2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)

Exploring Augmented Reality's Role in Enhancing Spatial Perception for Building Facade Retrofit Design for Non-experts

John Sermarini[*] Robert A. Michlowitz[†] Joseph J. LaViola Jr.[‡] Lori C. Walters[§] Roger Azevedo[¶]
Joseph T. Kider Jr.[‖]

University of Central Florida

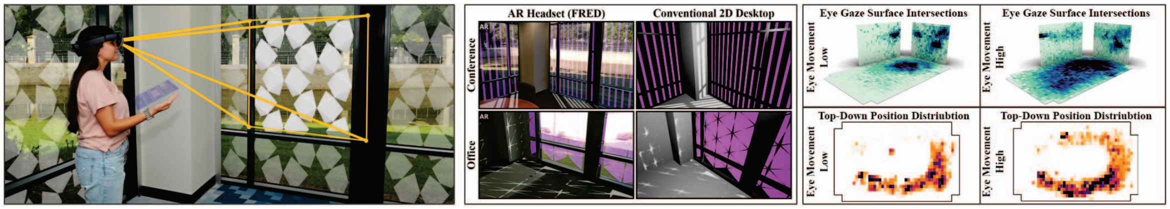


Figure 1: (Left) Our approach enables designers to see the impact that different facade designs will have over time and space in the built environment by displaying the design's daylighting, energy, and aesthetics in an embodied 3D environment. (Middle) Shows our experimental conditions and test conditions. (Right) Shows the distributions of eye gaze-surface intersections and top-down positions of our designers, separated by quantity of eye movement utilized.

ABSTRACT

Augmented Reality (AR) tools have demonstrated considerable promise to enhance creative architectural design and support the retrofitting problem-solving processes through on-site daylighting visualization. AR's capacity to integrate embodied motion enhances the non-expert's understanding of the spatial characteristics and design ramifications within the built environment for complex facade design. Motion provides insights and increases the accessibility of retrofitting, encouraging more energy-efficient rework as opposed to complete building reconstruction. This study investigates the decision-making outcomes and cognitive-physical load implications of integrating a Building Information Modeling-driven AR system into the retrofitting design process and how movement is best leveraged to understand daylighting impacts. We conducted a study with 128 non-expert participants, who were asked to choose a window facade retrofit to improve an interior space. We analyze the effects of head movement, head rotations, and eye movements to understand how embodied motion improves overall objective performance across several daylighting and energy design metrics. We found no significant difference in the overall decision-making outcome between those who used an AR tool or a conventional desktop approach and that greater eye movement in AR was related to non-experts better balancing the complicated impacts facades have on daylight, aesthetics, and energy. This study indicates future expansion of AR retrofitting tools should encourage more eye movement.

Keywords: Augmented Reality, Building Information Modeling, Daylighting, Human Computer Interaction

Index Terms: Applied computing—Arts and humanities—Architecture (buildings)—Computer-aided design; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality Human-centered computing—Visualization—Visualization application domains—Information visualization

1 INTRODUCTION

Retrofitting is a process where a component or feature of a structure that was not included in its initial construction is upgraded. Often, this is viewed as upgrading to new codes to benefit potential structural damage such as earthquakes and hurricanes. However, many architecturally unique or historically significant buildings are also being retrofitted. By choosing to retrofit over complete reconstruction, buildings can be repurposed and achieve new energy and occupant comfort goals while minimizing environmental waste [20] and preserve an area's cultural identity [44]. While retrofitting is important, designing retrofits is a delicate balance of improving functionality and maintaining original aesthetic form. Meyer [46] defines aesthetics as "the art and science of sense perception and cognition". These aesthetics communicate information and emotion to occupants. This also functions as an extension of Gibson's [25] concept of affordances, communicating experiences and feelings in addition to functional capabilities. Therefore, preserving the design is vital while increasing performance is crucial during the retrofitting process [45]. This adds extra complexity to the design process which is often compounded by limited budgeting (or more accurately, lengthy payback periods) [32] and insufficient technological innovation to support the problem-solving process [12].

By using Augmented Reality (AR) and Building Information Modeling (BIM), designers are able to situate design concepts in their real-life environments and visualize them using head-mounted displays (HMDs) or handheld systems. The designer can then interact with the visualizations by physically moving about their three-dimensional environment instead of being constrained to a two-dimensional screen. This permits a greater spatial understanding of the designs [7] and forms deep links between the mind and body by experiencing the human-centered impacts of the architecture through movement [64]. However, despite 90% of buildings in the United States being constructed before 1990 [70], the majority of BIM and mixed reality technology is focused on new construc-

---

[*] e-mail: john.sermarini@ucf.edu
[†] e-mail: robert.michlowitz@ucf.edu
[‡] e-mail: jlaviola@ucf.edu
[§] e-mail: lori.walters@ucf.edu
[¶] e-mail: roger.azevedo@ucf.edu
[‖] e-mail: joseph.kider@ucf.edu

2642-5254/24/$31.00 ©2024 IEEE
DOI 10.1109/VR58804.2024.00046

tion [2]. Existing building stock needs technological support so both experts and non-experts can have agency and play a role. For non-experts in particular, presenting BIM data within augmented reality technology is a promising method of improving design understandability and decision-making without extensive technical and BIM training [11, 71].

The objective of this study is to investigate how AR enables participants to leverage movement effectively while engaging in BIM-driven retrofitting design tasks. To achieve this, we leverage a BIM-enabled AR Facade Retrofitting Embodied Design (*FRED*) system [58, 59] to examine retrofit window facade designs *in situ* (Figure). Window facades refer to external structures and interventions installed over existing windows designed to enhance a building's energy efficiency and the comfort of its occupants. Installing them is a relatively inexpensive building retrofit, but their design requires balancing trade-offs with conflicting design variables [34]. We evaluate how individuals address these conflicting design goals using a human subject study where participants' movement and decision-making are tracked and evaluated against a conventional desktop-based design tool. This study primarily targets non-experts who need to understand these retrofits for an established commercial or private space and also serves as a foundation for replication with Architecture, Engineering, and Construction (AEC) industry professionals.

Our results found no significant difference in the overall decision-making outcome between these two design tools when displaying simulation-driven data to non-experts. They also suggest that, depending on the size and configuration of the wearer's physical environment, quantity of eye movement can be an effective indicator of objective decision-making performance and perceived physical load in understanding retrofitting impacts. These results will help guide future AEC applications to encourage movement to help decision-making outcomes for future AR retrofitting applications.

2 RELATED WORK

2.1 Facade Design

Design details in the built environment have both major and subtle impacts on occupant emotional and physical responses [38]. Because of this, when evaluating facade designs, one should consider their implications on both building energy performance and occupant well-being. Accessible exterior views reduce reliance on electrical lighting while simultaneously improving occupant physical health, reducing tiredness and stress [6, 65]. However, too much natural lighting can lead to discomfort and reduced productivity. The balance of possible retrofitting outcomes generally relies on a large set of design parameters, and visualizing these parameters to understand their impacts can be a demanding process for the designer. Parametric modeling [19] allows architects to optimize building elements, such as facades, by automatically altering parameters such as fenestration, rotation, quantity, and location. Applying a parametric modeling framework to facade design allows designers to achieve an optimal balance of comfort, solar radiation, and energy efficiency in a shorter period of time. Frequently, parametric modeling neglects occupant needs when analyzing vast amounts of data for multi-objective problem spaces [48]. The integration of AR+BIM tools will facilitate onsite evaluation of daylighting (natural sunlight) into the design process, which can enhance spatial cognition and lead to more favorable design outcomes.

2.2 AR-enabled Movement in Design

Contemporary views on cognitive science tend to embrace 4E perspectives that unify the roles of the body, brain, and environment in the cognitive process [51]. Clark & Chalmers [14] present the highly influential view that kinesthetic sensorimotor body interactions fundamentally shape cognition. These same interactions allow humans to utilize the environment to extend the problem-solving capabilities of their overall cognitive system through physical motions [50, 63], environmental offloading [36], and distributing task-load [31]. Related fields of thought position the cognitive process as situated, meaning it is intrinsically tied to the context in which it occurs [68]. These views are particularly interesting when applied to architectural design. Mallgrave [43] defines the idea that architecture's aesthetic, cultural, emotional, and experience are intrinsically tied to the embodied condition of immersion and enactive motion in the situated space. When evaluating designs, proper scaling [61] and a collection of view perspectives [57] are required to understand its impacts on occupants fully. Therefore, it is logical that experiencing architecture in an immersive and controllable manner would grant the most accurate interpretation of these factors. Robinson & Pallasmaa [55] further discussed how architecture and design link the body and mind.

AEC professionals rely on design visualization to accurately understand design implications for occupants of the built environment. Because of these developments in cognitive science, however, it is no longer sufficient to rely exclusively on systems and design frameworks that overemphasize the brain component of this relationship. Recently, Augmented Reality (AR) and Virtual Reality (VR) design systems have been investigated as a means to address this inefficiency [2]. The embodied perspective they provide enhances creativity and understanding of spatial relationships [1]. HMD-based systems better support stereoscopic vision and accommodate the natural depth cues human perception is optimized for [23]. Movement and inputs into these systems correspond with the body's natural movements, as opposed to peripheral-driven keyboard, mouse, and touchscreen methods [22, 60]. This also can enable the use of the designer's hands for cognitive gesturing [50] or visualization manipulation [10]. These factors collectively contribute to the potential of AR and VR in aiding designers to empathize with the experiences of future occupants and users of their designs; however, whether this empathy directly translates into improved design decision-making needs further study.

2.3 BIM-AR in AEC

Architects, engineers, construction workers, and facility managers are able to use and visualize expansive amounts of data with Building Information and Modeling (BIM) data, which encompasses all the information related to a building's lifecycle [2, 5]. The use of BIM can facilitate the organization and access of massive amounts of documentation, enabling data coordination, control systems, and maintaining simulations throughout the life cycle of a building [66]. The integration of BIM into the construction process has been effective in improving project costs, duration, stakeholder communication, and quality, but these impacts can be affected by software, hardware, and training issues [5, 9, 16, 26, 41].

AR and VR applications using BIM-sourced data have been explored to improve their visualization capabilities and accessibility [2]. General-purpose mixed reality applications exist at the industry level (for a brief summary, see Huang et al. [30]), however, the features provided by these applications are inconsistent at best. Consequently, problem-specific systems developed within modern simulation engines have become common-place [15, 69], and have been developed to support architectural concepts through design [67], construction [47], and maintenance [28]. Recent work focusing more specifically on design includes creating 3D geometry in AR [53], sketching 3D designs directly to support existing physical objects [39], freehand sketching buildings [42], evaluating interior thermal distribution [29], visualizing energy performance [62], design-based maintenance prevention [35], interior lighting design [49], and disaster prevention [72].

Despite the potential benefits AR and VR present to both visualizing BIM data and retrofitting, work combining all of these concepts effectively is considerably more sparse. A likely contributor to this is a lack of accessible digitized BIM data for older buildings [18,33]. Accordingly, Patil et al. [52] and Kumar et al. [37] present similar AR and VR systems for aiding in utilizing 3D point cloud data for retrofitting proposals. de Freitas & Ruschel [17] proposed a methodology for implementing BIM-driven AR into the post-occupancy evaluation process to assist in identifying retrofits. These systems can aid in preparing retrofitting proposals. Y. Liu et al. [40] presented a case study of showcasing BIM-data in VR during retrofitting design meetings, where it was found to be effective at improving design understanding. Fukuda et al. [24] showcased how indoor greening retrofit designs affect thermal conditions using BIM-sourced models in AR.

3 METHODOLOGY

The purpose of this study is to evaluate two primary research questions that are of interest when improving retrofit design workflow and outcomes:

> **RQ1**: Do the embodied movement benefits of increased head and eye movement by BIM-enabled AR permit better decision-making and workflows for retrofitting design?
>
> **RQ2**: Are these design decisions related to the quantity of movement in AR?

3.1 Participants

We recruited 131 participants (79 male, 51 female, 1 agender, age 18-45, *M* = 20.89, *SD* = 4.79), who were primarily students and other individuals affiliated with our university. Each volunteer received course credit or $10 compensation. Three (3) participants' data were excluded from the study due to software malfunctions or their inability to follow the study administrators instructions (2 male, 1 female). All participants had normal or corrected-to-normal vision and reported no previous VR-related motion sickness, or relevant motor or sensory deficits. The experimental protocol was approved by the university IRB, and each participant consented after being informed of the study's purpose. Participants were treated in accordance with the ethical guidelines of the American Psychological Association and the Declaration of Helsinki.

3.2 Study Design

We developed a 2 x 2 between-subjects design for evaluating the impacts of augmentation type and environmental complexity when evaluating retrofitting design decisions. Participants were evenly distributed across the four conditions (*n* = 32 each) and asked to explore the given facade design options to optimize three design variables: view factor, daylighting, and brightness discomfort.

Augmentation Type - AR Headset (*FRED*): A Microsoft *HoloLens 2* AR HMD with custom-designed software will provide participants with a user interface, facade visualizations, and daylighting feedback *in situ* of the design space in 3D augmented reality with virtual components (augmented facades and augmented daylighting).

Augmentation Type - Conventional 2D Desktop: A conventional desktop computer with custom-designed software will provide participants with a user interface, facade visualizations, and daylighting feedback on a 32 inch 2D screen.

Complexity Type - Conference Room (High): A conference room scenario where participants are asked to judge window facade retrofit designs on two walls windows to optimize impacts of usable daylighting, visibility out of the windows, and potential daylighting discomfort. The conference room is 14.64 ft x 23.23 ft with 1.46 ft x 1.46 ft columns in each corner that extend to the ceiling. The two wall-sized windows are 10.01 ft x 8.70 ft and 12.70 ft x 8.70 ft.

Complexity Type - Office (Low): An office scenario where participants are asked to judge window facade retrofit designs on a single surface window to optimize impacts of usable daylighting, visibility out of the windows, and potential daylighting discomfort. The office is a 12.17 ft x 9.76 ft room with a single wall-sized window that is 12.46 ft x 8.42 ft.

3.2.1 Design Environments

Two rooms were selected for study conditions: 1) a standard second-floor office space with one wall of floor-to-ceiling windows and 2) a larger first-floor conference room with two walls of floor-to-ceiling windows (Figure 2). The windows of the selected rooms face north and east and do not experience direct sunlight for the majority of the day, so lighting simulations were performed instead as if these windows were south-facing. This permits more dramatic daylighting effects for participants to consider.

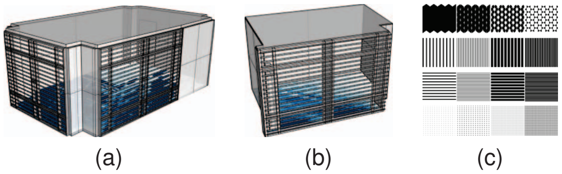


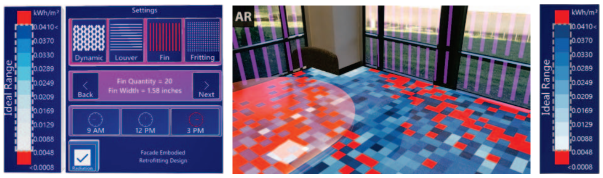
Figure 2: The larger complex conference room (a), smaller simple office room (b), and sixteen facade design options (c). Daylighting simulation results are visible on the floor of both rooms.

Because BIM data for these spaces were unavailable, they were recreated using on-site measurements for use in both systems using *Autodesk Revit* and *Rhinoceros 6*. Sixteen possible facade configurations were developed with *Rhinoceros 6* as well and were positioned and scaled to fit the window surfaces in both rooms. Following this, daylighting simulations were calculated in *Rhinoceros 6* using *Ladybug* and *Honeybee* [56] within the *Grasshopper* visual scripting environment. This generates a "solar radiation heatmap" which displays the the lighting levels in the room at the given time of day as both a CSV data file and floor-aligned texture. The texture values are rendered as a 0.1m x 0.1m grid aligned to the window surfaces and floor (Figure 3b).

These simulation data textures were imported into our development environment in *Unreal Engine 4.25* using the native *Datasmith* plug-in. This environment was the basis for both the AR and desktop systems. Real-time BIM data limitations in Datasmith resulted in the need for these daylighting simulations to be pre-computed in Rhinoceros 6, however aesthetic lighting is rendered in real-time using Unreal Engine and the physical location's real-life latitude, longitude, date, and north-facing direction. Participants can toggle the visibility of the solar radiation heatmap texture at will and are instructed to use it as they please. An annotated legend that outlines the ideal lighting range for a typical office building is available as well (Figure 3c). This system can be extended to other facade designs and environments, and the process, along with all custom software, has been released open-source on GitHub for replication and adaptation by other researchers (see Supplemental Materials).

3.2.2 Design Goals

Four common facade classes with four variations each were utilized in this study for a total of sixteen options: kinetic folding triangles, fins, louvers, and glass fritting (Figure 2c). These variations could be freely explored with no time limit imposed. Participants were asked to evaluate these options across three metrics: *view factor*, *daylighting*, and *brightness discomfort*. For view factor, participants are asked to minimize the impact of the facade on the window's exterior view. This value doubles as a partial stand-in for aesthetics and is the percentage of window surface area that is unobscured. For daylighting, participants are asked to maximize the optimal daylighting levels in the room throughout the course of the day. For brightness discomfort, participants are asked to minimize the area



(a) (b) (c)

Figure 3: The user interface used by participants (a). Environmentally embedded daylighting lighting simulation results rendered for the participant in AR (b). The corresponding legend with non-ideal lighting levels is displayed as red grid cells (c).

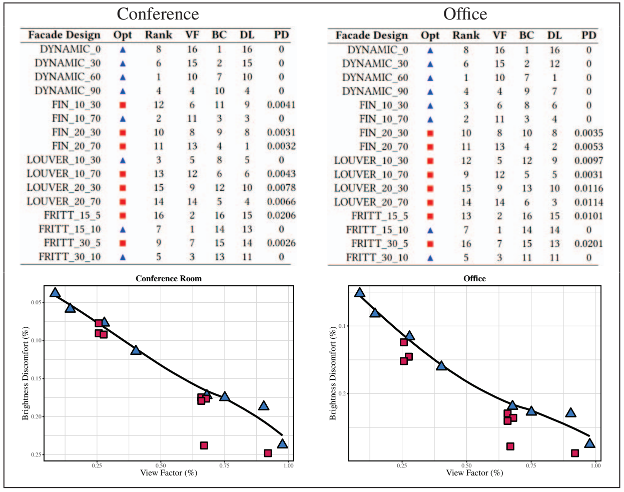
of the room outside of the ideal daylighting range. These lighting-based values were calculated using the Ladybug lighting simulation results and measuring the percentage of the floor that was within (for daylighting) or greater than (for brightness discomfort) the designated optimal lighting range of 100 lux to 5000 lux. This range encompasses a variety of "office-style" work in interior environments [3, 21], and allows for greater levels on the higher end due to greater occupant acceptance of excessive natural daylighting [54]. These results are calculated for each possible time of day presented (9 AM, 12 PM, and 3 PM) and the mean for each facade option is used.

An ideal window facade retrofit would optimize all three design variables, but their coupled nature makes this unattainable. For example, increasing the view factor allows more light into the room and can potentially lead to greater glare and brightness discomfort. In multi-objective scenarios such as this, the *Pareto optimality* is a measure that can approximate what the most "optimal" decision would be [13]. By charting these options, the *Pareto frontier* visualizes optimal solutions for the Conference room and Office room (Table 1). The sixteen points on each plot represent each facade option provided to participants, where blue triangles indicate optimal solutions and red squares indicate relatively poor solutions. The Y-axis is the facade's brightness discomfort percentage [0.0 (most optimal) – 1.0 (least optimal)] and the X-axis is the facade's view factor [0.0 (least optimal) – 1.0 (most optimal)]. Facades closer to the bottom-left corner (under the line of best fit Pareto frontier approximation) represent a less desirable balance or fail to maximize one of these two factors. For optimal facades, the third variable (daylighting) is then considered as a tie-breaker to rank them from one (most optimal) to eight. For the remaining non-optimal solutions, distance to the Pareto frontier line ranks them from nine (closest) to sixteen (furthest and least optimal). This ranking methodology examines all three considered variables while prioritizing aesthetics and occupant comfort (Table 1). These rankings and facade optimality classifications can be used to evaluate participant decision-making for retrofitting facade design.

3.2.3 Design Systems

Participants assigned to an AR condition were brought to the location of their room and then used the HoloLens 2 (FOV: 52 degrees diagonal, screen resolution: 1440 x 936 per eye, refresh rate: 60 Hz) to visualize the facades and their resultant lighting *in situ*. Those assigned to a desktop condition were situated in an isolated lab space in front of the desktop computer and single monitor. Both groups utilized a user interface to control which facade class and variant is active, change the simulated time of day, and toggle the solar radiation heatmap in an accessible fashion (Figure 3a). For AR participants, this user interface is constrained to the right of the participant's left hand and is visible only when their palm is flat, facing up toward the AR headset. When visible, the interface buttons are selectable by tapping them with the participant's right hand. For desktop participants, the interface is constrained to the left one-third of the screen, and its visibility is toggled using a keyboard input. Toggling visibility of the solar radiation heatmap is handled with the "Radiation" checkbox in the bottom left. Using this, participants can freely alternate between aesthetic rendered lighting and simulation-driven embedded lighting data. The solar radiation heatmap legend is constrained to the left side of this interface and is visible only when the solar radiation heatmap is toggled on. As the participant selects a facade configuration and simulation time of day, the system alters the virtual lighting and solar radiation heatmap to reflect those changes throughout their surrounding environment (Figure 3b).

Table 1: Ranking data and visualizations of the facade option simulation results. Designs considered optimal are marked as blue triangles, while non-optimal options are marked as red squares. **Opt** is a binary value representing optimal or not. **VF**, **BC**, and **DL** are the facade's ranking for view factor, brightness discomfort, and daylighting (lower = more optimal). **PD** is the facade's distance from the Pareto frontier when plotted. Points rendered closer to the top-right of the graphic represent more optimal design options, and those below the best-fit Pareto frontier are non-optimal.



Conference

Facade Design	Opt	Rank	VF	BC	DL	PD
DYNAMIC_0	▲	8	16	1	16	0
DYNAMIC_30	▲	6	15	2	15	0
DYNAMIC_60	▲	1	10	7	10	0
DYNAMIC_90	▲	4	4	10	4	0
FIN_10_30	■	12	6	11	9	0.0041
FIN_10_70	▲	2	11	3	3	0
FIN_20_30	■	10	8	9	8	0.0031
FIN_20_70	■	11	13	4	1	0.0032
LOUVER_10_30	▲	3	5	8	5	0
LOUVER_10_70	■	13	12	6	6	0.0043
LOUVER_20_30	■	15	9	12	10	0.0078
LOUVER_20_70	■	14	14	5	4	0.0066
FRITT_15_5	■	16	2	16	15	0.0206
FRITT_15_10	▲	7	1	14	13	0
FRITT_30_5	■	9	7	15	14	0.0026
FRITT_30_10	▲	5	3	13	11	0

Office

Facade Design	Opt	Rank	VF	BC	DL	PD
DYNAMIC_0	▲	8	16	1	16	0
DYNAMIC_30	▲	6	15	2	12	0
DYNAMIC_60	▲	1	10	7	1	0
DYNAMIC_90	▲	4	4	9	7	0
FIN_10_30	▲	3	6	8	6	0
FIN_10_70	▲	2	11	3	4	0
FIN_20_30	■	10	8	10	8	0.0035
FIN_20_70	■	11	13	4	2	0.0053
LOUVER_10_30	■	12	5	12	9	0.0097
LOUVER_10_70	■	9	12	5	5	0.0031
LOUVER_20_30	■	15	9	13	10	0.0116
LOUVER_20_70	■	14	14	6	3	0.0114
FRITT_15_5	■	13	2	16	15	0.0101
FRITT_15_10	▲	7	1	14	14	0
FRITT_30_5	■	16	7	15	13	0.0201
FRITT_30_10	▲	5	3	11	11	0

3.3 Procedure

Upon arriving, participants read and signed the IRB-approved consent form explaining the study's purpose and goals. They were then assigned a participant ID number (unassociated with the individual), assigned a study condition, screened for motor or sensory limitations using an *Apple iPad* and *Qualtrics* survey software, and asked to leave if any critical issues were found. Those assigned to an AR condition then completed the HoloLens 2 system eye calibration. Participants then again used the *iPad* to input their demographic information and complete a questionnaire gauging their experience with the relevant technology and retrofitting procedures. Next, they watched a three-minute video explaining their goal during the study and the variables they were to consider.

Participants were then asked to review the facade design options and make an a priori selection of which facade they felt would best optimize the design variables. Following this, they were asked to put on the AR device (if assigned to an AR condition) or take a seat at the desktop computer (if assigned to a desktop condition). Individuals in both conditions then completed a short training process where the system controls were explained to them and were asked to select a specific randomly-selected facade-facade parameter-time of day combination and toggle on and off the solar radiation heatmap. This training process occurred in the same environment as the main study task, but participant movement data tracking did not begin until its completion. They were then directed to freely explore the design environment and inform the study administrator when they decided which facade they believed was best. A post-study questionnaire was then completed to record information about the participants' decision-making process and their experience with the system.

3.4 Measures

For all participants, the overall ranking, optimality, and individual design variable rankings for both the a priori and final facade selected are recorded (see Table 1). Using the final facade selections, participants are *post hoc* categorized as high (selected an optimal facade) or low-performers (selected a relatively poor facade).

During the design exploration phase of the study, AR participants' 3D head position coordinates, head rotation vectors, and eye gaze vectors, as well as desktop participants' virtual character 3D position coordinates and rotation vectors, are logged at a rate of 5 Hz. All user interface interactions and simulation setting changes are logged for later simulation reconstruction as well.

These transform values and timestamps are then used to calculate new metrics for categorizing participants' movements. *Movement Rating* is the total distance (in centimeters) traveled between every pairing of sequential points divided by the total time (in seconds). *Rotation Rating* is the total angle (in radians) between every pairing of sequential head rotation vectors divided by the total time. *Gaze Rating* is similar, however, with eye rotation vectors. By including only points where optimal facades are visible, we created three new values: *Optimal Movement Rating*, *Optimal Rotation Rating*, and *Optimal Gaze Rating* to estimate nonconscious cognition occurring. This process is repeated with relatively poor facades for *Poor Movement Rating*, *Poor Rotation Rating*, and *Poor Gaze Rating*. For each of these new metrics, participants are post hoc categorized as high (the top half of participants, *n* = 16 per condition) or low (the bottom half of participants, *n* = 16 per condition) within their assigned study condition. These newly created classifications are called *Movement Classification*, *Rotation Classification*, and *Gaze Classification*, and each has poor/optimal counterparts as well. Additional time-related metrics are created as well. For each participant, the total time spent with optimal facades visible is divided by the total design exploration time to create the *Optimal Time Rating* value. The same process is done to create the *Poor Time Rating* value.

Demographic information and levels of experience with AR, VR, computer work, and construction and retrofitting procedures are collected as part of the pre-questionnaire that is completed. Participants were asked to rate these levels using a 1 (very low) to 7 (very high) Likert scale. Two additional post-questionnaires were completed by participants as well. The *System Usability Scale* (SUS) is a ten-question test that measures the participant's opinion on the usability of the system using a 1 (very low) to 5 (very high) Likert scale [8]. The *NASA Task Load Index* (TLX) measures participants' opinions on the demands of completing the task across five metrics: mental demand, physical demand, temporal demand, effort demand, and performance demand [27]. Each metric is measured using a single question that is scored using a 1 (very low) to 21 (very high) Likert scale. The SUS and TLX scores were normalized to convert them to a one-to-one-hundred scale.

4 RESULTS

First, we compare the decision-making performance and process among the augmentation groups. Second, we examine movement's impact on high-performers and low-performers in AR. Third, we study the cognitive and physical loads imposed on participants by the task and system using data derived from our post-study questionnaire. All statistical tests were performed at a significance level of 0.05, with no statistical correction applied. For each test, 95% confidence intervals were included. The ordinal nature of most of our dependent variables necessitates non-parametric tests to be performed in many cases, and each section will indicate its corresponding test and effect size metric.

4.1 Decision-Making Performance

Participant facade selection serves as the primary metric for evaluating design decisions throughout the study. To analyze these decisions, the overall ranking and individual factor rankings were compared across participant groups for the final selections and, second, the differences in ranking between the final and a priori selections.

4.1.1 Final Selections

Figure 4 presents box plot summaries of the overall rankings of the final facade selections. Because the assumption of normality was rejected using a Shapiro-Wilks test, an unpaired two-tailed Mann-Whitney U test was performed for each study room to compare performance between AR and non-AR participants. Reported alongside this is the effect size, *r*. For the conference room, there was no significant difference between the overall rankings of the final facade selections of participants using AR (*n* = 32, *Mdn* = 8.000, *SD* = 5.099) and participants using a conventional desktop system (*n* = 32, *Mdn* = 10.500, *SD* = 5.454), *U* = 486, *p* = 0.731, 95% CI [-3.000, 1.000], *r* = -0.0439 (negligible effect). For the office room, there was also no significant difference between the overall rankings of the final facade selections of participants using AR (*n* = 32, *Mdn* = 6.000, *SD* = 4.833) and participants using a conventional desktop system (*n* = 32, *Mdn* = 9.500, *SD* = 4.856), *U* = 429, *p* = 0.265, 95% CI [-3.000, 1.000], *r* = -0.14 (small effect).

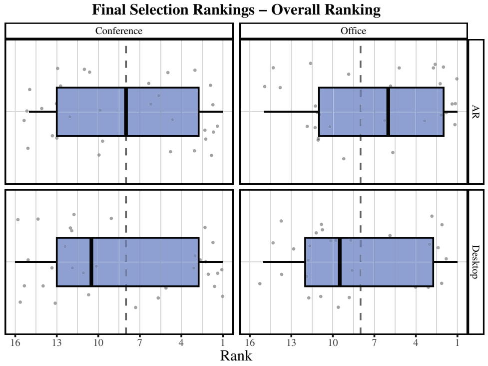


Figure 4: Distributions of the objective overall ranking. Values to the right of the dashed vertical line indicate comparatively optimal selections.

4.1.2 A Priori Vs. Final Selection

Comparisons between participants' a priori facade selections and those made after using the provided system can reveal insights into their effectiveness with using the system and what design variables they prioritized. Figure 5 demonstrates the distribution of differences between the final and initial ranking for each condition, with right-weighted distributions indicating improvement after using the system. A set of two-tailed one-sample Mann-Whitney U tests were performed to compare this difference against an expected mean difference of 0. Reported alongside this is Cohen's D effect size. For AR participants in the conference room (*n* = 32, *Mdn* = 0.000, *SD* = 6.836), no significant difference was reported compared to the expected mean of 0, *U* = 155.5, *p* = 0.861, 95% CI [-4.000, 3.000], *d* = -0.014 (negligible difference). AR participants in the office room (*n* = 32, *Mdn* = -3.000, *SD* = 5.382) appeared to exhibit the most drastic difference, and a significant difference was reported,

U = 124.5, p = 0.016, 95% CI [-5.000, -0.500], d = -0.470 (small difference). For desktop participants in the office room (n = 32, Mdn = 0.000, SD = 6.495), no significant difference was reported, U = 211.5, p = 0.855, 95% CI [-2.500, 3.500], d = 0.067 (negligible difference). For desktop participants in the conference room (n = 32, Mdn = 0.000, SD = 6.858), no significant difference was reported, U = 230.5, p = 0.786, 95% CI [-3.000, 3.000], d = 0.077 (negligible difference).

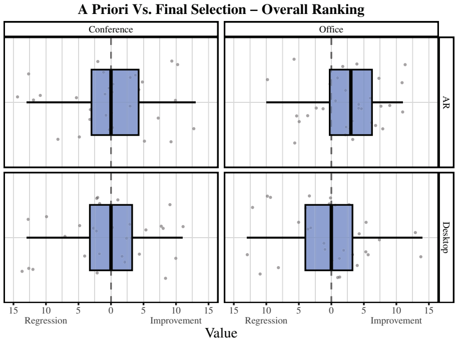


Figure 5: Distributions of the difference between the a priori selection and final selection in the objective overall ranking. Values to the right of the dashed vertical line indicate improvement after using the provided design system.

4.2 Movement in AR

Illustrating the distribution of movement classifications among the performance groups showcases where the ties between the categorizations are most clearly present. For each movement categorization (see Section 3.4), half the participants were categorized as high quantity or low quantity. If the number of high-performing participants is more significantly proportioned to either one of these movement quantities, that would indicate that this type of movement affects performance and, therefore, should be more heavily prioritized in future AR-based design systems. The transparent bars in the background of Figure 6 indicate the total quantity of participants that performed at a high level (blue bars on the right) or low level (red bars on the left). The darker shaded bars indicate the imbalance of those two values for each movement quantity. If the darker bar is blue and on the right side of the mid-point line, this indicates that the quantity of movement was more favored by high-performing participants. If this bar is red and to the left side, this movement level was favored by low-performing participants. Because of this, viewing the distance between these darker shaded bars efficiently identifies the discrepancy in performance between a high and low quantity of that style of movement.

Of all permutations of room and movement categorization, the most visibly drastic differences in performance were all within the conference room and were related to general eye movement, eye movement when optimal facades were visible, and rotations of the head when optimal facades were visible. These relationships were further evaluated using a set of chi-square tests of independence, and the strengths of significant relationships are reported as Cramer's V. Within the conference room, it was found that the relationship between the Gaze Classification metric (whether the participant used a relatively high amount of eye movement) and participant performance (whether the participant selected an optimal facade) was significant, X^2 (1, N = 32) = 4.500, p = 0.034, V = 0.375 (medium association). Interestingly, the group of participants who exhibited high levels of general eye movement also exhibited high levels of eye movement when active facades were visible, therefore the reported chi-square results of the relationship between the Optimal Gaze Classification metric (whether the participant used a relatively high amount of eye movement when an optimal facade was currently visible) and participant performance is the same. The next visibly large gap is also in the conference room with the Optimal Rotation Classification metric, which indicated the quantity of head rotations when optimal facades were currently visible. However, the relationship between this value and performance was not significant, X^2 (1, N = 32) = 2.000, p = 0.157.

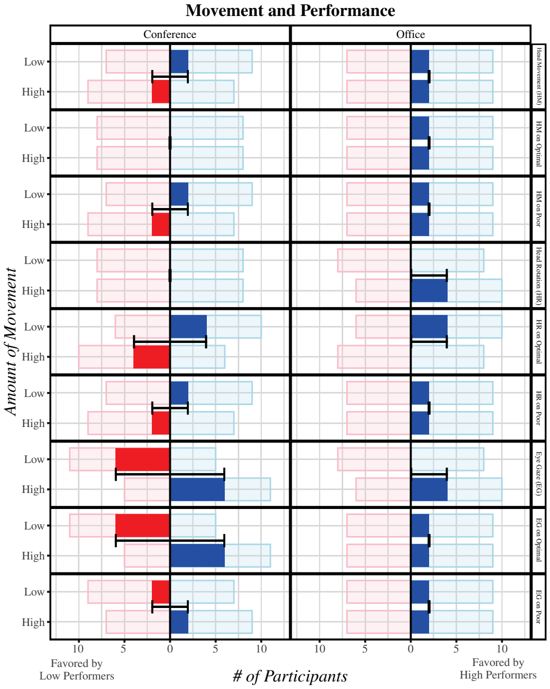
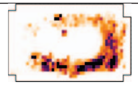
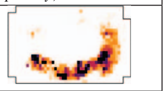
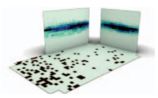
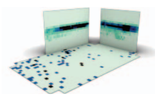
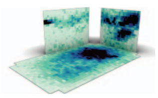
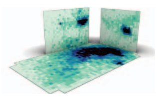
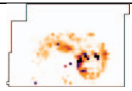
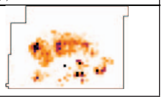
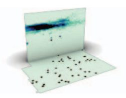
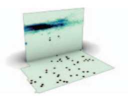
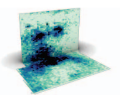
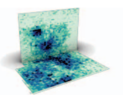


Figure 6: The total quantity of participants that performed at a high level (light blue bars on the right) or low level (light red bars on the left) in each movement categorization. The darker shaded bars indicate the imbalance of those two values for each movement quantity. Distance between the endpoints of these darker shaded bars indicates a larger discrepancy in how that type of movement was favored between high-performers and low-performers. The greatest differences were related to eye movement in the larger conference room.

Table 2 shows congregated movement data for AR participant groups. Areas, where those participant types most frequently stood, viewed, or gazed at, are presented as darker grid cells. The surfaces are normalized, meaning that one surface having fewer dark areas indicates that that surface had more concentrated areas. Head position shows a top-down view of where the participants were standing on the floor. Head rotation shows intersections of the head's forward vector and the floor and windows. Eye gaze shows the same but for the forward vector of the eyes. For both rooms, it appears that high-performing participants had less concentrated eye movement when looking at the ground, but more concentrated when looking at the windows.

A potential link between eye movement and decision-making warranted further investigation into whether eye movement influenced how individual factors were prioritized. A set of unpaired two-tailed Mann-Whitney U tests were performed to compare the final selected facade rankings of each of these factors between high and lower users of eye movement in the conference room. A significant difference in Daylighting ranking was found between high (n = 16, Mdn = 3.500, SD = 3.255) and low (n = 16, Mdn = 9.500, SD = 3.229), U =

Table 2: Congregated movement data for AR participants in both rooms separated by participant performance classification. Darker shaded grid cells indicate a heavier concentration of movement/forward vector intersections at that point.

	Low Performing Participants	**High Performing Participants**
Conference Room (High Complexity)		
Head Position		
Head Rotation Intersections		
Eye Gaze Intersections		
Office (Low Complexity)		
Head Position		
Head Rotation Intersections		
Eye Gaze Intersections		

61, $p = 0.0116$, 95% CI [-7.000, -1.000], $r = -0.449$ (medium effect). No significant difference was found between high ($n = 16$, $Mdn = 5.500$, $SD = 3.082$) and low ($n = 16$, $Mdn = 7.500$, $SD = 4.106$) in Brightness Discomfort, $U = 91$, $p = 0.166$, 95% CI [-5.000, 1.000], $r = -0.247$ (small effect). View Factor also exhibited no significant difference between high ($n = 16$, $Mdn = 11.000$, $SD = 3.324$) and low ($n = 16$, $Mdn = 10.000$, $SD = 3.631$), $U = 124$, $p = 0.894$, 95% CI [-2.000, 2.000], $r = -0.027$ (negligible effect).

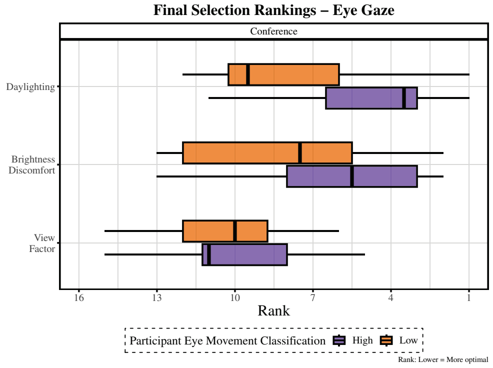
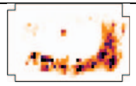
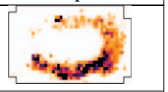
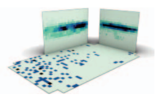
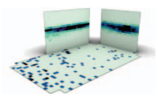
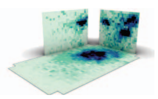
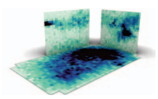


Figure 7: Distributions of individual facade factor rankings for each movement for both high and low users of eye movement in the conference room. Values to the right of the dashed vertical line indicate comparatively optimal selections.

This emphasis on eye movement spurred an investigation into its associations with other forms of movement (Table 3). Eye movement may not be the sole contributor, but instead, it may further engage other manners of movement in the body as a whole. It appears that those who use high eye movement tend to move around the room more, particularly around the oval-shaped table in the center of the room. Those who use high levels of eye movement also appear to rotate their heads more laterally rather than up-down. Two unpaired two-tailed Mann-Whitney U tests were computed to compare AR participants in the conference room's Gaze Classification with Movement Rating and then with Rotation Rating. If links were uncovered, it would indicate that these participants used these forms of movement in conjunction with each other. A significant difference in Movement Rating was found between those who exhibited high levels ($n = 16$, $Mdn = 35.538$, $SD = 9.170$) and low levels of eye movement ($n = 16$, $Mdn = 25.886$, $SD = 10.762$), $U = 191$, $p = 0.018$, 95% CI [1.205, 17.096], $r = 0.419$ (medium effect). A significant difference in Rotation Rating was also found between those who exhibited high levels ($n = 16$, $Mdn = 0.473$, $SD = 0.158$) and low levels of eye movement ($n = 16$, $Mdn = 0.290$, $SD = 0.199$), $U = 187$, $p = 0.027$, 95% CI [0.020, 0.289], $r = 0.392$ (medium effect).

Table 3: Congregated movement data for AR participants in the conference room separated by participant eye movement classification.

	Low Eye Movement Participants	**High Eye Movement Participants**
Head Position		
Head Rotation Intersections		
Eye Gaze Intersections		

4.3 Cognitive Impacts

Established assessment tools answered as part of the post-study questionnaire provided insight into the participants' experience using the system and the cognitive loads the design task imposed. Understanding the differences in responses among our participant groups aids in the evaluation of the *FRED* system and informs future retrofitting design processes. Scores from the NASA TLX and SUS scales were normalized to a scale of 0 (poor) to 100 (ideal).

4.3.1 Desktop vs. AR

The results of these usability metrics are reported in Figure 8, which represents more frequently reported values as having a wider width. Many of the reported distributions are quite homogeneous, although Physical Demand Score and SUS Score appear to more noticeably vary across the utilized systems. To more closely analyze these differences, a set of unpaired two-tailed Mann-Whitney U tests were performed on AR and non-AR participants in each study room. In the conference room, there was a significant difference between the self-reported SUS Scores of AR ($n = 32$, $Mdn = 72.500$, $SD = 11.824$) and desktop ($n = 32$, $Mdn = 82.500$, $SD = 11.548$) participants, $U = 283$, $p = 0.002$, 95% CI [-15.000, 5.000], $r = -0.386$ (medium effect). A significant difference in Physical Demand Scores required to complete the task was also found between AR ($n = 32$, $Mdn = 85.000$, $SD = 22.070$) and desktop participants ($n = 32$, $Mdn = 100.000$, $SD = 5.921$), $U = 213$, $p < 0.001$, 95% CI [-20.000, -5.000], $r = -0.521$ (large effect). In the office room, similar results followed. A significant difference in SUS Score was found between AR ($n = 32$, $Mdn = 72.500$, $SD = 11.650$) and desktop ($n = 32$, $Mdn = 82.500$, $SD = 15.264$) participants, $U = 241.5$, $p < 0.001$, 95% CI [-15.000, -5.000], $r = -0.455$ (medium effect). The difference between Physical

Demand Scores was also found to be significantly different between AR (n = 32, Mdn = 85.000, SD = 21.116) and desktop (n = 32, Mdn = 100.00, SD = 10.933) participants, U = 233, $p < 0.001$, 95% CI [-20.000, -5.00], r = -0.488 (medium effect). For both of these rooms, desktop participants were universally the ones who indicated that their process was less physically demanding and their system was more usable. The remaining relationships were not found to be significant (for more information, see Supplemental Materials).

Because data analysis indicated a gap between perceived system usability and that of AR and desktop participants in both rooms, further investigation of the links between a participant's self-reported level of previous AR experience (n = 64, Mdn = 2.000, SD = 1.662, 1 = very inexperienced, 7 = very experienced) and SUS Score (n = 64, Mdn = 72.500, SD = 11.680) was warranted. Spearman's rank correlation was calculated to determine if some level of correlation existed between these two metrics in AR participants, however, no such significant relationship was found, r(62) = -0.030, p = 0.812.

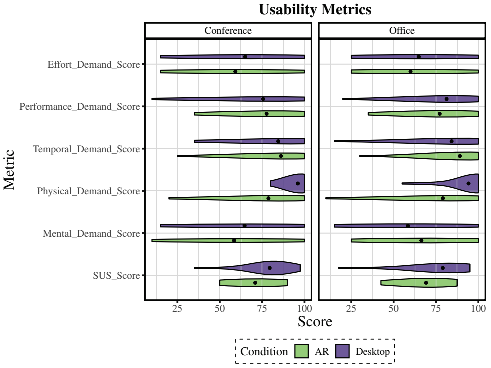


Figure 8: Distributions of the post-questionnaire subjective metrics. The black dot indicates median. Scores are normalized to a 0 (poor) to 100 (ideal) scale.

4.3.2 Movement within AR

These differences in usability and physical demands indicate that the designed *FRED* system may have been ineffective in making the retrofitting process less taxing on the designer. However, investigating differences in reporting within the AR groups may highlight how to take full advantage of its benefits (Figure 9). The reported SUS Scores and Physical Demand Scores between participants who were categorized as high and low users of the three discussed movement types, head movement, head rotations, and eye movement were investigated using a set of unpaired two-tailed Mann Whitney U tests in each room. Of these tested values, only significant links between Gaze Classification (i.e., eye movement) and Physical Demand Score within the office were uncovered. Participants who utilized high levels (n = 16, Mdn = 90.000, SD = 16.073) of eye movement reported significantly higher Physical Demand Scores than those who utilized lower levels (n = 16, Mdn = 77.500, SD = 23.514), U = 184.5, p = 0.033, 95% CI [0.000, 2.500], r = 0.380 (medium effect). This suggests that utilizing more eye movement in smaller design environments may be key in reducing physical load. Uncovered links between eye movement and performance within the conference room also necessitated interest in the relationship between Mental Demand Score and Gaze Classification, however the difference between high (n = 16, Mdn = 45.000, SD = 25.356) and low (n = 16, Mdn = 72.500, SD = 24.732) users of eye movement was not found to be significant, U = 88.5, p = 0.139, 95% CI [-40.000, 5.000], r = -0.265 (small effect). The remaining relationships were not found to be significant (for more information, see Supplemental Materials).

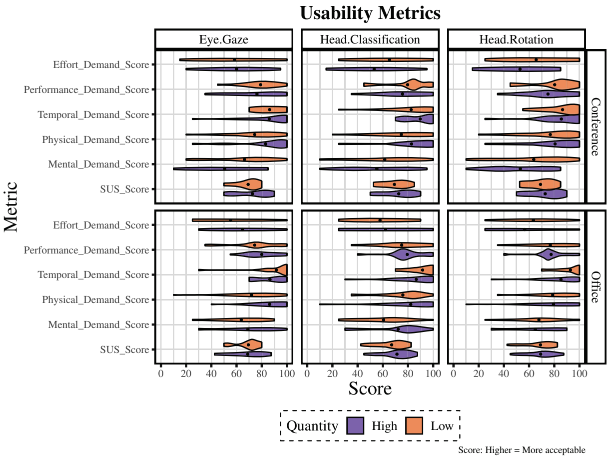


Figure 9: Distributions of the post-questionnaire subjective metrics for each movement type in AR. The black dot indicates median. Scores are normalized to a 0 (poor) to 100 (ideal) scale.

5 DISCUSSION

Our discussion will be split into three sections, one for each research question and then a short discussion of our limitations and future related work.

5.1 RQ1 - Embodied Movement Benefits

Looking at the objective values of the participant's design decisions, statistical testing uncovered no significant difference in final performance between those assisted by the BIM-AR system and those assisted by the desktop system in either room, nor was any non-small effect size reported between study condition and the final ranking. While at first review, this may appear as if incorporating AR did little to improve things, these findings should not be seen as discouraging for work in this area, given the technology's scope for improvement. Research on how to best leverage these systems in this context is relatively early, as is the available hardware's actual rendering capabilities and technical specifications. The headset that *FRED* was designed to run on, the HoloLens 2, provides a 52-degree Field of View (FOV), considerably less than the typical human FOV of 200 degrees [4]. The desktop system provides the Unreal Engine default 90-degree horizontal FOV (108 degrees diagonal at a 16:9 aspect ratio) on a 1920 x 1080 monitor, which was selected as it is what most simulation and gaming first-person environments utilize. Those in the smaller office were the only group who demonstrated significant improvement between their initial and final selections. If participants were situated in a corner of this room, they were almost able to fully access all embedded data, which was not possible in the larger conference room. Sacrifices to the AR system's FOV are necessary to achieve sufficient graphical fidelity, and it is likely that improvements to hardware will minimize these costs. Whether this translates to further improvements will necessitate future study replication. For now, it is likely that these systems are best suited for smaller environments to maximize simultaneous access of information.

Our results on AR's impact on the retrofitting design workflow were mixed. The desktop system was perceived as more usable to participants, which was mildly expected. One possible explanation is the relative novelty of AR for our study sample. While no significant relationship between previous AR experience and SUS Score was present in our sample, the reported levels of experience were exceptionally weighted towards the minimum (M = 2.516, Mdn

= 2.000, *SD* = 1.662), and therefore replication with more experienced groups may be of interest. Concerning system usability and design workflow, the other major discrepancy between systems was with the self-reported physical load. Within the office, there was a medium effect between the system utilized and the Physical Demand Score, while in the larger conference room, this effect was large. The stationary desktop systems outperformed the mobile AR system, and this effect seemed to scale with room size. The extent of this scaling is likely exacerbated by the previously discussed limited FOV requiring more movement to compensate. Still, AR performed admirably on this metric, and major contributors to the effect sizes are how truly non-taxing the desktop system was reported to be, with the median reported score in both rooms being the maximum (i.e., minimally difficult) value.

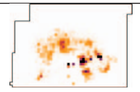
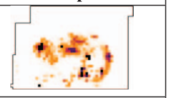
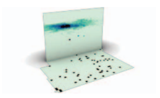
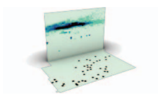
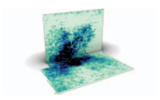
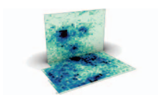
5.2 RQ2 - Cognition and Movement

Understanding how to best utilize and design AR systems to mitigate existing technical limitations can maximize the operator's potential while reducing any negatives the format introduces. It can also help inform what aspects of physical movement should be prioritized in future AR system development. Using *FRED*, our study identified comparatively high usage of eye motion as a potential area of interest in optimizing decision-making. Those who used higher levels tended to select optimal facades, and those who used lower levels tended to select non-optimal facades. However, this effect was only present in the larger conference room. This is especially interesting, as the total usage of eye movement did not significantly differ between rooms, while movements and rotations of the head did (for additional details, see Supplemental Materials). Naturally, it is expected that more frequent head rotations would be needed to compensate for limited FOV concerns in the conference room, which did appear to happen.

As shown in Table 2, those who performed well in this space tended to concentrate their gazes on the window to a specific point, whereas those who performed poorly tended to distribute their gaze to different points. These concentrated points likely provided a well-enough view of the entire window to gauge view factor/aesthetics, while the bulk of their eye movement was concentrated on the ground-shadows and the embedded solar radiation heatmap. The eye gaze surface intersections of high-performing participants on the ground appear to be less concentrated and more spread, often reaching the corners of the room, hinting at a more thorough consideration of the heatmap and solar radiation levels within the room. This view was supported by significant improvement in the final selections' average Daylighting ranking by these participants compared to their lower eye movement contemporaries.

How participants utilize movement may play a role in minimizing the physical demands that AR imposes compared to stationary desktop systems. In our data, this effect was evident with eye movement, but only within the office room. This effect may be a result of position within the room. Those with low movement congregated towards a corner of the room away from the window and focused the majority of their attention towards the opposite corner by the window (see Table 4). This contrasts with the higher users of eye movement, which appeared to have much less clear direct areas of focus as a whole. The band of head forward vector intersections appears taller in lower users of eye movement, and this band is located above the heaviest area of eye concentration. This indicates more frequent use of subtle up-and-down head movements by the participants in this area, as opposed to scanning the environment with their eyes. This may be a contributor to the higher reported physical demands by these participants, and it is clear that AR design systems should cue those who use it towards eye movements.

Table 4: Congregated movement data for AR participants in the office separated by participant eye movement classification.

	Low Eye Movement Participants	**High Eye Movement Participants**
Head Position		
Head Rotation Intersections		
Eye Gaze Intersections		

5.3 Limitations and Future Research

For this study, movement data of the virtual camera in Unreal Engine was recorded for participants using the desktop system. While analysis of movement was ultimately restricted to those within AR, analyzing desktop movement data in conjunction with eye tracking data of participants using the desktop system would provide a more in-depth comparison between the desktop and AR systems. This view extends to more extensive collections of AR movement data, such as full-body tracking, however, this analysis is only as useful at addressing the core retrofitting issue as accessible the technology would be *in situ*.

This study was primarily aimed at non-experts, and the study population reflected this. This information is useful, and the extension of this study to industry professionals is the logical next step, as it is likely that their interactions with the systems would differ. Both groups will likely play a role in future retrofitting, however, experts will likely perform the majority of work in commercial and industrial areas. Future extensions of this work should target professional groups and companies, as well as test other types of structures and different retrofitting design tasks. Additionally, our design task was relatively restricted to fit the needs of the study topic. Expansion into other design components of facades, such as economic or audio insulation impacts, as well as more free-form design or additional facade concepts, would increase study complexity and more rigorously test the systems.

6 CONCLUSION

We presented a human-subject study that evaluates how different types of body movement affect retrofit design decision-making. Our findings suggest that these BIM-enabled AR systems perform comparatively to conventional desktop-based approaches for non-expert designers, although because of current AR hardware and rendering limitations, the size and shape of the design space may heavily impact the AR system's perceived usability. That does not suggest there is no value, however. Further emphasis on embedding data within the environment and greater utilization of the AR wearer's eye movement can be effective counters to these limitations until improved hardware is widely available. This study presents a baseline comparison of these systems to standard approaches, and from this perspective, the potential for BIM-enabled AR in improving the accessibility of retrofitting is bright.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1917728. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] W. A. Abdelhameed. Creativity and VR Use. pp. 719–728. Kyoto, Japan, 2014. doi: 10.52842/conf.caadria.2014.719 2

[2] S. Alizadehsalehi, A. Hadavi, and J. C. Huang. From BIM to extended reality in AEC industry. *Automation in Construction*, 116:103254, Aug. 2020. doi: 10.1016/j.autcon.2020.103254 2

[3] ANSI/ASHRAE/IES. ANSI/ASHRAE/IES Standard 90.1-2019 – Energy Standard for Buildings Except Low-Rise Residential Buildings, 2020. Published: https://www.ashrae.org/technical-resources/bookstore/standard-90-1. 4

[4] K. W. Arthur. *Effects of field of view on performance with head-mounted displays*. PhD thesis, The University of North Carolina, Chapel Hill, 2000. 8

[5] S. Azhar. Building Information Modeling (BIM): Trends, Benefits, Risks, and Challenges for the AEC Industry. *Leadership and Management in Engineering*, 11(3):241–252, July 2011. doi: 10.1061/(ASCE)LM.1943-5630.0000127 2

[6] A. Borisuit, F. Linhart, J.-L. Scartezzini, and M. Münch. Effects of realistic office daylighting and electric lighting conditions on visual comfort, alertness and mood. *Lighting Research & Technology*, 47(2):192–209, Apr. 2015. doi: 10.1177/1477153514531518 2

[7] D. A. Bowman and R. P. McMahan. Virtual Reality: How Much Immersion Is Enough? *Computer*, 40(7):36–43, July 2007. doi: 10.1109/MC.2007.257 1

[8] J. Brooke. SUS - A quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996. Publisher: London, England. 5

[9] D. Bryde, M. Broquetas, and J. M. Volm. The project benefits of Building Information Modelling (BIM). *International Journal of Project Management*, 31(7):971–980, Oct. 2013. doi: 10.1016/j.ijproman.2012.12.001 2

[10] G. Carmichael, R. Biddle, and D. Mould. Understanding the Power of Augmented Reality for Learning. In *Proceedings of E-Learn 2012–World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 1*, pp. 1761–1771. Association for the Advancement of Computing in Education (AACE), Montréal, Quebec, Canada, 2012. 2

[11] F. Castronovo, D. Nikolic, S. Mastrolembo Ventura, V. Shroff, A. Nguyen, N. Dinh, S. Yilmaz, R. Akhavian, and C. Gaedicke. Design and Development of a Virtual Reality Educational Game for Architectural and Construction Reviews. In *2019 ASEE Annual Conference & Exposition Proceedings*, p. 32592. ASEE Conferences, Tampa, Florida, June 2019. doi: 10.18260/1-2--32592 2

[12] D. Ceglia, M. C. S. d. Abreu, and J. C. L. Da Silva Filho. Critical elements for eco-retrofitting a conventional industrial park: Social barriers to be overcome. *Journal of Environmental Management*, 187:375–383, Feb. 2017. doi: 10.1016/j.jenvman.2016.10.064 1

[13] Y. Censor. Pareto optimality in multiobjective problems. *Applied Mathematics & Optimization*, 4(1):41–59, Mar. 1977. doi: 10.1007/BF01442131 4

[14] A. Clark and D. Chalmers. The Extended Mind. *Analysis*, 58(1):7–19, 1998. Publisher: JSTOR. 2

[15] J. Cook, M. Brown, M. Sellwood, C. Campbell, P. Kouppas, and P. Poronnik. XR game development as a tool for authentic, experiential, and collaborative learning. *Biochemistry and Molecular Biology Education*, 49(6):846–847, Nov. 2021. doi: 10.1002/bmb.21572 2

[16] T. Crosbie, N. Dawood, and S. Dawood. Improving the energy performance of the built environment: The potential of virtual collaborative life cycle tools. *Automation in Construction*, 20(2):205–216, Mar. 2011. doi: 10.1016/j.autcon.2010.09.018 2

[17] M. R. de Freitas and R. C. Ruschel. Augmented reality supporting building assesment in terms of retrofit detection. In *Proceedings of the CIB W078 International Conference, Beijing, China*, pp. 9–12, 2013. 3

[18] L. Dias Pereira, V. Tavares, and N. Soares. Up-To-Date Challenges for the Conservation, Rehabilitation and Energy Retrofitting of Higher Education Cultural Heritage Buildings. *Sustainability*, 13(4):2061, Feb. 2021. doi: 10.3390/su13042061 3

[19] A. Eltaweel and Y. Su. Parametric design and daylighting: A literature review. *Renewable and Sustainable Energy Reviews*, 73:1086–1103, June 2017. doi: 10.1016/j.rser.2017.02.011 2

[20] Environmental Protection Agency. Advancing Sustainable Materials Management: 2018 Fact Sheet. Technical report, United States Environmental Protection Agency, 2020. 1

[21] D. Erbe, L. L. Sciarra, J. Glazer, K. Seeger, J. Boldt, R. Hart, and M. I. Rosenberg. A Conversation on Standard 90.1-2019. *ASHRAE Journal*, 62(PNNL-SA-152341), 2020. Publisher: Pacific Northwest National Lab.(PNNL), Richland, WA (United States). 4

[22] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, 108(1-2):52–73, Oct. 2007. doi: 10.1016/j.cviu.2006.10.012 2

[23] H. Fuchs, M. A. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J. R. Crawford, P. Rademacher, S. H. Drake, and A. A. Meyer. Augmented reality visualization for laparoscopic surgery. In W. M. Wells, A. Colchester, and S. Delp, eds., *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*, vol. 1496, pp. 934–943. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998. Series Title: Lecture Notes in Computer Science. doi: 10.1007/BFb0056282 2

[24] T. Fukuda, K. Yokoi, N. Yabuki, and A. Motamedi. An indoor thermal environment design system for renovation using augmented reality. *Journal of Computational Design and Engineering*, 6(2):179–188, Apr. 2019. doi: 10.1016/j.jcde.2018.05.007 3

[25] J. J. Gibson. The theory of affordances. *Hilldale, USA*, 1(2):67–82, 1977. 1

[26] N. Gu and K. London. Understanding and facilitating BIM adoption in the AEC industry. *Automation in Construction*, 19(8):988–999, Dec. 2010. doi: 10.1016/j.autcon.2010.09.002 2

[27] S. G. Hart. NASA Task Load Index (TLX), 1986. 5

[28] S. Henderson and S. Feiner. Exploring the Benefits of Augmented Reality Documentation for Maintenance and Repair. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1355–1368, Oct. 2011. doi: 10.1109/TVCG.2010.245 2

[29] M. Hosokawa, T. Fukuda, N. Yabuki, T. Michikawa, and A. Motamedi. Integrating CFD and VR for indoor thermal environment design feedback. In *Proceedings of the 21st International Conference of the Association for Computer-Aided Architectural Design Research in Asia CAADRIA 2016*, pp. 663–672. Hong Kong, 2016. Publisher: CUMINCAD. doi: 10.52842/conf.caadria.2016.663 2

[30] Y. Huang, S. Shakya, and L. Zhu. Capabilities of Mixed Reality Applications for Architecture and Construction: A Comparative Review with HoloLens. In *Computing in Civil Engineering 2019*, pp. 162–168. American Society of Civil Engineers, Atlanta, Georgia, June 2019. doi: 10.1061/9780784482421.021 2

[31] E. Hutchins. *Cognition in the Wild*. Number 1995. MIT press, 1995. 2

[32] A. Jafari and V. Valentin. Decision-making life-cycle cost analysis model for energy-efficient housing retrofits. *International Journal of Sustainable Building Technology and Urban Development*, 6(3):173–187, July 2015. doi: 10.1080/2093761X.2015.1074948 1

[33] L. Joblot, T. Paviot, D. Deneux, and S. Lamouri. Literature review of Building Information Modeling (BIM) intended for the purpose of renovation projects. *IFAC-PapersOnLine*, 50(1):10518–10525, July 2017. doi: 10.1016/j.ifacol.2017.08.1298 3

[34] C. Kasinalis, R. Loonen, D. Cóstola, and J. Hensen. Framework for assessing the performance potential of seasonally adaptable facades using multi-objective optimization. *Energy and Buildings*, 79:106–113, Aug. 2014. doi: 10.1016/j.enbuild.2014.04.045 2

[35] I. A. Khalek, J. M. Chalhoub, and S. K. Ayer. Augmented Reality for Identifying Maintainability Concerns during Design. *Advances in Civil Engineering*, 2019:1–12, Mar. 2019. doi: 10.1155/2019/8547928 2

[36] D. Kirsh and P. Maglio. On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, 18(4):513–549, Oct. 1994. doi: 10.1207/s15516709cog1804_1 2

[37] Kumar, Patil, Kang, and Chai. Sensor Fusion Based Pipeline Inspection for the Augmented Reality System. *Symmetry*, 11(10):1325, Oct. 2019. doi: 10.3390/sym11101325 3

[38] S. Lee, W. Shin, and E. J. Park. Implications of neuroarchitecture for the experience of the built environment: a scoping review. *Archnet-IJAR: International Journal of Architectural Research*, 16(2):225–244, June 2022. doi: 10.1108/ARCH-09-2021-0249 2

[39] J. Li, J. Kim, and X. A. Chen. Robiot: A Design Tool for Actuating

Everyday Objects with Automatically Generated 3D Printable Mechanisms. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 673–685. ACM, New Orleans LA USA, Oct. 2019. doi: 10.1145/3332165.3347894 2

[40] Y. Liu, J. Lather, and J. Messner. Virtual Reality to Support the Integrated Design Process: A Retrofit Case Study. In *Computing in Civil and Building Engineering (2014)*, pp. 801–808. American Society of Civil Engineers, Orlando, Florida, United States, June 2014. doi: 10.1061/9780784413616.100 3

[41] Y. Liu, S. van Nederveen, and M. Hertogh. Understanding effects of BIM on collaborative design and construction: An empirical study in China. *International Journal of Project Management*, 35(4):686–698, May 2017. doi: 10.1016/j.ijproman.2016.06.007 2

[42] Z. Liu, F. Zhang, and Z. Cheng. BuildingSketch: Freehand Mid-Air Sketching for Building Modeling. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 329–338. IEEE, Bari, Italy, Oct. 2021. doi: 10.1109/ISMAR52148.2021.00049 2

[43] H. F. Mallgrave. *Architecture and Embodiment The Implications of the New Sciences and Humanities for Design*. Routledge, 1st ed., 2013. 2

[44] S. K. Meeks, D. J. Brown, J. Lindberg, M. Powe, M. O'Neal, C. Hartmann, S. Magill, and C. Crothers. Reuse and Revitalization in Jacksonville: Discovering the value of older buildings and blocks. Technical report, Preservation Green Lab, May 2017. 1

[45] S. K. Meeks, D. J. Brown, T. Richey, M. Huppert, J. Lindberg, and M. Powe. Older, Smaller, Better: Measuring how the character of buildings and blocks influences urban vitality. Technical report, Preservation Green Lab, May 2014. 1

[46] E. K. Meyer. Beyond "Sustaining Beauty": Musings on a Manifesto. In *Values in Landscape Architecture and Environmental Design: Finding Center in Theory and Practice*, pp. 30–53. Louisiana State University Press, Baton Rouge, LA, 2015. Publisher: Baton Rouge, LA: LSU Press. 1

[47] D. Mitterberger, K. Dörfler, T. Sandy, F. Salveridou, M. Hutter, F. Gramazio, and M. Kohler. Augmented bricklaying: Human–machine interaction for in situ assembly of complex brickwork using object-aware augmented reality. *Construction Robotics*, 4(3-4):151–161, Dec. 2020. doi: 10.1007/s41693-020-00035-8 2

[48] A. S. Montoya-Olsson. *Parametric Modeling and Optimization of External Braced Frames for Embodied Energy and Daylight Availability*. PhD thesis, 2020. 2

[49] W. Natephra, A. Motamedi, T. Fukuda, and N. Yabuki. Integrating building information modeling and virtual reality development engines for building indoor lighting design. *Visualization in Engineering*, 5(1):19, Oct. 2017. doi: 10.1186/s40327-017-0058-x 2

[50] M. J. Nathan, K. E. Schenck, R. Vinsonhaler, J. E. Michaelis, M. I. Swart, and C. Walkington. Embodied geometric reasoning: Dynamic gestures during intuition, insight, and proof. *Journal of Educational Psychology*, 113(5):929, 2021. Publisher: American Psychological Association. doi: 10.1037/edu0000638 2

[51] A. Newen, S. Gallagher, and L. De Bruin. 4E Cognition: Historical Roots, Key Concepts, and Central Issues. In A. Newen, L. De Bruin, and S. Gallagher, eds., *The Oxford Handbook of 4E Cognition*, pp. 2–16. Oxford University Press, Sept. 2018. doi: 10.1093/oxfordhb/9780198735410.013.1 2

[52] A. K. Patil, G. A. Kumar, T.-H. Kim, and Y. H. Chai. Hybrid approach for alignment of a pre-processed three-dimensional point cloud, video, and CAD model using partial point cloud in retrofitting applications. *International Journal of Distributed Sensor Networks*, 14(3):155014771876645, Mar. 2018. doi: 10.1177/1550147718766452 3

[53] H. Peng, J. Briggs, C.-Y. Wang, K. Guo, J. Kider, S. Mueller, P. Baudisch, and F. Guimbretière. RoMA: Interactive Fabrication with Augmented Reality and a Robotic 3D Printer. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, Montreal QC Canada, Apr. 2018. doi: 10.1145/3173574.3174153 2

[54] C. F. Reinhart, J. Mardaljevic, and Z. Rogers. Dynamic Daylight Performance Metrics for Sustainable Building Design. *LEUKOS*, 3(1):7–31, July 2006. doi: 10.1582/LEUKOS.2006.03.01.001 4

[55] S. Robinson and J. Pallasmaa, eds. *Mind in Architecture: Neuroscience, Embodiment, and the Future of Design*. The MIT Press, 2015. doi: 10.7551/mitpress/10318.001.0001 2

[56] M. S. Roudsari, M. Pak, A. Smith, and others. Ladybug: A Parametric Environmental Plugin For Grasshopper To Help Designers Create An Environmentally-conscious Design. In *Proceedings of the 13th international IBPSA conference held in Lyon, France Aug*, pp. 3128–3135. Chambery, France, 2013. doi: 10.26868/25222708.2013.2499 3

[57] K. Sauvé, D. Potts, J. Alexander, and S. Houben. A Change of Perspective: How User Orientation Influences the Perception of Physicalizations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, Honolulu HI USA, Apr. 2020. doi: 10.1145/3313831.3376312 2

[58] J. Sermarini, J. J. LaViola, R. Azevedo, L. C. Walters, R. Michlowitz, and J. T. Kider. Embodied facade retrofit design with Augmented Reality in the built environment. In *Proceedings of the 17th IBPSA Conferences*, pp. 2844–2851. IBPSA, Bruges, Belgium, 2021. doi: 10.26868/25222708.2021.30984 2

[59] J. Sermarini, R. A. Michlowitz, J. J. LaViola, L. C. Walters, R. Azevedo, and J. T. Kider Jr. BIM Driven Retrofitting Design Evaluation of Building Facades. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, pp. 1–10. ACM, Online CA USA, Dec. 2022. doi: 10.1145/3565970.3567688 2

[60] B. E. Shelton and N. R. Hedley. Exploring a Cognitive Basis for Learning Spatial Relationships with Augmented Reality. *Technology, Instruction, Cognition and Learning*, 1(4):323–357, 2003. Publisher: Old City Publishing. 2

[61] A. Shemesh, R. Talmon, O. Karp, I. Amir, M. Bar, and Y. J. Grobman. Affective response to architecture – investigating human reaction to spaces with different geometry. *Architectural Science Review*, 60(2):116–125, Mar. 2017. doi: 10.1080/00038628.2016.1266597 2

[62] D. Siegele, P. Penna, and M. Riedl. Visualizing Building Energy Measurement Data in Mixed Reality Applying BIM. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pp. 247–253. Springer, 2021. 2

[63] M. Stel, E. v. Dijk, P. K. Smith, W. W. v. Dijk, and F. M. Djalal. Lowering the Pitch of Your Voice Makes You Feel More Powerful and Think More Abstractly. *Social Psychological and Personality Science*, 3(4):497–502, July 2012. doi: 10.1177/1948550611427610 2

[64] A. Sussman and J. B. Hollander. *Cognitive Architecture: Designing For How We Respond to the Built Environment*. Routledge, 2021. 1

[65] R. S. Ulrich, R. F. Simons, B. D. Losito, E. Fiorito, M. A. Miles, and M. Zelson. Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, 11(3):201–230, Sept. 1991. doi: 10.1016/S0272-4944(05)80184-7 2

[66] R. Volk, J. Stengel, and F. Schultmann. Building Information Modeling (BIM) for existing buildings — Literature review and future needs. *Automation in Construction*, 38:109–127, Mar. 2014. doi: 10.1016/j.autcon.2013.10.023 2

[67] J. Whyte. *Virtual Reality and the Built Environment*. Routledge, 1st ed., 2002. 2

[68] B. G. Wilson and K. M. Myers. Situated Cognition in Theoretical and Practical Context. In *Theoretical Foundations of Learning Environments*, pp. 57–88. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 2000. 2

[69] W. Yan, C. Culp, and R. Graf. Integrating BIM and gaming for real-time interactive architectural visualization. *Automation in Construction*, 20(4):446–458, July 2011. doi: 10.1016/j.autcon.2010.11.013 2

[70] X. Yang, S. Ergan, and K. Knox. Requirements of Integrated Design Teams While Evaluating Advanced Energy Retrofit Design Options in Immersive Virtual Environments. *Buildings*, 5(4):1302–1320, Dec. 2015. doi: 10.3390/buildings5041302 1

[71] R. Zaker and E. Coloma. Virtual reality-integrated workflow in BIM-enabled projects collaboration and design review: a case study. *Visualization in Engineering*, 6(1):4, Dec. 2018. doi: 10.1186/s40327-018-0065-6 2

[72] F. Zhang, Z. Xu, Y. Yang, M. Qi, and H. Zhang. Virtual reality-based evaluation of indoor earthquake safety actions for occupants. *Advanced Engineering Informatics*, 49:101351, Aug. 2021. doi: 10.1016/j.aei.2021.101351 2