time (s). The criteria for evaluating pipe maintenance accuracy were as followed:

1. For each maintenance step, if the participant did not touch the valve, the action was considered a failure.
2. For each maintenance step, if the participant touched the valve in the wrong sequence, the action was considered a failure.
3. For each maintenance step, if the participant touched the valve in the right sequence, the action was considered a success.

We performed a one-way ANOVA with pipe maintenance accuracy as the DV and group as the factor. In order to move away from Null Hypothesis Significance Testing (NHST) we focused our analysis on effect size and confidence interval estimates. The effect size for the omnibus ANOVA was substantial ($\eta^2 = 0.339$, $\omega^2 = 0.320$) [28]. To obtain a 95% CI for this effect size we used R's *car* and *sjstats* packages to bootstrap a distribution of effect size estimates by sampling with replacement. Because η^2 is biased, and cannot be negative, we used ω^2 to infer whether the 95% CI for effect size did not include zero. This yielded a 95% CI for ω^2 of [0.200, 0.483] which was well above zero, indicating a strong effect of group on pipe maintenance accuracy [28]. We next compared the group means and 95% CIs for each group. To create 95% CIs we multiplied the standard error of the mean by 2.045, the critical t-value at $\alpha = 0.05$ and $df = 29$, given samples sizes of 30 per group. We inferred the groups to be significantly different from one another if one group's mean did not fall within the other group's 95% CI. The 2D-complex group (M = 0.433, SD = 0.355, 95% CI: 0.301, 0.566) had the lowest mean accuracy. The mean for the 2D-simple group (M = 0.633, SD = 0.345, 95% CI: 0.505, 0.762) was not within the 95% CI for the 2D-complex group, indicating that participants in the

Table 2
10-Steps for the pre-start-up maintenance sequence.

Hot water side	Cold water side
Step 1: Close hot side pump isolation valve (v1)	Step 6: Close cold side pump isolation valve (v9)
Step 2: Close hot drain valves (v2 and then v5)	Step 7: Close cold drain valves (v8 and then v14)
Step 3: Close hot isolation valves (v3 and then v4)	Step 8: Close cold isolation valves (v10 and then v11)
Step 4: Open hot vent valve (v6)	Step 9: Open cold vent valve (v12)
Step 5: Close hot feed valve (v7)	Step 10: Close cold feed valve (v13)

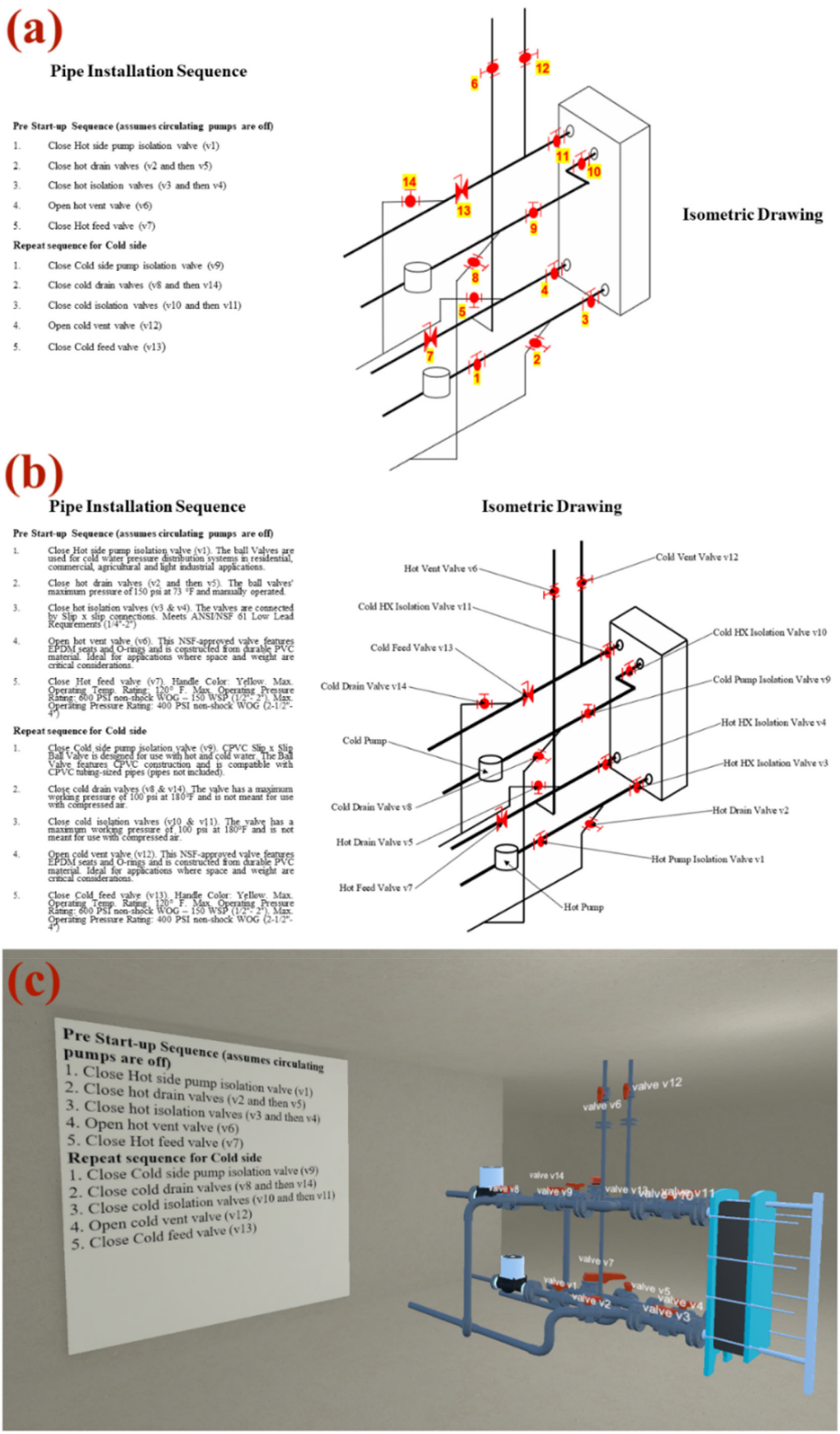

Fig. 5. Four types of operating instructions: (a) 2D-simple; (b) 2D-complex; (c) 3D and VR.

Table 3
Summary of participants' demographic information.

Demographic factors	Response range	Mean or percentage	Median
Gender	Male/female	56.67% male	–
Age	18–45	24.29	23
Degree level	Undergraduate/graduate	55.83% graduate student	–
Game experience	1–10	5.81	6
VR experience	1–10	3.18	2
The knowledge of HVAC system	1–10	2.3	1
The ability of understanding text instructions	1–10	8.37	9
The ability of understanding 2D drawing	1–10	8.13	8.5
The ability of understanding 3D model	1–10	7.53	8

Table 4
Details of participants' task performance.

	2D group		2D-complex		3D group		VR group	
Indicators	Operation accuracy (%)	Operation time (s)	Operation accuracy (%)	Operation time (s)	Operation accuracy (%)	Operation time (s)	Operation accuracy (%)	Operation time (s)
Mean	63.3%	96.684	43.3%	124.787	92.3%	68.473	93.7%	65.835
SD	34.5%	54.627	35.5%	65.905	10%	27.927	11%	31.14
Lower 95% CL	50.5%	76.287	30.1%	100.178	88.6%	58.045	89.1%	53.516
Upper 95% CL	76.2%	117.083	56.6%	149.396	96.1%	78.901	98.2%	78.154

2D-simple group were significantly more accurate than participants in the 2D complex group. The means for the 3D group (M = 0.923, SD = 0.1, 95% CI: 0.886, 0.961) and the VR group (M = 0.937, SD = 0.11, 95% CI: 0.891, 0.982) were not significant different from each other, and they were significantly above the upper bounds for the 95% CIs of the 2D-simple and 2D-complex groups. Thus, the VR and 3D groups had significantly higher accuracy than the 2D-group, and the 2D-complex group had significantly lower accuracy than all other groups. The details of pipe maintenance accuracy are shown in Table 4.

We next performed a similar ANOVA with operation time as the DV and group as the factor. The effect size for the omnibus ANOVA was large ($\eta^2 = 0.188$, $\omega^2 = 0.166$). A bootstrapped 95% CI for ω^2 did not include zero [0.070, 0.323], thus indicating a non-zero effect. The average operation time for the 2D-complex group (M = 124.787, SD = 65.905, 95% CI: 100.178, 149.396) was significantly longer than the operation time of the 2D-simple group (M = 96.685, SD = 54.627, 95% CI: 76.287, 117.083), and the means for each group were outside the bounds of the 95% CI of the other group. Average operation times were fastest for the 3D (M = 65.835, SD = 31.14, 95% CI: 53.516, 78.154) and VR groups (M = 70.950, SD = 34.37, 95% CI: 59.313, 84.411). These 95% CIs for these groups largely overlap, and they do not include the means for the 2D groups. Thus the pattern of average operation times among the four groups largely mirrors the pattern we observed for accuracy, with the 3D and VR groups performing faster and more accurately than the 2D-simple group, which is still faster and more accurate than the 2D-complex group. Further details of operation time data are shown in Table 4 (Fig. 6).

We also computed the geometric mean for operation time for participants in each condition. The geometric mean has been shown to more accurately estimate the population median than the arithmetic mean; it is also less biased than the sample median which tends to overestimate the population median [78,79]. The geometric mean is computed by taking the natural log of each value, computing the mean of these log-values, and then exponentiating this mean. Fig. 7 plots the geometric mean for each group along with 95% confidence intervals. The confidence intervals for the 3D and VR conditions did not include the means for the 2D conditions, which suggests that these groups reliably differ in operation time. As expected, the natural logarithms of each participant's operation time were more normally distributed than raw operation times.

5.3. Cognitive load in information encoding phase

Next, we evaluated cognitive load, as measured by pupil dilation, in the training or information-encoding phase. Similar to our analyses above, we performed a one-way ANOVA with group as a factor and average dilation in the training phase as the DV. This yielded a large effect size ($\eta^2 = 0.280$, $\omega^2 = 0.260$) with a bootstrapped 95% CI for ω^2 that did not include zero [0.132, 0.452]. Average pupil dilations were the lowest for the 2D-simple (M = 0.06, SD = 0.06, 95% CI: 0.037, 0.083) and 2D complex groups (M = 0.050, SD = 0.058, 95% CI: 0.029, 0.072), and these groups did not significant differ. The 3D (M = 0.126, SD = 0.069, 95% CI: 0.101, 0.152) and VR groups (M = 0.138, SD = 0.063, 95% CI: 0.114, 0.161) did not significantly differ from each other, and both groups had significantly larger average pupil dilations than the 2D groups. These results reveal that the participants in the 3D and VR groups had larger pupil dilations, indicating higher cognitive load levels during training. Moreover, across all conditions pupil dilation scores were positively correlated with pipe operation accuracy ($r = 0.34$, $p < .001$).

We further examined the aggregated pupil dilation in the information-encoding phase as shown in Fig. 8(b). A one-way ANOVA yielded a very large effect size ($\eta^2 = 0.533$, $\omega^2 = 0.519$) with a bootstrapped 95% CI for ω^2 that did not include zero [0.414, 0.639]. Aggregate pupil dilations were the lowest for the 2D-simple (M = 926, SD = 2105, 95% CI: 140, 1712) and 2D complex groups (M = 760, SD = 1621, 95% CI: 155, 1366), and these groups did not significantly differ. The 3D (M = 7569, SD = 4795, 95% CI: 5778, 9359) and VR groups (M = 10,128, SD = 5565, 95% CI: 8050, 12,205) did not significantly differ from each other, and both groups had significantly larger average pupil dilations than the 2D groups; means for the 3D and VR groups were well outside the 95% CIs for the 2D groups. Across all conditions aggregated pupil dilation was positively correlated with pipe operation accuracy ($r = 0.39$, $p < .001$).

5.4. Cognitive load in information retrieval phase

We also evaluated pupil dilation, as a proxy for cognitive load, in the performance phase where participants were attempting to execute the pipe maintenance operation. Average pupil dilation was found to be normally distributed according to the Shapiro-Wilk test of normality, and therefore a one-way ANOVA was used to compare the average pupil

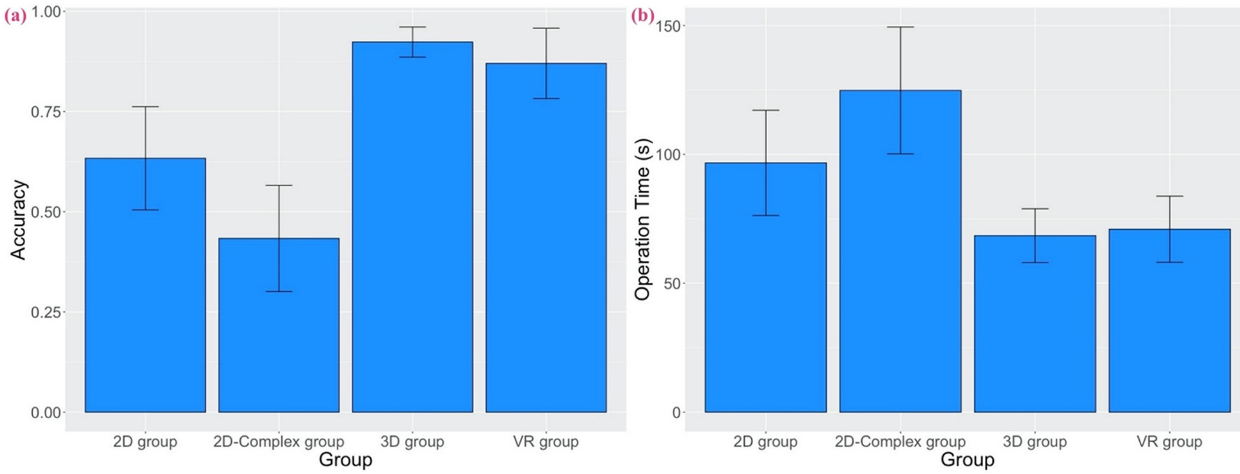

Fig. 6. Task performance of four groups: (a) pipe operation accuracy; (b) operation time. The error bars represent 95% CIs.

dilations across different groups ($\alpha = 0.05$). The effects size from this ANOVA was very small ($\eta^2 = 0.016$, $\omega^2 = -0.009$), even less than zero for ω^2. The 95% CI for ω^2 included zero [−0.022, 0.087], which suggests that there was little to no effect of group on pupil dilation during the retrieval phase. However, it should be noted that because the time used by each participant in the task operation phase was significantly different, the average pupil dilation percentage might not properly reflect the total cognitive costs in the task operation phase. For example, Fig. 9 shows the pupil changes for two subjects in the operation phase. Although subject (a) presented a lower average pupil dilation, because she used more time for the task, the total cognitive cost was still higher than subject (b). Thus, the aggregated pupil dilation is a more suitable method of capturing a participant's total cognitive cost in information recalling. As a result, we compared the aggregated pupil dilation with a one-way ANOVA test (normality test passed). These results are illustrated in Fig. 10(b). The effect size for the ANOVA was larger than for average dilation during test, but still only small to moderate in size ($\eta^2 = 0.059$, $\omega^2 = 0.035$). The bootstrapped 95% CI for ω^2 included zero [−0.007, 0.153]. This suggests that while there was a larger effect for group for aggregate than for average pupil dilation during test, the effect was nevertheless small and not significantly different from zero.

5.5. Cognitive load validation

Finally, we validated the findings of pupil size analysis with a cognitive load questionnaire at the end of the experiment.

For the results of intrinsic cognitive load, a one-way ANOVA yielded a very large effect size ($\eta^2 = 0.302$, $\omega^2 = 0.282$) with a bootstrapped 95% CI for ω^2 that did not include zero [0.171, 0.0.427] as illustrated in Fig. 11(a). Intrinsic cognitive load was the lower for the 3D (M = 2.46, SD = 1.36, 95% CI: 1.95, 2.97) and VR groups (M = 3.35, SD = 2.02, 95% CI: 2.6, 4.1), and these groups significantly differ. The 2D-simple (M = 4.17, SD = 1.76, 95% CI: 3.51, 4.82) and 2D-complex groups (M = 5.72, SD = 2.19, 95% CI: 4.91, 6.54) significantly differ from each other, and both groups had significantly larger Intrinsic cognitive load than the 3D and VR groups; means for the 3D and VR groups were well outside the 95% CIs for the 2D groups. These results indicate that the participants in the 3D and VR groups reported the lowest intrinsic

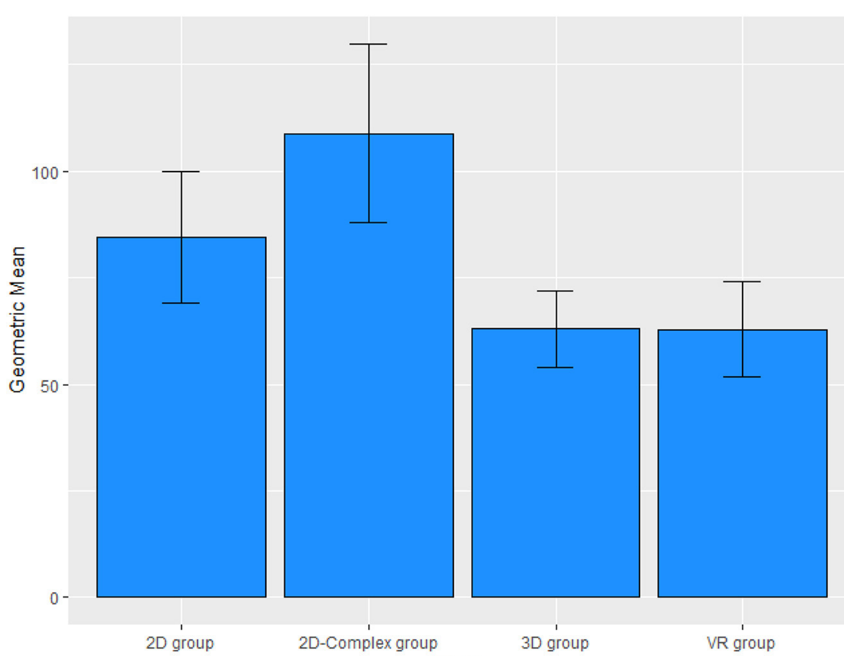

Fig. 7. The geometric mean of operation time across different groups. The error bars are 95% confidence intervals.

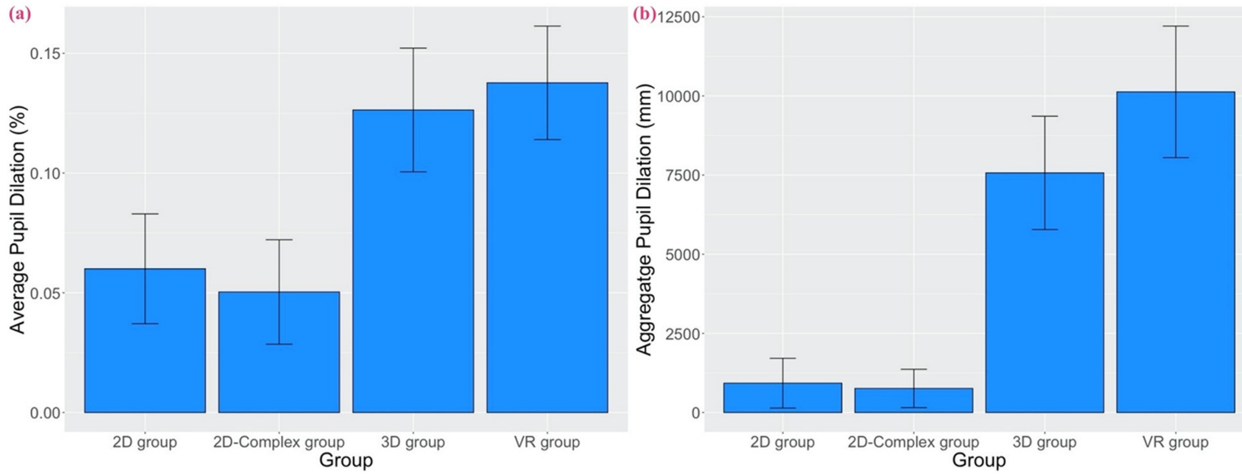

Fig. 8. Pupil dilation in information encoding: (a) average pupil dilation percentage; (b) aggregated pupil dilation. Error bars represent 95% CIs.

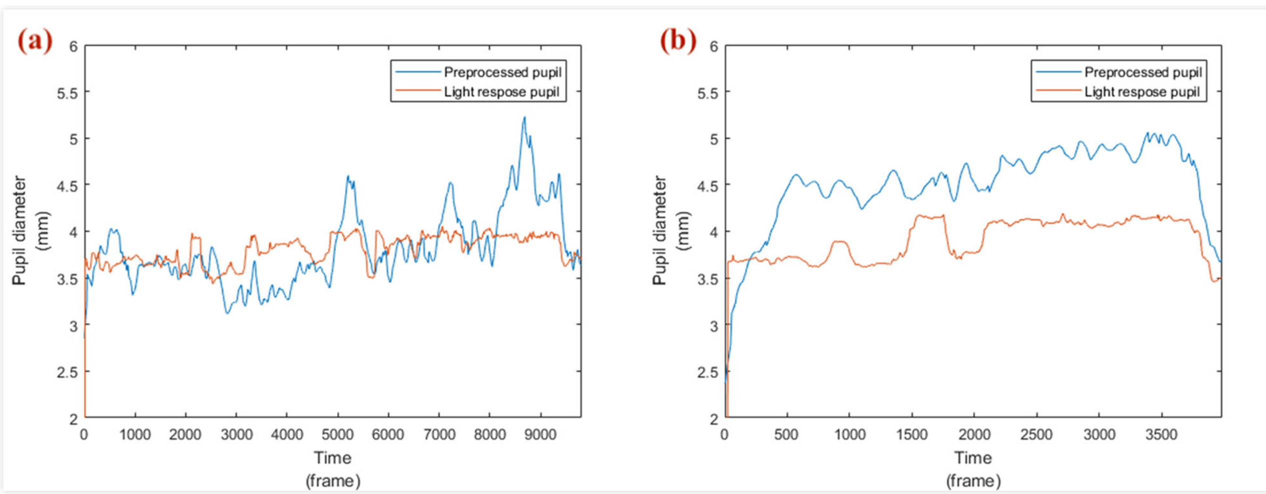

Fig. 9. Pupil variations in information retrieval phase (a) mixed pupil dilate and pupil constrict; (b) pupil dilation dominant.

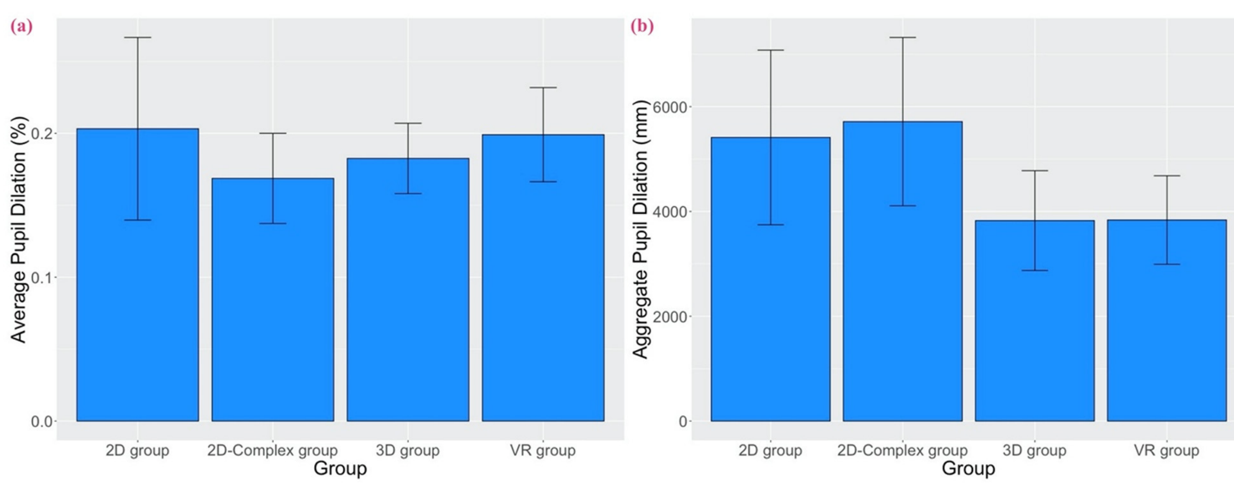

Fig. 10. Pupil dilation in the information-recalling phase: (a) average pupil dilation percentage; (b) aggregated pupil dilation. Error bars represent 95% CIs.

cognitive load, and the 2D-complex group reported the highest levels.

For the results of extraneous cognitive load, a one-way ANOVA yielded a very large effect size (η^2 = 0.274, ω^2 = 0.254) with a bootstrapped 95% CI for ω^2 that did not include zero [0.137, 0.0.411] as illustrated in Fig. 11(b). Extraneous cognitive load was the lower for the 3D (M = 1.26, SD = 1.1, 95% CI: 0.847, 1.67) and VR groups (M = 1.46, SD = 1.03, 95% CI: 1.07, 1.84), and these groups did not significantly differ. The 2D-simple (M = 2.81, SD = 1.72, 95% CI: 2.17, 3.45) and 2D-complex groups (M = 3.76, SD = 2.49, 95% CI: 2.83, 4.69) significantly differ from each other, and both groups had

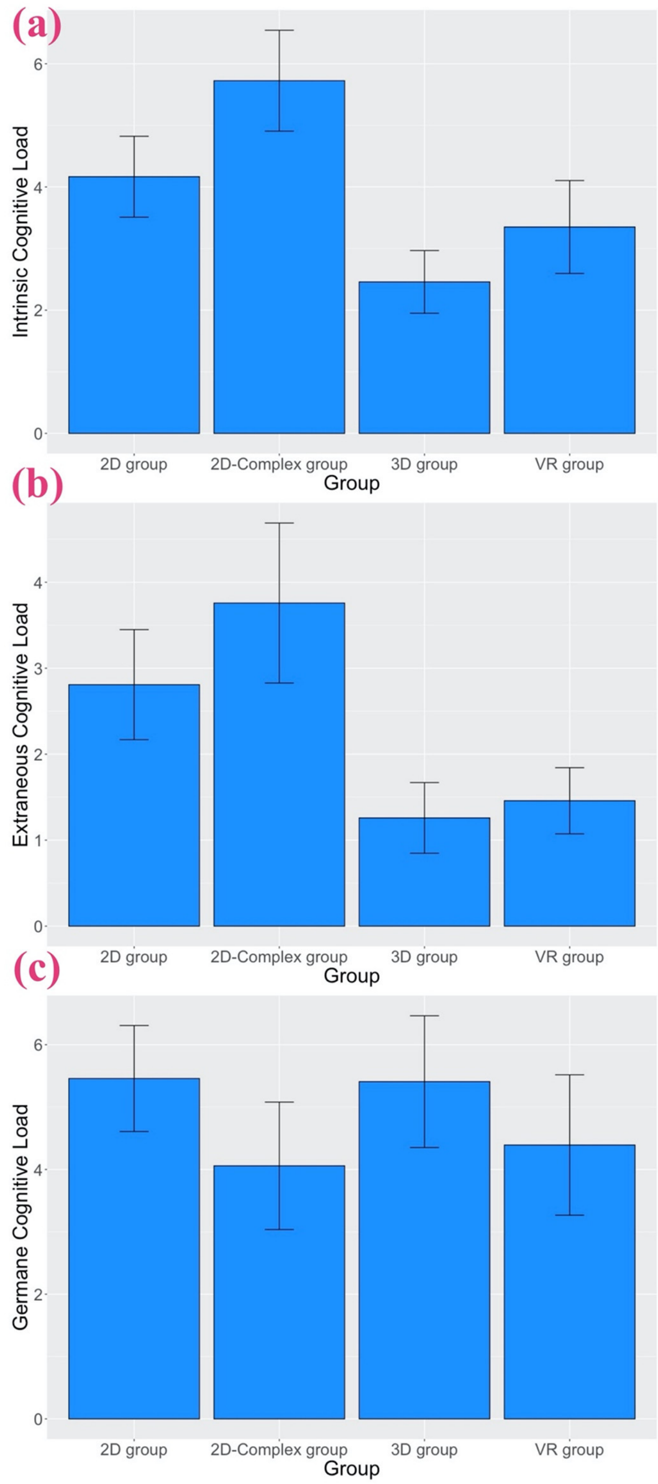

Fig. 11. Results of cognitive load surveys: (a) Intrinsic cognitive load; (b) Extraneous cognitive load; (c) Germane cognitive load.

significantly larger extraneous cognitive load than the 3D and VR groups; means for the 3D and VR groups were well outside the 95% CIs for the 2D groups. These results indicate that the participants in the 3D and VR groups reported the lowest extraneous cognitive load, and the two 2D groups reported the highest extraneous cognitive load levels. Since the extraneous cognitive load is closely related to how information is presented, this result is consistent with the differences of cognitive load between encoding and retrieval phase based on pupil dilation data as shown in Fig. 10. Finally, we did not find any significant difference in germane cognitive load between different groups ($\eta^2 = 0.05$, $\omega^2 = 0.025$) with a bootstrapped 95% CI for ω^2 that include zero [−0.01, 0.0.153], as illustrated in Fig. 11(c). Because the germane cognitive load is related to the previous experience of the task, the result indicates that participants in each group had similar levels of previous knowledge about the task.

In general, the post-experiment cognitive survey supports the findings regarding pupil size analysis that the 3D and VR groups tended to demonstrate lower cognitive load levels in the information-recalling phase. This could have driven better performance for these two groups. However, the two groups also demonstrated a higher cognitive load in the information encoding phase or the development phase of working memory. This represents a potentially complex cognitive phenomenon that deserves further discussion.

6. Discussion

These experimental results reveal several important implications regarding how engineering information formats affect learning and retention of pipe maintenance tasks. First, it indicates that more immersive information display and visualization during the review session does improve final task performance. When given the same amount of time for engineering review, the 3D and VR groups outperformed the 2D-simple and 2D-complex groups in both operation time and accuracy. These results echo recent findings in the construction informatics literature that confirm the benefits of emerging visualization technologies such as BIM and mixed reality [3,69,77,89]. We also found evidence that greater pupil dilation during the training phase predicted better performance at test, and that the 3D and VR groups had larger average pupil dilation than the 2D groups. Given the previous connection between pupil dilation and successful working memory encoding it seems possible that the pupil dilation we observed was a physiological indicator of successful working memory encoding [32]. Meanwhile, participants from the 3D and VR groups reported lower extraneous cognitive load compared to 2D groups based on cognitive load survey results. This result is supported by Wickens' multiple resource model [103,104]; participants in the 2D groups were required to process more of the same type of (phonological) information that may have been taxing the same sensory process to increase the cognitive load, compared to participants in the 3D and VR groups who were able to process information presented in additional sensory dimensions. This likely allowed for information to be processed by more sensory channels which reduced self-reported extraneous cognitive load.

However, it should be acknowledged that the better performance of the 3D and VR groups may be mediated by a different mechanism than enhanced working memory encoding; instead, it may represent a more complex cognitive process related to both working memory development (encoding) and working memory retrieval (recalling). For example, the psychology literature discusses the cognitive phenomenon called “state-dependent memory” and “transfer appropriate processing (TAP)”, whereby memory retrieval is most efficient when an individual is in the same state of consciousness as they were when the memory was formed [25,94]. TAP also suggests that memory performance is not only determined by the levels-of-processing required, but also by how information is initially encoded and later retrieved [14,33,58]. Because 3D and VR models can best reproduce scenes that will be seen in the operation phase, memory retrieval may be easier for participants. The higher cognitive load for 2D during retrieval can still be explained by matching between learning and retrieval. Notably, the 2D groups may show increased cognitive load because they are mentally trying to translate information they learned to a new format for retrieval. In other words, the better performance of the 3D and VR groups in task performance does not necessarily mean that learning, or information encoding, was enhanced during the review session; rather, it may also indicate easier retrieval of working memory in the operation phase. To evaluate the impact of 3D and VR on operation task performance, a

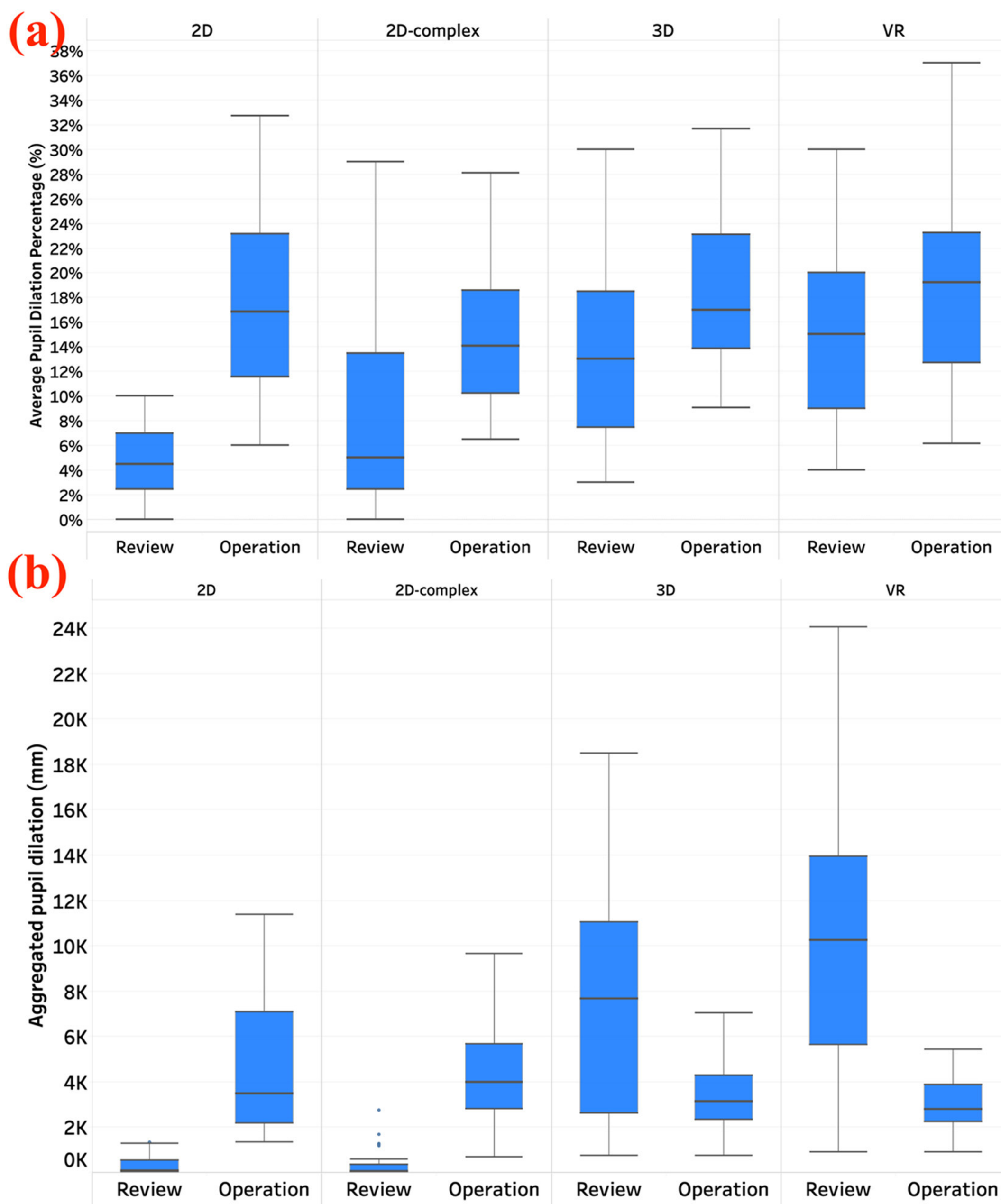

Fig. 12. Pupil dilation between review and operation phases across four groups: (a) average pupil dilation percentage; (b) aggregated pupil dilation.

thorough investigation into the cognitive processes of both the information encoding and recalling phases is needed.

Therefore, we further examined the cognitive costs across the four groups in both the review session and the operation session. Prior work has found that pupil dilation is a strong indicator of an increased cognitive load [18,34,80]. Our analysis on pupil data found that the 3D and VR groups demonstrated higher frequency and magnitude of pupil dilations in the training session (in both average pupil dilation percentage and aggregated pupil dilation) as shown in Fig. 12. This indicates that the cognitive costs of both 3D and VR groups in working memory development (or information encoding) were much higher than those of the 2D-simple and 2D-complex groups. The pupil dilation data also suggest that the 3D and VR groups have the smaller cognitive increase differences in average pupil dilation percentage between review and operation phases compared to the 2D groups as shown in Fig. 12(a). The 3D and VR groups have cognitive decrease differences in aggregated pupil dilation between review and operation phases compared to the 2D group as shown in Fig. 12(b). These results suggest that 3D and VR groups have smaller cognitive costs in memory retrieval. In other words, in terms of cognition, it uses more cognitive load for the 3D and VR groups to digest and memorize information in the review session, but once the information was encoded, it was much easier for these two groups to retrieve or recall information. Based on an overall assessment of the results, we derived a theoretical framework for explaining why different formats of engineering information affect working memory-based tasks in distinct ways, and how learning or training can be optimized, as illustrated in Fig. 13.

The proposed framework shows that the advantage of a certain format of engineering information in working memory-based tasks should be evaluated in two parts: the cognitive cost for encoding working memory α_e, and the cognitive cost for recalling working memory α_r. The overall efficiency of a particular information format, with respect to the performance of working memory-based tasks, shall represent a lower recalling-to-encoding conversion ratio: $\beta = \alpha_r/\alpha_e$,

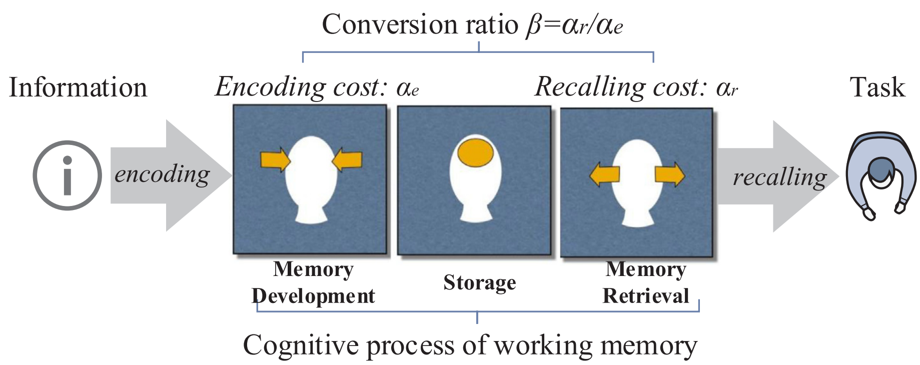

Fig. 13. Theoretical framework for working memory use efficiency assessment.

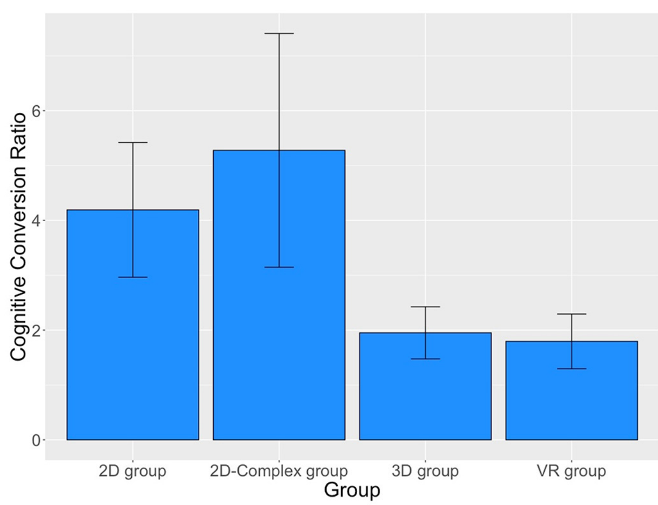

Fig. 14. Recalling-to-encoding cognitive conversion ratios of four groups.

that is, recalling (of working memory) shall be easier given the same level of difficulty in the encoding (of working memory). Using our data as an example, a one-way ANOVA yielded a very large effect size ($\eta^2 = 0.163$, $\omega^2 = 0.14$) with a bootstrapped 95% CI for ω^2 that did not include zero [0.079, 0.0.267]. Cognitive conversion ratio were the lowest for the 3D (M = 1.95, SD = 1.27, 95% CI: 1.48, 2.43) and VR groups (M = 760, SD = 1621, 95% CI: 155, 1366), and these groups did not significantly differ. The 2D-simple (M = 4.19, SD = 3.29, 95% CI: 2.96, 5.42) and 2D-complex groups (M = 5.28, SD = 5.71, 95% CI: 3.15, 7.41) did not significantly differ from each other, and both groups had significantly larger cognitive conversion ratio than the 3D and VR groups; means for the 3D and VR groups were well outside the 95% CIs for the 2D groups. The cognitive conversion ratio was negatively correlated with accuracy across all groups ($r = -0.29, p < .01$) (Fig. 14).

Using this framework, we found that a more immersive information visualization (e.g., 3D and VR) that resembles the scenes in the task operation phase actually requires a higher mental engagement in information encoding, but retrieval requires less cognitive process. The overall conversion ratio β is low and thus the 3D and VR formats lead to better performance. The 2D-complex group represented the worst performance in the experiment because the recalling-to-encoding ratio was significantly higher than for the other groups. To be noted, although not the main focus of this study, our results also suggest that information complexity affects working memory development. Specially, we found that complex display of the same information (including many contents not directly relevant to the task) impacted task performance (both time and accuracy) in a negative way. The survey results show that providing more semantic content in the engineering information increases intrinsic cognitive load in the operation phase. Intrinsic cognitive load has been found to be related to the comprehension and interpretation of information [16,87,88]; therefore, it is reasonable that test subjects demonstrated a higher mental load when recalling the information. This study contributes to methods and theories pertaining to construction productivity assessment and improvement. The findings of this study will help inspire the design of cognition-driven performance prediction and early intervention systems for complex construction operations that rely on working memory. Specifically, as the clear relationship between pupillary changes and working memory development and retrieval quality was found in this study, it supports a promising direction of applying eye tracking in construction tasks for real-time memory quality assessment. In addition, our analysis showed that there was a close correlation between working memory metrics and task performance, especially the accuracy, suggesting that eye tracking, in addition to other existing psychological measurement methods, has the potential to be deployed on-site for individual level performance assessment, prediction and intervention.

7. Conclusions

Given the increasing complexity of construction operations and the growing work scenarios in confined workplaces, it is critical to examine the role cognitive mechanisms such as working memory (i.e., the short term and temporary storage of information), as these types of mechanisms likely play a critical role in construction projects. Field workers often need to recall information from memory, instead of querying information on site, to ensure a seamless execution of construction tasks. Recently, researchers have shown a great interest in understanding the implications of different engineering information formats on the encoding and retention of information in construction tasks, driven by the rapid development of new information and visualization technologies such as VR. However, there is still an obvious theoretical disagreement and a lack of investigation into the mechanisms that lead to different information formats having different effects on learning. This study proposes that different information formats and display methods trigger varying levels of cognitive load, particularly in the working memory encoding phase. In the tentative model we developed, the overall efficiency of working memory use is defined as a recalling-to-encoding conversion ratio to represent the level of cognitive difficulty in recalling information, given the same level of difficulty in encoding the information. Therefore, a smaller conversion ratio represents more efficient use of working memory resources and is associated with improved performance in working memory-based tasks. To prove the concept, we performed a human-subject experiment where test subjects (n = 120) were required to review the information for a pipe maintenance task for a short period of time, and then perform the task based on working memory. Depending on what format of information was given, test subjects were divided into four groups: 2D-simple, 2D-complex, 3D, and VR.

In summary, this research study contributes to the construction science field in the following ways: First, we developed a pupil dilation analysis workflow that rules out the influence of light flex generated by the monitor or lens of the HMD to assess user's cognitive status in real time. This approach can greatly help safety managers to detect construction workers' cognitive status during the construction operations. Second, our results showed that the 3D and VR groups outperformed the 2D-simple and 2D-complex groups in both task time and accuracy. The additional cognitive load analysis based on surveys and pupil dilation analyses found that the 3D and VR groups showed lower cognitive load levels in the operation phase, but higher cognitive load levels in the review phase. The better performance for these two groups may therefore be a result of a complicated cognitive process related to both information encoding and decoding. Third, we further proposed a framework for explaining how information format affects working memory-based tasks. The findings are expected to provide more evidence about the interplay between information formats and working memory functioning that should help resolve the current theoretical disagreement in the construction literature and inspire the design of cognition-driven information systems that lead to the most efficient learning for construction workers.

Several research limitations still need to be addressed in the future research agenda. First, this study was conducted in a well-controlled laboratory environment. In real world, construction sites and construction operations are more complex and unpredictable. Thus, more complex and dynamic scenarios should be tested in future research. Second, in this study, we used pupil dilation to assess participants' cognitive status. In our future research, more physiological sensors such as Electrocardiogram (ECG), Electroencephalogram (EEG), and functional near-infrared spectroscopy (fNIRS) will be implemented to cross validate the physiological and psychological data and more accurately predict individual's cognitive status. Last but not the least, the construction operation task in this study was tested in a normal situation. According to the previous literature, stress can greatly weaken individual's brain networks. Therefore, we will test the construction operation tasks under different cognitive status. Despite these limitations, this study provides evidence that 3D and VR information formats can lead to better learning of construction maintenance operations due to reduced cognitive load from the presentation of information in multiple sensory modalities.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to acknowledge Xin Liu and Baiyu Chen, students from Texas A&M University and City University of Hong Kong for facilitating the experiment. This material is supported by the National Science Foundation (NSF) under Grant 1937878, as well as National Institute of Standards and Technology (NIST) under Grant 60NANB18D152. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not reflect the views of the NSF and NIST.

References

[1] AlfaLaval, Instruction Manual - Plate Heat Exchangers, 2016. Available on-line via https://www.alfalaval.com/globalassets/documents/products/heat-transfer/plate-heat-exchangers/gasketed-plate-and-frame-heat-exchangers/industrial/instruction-manual-gphe-small-en.pdf, Accessed date: March 2020.

[2] S.K. Ayer, J.I. Messner, C.J. Anumba, Augmented reality gaming in sustainable design education, J. Archit. Eng. 22 (1) (2016) 04015012, , https://doi.org/10.1061/(ASCE)AE.1943-5568.0000195.

[3] F. Bacim, E. Ragan, S. Scerbo, N.F. Polys, M. Setareh, B.D. Jones, The effects of display fidelity, visual complexity, and task scope on spatial understanding of 3D graphs, Proceedings of Graphics Interface 2013, Canadian Information Processing Society, 2013, pp. 25–32, https://dl.acm.org/doi/10.5555/2532129.2532135.

[4] A. Baddeley, The fractionation of working memory, Proc. Natl. Acad. Sci. 93 (24) (1996) 13468–13472, https://doi.org/10.1073/pnas.93.24.13468.

[5] A. Baddeley, Working memory, Science 255 (5044) (1992) 556–559, https://doi.org/10.1126/science.1736359.

[6] A. Baddeley, Working memory: looking back and looking forward, Nat. Rev. Neurosci. 4 (10) (2003) 829, https://doi.org/10.1038/nrn120.

[7] A. Baddeley, S. Gathercole, C. Papagno, The phonological loop as a language learning device, Psychol. Rev. 105 (1) (1998) 158, https://doi.org/10.1037/0033-295X.105.1.158.

[8] A. Baddeley, G. Vallar, B. Wilson, Sentence comprehension and phonological memory: some neuropsychological evidence, in M. Coltheart (Ed.), Attention and Performance 12: The Psychology of Reading. Lawrence Erlbaum Associates, Inc. (1987) pp. 509–529. Available on-line via https://psycnet.apa.org/record/1987-98557-023. Accessed date: March 2020.

[9] J. Barbur, A. Harlow, A. Sahraie, Pupillary responses to stimulus structure, colour and movement, Ophthalmic Physiol. Opt. 12 (2) (1992) 137–141, https://doi.org/10.1111/j.1475-1313.1992.tb00276.x.

[10] D. Bawden, C. Holtham, N. Courtney, Perspectives on information overload, Aslib Proceedings, Vol. 51, MCB UP Ltd, 1999, pp. 249–255, doi:https://doi.org/10.1108/EUM0000000006984.

[11] R.M. Baños, C. Botella, I. Rubió, S. Quero, A. García-Palacios, M. Alcañiz, Presence and emotions in virtual environments: the influence of stereoscopy, CyberPsychology & Behavior 11 (1) (2008) 1–8, https://doi.org/10.1089/cpb.2007.9936.

[12] O. Bergamin, R.H. Kardon, Latency of the pupil light reflex: sample rate, stimulus intensity, and variation in normal subjects, Invest. Ophthalmol. Vis. Sci. 44 (4) (2003) 1546–1554, https://doi.org/10.1167/iovs.02-0468.

[13] C. Blackie, H. Howland, An extension of an accommodation and convergence model of emmetropization to include the effects of illumination intensity, Ophthalmic Physiol. Opt. 19 (2) (1999) 112–125, https://doi.org/10.1046/j.1475-1313.1999.00436.x.

[14] T.A. Blaxton, Investigating dissociations among memory measures: support for a transfer-appropriate processing framework, Journal of Experimental Psychology: Learning, Memory, and Cognition 15 (4) (1989), pp. 657, doi:https://doi.org/10.1037/0278-7393.15.4.657.

[15] J.P. Bliss, P.D. Tidwell, M.A. Guest, The effectiveness of virtual reality for administering spatial navigation training to firefighters, Presence Teleop. Virt. 6 (1) (1997) 73–86, https://doi.org/10.1162/pres.1997.6.1.73.

[16] P. Chandler, J. Sweller, Cognitive load theory and the format of instruction, Cogn. Instr. 8 (4) (1991) 293–332, https://doi.org/10.1207/s1532690xci0804_2.

[17] H. Chen, A. Dey, M. Billinghurst, R.W. Lindeman, Exploring pupil dilation in emotional virtual reality environments (2017), https://doi.org/10.2312/egve.20171355.

[18] S. Chen, J. Epps, Automatic classification of eye activity for cognitive load measurement with emotion interference, Comput. Methods Prog. Biomed. 110 (2) (2013) 111–124, https://doi.org/10.1016/j.cmpb.2012.10.02.

[19] A. Cockburn, Revisiting 2D vs 3D implications on spatial memory, Proceedings of the Fifth Conference on Australasian User Interface - Volume 28, Inc, Australian Computer Society, 2004 pp. 25-31 https://dl.acm.org/doi/abs/10.5555/976310.976314.

[20] G.B. Dadi, P.M. Goodrum, T.R. Taylor, C.M. Carswell, Cognitive workload demands using 2D and 3D spatial engineering information formats, J. Constr. Eng. Manag. 140 (5) (2014) 04014001, , https://doi.org/10.1061/(ASCE)CO.1943-7862.0000827.

[21] J. Du, Y. Shi, C. Mei, J. Quarles, W. Yan, Communication by interaction: a multiplayer VR environment for building walkthroughs, Construction Research Congress 2016 (2016) 2281–2290, https://doi.org/10.1061/9780784479827.227.

[22] J. Du, Y. Shi, Z. Zou, D. Zhao, CoVR: cloud-based multiuser virtual reality headset system for project communication of remote users, J. Constr. Eng. Manag. 144 (2) (2017) 04017109, , https://doi.org/10.1061/(ASCE)CO.1943-7862.0001426.

[23] J. Du, Z. Zou, Y. Shi, D. Zhao, Zero latency: real-time synchronization of BIM data in virtual reality for collaborative decision-making, Autom. Constr. 85 (2018) 51–64, https://doi.org/10.1016/j.autcon.2017.10.009.

[24] A. Dünser, K. Steinbügl, H. Kaufmann, J. Glück, Virtual and augmented reality as spatial ability training tools, Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction: Design Centered HCI, ACM, 2006, pp. 125–132, doi:https://doi.org/10.1145/1152760.1152776.

[25] E. Eich, Theoretical issues in state dependent memory, Varieties of memory and consciousness: essays in honour of Endel Tulving. Lawrence Erlbaum Associates, Inc. (1989) 331–354. Available on-line via https://psycnet.apa.org/record/1989-97876-017. Accessed date: March 2020.

[26] L. Enochsson, B. Isaksson, R. Tour, A. Kjellin, L. Hedman, T. Wredmark, L. Tsai-Felländer, Visuospatial skills and computer game experience influence the performance of virtual endoscopy, J. Gastrointest. Surg. 8 (7) (2004) 874–880, https://doi.org/10.1016/j.gassur.2004.06.015.

[27] M.J. Eppler, J. Mengis, The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines, The Information Society 20 (5) (2004) 325–344, https://doi.org/10.1080/01972240490507974.

[28] W.H. Finch, B.F. French, A comparison of methods for estimating confidence intervals for omega-squared effect size, Educ. Psychol. Meas. 72 (1) (2012) 68–77,

https://doi.org/10.1177/0013164411406533.
[29] K. Fletcher, A. Neal, G. Yeo, The effect of motor task precision on pupil diameter, Appl. Ergon. 65 (2017) 309–315, https://doi.org/10.1016/j.apergo.2017.07.010.
[30] C.K. Foroughi, C. Sibley, J.T. Coyne, Pupil size as a measure of within-task learning, Psychophysiology 54 (10) (2017) 1436–1443, https://doi.org/10.1111/psyp.12896.
[31] S.E. Gathercole, Cognitive approaches to the development of short-term memory, Trends Cogn. Sci. 3 (11) (1999) 410–419, https://doi.org/10.1016/S1364-6613(99)01388-1.
[32] S.D. Goldinger, M.H. Papesh, Pupil dilation reflects the creation and retrieval of memories, Current Directions in Psychological Science 21 (2) (2012), pp. 90–95, doi:https://doi.org/10.1177/0963721412436811.
[33] P. Graf, L. Ryan, Transfer-appropriate processing for implicit and explicit memory, Journal of Experimental Psychology: Learning, Memory, and Cognition 16 (6) (1990), pp. 978, doi:https://doi.org/10.1037/0278-7393.16.6.978.
[34] E. Granholm, R.F. Asarnow, A.J. Sarkin, K.L. Dykes, Pupillary responses index cognitive resource limitations, Psychophysiology 33 (4) (1996), pp. 457–461, doi:https://doi.org/10.1111/j.1469-8986.1996.tb01071.x.
[35] R.P. Heitz, J.C. Schrock, T.W. Payne, R.W. Engle, Effects of incentive on working memory capacity: behavioral and pupillometric data, Psychophysiology 45 (1) (2008) 119–129, https://doi.org/10.1111/j.1469-8986.2007.00605.x.
[36] A. Heydarian, J.P. Carneiro, D. Gerber, B. Becerik-Gerber, T. Hayes, W. Wood, Immersive virtual environments versus physical built environments: a benchmarking study for building design and user-built environment explorations, Autom. Constr. 54 (2015) 116–126, https://doi.org/10.1016/j.autcon.2015.03.020.
[37] A. Heydarian, E. Pantazis, A. Wang, D. Gerber, B. Becerik-Gerber, Towards user centered building design: identifying end-user lighting preferences via immersive virtual environments, Autom. Constr. 81 (2017) 56–66, https://doi.org/10.1016/j.autcon.2017.05.003.
[38] L. Hou, Y. Wang, X. Wang, N. Maynard, I.T. Cameron, S. Zhang, Y. Jiao, Combining photogrammetry and augmented reality towards an integrated facility management system for the oil industry, Proc. IEEE 102 (2) (2014) 204–220, https://doi.org/10.1109/JPROC.2013.2295327.
[39] HTC, HTC VIVE PRO, Vol. 2019, 2018. Available on-line via https://www.vive.com/us/product/vive-pro/. Accessed date: March 2020.
[40] D. Kahneman, J. Beatty, Pupil diameter and load on memory, Science 154 (3756) (1966) 1583–1585, https://doi.org/10.1126/science.154.3756.1583.
[41] B.H. Kantowitz, J.L. Knight Jr., Testing tapping timesharing, II: Auditory secondary task, Acta Psychologica 40 (5) (1976) 343–362, https://doi.org/10.1016/0001-6918(76)90016-0.
[42] R. Kardon, S.C. Anderson, T.G. Damarjian, E.M. Grace, E. Stone, A. Kawasaki, Chromatic pupil responses: preferential activation of the melanopsin-mediated versus outer photoreceptor-mediated pupil light reflex, Ophthalmology 116 (8) (2009) 1564–1573, https://doi.org/10.1016/j.ophtha.2009.02.007.
[43] E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra, Locally adaptive dimensionality reduction for indexing large time series databases, ACM SIGMOD Rec. 30 (2) (2001) 151–162 https://dl.acm.org/doi/abs/10.1145/375663.375680.
[44] J. Klingner, B. Tversky, P. Hanrahan, Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks, Psychophysiology 48 (3) (2011) 323–332, https://doi.org/10.1111/j.1469-8986.2010.01069.x.
[45] C. Koester, We must change the statistics of confined space injuries and fatalities, 2019. Available on-line via https://ohsonline.com/articles/2018/08/01/we-must-change-the-statistics-of-confined-space-injuries-and-fatalities.aspx. Accessed date: March 2020.
[46] M.T. Kucewicz, J. Dolezal, V. Kremen, B.M. Berry, L.R. Miller, A.L. Magee, V. Fabian, G.A. Worrell, Pupil size reflects successful encoding and recall of memory in humans, Scientific Reports 8 (1) (2018), pp. 4949m, doi:https://doi.org/10.1038/s41598-018-23197-6.
[47] J. Leppink, F. Paas, C.P. Van der Vleuten, T. Van Gog, J.J. Van Merriënboer, Development of an instrument for measuring different types of cognitive load, Behav. Res. Methods 45 (4) (2013) 1058–1072, https://doi.org/10.3758/s13428-013-0334-1.
[48] J. Leppink, F. Paas, T. Van Gog, C.P. van Der Vleuten, J.J. Van Merrienboer, Effects of pairs of problems and examples on task performance and different types of cognitive load, Learn. Instr. 30 (2014) 32–42, https://doi.org/10.1016/j.learninstruc.2013.12.001.
[49] J. Lin, E. Keogh, S. Lonardi, B. Chiu, A symbolic representation of time series, with implications for streaming algorithms, Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, ACM, 2003, pp. 2–11, doi:https://doi.org/10.1145/882082.882086.
[50] Y. Ling, W.-P. Brinkman, H.T. Nefs, C. Qu, I. Heynderickx, Effects of stereoscopic viewing on presence, anxiety, and cybersickness in a virtual reality environment for public speaking, Presence Teleop. Virt. 21 (3) (2012) 254–267, https://doi.org/10.1162/PRES_a_00111.
[51] S.J. Luck, E.K. Vogel, K.L. Shapiro, Word meanings can be accessed but not reported during the attentional blink, Nature 383 (6601) (1996) 616, https://doi.org/10.1038/383616a0.
[52] S. Mathôt, A simple way to reconstruct pupil size during eye blinks, Retrieved from 10 (2013) m9. Available on-line via https://www.researchgate.net/profile/Sebastiaan_Mathot/publication/236268543_A_simple_way_to_reconstruct_pupil_size_during_eye_blinks/links/00b7d51779b2e5e567000000/A-simple-way-to-reconstruct-pupil-size-during-eye-blinks.pdf. Accessed date: March 2020.
[53] S. Mathôt, J. Fabius, E. Van Heusden, S. Van der Stigchel, Safe and sensible preprocessing and baseline correction of pupil-size data, Behav. Res. Methods 50 (1) (2018) 94–106, https://doi.org/10.3758/s13428-017-1007-2.
[54] G. Matthews, W. Middleton, B. Gilmartin, M. Bullimore, Pupillary diameter and cognitive load, Journal of Psychophysiology (1991). Available on-line via https://psycnet.apa.org/record/1992-11805-001. Accessed date: March 2020.
[55] R. Mayer, R.E. Mayer, The Cambridge Handbook of Multimedia Learning, Cambridge University Press, 2005, https://doi.org/10.1017/CBO9781139547369.
[56] R.E. Mayer, Multimedia learning, Psychology of Learning and Motivation, Vol. 41, Elsevier, 2002, pp. 85–139, doi:https://doi.org/10.1016/S0079-7421(02)80005-6.
[57] S. Moresi, J.J. Adam, J. Rijcken, P.W. Van Gerven, H. Kuipers, J. Jolles, Pupil dilation in response preparation, Int. J. Psychophysiol. 67 (2) (2008) 124–130, https://doi.org/10.1016/j.ijpsycho.2007.10.011.
[58] C.D. Morris, J.D. Bransford, J.J. Franks, Levels of processing versus transfer appropriate processing, J. Verbal Learn. Verbal Behav. 16 (5) (1977) 519–533, https://doi.org/10.1016/S0022-5371(77)80016-9.
[59] N.V. Nayak, J.E. Taylor, Offshore outsourcing in global design networks, J. Manag. Eng. 25 (4) (2009) 177–184, https://doi.org/10.1061/(ASCE)0742-597X(2009)25:4(177).
[60] OSHA, Paperwork Reduction Act 1995 Submission supporting statement for permit-required confined spaces, Vol. 2019, 1999. Available on-line via https://www.osha.gov/laws-regs/federalregister/2005-08-31. Accessed date: March 2020.
[61] F. Paas, A. Renkl, J. Sweller, Cognitive load theory and instructional design: recent developments, Educ. Psychol. 38 (1) (2003) 1–4, https://doi.org/10.1207/S15326985EP3801_1.
[62] O. Palinko, A.L. Kun, A. Shyrokov, P. Heeman, Estimating cognitive load using remote eye tracking in a driving simulator, Proceedings of the 2010 Symposium on Eye-tracking Research & Applications, ACM, 2010, pp. 141–144, doi:https://doi.org/10.1145/1743666.1743701.
[63] M.H. Papesh, S.D. Goldinger, M.C. Hout, Memory strength and specificity revealed by pupillometry, Int. J. Psychophysiol. 83 (1) (2012) 56–64, https://doi.org/10.1016/j.ijpsycho.2011.10.002.
[64] D.T. Payne, M.E. Parry, S.J. Harasymiw, Percentage of pupillary dilation as a measure of item difficulty, Percept. Psychophys. 4 (3) (1968) 139–143, https://doi.org/10.3758/BF03210453.
[65] R.K. Pearson, Y. Neuvo, J. Astola, M. Gabbouj, Generalized hampel filters, EURASIP Journal on Advances in Signal Processing 2016 (1) (2016) 87, https://doi.org/10.1186/s13634-016-0383-6.
[66] L. Phillips, V. Interrante, M. Kaeding, B. Ries, L. Anderson, Correlations between physiological response, gait, personality, and presence in immersive virtual environments, Presence 21 (2) (2012) 119–141, https://doi.org/10.1162/PRES_a_00100.
[67] P. Piu, V. Serchi, F. Rosini, A. Rufa, A cross-recurrence analysis of the pupil size fluctuations in steady scotopic conditions, Front. Neurosci. 13 (2019) 407, https://doi.org/10.3389/fnins.2019.00407.
[68] E.D. Ragan, D.A. Bowman, K.J. Huber, Supporting cognitive processing with spatial information presentations in virtual environments, Virtual Reality 16 (4) (2012) 301–314, https://doi.org/10.1007/s10055-012-0211-8.
[69] E.D. Ragan, R. Kopper, P. Schuchardt, D.A. Bowman, Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task, IEEE Trans. Vis. Comput. Graph. 19 (5) (2013) 886–896, https://doi.org/10.1109/TVCG.2012.163.
[70] E.D. Ragan, A. Sowndararajan, R. Kopper, D.A. Bowman, The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments, Presence Teleop. Virt. 19 (6) (2010) 527–543, https://doi.org/10.1162/pres_a_00016.
[71] M.A. Recarte, L.M. Nunes, Effects of verbal and spatial-imagery tasks on eye fixations while driving, J. Exp. Psychol. Appl. 6 (1) (2000) 31, https://doi.org/10.1037/1076-898X.6.1.31.
[72] D. Richards, M. Taylor, A comparison of learning gains when using a 2D simulation tool versus a 3D virtual world: an experiment to find the right representation involving the Marginal Value Theorem, Comput. Educ. 86 (2015) 157–171, https://doi.org/10.1016/j.compedu.2015.03.009.
[73] A.E. Richardson, D.R. Montello, M. Hegarty, Spatial knowledge acquisition from maps and from navigation in real and virtual environments, Mem. Cogn. 27 (4) (1999) 741–750, https://doi.org/10.3758/BF03211566.
[74] B. Roberta, P.L. Esther, O. Diana, Pupil responses to pitch deviants reflect predictability of melodic sequences, BioRxiv (2019), pp. 693382, doi:https://doi.org/10.1016/j.bandc.2019.103621.
[75] R. Sacks, R. Barak, Impact of three-dimensional parametric modeling of buildings on productivity in structural engineering practice, Autom. Constr. 17 (4) (2008) 439–449, https://doi.org/10.1016/j.autcon.2007.08.003.
[76] R. Sacks, A. Perlman, R. Barak, Construction safety training using immersive virtual reality, Constr. Manag. Econ. 31 (9) (2013) 1005–1017, https://doi.org/10.1080/01446193.2013.828844.
[77] A.Z. Sampaio, M.M. Ferreira, D.P. Rosário, O.P. Martins, 3D and VR models in Civil Engineering education: Construction, rehabilitation and maintenance, Autom. Constr. 19 (7) (2010) 819–828, https://doi.org/10.1016/j.autcon.2010.05.006.
[78] J. Sauro, J.R. Lewis, Average task times in usability tests: what to report?, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2010, pp. 2347–2350, doi:https://doi.org/10.1145/1753326.1753679.
[79] J. Sauro, J.R. Lewis, Quantifying the user experience: practical statistics for user research, Morgan Kaufmann, 2016. Available on-line via https://www.sciencedirect.com/book/9780128023082/quantifying-the-user-experience. Accessed date: March 2020.
[80] M. Schwalm, A. Keinath, H.D. Zimmer, Pupillometry as a method for measuring

mental workload within a simulated driving task, Human Factors for Assistance and Automation (1986) (2008) 1–13. Available on-line via http://www.hfes-europe.org/wp-content/uploads/2014/06/6-fa.pdf. Accessed date: March 2020.
[81] Y. Shi, J. Du, C.R. Ahn, E. Ragan, Impact assessment of reinforced learning methods on construction workers' fall risk behavior using virtual reality, Autom. Constr. 104 (2019) 197–214, https://doi.org/10.1016/j.autcon.2019.04.015.
[82] Y. Shi, J. Du, S. Lavy, D. Zhao, A multiuser shared virtual environment for facility management, Procedia Engineering 145 (2016) 120–127, https://doi.org/10.1016/j.proeng.2016.04.029.
[83] Y. Shi, J. Du, E. Ragan, K. Choi, S. Ma, Social influence on construction safety behaviors: a multi-user virtual reality experiment, Construction Research Congress 2018, pp. 174–183, doi:https://doi.org/10.1061/9780784481288.018.
[84] Y. Shi, J. Du, P. Tang, D. Zhao, Characterizing the role of communications in teams carrying out building inspection, Construction Research Congress 2018, pp. 554–564, doi:https://doi.org/10.1061/9780784481271.054.
[85] T.K. Srull, M. Lichtenstein, M. Rothbart, Associative storage and retrieval processes in person memory, Journal of Experimental Psychology: Learning, Memory, and Cognition 11 (2) (1985), pp. 316, doi:https://doi.org/10.1037//0278-7393.11.2.316.
[86] J. Sweany, P. Goodrum, J. Miller, Analysis of empirical data on the effects of the format of engineering deliverables on craft performance, Autom. Constr. 69 (2016) 59–67, https://doi.org/10.1016/j.autcon.2016.05.017.
[87] J. Sweller, Cognitive load during problem solving: effects on learning, Cogn. Sci. 12 (2) (1988) 257–285, https://doi.org/10.1016/0364-0213(88)90023-7.
[88] J. Sweller, Cognitive load theory: recent theoretical advances, Cognitive load theory. Cambridge University Press. (2010) (p. 29–47), doi:https://doi.org/10.1017/CBO9780511844744.004.
[89] M. Tavanti, M. Lind, 2D vs 3D, implications on spatial memory, IEEE Symposium on Information Visualization, 2001. INFOVIS 2001., IEEE, 2001, pp. 139–145, doi:https://doi.org/10.1109/INFVIS.2001.963291.
[90] Tobii, How do Tobii Eye Trackers work?, Vol. 2019, 2019. Available on-line via https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/. Accessed date: March 2020.
[91] Tobii, Tobii Pro SDK, Vol. 2019, 2019. Available on-line via http://developer.tobiipro.com/. Accessed date: March 2020.
[92] Tobii, Tobii Pro X3-120, Vol. 2019, 2019. Available on-line via https://www.tobiipro.com/product-listing/tobii-pro-x3-120/. Accessed date: March 2020.
[93] U.S. Bureau of Labor Statistics, Census of fatal occupational injuries summary, 2017, Vol. 2019, 2018. Available on-line via https://www.bls.gov/news.release/cfoi.nr0.htm. Accessed date: March 2020.
[94] C.G. Ucros, Mood state-dependent memory: a meta-analysis, Cognit. Emot. 3 (2) (1989) 139–169, https://doi.org/10.1080/02699938908408077.
[95] UK, Construction 2025: industrial strategy for construction - government and industry in partnership, Vol. 2019, 2013, Available on-line via https://www.gov.uk/government/publications/construction-2025-strategy. Accessed date: March 2020.
[96] M. Usoh, E. Catena, S. Arman, M. Slater, Using presence questionnaires in reality, Presence: Teleoperators & Virtual Environments 9 (5) (2000) 497–503, https://doi.org/10.1162/105474600566989.
[97] P. van der Wel, H. van Steenbergen, Pupil dilation as an index of effort in cognitive control tasks: a review, Psychon. Bull. Rev. 25 (6) (2018) 2005–2015, https://doi.org/10.3758/s13423-018-1432-y.
[98] A. Verghote, S. Al-Haddad, P. Goodrum, S. Van Emelen, The effects of information format and spatial cognition on individual wayfinding performance, Buildings 9 (2) (2019) 29, https://doi.org/10.3390/buildings9020029.
[99] S.P. Verney, E. Granholm, D.P. Dionisio, Pupillary responses and processing resources on the visual backward masking task, Psychophysiology 38 (1) (2001) 76–83, https://doi.org/10.1111/1469-8986.3810076.
[100] W3C, Working Draft-techniques for Accessibility Evaluation And Repair Tools, in: C. Ridpath (Ed.), Vol. 2019, 2000. Available on-line via https://www.w3.org/TR/AERT/. Accessed date: March 2020.
[101] X. Wang, P.E. Love, M.J. Kim, C.-S. Park, C.-P. Sing, L. Hou, A conceptual framework for integrating building information modeling with augmented reality, Autom. Constr. 34 (2013) 37–44, https://doi.org/10.1016/j.autcon.2012.10.012.
[102] L. Wei, N. Kumar, V.N. Lolla, E.J. Keogh, S. Lonardi, C. Ratanamahatana, Assumption-free anomaly detection in time series, SSDBM 5 (2005) 237–242 https://dl.acm.org/doi/abs/10.5555/1116877.1116907.
[103] C.D. Wickens, Multiple resources and mental workload, Hum. Factors 50 (3) (2008) 449–455, https://doi.org/10.1518/001872008X288394.
[104] C.D. Wickens, Multiple resources and performance prediction, Theor. Issues Ergon. Sci. 3 (2) (2002) 159–177, https://doi.org/10.1080/14639220210123806.
[105] C.D. Wickens, Processing resources and attention, Multiple-task Performance, 1991, pp. 3–34. Available on-line via https://books.google.com/books?hl=en&lr=&id=_KfaN4gKP8MC&oi=fnd&pg=PA3&ots=S0F2Em5als&sig=h8eu2G4ksdr7zap_LZCCY57VMbs#v=onepage&q&f=false (Accessed date: March 2020).
[106] C.D. Wickens, The effects of divided attention on information processing in manual tracking, J. Exp. Psychol. Hum. Percept. Perform. 2 (1) (1976) 1, https://doi.org/10.1037/0096-1523.2.1.1.
[107] C.D. Wickens, J.M. Flach, Information Processing, Elsevier, Human Factors in Aviation, 1988, pp. 111–155, https://doi.org/10.1016/B978-0-08-057090-7.50011-4.
[108] B.J. Wilhelm, H. Wilhelm, S. Moro, J.L. Barbur, Pupil response components: studies in patients with Parinaud's syndrome, Brain 125 (10) (2002) 2296–2307, https://doi.org/10.1093/brain/awf232.
[109] W. Wu, J. Hartless, A. Tesei, V. Gunji, S. Ayer, J. London, Design assessment in virtual and mixed reality environments: comparison of novices and experts, J. Constr. Eng. Manag. 145 (9) (2019) 04019049, , https://doi.org/10.1061/(ASCE)CO.1943-7862.0001683.
[110] S. Zhang, J. Teizer, J.-K. Lee, C.M. Eastman, M. Venugopal, Building information modeling (BIM) and safety: automatic safety checking of construction models and schedules, Autom. Constr. 29 (2013) 183–195, https://doi.org/10.1016/j.autcon.2012.05.006.