Building and Environment 214 (2022) 108962

Contents lists available at ScienceDirect

Building and Environment

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

Stereoscopic Images and Virtual Reality techniques in daylighting research: A method-comparison study

Claudia Moscoso [a,*], Marzieh Nazari [b], Barbara Szybinska Matusiak [b]

[a] SINTEF Community, Department of Mobility and Economics, S.P. Andersens vei 5, 7465, Trondheim, Norway
[b] Light & Colour Centre, Department of Architecture and Technology, Faculty of Architecture and Design, Norwegian University of Science and Technology (NTNU), Alfred Getz vei 3, 7034, Trondheim, Norway

A R T I C L E I N F O

Keywords:
Stereoscopic images
Virtual reality
Method comparison
Visualization techniques
Architectural space
Daylight

A B S T R A C T

Visualization and presentation techniques are experiencing a rapid development and application in the architectural profession and beyond. As such, Stereoscopic Images (SI) and Virtual Reality (VR), both advanced visualization techniques, have found their way in daylighting research. This paper explores the correspondence between these two techniques using both quantitative and qualitative methods. The study focused on two small rooms with the same dimensions, with alternately white and black surfaces, and three different windows sizes. Seven architectural qualities were studied: Pleasantness, Calmness, Interest, Excitement, Complexity, Spaciousness, and Amount of View. The attributes were evaluated by 20 participants using a Likert-type scale. The collected data was analyzed using Wilcoxon signed-rank tests and the Bland-Altman method. The results revealed that the attributes Pleasantness, Calmness, Interest, and Complexity in rooms with both white and black surfaces, and Excitement and Spaciousness in a room with white surfaces are evaluated similarly in both VR and SI. The attributes Excitement and Spaciousness in a room with black surfaces and the Amount of View in rooms with both white and black surfaces were not evaluated similarly with both methods. In addition, qualitative results indicated that the visualization technique affects how a space is perceived. Indeed, the Virtual Reality's nature as an immersive environment provides the feeling of 'presence' which does not apply to SI. The results of this paper can help researchers working with daylighting in buildings in selecting the appropriate visualization technique to reduce experimental and logistical constraints caused by varying daylight conditions.

1. Introduction

Due to the challenging process of conducting experiments in daylighting research, in which the weather conditions are uncontrollable, visualization methods can be used. Indeed, different simulation methods have been used for lighting research throughout the years, e.g. slides [1], computer simulated images [2], rendered images [3], stereoscopic imaging [18] and Virtual Reality [55,56]). The traditional approach for research studies usually includes a full-scale laboratory environment for repeatability and reliability of data [4]. Previous studies in this area have reported that the process of making full-scale environment is time-consuming and can be very expensive [3,4]. Moreover, due to the nature of daylight, in which there are constant variations in the light level, keeping stable daylighting conditions is even more challenging. Although there can be found studies making use of artificial lighting for mimicking daylit environments, when researchers attempt to use real daylight in their laboratory studies these challenges become more evident. Similarly, controlling other sources of light that might affect a scene can be a very complicated task in lighting studies as Villa and Labayrade [5] reported. Therefore, finding other tools and methods to conduct an experiment and evaluate a scene in an easy-access way in lighting research seems important and necessary.

The growing development of image-based techniques provides various options for researchers to conduct their research. These techniques can vary from evaluating the desired environment simply on a 2D screen to non-immersive 3D methods or more advanced techniques like immersive 3D tools. Stereoscopic Images (SI) and Virtual Reality (VR) are two visualization and presentation techniques based on three-dimensional vision which have been found to be able to overcome logistical challenges proper to these experimental studies.

Three-dimensional vision by Stereoscopic Images is based on the binocular stereopsis theory by Wheatstone [6]; who observed the fact

* Corresponding author.
E-mail addresses: Claudia.moscoso@sintef.no (C. Moscoso), Marzieh.nazari@ntnu.no (M. Nazari), Barbara.matusiak@ntnu.no (B.S. Matusiak).

https://doi.org/10.1016/j.buildenv.2022.108962
Received 16 December 2021; Received in revised form 28 February 2022; Accepted 1 March 2022
Available online 3 March 2022

that each eye sees the same object with a slightly different view. A mirror stereoscope was created based on this idea, presenting images in a way that the brain perceives it as a vivid 3D object [6]. Stereoscopic images allow the user a better understanding of the visual information presented, and it usually makes use of glasses, such as passive or active 3D glasses. In addition, using a satisfactory display screen for SI can provide perception of depth, improved perception of structure in visually complex scenes, improved perception of surface curves, improved motion judgment, and enhanced perception of surface material type [7]. Alternatively, Virtual Reality, which is also based on binocular stereopsis theory, and is an immersive 3D image-based tool, is quickly becoming a popular method for research. In contrast to SI's screen and glasses, VR often uses a headset which includes two different screens, one for each eye. Both the SI and VR methods can make use of either computer renderings or photographic images to create a virtual environment. The use of either visualization technique is dependent on a study's objective. Naturally, unbuilt environments can be simulated using only renderings, while existing built environments can be replicated using both renderings and photographs. In lighting research, examples of studies using photographs for replicating environments have been used for SI [8,18]; and VR [9,10]; [57,58].

Previous studies have indicated that, compared to real environments, SI are a valid tool for lighting studies. For example, [11] used tone-mapped 3D projections of HDR images, while [18] combined two photographs mimicking a left and right view to create 3D images. Both studies compared the subjective responses of daylit spaces presented in real environments and in projected 3D images (referred to SI in Moscoso et al., [18]). Moreover, both studies found that the projection of 3D images produced similar results to the evaluations of the real environments for the subjective evaluation of daylit spaces. In addition, the validity of VR in lighting studies has also been evaluated by different studies that compared immersive virtual environments with real environments, indicating that VR may be used as a surrogate to real spaces for specific applications in lighting research [9,10,56]. Moreover, while SI has been used in lighting research e.g. for the evaluation of the aesthetic perception of offices [12], VR has been used e.g. for the subjective responses to sunlight pattern geometry [55,56]), daylight-driven interest in rendered architectural scenes [13], perceptual impressions of daylit spaces [14], impact of viewing location on view perception [15], impact of design features and occupant lighting choices [57,58]), spatiotemporal information visualization for influencing occupant's choices including lighting [16] and regional differences in the perception of daylit scenes [17].

Although SI and VR seem promising and are being applied in lighting research, both methods differ greatly in their application and the scope they present. Natural differences are e.g. the equipment needed for each method and how many people can make use of each method at the same time. Whereas one method might be useful for a specific study/researcher, the other method might not. If either of these are to be used by the lighting community for perceptual studies, it becomes important to clearly determine the difference between both methods for perceptual evaluations. Moreover, a motivation for the study is to explore the advantages and disadvantages each method presents to provide enough information for their applicability in lighting research studies. To the authors' knowledge, there is no study comparing SI and VR as empirical research tools to be used in lighting research.

This paper presents the results of an experimental procedure designed to answer the following research question (RQ): *How do perceptual attributes for simulated daylit spaces differ from using stereoscopic images (SI) and virtual reality (VR)?* To answer the RQ, this study focuses on the comparison between SI and VR as tools for evaluating daylit architectural spaces. The analysis is focused on two main objectives: First, analyzing the correspondence between VR and SI techniques in daylighting research, that is, examining if both methods give similar results. Second, finding the main differences between VR and SI techniques that may affect the human perception of a daylit architectural space by evaluating the users' subjective experience when using VR and SI. To this end, the study hypothesizes that although two methods can be used for assessing a daylit space, each one may have different effects on the subjective impressions of it. A pragmatic statistical approach, Bland-Altman, for assessing agreement between two methods was used.

2. Method

2.1. Experimental setting

The spaces used for this study was based on a previous full-scale study used to evaluate the validity of SI as a research method [18]. In order to simulate spaces lit by daylight, two factors were considered: window size and room color, based on the work by Moscoso et al. [18]. Considering that the study focused on daylit spaces, different openings allowing the entrance of daylight depicting overcast sky conditions were represented in three window sizes producing different indoor light levels. Room color was used to study large differences in light reflective surfaces. To do this, two achromatic colors with a large difference in reflectance (i.e., black and white) were used for the room surfaces. These two factors allowed to evaluate different daylighting configurations present in spaces lit by daylight and not artificial light, in which the amount of light is based on the daylight openings and the room reflectance properties.

Two rooms of equal dimensions (3.0 m × 3.5 m, 2.5 m height) with black (BR=Black Room) and white (WR=White Room) surfaces were built with wall-bricks boxes (0.50 m × 0.50 m x 0.25 m) and preconstructed wall panels (width 0.60 m) in the Room Laboratory (Romlab) at the Norwegian University of Science and Technology (NTNU) (see Fig. 1). The windows that provide daylight to the room were oriented north-east in front of an existing glass wall in the Romlab facing the same direction. The glass wall of the laboratory was covered by white diffuse curtains, and the two rooms' windows were built using metal frames.

Three different sizes of the opening resembling windows, D1 (1.0 m × 1.0 m), D2 (2.0 m × 1.0 m), and D3 (2.0 m × 1.5 m) were used in the rooms. The furniture of the room was modelled abstractly (i.e. free of intricate details) using boxes (0.50 m × 0.50 m x 0.25 m). A chair, a bedding set, a lamp, and a table were in the same coordinates related to the corners to create a 'student room' feeling. The six rooms (three white and three black) were simulated in SI and VR serving as experimental stimuli.

2.2. Generation and presentation of stimuli using virtual reality (VR)

For generating the images to be used in VR, a specific workflow was followed based on the work by Chamilothori et al. [56]. The 3D models of spaces were simulated using the software Rhinoceros and exported using the DIVA-for-Rhino 4.0 toolbar to Radiance 5.2a. The scenes were based on the aforementioned room's dimensions, openings, furniture, and daylight characteristics.

The color and specularity of the surfaces in the experimental room were measured with the Spectroradiometer PR®-650 SpectraScan®, Photo Research, and a white Reflectance Standard (RS) model: RS-3 also from Photo Research. For each surface, a small evenly illuminated central area without shadows was selected. In addition, four points (corners of an imaginative square) were marked in the area of measurement at 20 cm distance between them. The RS was put over the points and a first set of measurements were taken on the RS. After completion, the RS was removed, and a new set of measurements were taken directly on the surface. This procedure was repeated 5 times at each point to collect data for calculating the average value for each point. In this way, 20 double measurements (RS and surface) were performed for each surface. The measurements were taken indoors under overcast sky conditions. The SpectraScan was fixed to a tripod which enabled to position it in a 1.5–2.5 m distance from the

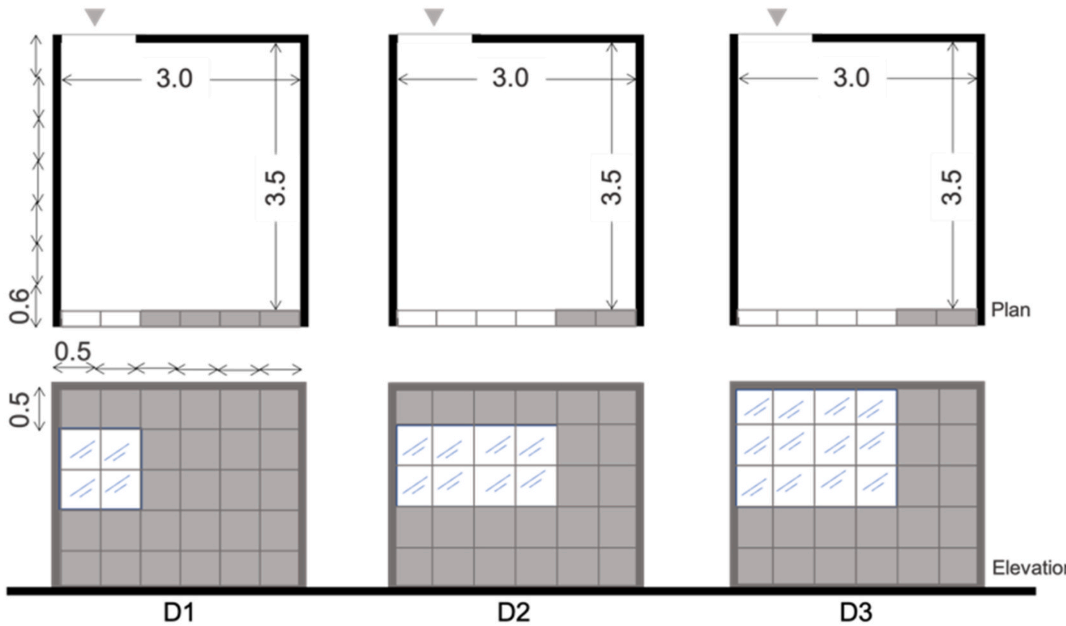

Fig. 1. Floor Plan and Schematic Elevation of the rooms with the three different opening sizes for daylight.

measurement points and to keep 45° of "incidence angle" in relation to the measured surface. To avoid shadowing the RS by the researcher's body or the equipment, the points were selected at the height of minimum 1 m over the floor. The measurements of the surfaces were then translated to Radiance material properties (see Table 1).

Two separate 360° images for each eye were rendered to create the perception of depth by projecting in VR headset. High dynamic range (HDR) scenes of 360 over-under stereo-equirectangular images were rendered using high accuracy parameters (see Table 2).

To have a uniform exposure multiplier of the HDR renderings, *pfilt* was used for *anti*-aliasing and scaling and sets the exposure to the correct average value by making two passes on the image [19]. A tone reproduction operator was selected to reliably map real-world luminance to display luminance. The tone-mapping operator *pfstmo_reinhard02* was used by *pfstools* for the White Room which is a photographic tone reproduction and generates a realistic image of the scene as described in Reinhard et al. [20]. However, the tone mapping operator by Reinhard changed drastically the contrast of the Black Room scene and did not match the perception of the scene. To overcome that challenge, the *pcond* tone-mapping operator by Larson et al. [21] was applied to the Black Room which adjust the adaptation mapping level base on the population of the luminance rather than on spatial location in the image. This algorithm preserves the visibility of black rooms scenes and consider the human visual limitations providing a realistic image for dark scenes in black rooms. To provide the application of identical settings to all the scenes, the images were converted to low dynamic range BMP files using *ra_bmp* with a gamma correction factor of 1.8 for the White Rooms and 0.5 for the Black Rooms.

Considering human anthropometric data [22,23] and height differences among the participants, four different viewpoints of an observer were used to create scenes differing in standing height (H1: up to 1.58 m, H2: 1.58–1.69 m, H3: 1.69–1.80 m, and H4 higher than 1.80 m). This was determined under the notion that when a viewpoint height is significantly different from one's height, there is a risk for a VR user to feel as 'floating above' or 'sinking in the ground'. The use of four different heights contribute to avoid such feelings and thus, providing a more correct spatial and immersive experience. These four viewpoints were based in relation to coordinate z, from which only one viewpoint was presented to each participant according to their height. The viewpoint in relation to coordinates x and y were X: 1.5, 7.5, 13.5; Y:2.9; Z: 1.427, 1.533, 1.637, 1.748 (see Fig. 2). Each viewpoint included two images, one for left-eye and another for right-eye. The mean interpupillary distance (IPD), which is the distance between the observer's left and right eyes for average adults was considered as 0.063 m according to Dodgson [24].

The position of the viewpoints (see Fig. 2) was purposedly selected to avoid large disparity distortions that could compromise the evaluation of the scenes. The viewpoints were selected based on the authors' expectation that the gaze direction of the participants might be focused towards the room window and furniture. Although the location of the viewpoints does not correspond with the other directions (e.g., when a participant turns the gaze right or left), those directions contained no

Table 1
Radiance material properties for the main surfaces.

Surface	Type	R	G	B	Reflectance %	Specularity	Roughness	T_{vis}
WR ceiling	Opaque plastic	0.894	0.979	0.965	96	.006	0	
BR ceiling	Opaque plastic	0.077	0.081	0.078	8	.005	0	
Chair	Opaque plastic	0.212	0.237	0.243	23	.007	.020	
WR floor	Opaque plastic	0.923	0.899	0.843	90	.006	.010	
BR floor	Opaque plastic	.063	.06	.048	6	.005	.01	
Furniture	Opaque plastic	.294	.307	.280	30	.004	.002	
Table Lamp	Opaque plastic	.03	.03	.03	.03	.030	0	
Textile	Opaque plastic	.091	.103	.097	10	0	.04	
WR walls	Opaque plastic	.960	.988	.924	98	.004	.002	
BR walls	Opaque plastic	.086	.088	.082	9	.003	.002	
Glazing	Glass	0.763	0.763	0.763				70%

Table 2
Radiance rpicts and rtrace parameters for view renderings.

ds	-aa	-ar	-ad	-as	-lr	-lw	-st	-dj	-ds	-dr	-dp
ambient bounces	ambient accuracy	ambient resolution	ambient divisions	ambient super-samples	Limit reflections	Limit the weight of each ray	specular sampling	direct jittering	direct sampling	number of relays	presampling density
6	0.01	32	16384	8192	8	0.000002	0.1	0.9	0.02	6	1024
Resolution: 17280 × 17280, scaled down to 4320 × 4320 using *pfilt*.											

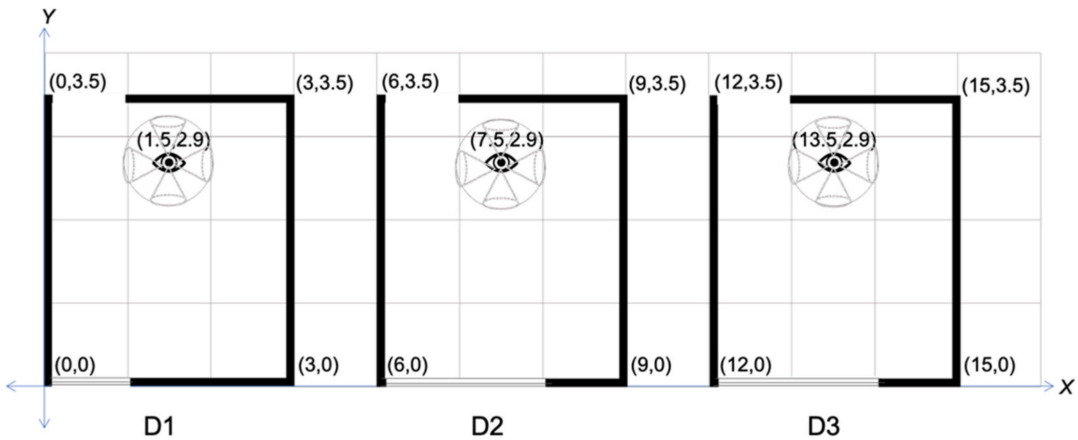

Fig. 2. Viewpoints in relation to coordinates x and y.

other elements than the room walls, granting perceptually imperceptible disparity distortions and thus, not affecting depth perception.

For the 24 simulated scenes (6 conditions - based on 3 window sizes and 2 room colors, see Fig. 3 - with 4 viewpoints differing in height), a total of 48 rendered images were generated using Idun High Performance Computers at NTNU (https://www.hpc.ntnu.no/idun) with a total rendering time of 6 days and 17 h for all 24 scenes. The equirectangular images were imported into the Unity game engine (version 2018.1.0f2) to create immersive scenes to be projected in VR. Each pair of equirectangular images was applied as textures on two spheres, each sphere simulating one eye following the principle of stereopsis. A virtual camera was located in the center of each sphere using the *OVRCameraRig* function in Unity, supporting the control of the camera using the Oculus head-tracking feature. This procedure admitted a correspondence between the viewpoint used for the generation of the renderings and the viewpoint used by the experiment participants. This allowed the participants to explore the scenes from an established viewpoint, experiencing them as three-dimensional and fully immersive. Finally, the fully immersive 360° stereoscopic scenes for VR (see Fig. 4) were presented via an Oculus Rift CV1 VR headset. The Oculus Rift CV1 VR headset uses a PenTile OLED display with a 2160 × 1200-pixel low persistence organic light emitting diode (resulting in a resolution of 1080 × 1200 pixels per eye), with a refresh rate up to 90 Hz, and a maximum luminance of 80 cd/m2.

2.3. Generation and presentation of stimuli using stereoscopic images (SI)

As pointed out previously, the stereoscopic images have been found to be a valid research tool for lighting studies by Moscoso et al. [18]. However, that particular study used photographs of real environments to be projected and presented. A pilot study from the authors of this study showed that the difference in representation between photographs in SI and rendered images in VR affected the subjective impressions of the participants. Aiming to reduce the potential effect of image type on the study's outcomes, a set of rendered images was selected to be used in the SI method instead of photographs. In this way, it was ensured that only the visualization technique (SI vs VR) and not the image type was evaluated.

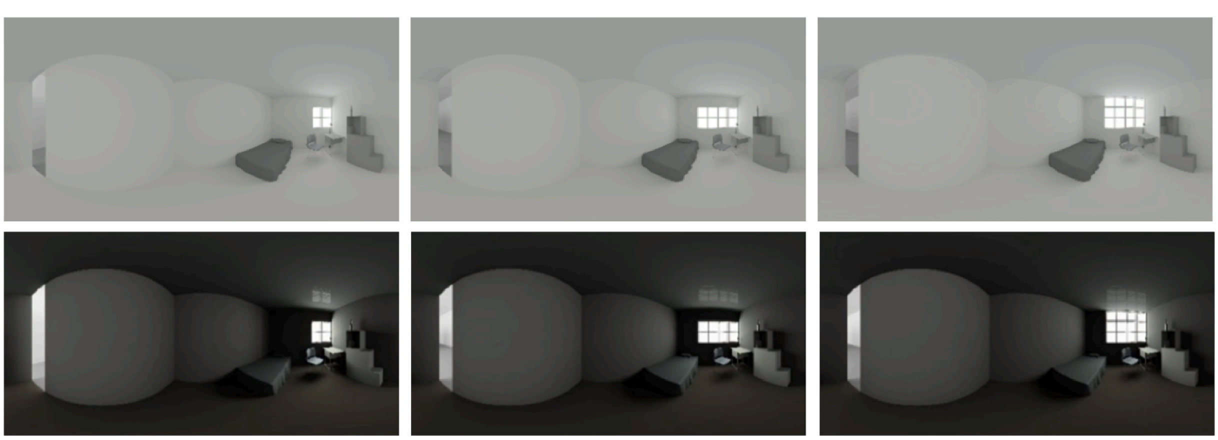

Fig. 3. *Equirectangular images of the white and black room with the three window sizes, used for the fully immersive 360° stereoscopic scenes presented using Virtual Reality.*

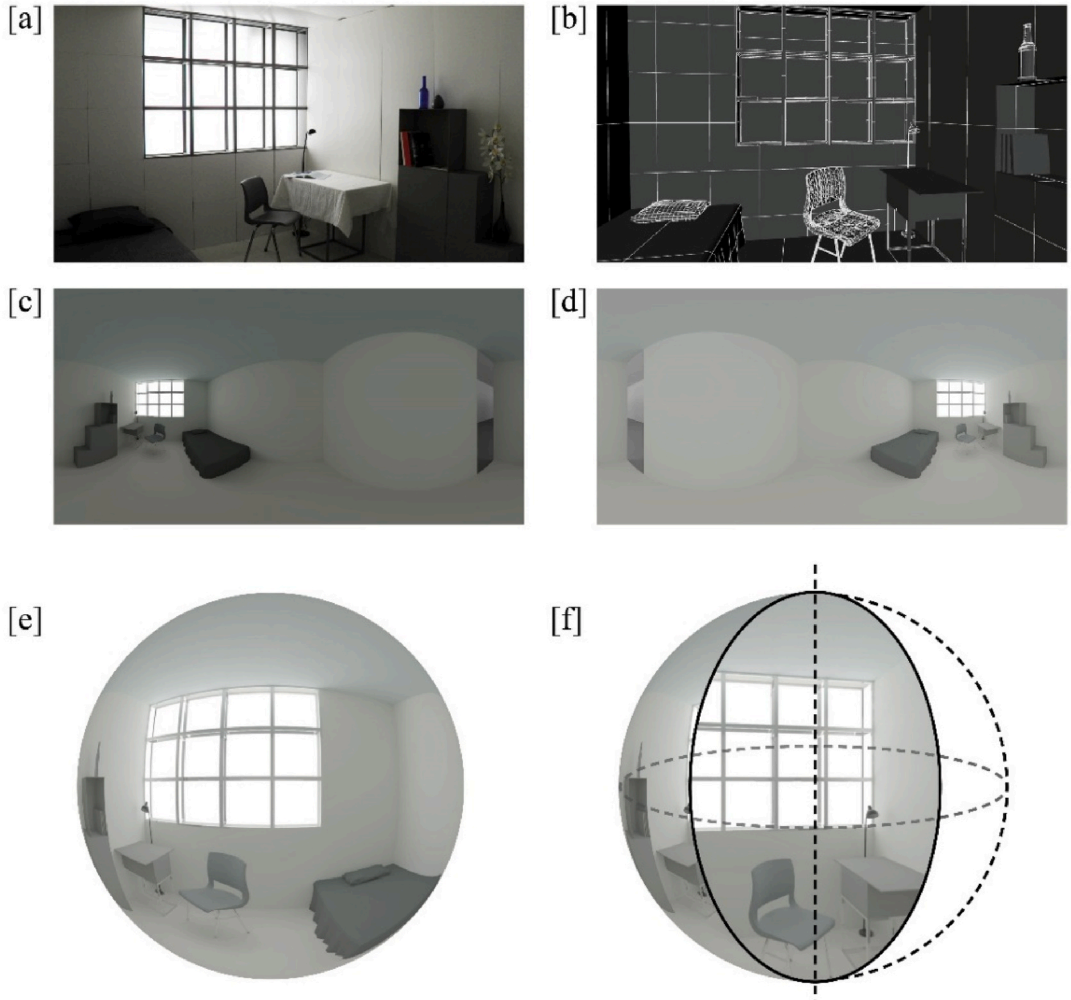

Fig. 4. *Illustration of the procedure for the generation of images to be used in VR. [a] Photograph of the real white room, [b] modelled room in Rhinoceros, [c] rendered equirectangular 360° HDR image, [d] tone-mapped equirectangular 360° image, [e] sphere with equirectangular image used as texture in Unity, and [f] representation of the degrees of freedom from the participants' viewpoint.*

Fig. 5. The renders for SI from the simulated model for SI presentation on the silver screen.

To this end, a series of high dynamic range (HDR) renders showing perspective images of the rooms by using *rpict* in Radiance were generated. The view of an observer at 1.65 m eye height was used, following statistical data indicating the average height for men aged 20 to 49 in Norway of 1.80 m [25]. Two rendered images were used, one for a “left-eye view’ and a second one for a “right-eye view” with an IPD of 0.063 m.

A similar workflow as the one described for the Virtual Reality images were also applied to these series of renders. The viewpoint in relation to coordinates x and y were X: 1.5, 7.5, 13.5; Y:2.9; Z: 1.65 (see Fig. 2). The material properties and render parameters for *rpict* were exactly the same as for VR (see Tables 1 and 2).

Based on the work by Moscoso et al. [18], the renders for SI (see Fig. 5) were projected superimposed on an Antipolarized Silver Screen using two Full High-Definition Projectors of more than 5000 ANSI Lumen, each through different circular polarized filters and by using Stereoscopic Player software on a PC. The participants assessed the 3D images on the screen from a distance no larger than 3 m, wearing circular polarized glasses during the experiment.

2.4. Dependent variables

The dependent variables were used based on previous studies [17]. Considering that an objective of the study was to evaluate how both methods perform for assessing spatial perception, seven perceptual attributes were selected to study. The selection was based on the Circumplex model of affect found in Russell’s emotion theory [26], as adapted from Boubekri et al. [27] for lighting research. Four attributes were then selected: *Pleasantness, Calmness, Interest*, and *Excitement*.

In addition, *Complexity*, based on the Kaplan’s framework of predictors of preference [28], which directly refers to the information richness of a scene was also selected. Although the visual complexity of the scenes might be considered low, this study argues that the lighting conditions associated with both the window size and the room reflectances might affect the amount of visual information received (e.g. room details) and thus the evaluation of *Complexity*. Moreover, considering that both methods differ in the amount of visual information provided to the participants (limited to one view with SI and a 360° view with VR), the evaluation of this particular attribute is thus highly relevant for the purpose of the study. Due to the focus of the methods for their use in lighting and architectural research, two other attributes strongly connected to the perception of a space were included: *Spaciousness* and *Amount of View*. Indeed, Spaciousness is undeniably related to how a person perceives the space and has been found to be affected by both window size and room color [29]. Amount of View has also been found to affect the perceived pleasantness of a space. As stated in Section 2.1, as white diffuse curtains were used in the original study [29], these were also modelled to be used in the VR and SI environments. Due to the use of these white diffuse curtains, the access to view out was restricted not allowing to register urban or natural views throughout the windows. Thus, allowing the focus of studying *Amount of View* (directly associated with the window size) rather than quality of view (associated with the visual content outside the windows).

Table 3
Dependent variables and their respective questionnaire items.

Attribute	Questionnaire item
Pleasantness	*How pleasant is this space?*
Calmness	*How calming is this space?*
Interest	*How interesting is this space?*
Excitement	*How exciting is this space?*
Complexity	*How complex is this space?*
Spaciousness	*How spacious is this space?*
Amount of View	*How satisfied are you with the amount of view?*

*Each question was complemented with the rating scale range, i.e. “*How [x] is this space, on a scale from 1 to 10, where 1 is not at all [x] and 10 is very [x]?*"

The seven perceptual attributes are presented in Table 3, with their corresponding questionnaire items. To avoid different interpretations of the attributes by the participants, known to become a possible bias in social research [30], laconic dictionary definitions were provided to them both verbally and in written form prior to the experimental sessions. While the evaluation of the scenes in SI could and were performed via a printed questionnaire, in which the participants were asked to rate the scenes using evaluation scales, this was not feasible using VR. Due to the use of the VR headset to achieve a fully immersive experience, the participants could not make use of a printed questionnaire. Instead, the experimenter asked the questions to the participants verbally, and the evaluations were also given in a verbal way, which the experimenter registered. For this reason, a 10-point Likert-type scale [31], commonly known to general population, was used. The use of scales with a large number of anchors (such as a 10-point scale) have been found to be preferred by participants as these offer them a ‘greater expression of feelings’ [32,33].

2.5. Participants

The participants were recruited via messages on social media sites, and through the university’s intranet site. The experiment was conducted in October 2020 with 20 participants (10 female and 10 male) presenting an average age of 25.9 years old (SD = 6.80). The youngest participant was 20 years old and the oldest was 47. Participants were from different nationalities, education, and career backgrounds.

As part of the eligibility criteria, the recruitment focused on volunteers who had previously experienced Virtual Reality. Indeed, most of the participants reported to had experience VR mostly through VR gaming. This criterion was included to make sure that excitement of experiencing a new technology for the first time would not affect the results.

To ensure that the participants did not have a vision impairment related to their depth perception that could compromise the collection of the data, the participants were tested using the Stereoscopic Vision Test, making use of the *Random Dot 2 Stereo Acuity Test*. The stereoscopic vision test was conducted prior to the experiments, and the results were given verbally to the participants. The participants with a minimum of 32 s of arc disparity in the stereoscopic vision test were eligible to participate in the study. No result of the vision test was recorded by the experimenter.

2.6. Experimental procedure

The experiment was conducted in October 2020 in the Room Laboratory (Romlab) at the Norwegian University of Science and Technology (NTNU). During the experiment, the rendered images presented in Fig. 3 were shown in Virtual Reality headset and the rendered images presented in Fig. 5 were projected on an antipolarized silver screen creating Stereoscopic Images of the simulated rooms.

The experimental procedure was maintained equal for all the participants. This means that while the experimenter ensure that the presentation of the stimuli followed a randomization principle, all the participants followed the different stages of the experiment (as indicated in Fig. 7). At the beginning of the experiment, the researcher welcomed and explained the procedure to the participants. To avoid language-related uncertainty, the dictionary definitions of the seven attributes were provided, as explained in Section 2.2. All the explanation and instructions were also presented to participants in a printed version. Then, the participants were tested for their stereoscopic vision, as stated in Section 2.5. After having the opportunity to ask questions about the procedure and agree to participate in the experiment, the participants signed a consent form. The consent form was approved and registered by NSD (The Norwegian Center for Research Data AS). The consent form included the participants rights (including withdrawal at any time) and contact information of the responsible researcher, the NTNUs data protection officer, and the NSD contact.

This study made use of a within-subjects design, in which all the 20 participants evaluated both the SI and VR methods. The presentation order of the methods was counterbalanced for all the participants. Due to the COVID-19 pandemic, strict infection control measures were implemented both at university and national level. These measures involved restrictions at the university, in which only one participant could attend the experiment at a time, and one experimenter could be present at the laboratory. The only exception allowed was the cohort, that is participants who were previously in close contact or lived together; in such a case two persons could attend the experiment simultaneously. In the case of two participants participating at the same time, one person was evaluating the SI images on the silver screen in the right side of the Romlab and the other participant assessed the VR scene in the other side of the room while both were totally aware that the scene presented to each person can be totally different from each other (see Fig. 6).

The participants who evaluated the SI wrote down their answer on the printed questionnaire provided by the researcher and the second person with VR headset answered different questions verbally. The use of both verbal and written questionnaires allowed for the experimenter to register the answers provided by the participants using the VR headset, while the participants evaluating the SI scenes could do it independently. Due to the nature of the SI method, in which the images displayed by the projectors could not be blocked, the position of the observers for the evaluation of the SI was set as close to the used viewpoint as possible for displaying the correct stereoscopic information. Although the participants could choose between two sides to find a comfortable position, the experimenter noted that all the participants preferred to sit on a chair to the right while answering the questionnaires. Only one participant reported lacking a 3D experience with the SI, and although the stereoscopic vision test was passed, the data for this participant was not included in the analysis.

In addition to maintaining social distance between the participants and the experimenter, other measures were taken: all the equipment was disinfected between each experimental session and disposable face masks for VR were also used. The participants were asked to evaluate the scenes in a randomized order. Seven questions related to each of the studied attributes in each presented scene were answered choosing an integer number from the 10-point scale according to their own evaluation. Both, the order of presentation mode, i.e. VR or SI, and the order of the images on VR environment and SI on the silver screen were presented to the participants in a random order. Participants were asked to avoid comparing the scenes and to evaluate each scene independently from the previous ones. To avoid any effect from the lighting situation the experiment was conducted mostly after sunset in a dark environment by covering the windows of the laboratory by blackout curtains. Moreover, the luminous conditions between the evaluations of VR and SI were kept as constant as possible. In addition to keeping the blackout rolling curtains down, the electrical light of the laboratory was kept off during the experimental sessions. This means that the Romlab was not illuminated by other form of light than the light from the projectors on the silver screen and from indirect light from the researcher's PC-screen, which was not in the field of view of the participants during the experiment. The perceived luminance differences between method evaluations were not higher than the differences between the white and the black room, leading to a minimal effect of the laboratory's luminous conditions on the perceptual evaluations. At the end of the experiment and after observation and assessment of the scenes, a debriefing interview about participant's experience of SI and VR was performed for all the participants. The experimental session for each participant lasted for an average of 45 min. The participation was voluntary and was rewarded at the end of the experimental session with a gift card of 150 NOK.

3. Results

3.1. Analysis strategy

In order to perform a method-comparison and to find the agreement between the two methods, i.e. VR and SI, the study made use of two statistical approaches. First, considering the repeated measures design of the study and after verifying the non-normal distribution of the data, Wilcoxon signed-rank tests were used to indicate whether the differences of the methods were statistically significant. Further, the Bland-

Fig. 6. Experimental setting: one participant evaluating SI in front of the silver screen while wearing a circular polarized glass and the other one evaluating a different scene in VR using an Oculus headset.

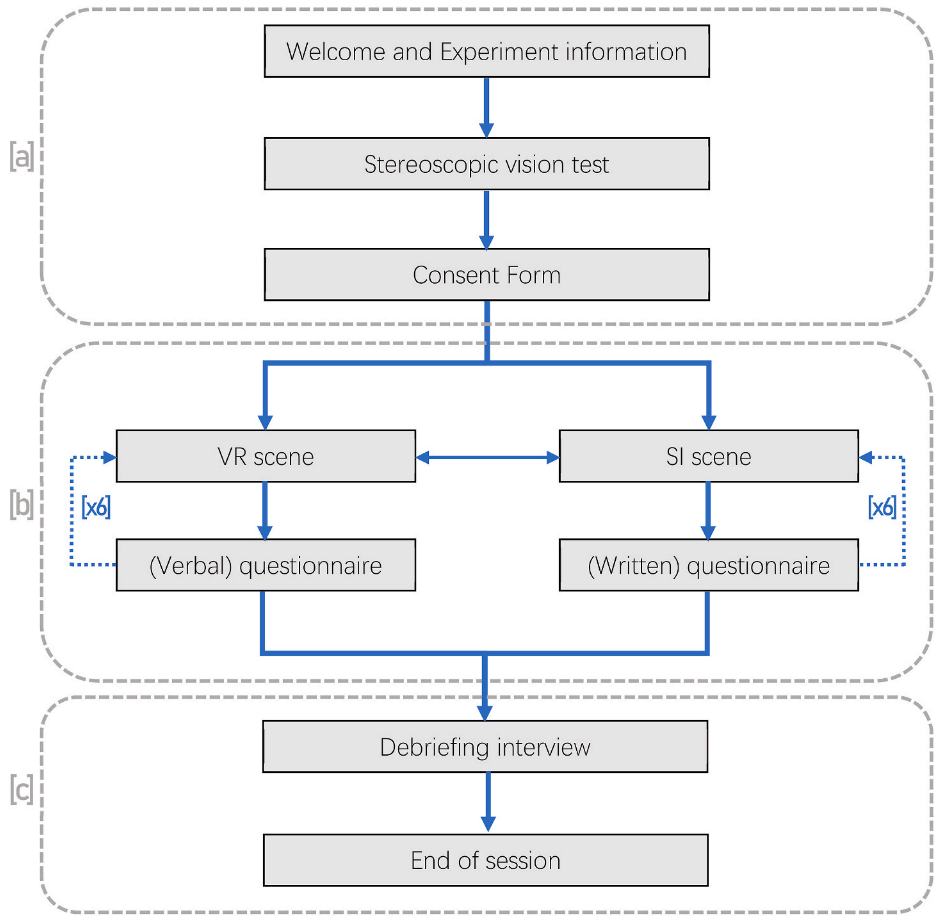

Fig. 7. Diagram of the experimental procedure used in the study for each participant, divided in [a] Information of the experiment, [b] Experimental phase, and [c] Debriefing phase and End of experimental session.

Altman method was applied [34] to reveal the extent of such differences.

For the Wilcoxon signed-rank tests, a significance alpha level of 0.05 was used. For the purpose of the study, no statistical differences represented by a *p*-value higher than 0.05 indicate the similarity of results obtained with the VR and SI. Moreover, the calculation of the effect sizes was used to evaluate the relative magnitude of the observed differences. As suggested by Rosenthal [35]; a more appropriate effect size calculation for the Wilcoxon signed-rank tests is converting a Z-score into the effect size estimate *r*. To do this, the Z value is divided by the square root of N, in which N is the number of total observations, not the number of cases. The interpretation of the effect sizes was carried out using Cohen's thresholds, in which a value of 0.2 is considered a small effect [36,37]. For the purpose of this study, an effect size equal or lower than 0.2 indicates a similarity between both tested methods.

Furthermore, and as mentioned earlier, the Bland-Altman statistical approach was also used. Naturally, the likeliness of two methods agreeing exactly on every rating given by the participants of a study is very low. The Bland-Altman method acknowledges this, and thus, its approach was developed to reveal the extension of the difference between ratings of two methods. The Bland-Altman method is a paired value analyzing approach, which examines the data patterns through assessing a scatter diagram of difference between values for each participant against its mean which is known as the Bland-Altman plot. This plot analysis is a simple visual way to evaluate possible bias between the mean differences, and to estimate an agreement interval, in which 95% of the differences of the methods lays [38]. The analyses are based on the visual examination of data patterns, statistical quantification of Bias (the difference between two methods) and the precision of Limit of Agreement of the difference (LoA precision).

The numerical calculations and graphical techniques of the Bland-Altman method were applied to all the data obtained from both studied methods (VR and SI) in all rooms with White surfaces (WR) and Black surfaces (BR), different window size (D1, D2 & D3) and for all seven attributes (as described in 2.2). Both rooms were evaluated separately on all seven spatial attributes. Considering the focus of the paper on method comparison, the data were divided into two groups to simplify the evaluation to two levels, i.e., WR and BR. The three levels of the window size were treated as 60 different observations (20 participants assessed 3 window sizes) for each room. This approach granted the use of a replicated measurements analysis [39] and allowed for an eventual comparison with a previous study in which SI and real spaces were also studied [18]. The analyses and graphical plots are thus presented for the white room and the black room. However, the graphical plots for each of the six room alternatives (including each window size) are available from the authors upon request. The authors considered a priori that the benchmark Bias <1, in view of the 10-point Likert-type scale, is acceptable for the aim of this study.

3.2. *Wilcoxon signed-rank tests*

The statistical results of the Wilcoxon signed-rank tests indicate no significant differences between the responses obtained in VR and in SI for most of the studied attributes. Specifically, there were no significant differences in six of the studied attributes in the WR (i.e. Pleasantness, Calmness, Interest, Excitement, Complexity and Spaciousness), and no significant differences in four attributes in the BR (i.e. Pleasantness, Calmness, Interest, and Complexity). Furthermore, the calculated effect sizes showed values lower than 0.2, indicating small effects in such attributes.

On the other side, the Wilcoxon signed-rank tests showed statistically significant differences between VR and SI for Amount of View in the WR ($Z = -3.384$, $p < 0.001$; $r = 0.31$). In the BR, the Wilcoxon signed-rank tests showed statistically significant differences for Excitement ($Z = -2.655$, $p = 0.008$; $r = 0.24$), Spaciousness ($Z = -2.562$, $p = 0.01$; $r = 0.23$), and Amount of View ($Z = -3.285$, $p = 0.001$; $r = 0.30$). As indicated, the analysis of the effect sizes for the significant results revealed values over 0.20. Both the results of the Wilcoxon signed-rank tests and the calculation of the effect sizes indicate that the evaluation of Amount of View in the WR, and Excitement, Spaciousness and Amount of View in the BR is different with the use of VR and the use of SI.

3.3. Bland-Altman method with multiple observations

The following paragraphs present the findings related to the evaluation of the seven studied spatial attributes (Pleasantness, Calmness, Interest, Excitement, Complexity, Spaciousness, and Amount of View).

The numerical results of Bland-Altman's method-comparison approach is presented in Table 4. This table indicates: Bias as the mean difference in values obtained in VR and SI methods; 95% Confidence Interval (CI - repeatability) indicating the cluster values around the mean which the same method produces the same results on repeated measurements; SD (Standard Deviation) as a measure of variability of the individual differences; computed Lower Limit of Agreement (LoA) as Bias −1.96 SD and Upper LoA is Bias +1.96 SD; Confidence Limit provide a measure of the precision of the values which is the difference between the Upper LOA and Lower LoA. The percentage (%) Error is the proportion between the magnitude of measurement and the error in measurement.

The results of the calculation of the attributes' bias show in general positive mean differences between VR and SI. The results imply that the participants evaluated attributes slightly higher with Virtual Reality than with the Stereoscopic Image presentation. The attribute Calmness is an exception for this, with a negative mean difference in the black room which means the scenes were assessed at a slightly higher value in SI than VR. The attributes Pleasantness in both BR and WR has the lowest bias scores, indicating that the participants evaluated this attribute in the scenes similarly in VR and SI methods. The results indicate that all the attributes have a mean difference lower than a priori acceptance criterion of Bias <1. Nevertheless, the 95% Confidence Intervals for the mean difference in the dataset of the attributes Excitement in BR are 0.2 and 1.2, and for the Amount of View in white and black rooms, are 0.4–1.3 and 0.3–1.2 respectively. Considering the a priori acceptance criterion established by the authors, VR environment and SI presentation are not in agreement to evaluate Amount of View (in both WR and BR) and Excitement (in BR), while the agreement between the SI and VR in the attribute Spaciousness in BR with the 95% Confidence Interval 0.1–1.0 warrants further analyses.

Moreover, [40] noted that there is a possibility of hidden proportional error in the estimation of the bias and repeatability because they are computed across all the data points. To avoid this problem, he suggests the calculation of the percentage error by dividing the limits of agreement by the mean value of the measurements established methods. Since the true values of each attribute in a real scenario are unknown, the percentage of error in this study implies the error of the data in VR from SI. The percentage error column in Table 4 shows Pleasantness and Interest in WR with 1% have the lowest values. While Excitement in BR with 19%, Amount of view with 16% in WR and 15% in BR have the highest percentage of error. This confirms the disagreement between Excitement in BR and Amount of View in both WR and BR. The percentage of error of Spaciousness in BR is 14% and considering its 95%CI 0.1–1.0 indicates that this attribute also does not support enough evidence for agreement between VR and SI. Therefore, the results of the method-comparison between VR and SI for daylighting studies suggest that these two methods are similar to assess Pleasantness, Calmness, Interest, and Complexity of a small room with both white and black surfaces and to evaluate Excitement and Spaciousness of a room with white surfaces only.

Furthermore, to investigate any possible relationship between the values and determine agreement between two methods, [38] recommended the use of plots with bias and precision statistics. The calculated data are presented in scatter plots of difference against mean with Bias and Limit of Agreements (LoA) lines and 95% CI in Fig. 8.

The position of the bias line allows to analyze the plot. As expected, the bias line lies closer to the 0 point for almost all the attributes, showing also reasonable confidence intervals, except for Amount of View (BR & WR) and Excitement & Spaciousness (BR). A narrow agreement interval (the area between Upper LoA and Lower LoA in the plots of average against difference) demonstrates the agreement between the two methods. The distribution of data in the diagrams in Fig. 8 show the agreement interval that is narrow for attributes Pleasantness and Excitement in WR, and Calmness in BR, compared to the other attributes' agreement interval. Complexity (BR) and Interest (WR) have the widest agreement interval and interpret the lowest precision agreement between the two methods.

Finally, [41] emphasizes that if too many points, more than about 5% of the data, lie outside and are dramatically far from the LoAs, it means

Table 4
Numerical results from Bland-Altman method for each attribute in the White Room (WR) and Black Room (BR).

Attribute	Room	Bias	95% CI (mean difference)		SD	Lower LoA	95% CI of L-LoA		Upper LoA	95% CI of U-LoA		Confidence Limit/Precision (Lower LoA -Upper LoA)	% error
Pleasantness	WR	0.03	−0.4	0.4	1.47	−2.85	−3.5	−2.2	2.92	2.3	3.6	−5.76	1%
	BR	0.10	−0.4	0.6	1.92	−3.66	−4.5	−2.8	3.86	3.0	4.7	−7.51	2%
Calmness	WR	0.30	−0.2	0.8	1.84	−3.30	−4.1	−2.5	3.90	3.1	4.7	−7.21	6%
	BR	−0.12	−0.5	0.3	1.64	−3.32	−4.1	−2.6	3.09	2.4	3.8	−6.41	−3%
Interest	WR	0.07	−0.6	0.6	2.06	−3.98	−4.9	−3.1	4.11	3.2	5.0	−8.09	1%
	BR	0.18	−0.3	0.7	1.81	−3.36	−4.2	−2.6	3.73	2.9	4.5	−7.08	4%
Excitement	WR	0.12	−0.3	0.5	1.65	−3.11	−3.9	−2.4	3.34	2.6	4.1	−6.46	3%
	BR	0.70	0.2	1.2	1.87	−2.97	−3.8	−2.1	4.37	3.5	5.2	−7.35	19%
Complexity	WR	0.20	−0.2	0.7	1.80	−3.33	−4.1	−2.5	3.73	2.9	4.5	−7.06	6%
	BR	0.23	−0.3	0.8	2.16	−3.99	−5.0	−3.0	4.46	3.5	5.4	−8.45	6%
Spaciousness	WR	0.10	−0.4	0.6	1.84	−3.50	−4.3	−2.7	3.70	2.9	4.5	−7.21	2%
	BR	0.57	0.1	1.0	1.74	−2.84	−3.6	−2.1	3.98	3.2	4.8	−6.82	14%
Am. Of View	WR	0.85	0.4	1.3	1.75	−2.58	−3.4	−1.8	4.28	3.5	5.1	−6.85	16%
	BR	0.73	0.3	1.2	1.81	−2.82	−3.6	−2.0	4.29	3.5	5.1	−7.11	15%

Table notes: 95%–95% Confidence Interval (CI); SD, Standard Deviation; L-LoA, Lower Limit of Agreement; U-LoA, Upper Limit of Agreement. *In the Bland-Altman method, the accuracy of any given method in comparison to another method is given by the Mean Difference, referring to the average of the difference in values obtained with the two methods. The precision of any tested method is given by the Limits of Agreement (LoA), referring to the limits within which 95% of all the scores fall on either side of the Mean Difference.

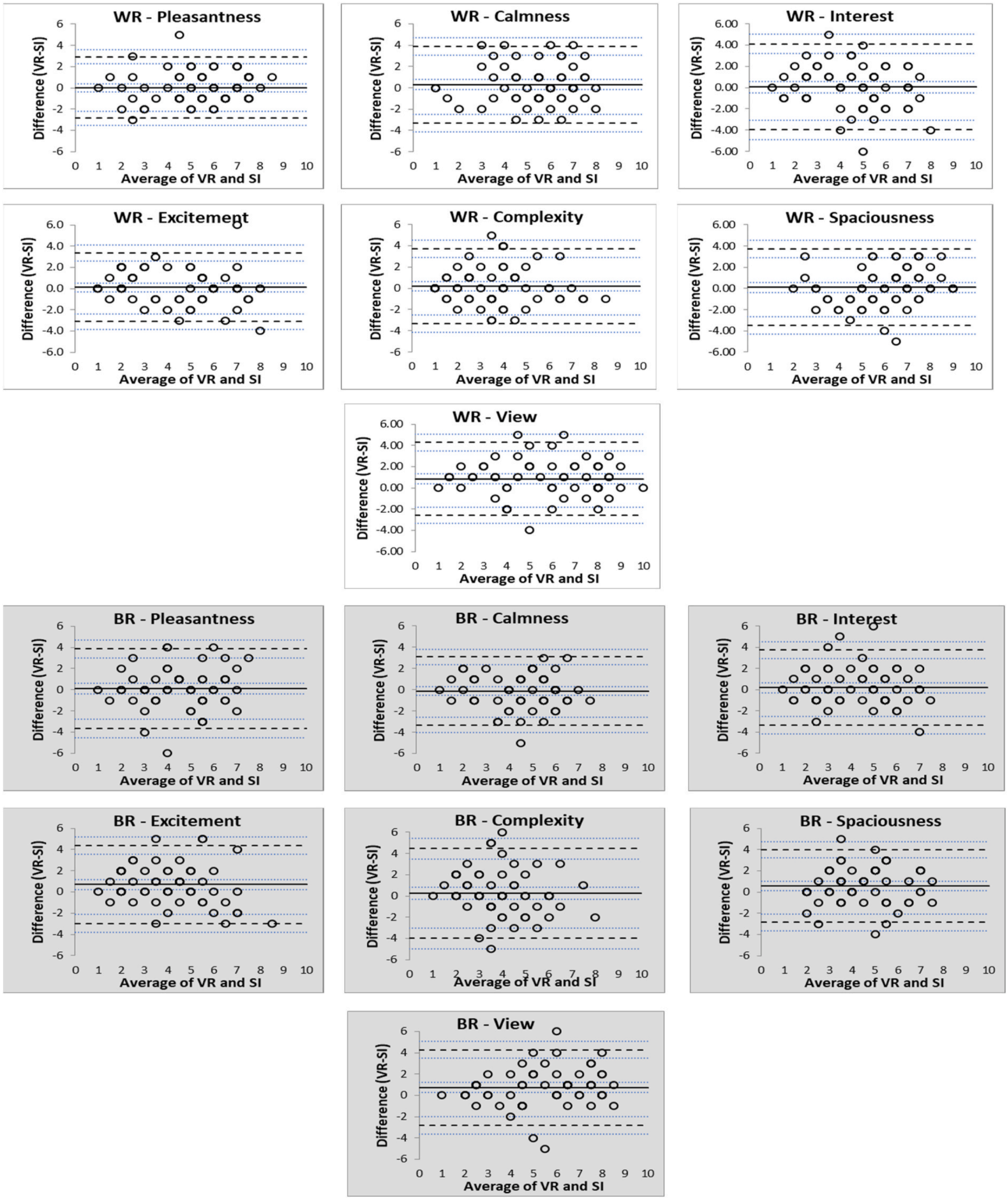

Fig. 8. Plots of difference against mean with the line of Bias, 95%CI, Upper and Lower Limit of Agreement. The X-axis represents the average of paired values from the Virtual Reality (RE) and Stereoscopic Imaging (SI) methods. The Y-axis represents the difference between the values of the VR and SI methods. The Mean Difference is expressed by the solid line, and the Limits of Agreement (both Upper and Lower) are represented by the dashed lines. The 95% CI lines are presented with dotted blue lines. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 5
The percentage of outliers of data for each attribute.

Pleasantness		Calmness		Interest		Excitement		Brightness		Complexity		Spaciousness		View	
WR	BR	WR	BR	WR	BR	WR	BR	WR	BR	WR	BR	WR	BR	WR	BR
3%	3%	0%	2%	3%	7%	3%	5%	5%	3%	7%	7%	3%	3%	5%	5%

that the data are erratically variable, and the study needs a larger sample size. As it is visible in the diagrams of Fig. 8, there are some points outside of Confidence Limit, below of Lower LoA and above Upper LoA. Table 5 shows the percentage of these data for all the attributes. Although Interest (BR) and Complexity (WR and BR) have 7% outliers, and are not far from the 95% LoA, the interpretations for these two attributes should be taken with caution.

4. Discussion

4.1. Limitations and further research

The present study focused on investigating the differences between two visualization and presentation methods, i.e., VR and SI. Although the findings suggest similarity between these two methods, it becomes important to be aware of the difference between these methods and real environments studies. To the authors' knowledge, there is currently no level of simulation that renders equal stimuli as a real environment provides. For instance, background sound, environmental interaction, temperature, and visual conditions are examples of stimuli that are inherent to real spaces. Much as technological advancements can mimic many of these environmental parameters in laboratory studies, it seems rarely possible to achieve high precision using simulated environments in social research. However, the challenging process of conducting daylighting research calls for visualization methods capable to streamline initial experimental procedures to be used as basis for further research. In that vein, both methods presented in this paper have been previously examined to reveal their adequacy as research tools in daylighting research. While stereoscopic images have been found to be an acceptable representation of real environments [18], virtual reality have also been adequate to be used as a surrogate for real daylit spaces [55,56]. Yet, more research efforts are needed to confirm the suitability of visualization methods for overcoming challenges associated with studies in real environments.

Image fidelity and the luminance range of the devices are examples of limitations that inhibit a real experience in simulated environments. In the present study, the luminance range of the scenes was restricted, i.e., the VR headset had a maximum luminance of 80 cd/m^2 while the silver screen presented a maximum luminance of 40 cd/m^2. The luminance decreased to 10 cd/m^2 when measured through the polarized glasses that the participants used while evaluating the SI. This, together with the tone-mapping operator used for the study, which does not allow a dynamic tone-mapping while an observer sees around a scene, presented a limitation to the study. Indeed, the overall brightness displayed to the participants in both methods present a crucial difference not only between them, but also compared to real spaces. Further research is thus needed to test the robustness of the results. Studies using a dynamic and gaze-responsive tone-mapping operator might address this limitation concerning brightness.

4.2. Discussion of the findings

Considering the focus of the present paper, the analysis of the results focused on a method-comparison study, answering the research question: *How do perceptual attributes for simulated daylit spaces differ from using stereoscopic images (SI) and virtual reality (VR)?* The discussion of the findings will be presented following both study's objectives. The first objective was to analyze the correspondence between VR and SI, i.e., if both experimental methods obtained similar results and if so, to what degree these are similar. To address the first objective, the discussion is based on the statistical and graphical results obtained using the Wilcoxon signed-rank tests and the Bland-Altman method. The second study objective was set to finding the main differences between VR and SI, analyzed from the users' subjective experience.

4.2.1. Statistical and graphical results

The statistical analyses by using the Wilcoxon signed-rank tests and the Bland-Altman method-comparison approach revealed that the attributes *Pleasantness, Calmness, Interest*, and *Complexity* in a room with white or black surfaces, and *Excitement* and *Spaciousness* in a room with white surfaces, are evaluated similarly in both VR and SI methods. The attributes *Excitement* in a room with black surfaces and *Amount of View* in a room with both white or black surfaces were found to be evaluated different in VR and SI, showing bias values over the 1.0 a priori criterion and a high percentage of error. For the attribute *Spaciousness* in the black room, despite obtaining a bias value of 1.0, the high percentage of error suggests little statistical support to validate the similarity of the results in SI and VR. The attribute *Calmness* presented unique results, distinguishing from the other studied attributes. Whereas the rest of the attributes were evaluated higher with VR than with the SI, *Calmness* was rated higher with SI than with VR.

To aid readability, the results are summarized in Table 6. It is important to notice that the results were equal using both the traditional statistical approach (i.e. Wilcoxon signed-rank test) and the pragmatic statistical approach (i.e. Bland-Altman method), as such Table 6 offers the unified results based on the statistical results of both approaches. The check mark refers to similar results for a particular attribute in both VR and SI. The attributes in which the number of outliers is high are marked with a superscript a. Attributes without any symbol have not presented similar results between both tested methods.

The statistical analyses and findings of the experimental work presented in this article suggests that a daylit small room with white or black surfaces can be evaluated for most studied attributes in SI presentation and VR environment similarly, but the researcher needs to be very careful with interpreting the data especially in rooms with dark surfaces or low level of luminance due to the wide range of agreement between the data in these spaces.

The applied technique for evaluating the space in VR or SI can deliver a very different spatial perception, especially when light level is low. With the increase in the level of light inside the space, these two methods might be more similar, and the results may be more accurate when the test case is not a very dark space. These results are in line with the previous study by [18], who indicated that the room with white surfaces would be a more suitable setting to be studied with SI than the black room.

Special attention should be given to the results for the attributes *Excitement* and *Spaciousness,* which results differed between room colors, presenting values over the a priori study criterion and higher percentage of error in the black room. For both attributes, the results indicated that the spaces were rated more exciting and more spacious in the black room than in the white room when using VR. This could be caused by the level of visualization obtained with VR, in which the participants could observe in all directions of the space using VR, acquiring more spatial information, and thus experiencing a more exciting room compared to the limited vision and experience obtained with the SI. This is supported by previous studies suggesting that spatial judgements are improved by the addition of higher fidelity system features, such as head-tracked rendering [42]. Moreover, VR seems to provide more visual

Table 6
Summarized findings based on the statistical results divided by attributes and room color.

Attribute	WR	BR
Pleasantness	✓	✓
Calmness	✓	✓
Interest	✓	✓[a]
Excitement	✓	
Complexity	✓[a]	✓[a]
Spaciousness	✓	
Amount of view		

[a] *Results should be taken with caution.*

information regarding distance perception compared to the SI. Indeed, previous research has shown that adding pictorial depth cues and presenting virtual replicas of actual rooms enhance a sense of presence in virtual environments providing a more accurate distance perception [43]. Distance perception can be considered a factor closely related to the evaluation of spaciousness, thus yielding higher ratings for the black room with VR than with SI.

Although the precision of true values (represented by the LoA) was found to be low, we can be certain about the accuracy of the methods to study the listed attributes under certain settings. Additionally, the evaluation process of any real study in social sciences, in which people decide their own ratings, usually has a risk of imprecision. Thus, the precision of SI and VR, as simulated environments, cannot expect to have the precision that real environments do not present. Regarding the sample size, considering that each participant evaluated 3 different window sizes (i.e., D1, D2, and D3) in both white and black room, the sample size was increased from 20 to 60 in the experiment. Although a larger sample size might have increased the precision of the methods, the accuracy would have remained the same, thus supporting the results that both methods are accurate for studying spatial perception of small daylit rooms.

4.2.2. Subjective experience of both studied methods

The second objective of the study was to identify the main differences between these two methods that may affect the visual perception of a daylit architectural space. There are some challenges for conducting research experiments by using both studied methods. Some of these challenges seem necessary to be discussed to provide a better understanding about the experience of participants and researchers during the study.

The results of the debriefing interviews at the end of the experiment with each participant indicate that the nature of the Virtual Reality environment inducing a feeling of 'presence' or 'being inside the space' made VR the favorite method for evaluating the scene in the opinion of most participants. Thus, suggesting that the difference between the SI as a non-immersive and VR as an immersive environment has a large impact on the way that space is perceived. These results are in line with the findings of [44] who found that presence is affected by field-of-view and display type, and with those from Higuera-Trujillo et al. [45]; who found that the feeling of presence is an important predictor for psychological and physiological responses in simulated environments.

This observation is confirmed by the placement of the shifted bias line to the positive side of the plots in Fig. 8 for some attributes. As previously stated, this means that the participants ranked higher values for a small daylit space in the Virtual Reality environment than Stereoscopic Images presentation. This difference might be biased by the nature of the VR as an immersive environment and the excitement that this experience can provoke in some people. Indeed, the experimenter observed a difference of interest between the two methods. Whereas the participants took their time to see each scene carefully in VR, most of them were impatient and shown less interest when evaluating the scenes using SI. This could be linked to the more unique experience that VR provides, thus creating a 'novelty effect', as discussed by Bardo et al. [46].

The feeling of presence inside of a space in VR as an immersive 3D environment has been the topic of some studies [47,48]. This matter was specifically discussed during the debriefing interview with participants in this study. Most of the interviewed participants reported that the feeling of presence and being inside the space had an impact on how they perceive the space. However, not all VR experience was considered positive. Although most participants reported to have had a positive experience using VR, some of them felt dizzy by using the headset. Indeed, dizziness is a known symptom of motion sickness associated to the use of VR and found in lighting studies using VR [9]. Recent studies link the feeling of presence in VR with sickness [49], while others indicate that multiple factors (including hardware, content and human factors) cause such VR sickness [50]. Although VR sickness symptoms often dissipate after the virtual experience is terminated, further research efforts are needed for a better estimation of VR sickness and its potential effect on experimental studies. Furthermore, one participant who attended the experiment alone mentioned feeling completely disconnected from the real surrounded environment during the experiment, producing feelings of nervousness to be in an experimental room with some other people while the eyes were covered by the headset. During the evaluation of the black room with a small window (D1), another participant expressed feeling claustrophobic in the virtual room and experienced not feeling comfortable when evaluating the SI for the same scene.

From the perspective of the experimenter, conducting this experiment during a global pandemic was a challenging task, due to the risk of spreading the COVID-19 virus by using the VR headset for all participants. Indeed, conducting the experiment needed very strict infection control measures. These measures included: limiting the presence of multiple participants at the same time in the laboratory, introducing a time gap (>3 h) between each use of the same headset, and high precaution for the disinfection process of all equipment used by the participant for the next one. All these made the process of conducting the experiment a demanding task for VR, while SI did not present most of these struggles as the SI does not require any physical contact and share of equipment between participants.

The two studied methods differ also regarding logistics. While SI demands use of a laboratory with silver screen and a good control of light sources (daylight and electric light), the VR headset can theoretically be used at any location. Still, other factors such as safety when the participant is cut off from the real room during the immersive experience, acoustic conditions, or temperature should be considered carefully. It needs also to be mentioned that the experience of the room differs between those two methods regarding co-experience. While the SI projection enables simultaneous experience by a group of people, the VR headset does not allow this, making it a solitary experience. The possibility of simultaneous observation of a new design of architectural space gives the possibility for discussion about the choices made by architects or other involved partners. It is also an important advantage in teaching and in different dissemination tasks. Table 7 describes the main differences of both studied methods.

Finally, the results of this study seem to be in line with the findings from Patrick et al. [51]; who found no statistically significant difference between head-mounted displays and large projection screen conditions for investigating spatial knowledge. As discussed by de Kort et al. [59], not only is the study of virtual environments an interesting research subject, but also a necessary research endeavor. Future research in this field is encouraged on the use of VR or SI and the perceived quality of the space and how the use of this visualization techniques affects perception in daylighting studies.

5. Conclusions

Findings showed that the SI presentation and VR environment provide similar evaluations from a small daylit room for six of the studied spatial attributes. Specifically, the attributes *Pleasantness, Calmness, Interest*, and *Complexity* in a room with white or black surfaces; and the attributes *Excitement* and *Spaciousness* in a room with white surfaces present similar results using both methods.

Although the results suggest that both methods are valid and can be used interchangeably in empirical studies, each method presents benefits and disadvantages proper to each study, warranting further research efforts to determine the adequacy of the methods for lighting and architectural research. Researchers working with daylight aiming to use either of these visualization methods are thus encouraged to carry a validation study prior to their experiment and to assess the competence of each method, including the conduction of experiments in real environments. The findings of this study contribute to the discussion of the

Table 7
Differences between VR and SI methods.

	Virtual Reality (VR)	Stereoscopic Images (SI)
Experimental environment	Observer is exposed to a small monitor with a light source in a VR headset.	Observer is in a dark room to see the image projected on a large screen.
Interaction	Dynamic view.	Static view.
Co-experience	Only one person can see the room image at a time, which means a solitary experience. This can also increase the total experimental time.	A group of people can observe a stereoscopic image simultaneously, which is advantageous e.g., in teaching, to reduce experimental time, and in all other situations where discussion about the stimuli is important.
Opportunities	Feeling of experiencing a brand-new technology. Observers reported they feel themselves inside the space and it makes the evaluation easier. Feelings of 'presence' or 'being inside the space'.	Observer reported that they feel more comfortable because they do not have to wear any gadget.
Obstacles	Some observer report feeling dizzy or nauseous while using the headset. Other reported feeling claustrophobic. Conducting the experiment during a global pandemic and health related risk for spreading the virus by using a wearable gadget.	There might be difficulties to perceive the image as a 3D environment.
Logistics	There is no need for a laboratory to use a VR headset, which gives researchers a possibility to carry out studies in any location at any time.	The study must be carried out in a laboratory with a large silver screen and good control of daylight and electric light.
Experimental equipment	VR headset, computer with a powerful graphics card, software (Rhinoceros, DIVA-for-Rhino, Radiance, Unity)	Antipolarized Silver Screen, circular polarized filters, two high-definition projectors, circular polarized glasses for the participants, software (Rhinoceros, DIVA-for-Rhino, Radiance, Stereoscopic Player)

use of simulated environments in daylighting research, and by providing researchers information about using either of the studied methods in their research studies.

CRediT authorship contribution statement

Claudia Moscoso: Writing – original draft, Visualization, Resources, Methodology, Funding acquisition, Conceptualization, Writing – review & editing. **Marzieh Nazari:** Formal analysis, Investigation, Software, Visualization, Writing – original draft, Writing – review & editing. **Barbara Szybinska Matusiak:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

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.

Acknowledgements

This work was supported by the VELUX Stiftung Foundation [Project 1022: "Identifying the Impact of Regional Differences on the Perceived Quality of Daylit Architectural Spaces: A Comparison Study across Different Latitudes."] and by the Norwegian University of Science and Technology (NTNU). The authors gratefully acknowledge the contribution of Krzysztof Orleanski for his assistance in preparing the experimental equipment, and of Kine Angelo and Marie Boucher for their assistance in recruiting the participants. We are truly grateful to Jan Christian Meyer, member of the IT Development Section who generously helped us in the process of rendering with High Performance Computers at NTNU. We are also thankful all the unknown participants in the experiment who helped us through this project.

References

[1] C. Hendrick, O. Martyniuk, T. Spencer, J.E. Flynn, Procedures for investigating the effect of light on impression: simulation of a real space by slides, Environ. Behav. 9 (4) (1977) 491–510, https://doi.org/10.1177/001391657794003.
[2] B. Matusiak, B. Sudbø, Width or height? Which has the strongest impact on the size impression of rooms? Results from full-scale studies and computer simulations, Architect. Sci. Rev. 51 (2) (2008) 165–172, https://doi.org/10.3763/asre.2008.5120.
[3] G.R. Newsham, C. Richardson, C. Blanchet, J.A. Veitch, Lighting quality research using rendered images of offices, Light. Res. Technol. 37 (2005) 93–112.
[4] I. Sprow, Z. Baranczuk, T. Stamm, P. Zolliker, Web-based psychometric evaluation of image quality, in: Image Quality and System Performance VI, International Society for Optics and Photonics, 2009, 72420A.
[5] C. Villa, R. Labayrade, Validation of an online protocol for assessing the luminous environment, Light. Res. Technol. 45 (2013) 401–420, https://doi.org/10.1177/1477153512450452.
[6] C. Wheatstone, XVIII. Contributions to the Physiology of vision.—Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision, Philosophical transactions of the Royal Society of London, 1838, pp. 371–394.
[7] G.R. Jones, D. Lee, N.S. Holliman, D. Ezra, Controlling perceived depth in stereoscopic images, in: Stereoscopic Displays and Virtual Reality Systems VIII, International Society for Optics and Photonics, 2001, pp. 42–53.
[8] C. Cauwerts, Influence of presentation modes on visual perceptions of daylit spaces, Doctoral thesis, Université catholique de Louvain, Louvain-la-Neuve, 2013. November 2013.
[9] F. Abd-Alhamid, M. Kent, C. Bennett, J. Calautit, Y. Wu, Developing an innovative method for visual perception evaluation in a physical-based virtual environment, Build. Environ. 162 (2019), https://doi.org/10.1016/j.buildenv.2019.106278.
[10] Y. Chen, Z. Cui, L. Hao, Virtual reality in lighting research: comparing physical and virtual lighting environments, Light. Res. Technol. 51 (6) (2019) 820–837, https://doi.org/10.1177/1477153518825387.
[11] C. Cauwerts, M. Bodart, Investigation of 3D projection for qualitative evaluation of daylit spaces, in: PLEA 2011 "Architecture and Sustainable Development, 2011.
[12] C. Moscoso, B. Matusiak, Aesthetic perception of a small office with different daylighting systems, Indoor Built Environ. 27 (2018) 1187–1202, https://doi.org/10.1177/1420326X17711490.
[13] S.F. Rockcastle, K. Chamilothori, M. Andersen, An experiment in virtual reality to measure daylight-driven interest in rendered architectural scenes, in: Proceedings of Building Simulation 2017, Building Simulation 2017, San Francisco, California, USA, 2017, pp. 7–9. August.
[14] M. Hegazy, K. Yasufuku, H. Abe, Evaluating and visualizing perceptual impressions of daylighting in immersive virtual environments, J. Asian Architect. Build Eng. 20 (6) (2020) 768–784, https://doi.org/10.1080/13467581.2020.1800477.
[15] F. Abd-Alhamid, M. Kent, J. Calautit, Y. Wu, Evaluating the impact of viewing location on view perception using a virtual environment, Build. Environ. 180 (2020), https://doi.org/10.1016/j.buildenv.2020.106932.
[16] J.P. Carneiro, A. Aryal, B. Becerik-Gerber, Influencing occupant's choices by using spatiotemporal information visualization in Immersive Virtual Environments, Build. Environ. 150 (2019) 330–338, https://doi.org/10.1016/j.buildenv.2019.01.024.
[17] C. Moscoso, K. Chamilothori, J. Wienold, M. Andersen, B. Matusiak, Regional Differences in the Perception of Daylit Scenes across Europe Using Virtual Reality. Part I: Effects of Window Size, vols. 1–22, LEUKOS, 2021, https://doi.org/10.1080/15502724.2020.1854779.
[18] C. Moscoso, B. Matusiak, U.P. Svensson, K. Orleanski, Analysis of stereoscopic images as a new method for daylighting studies, Trans. Appl. Percept. 11 (2015), https://doi.org/10.1145/2665078.
[19] G. Ward, RADIANCE Pfilt Manual, RADSITE, 1996. Retrieved Nov. 2021 from, https://www.radiance-online.org/learning/documentation/manual-pages/pdfs/pfilt.pdf/view.
[20] E. Reinhard, M. Stark, P. Shirley, J. Ferwerda, Photographic tone reproduction for digital images, in: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, 2002, pp. 267–276.
[21] G. Larson, H. Rushmeier, C. Piatko, A visibility matching tone reproduction operator for high dynamic range scenes, IEEE Trans. Visual. Comput. Graph. 3 (1997) 291–306.
[22] S. Chapple, M. Ladaique, Society at a Glance 2009: OECD Social Indicators, Retrieved from Organisation for Economic Co-operation and Development, 2009.
[23] R.L. Huston, Fundamentals of Biomechanics, CRC Press, 2013.
[24] N.A. Dodgson, Variation and extrema of human interpupillary distance, in: Stereoscopic Displays and Virtual Reality Systems XI International Society for Optics and Photonics, 2004, pp. 36–46.
[25] OECD, Society at a Glance 2009 OECD Social Indicators, OECD Publishing, 2009.
[26] J.A. Russell, A circumplex model of affect, J. Pers. Soc. Psychol. 39 (1980) 1161.

[27] M. Boubekri, R.B. Hull, L.L. Boyer, Impact of window size and sunlight penetration on office workers' mood and satisfaction: a novel way of assessing sunlight, Environ. Behav. 23 (1991) 474–493.
[28] S. Kaplan, Aesthetics, affect, and cognition: environmental preference from an evolutionary perspective, Environ. Behav. 19 (1987) 3–32.
[29] C. Moscoso, B. Matusiak, U.P. Svensson, Impact of window size and room reflectance on the perceived quality of a room, J. Architect. Plann. Res. (2015) 294–306.
[30] R. Rosenthal, R.L. Rosnow, Essentials of Behavioral Research: Methods and Data Analysis, 2008.
[31] R. Likert, A Technique for the Measurement of Attitudes, Archives of psychology, 1932.
[32] C.C. Preston, A.M. Colman, Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences, Acta Psychol. 104 (2000) 1–15, https://doi.org/10.1016/S0001-6918(99)00050-5.
[33] S.Y. Chyung, K. Roberts, I. Swanson, A. Hankinson, Evidence-based survey design: the use of a midpoint on the Likert scale, Perform. Improv. 56 (2017) 15–23, https://doi.org/10.1002/pfi.21727.
[34] J.M. Bland, D. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, Lancet 327 (1986) 307–310.
[35] R. Rosenthal, Parametric measures of effect size, in: H. Cooper, L.V. Hedges (Eds.), The Handbook of Research Synthesis, Russell Sage Foundation, 1994, pp. 231–244.
[36] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, second ed., Lawrence Erlbaum, New Jersey, 1988.
[37] J. Cohen, Statistical power analysis, Curr. Dir. Psychol. Sci. 1 (3) (1992), https://doi.org/10.1111/1467-8721.ep10768783.
[38] J.M. Bland, D.G. Altman, Measuring agreement in method comparison studies, Stat. Methods Med. Res. 8 (1999) 135–160.
[39] J.M. Bland, D.G. Altman, Agreement between methods of measurement with multiple observations per individual, J. Biopharm. Stat. (2007) 571–582, https://doi.org/10.1080/10543400701329422.
[40] S.K. Hanneman, Design, analysis, and interpretation of method-comparison studies, AACN Adv. Crit. Care 19 (2008) 223–234, https://doi.org/10.1097/01.AACN.0000318125.41512.a3.
[41] R. Riffenburgh, Statistics in Medicine, 2012. London.
[42] 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 judgement task, IEEE Trans. Visual. Comput. Graph. 19 (2012) 886–896, https://doi.org/10.1109/TVCG.2012.163.
[43] R.S. Renner, B.M. Velichovsky, J.R. Helmert, The perception of egocentric distances in virtual environments – a review, ACM Comput. Surv. 46 (2) (2013) 40, https://doi.org/10.1145/2543581.2543590. Article 23 (November 2012).
[44] A. Alshaer, H. Regenbrecht, D. O'hare, Immersion Factors Affecting Perception and Behavior in a Virtual Reality Power Wheelchair Simulator, 2017.
[45] J.L. Higuera-Trujillo, J. Lopez-Tarruela Maldonado, C. Llinares Milan, Psychological and physiological human responses to simulated and real environments: a comparison between photographs, 360° panoramas and virtual reality, Appl. Ergon. 65 (2017) 398–409, https://doi.org/10.1016/j.apergo.2017.05.006.
[46] M.T. Bardo, R.L. Donohew, N.G. Harrington, Psychobiology of novelty seeking and drug seeking behavior, Behav. Brain Res. 11 (1–2) (1996) 23–43.
[47] M.J. Schuemie, P. Van der Straaten, M. Krijn, C.A. Van der Mast, Research on presence in virtual reality: a survey, Cyberpsychol. Behav. 4 (2001) 183–201, https://doi.org/10.1089/109493101300117884.
[48] B.G. Witmer, M.J. Singer, Measuring presence in virtual environments: a presence questionnaire, Presence 7 (1998) 225–240.
[49] S. Weech, S. Kenny, M. Barnett-Cowan, Presence and Cybersickness in Virtual Reality are negatively related: a review, Front. Psychol. 10 (158) (2019), https://doi.org/10.3389/fpsyg.2019.00158.
[50] E. Chang, H.T. Kim, B. Yoo, Virtual reality sickness: a review of causes and measurements, Int. J. Hum. Comput. Interact. 36 (17) (2020) 1658–1682, https://doi.org/10.1080/10447318.2020.1778351.
[51] E. Patrick, D. Cosgrove, A. Slavkovic, J.A. Rode, T. Verratti, G. Chiselko, Using a large projection screen as an alternative to head-mounted displays for virtual environments, in: CHI '00: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. April 2000, 2000, pp. 478–485, https://doi.org/10.1145/332040.332479.
[55] K. Chamilothori, G. Chinazzo, J. Rodrigues, E. Dan-Glauser, J. Wienold, M. Andersen, Subjective and physiological responses to façade and sunlight pattern geometry in virtual reality, Build. Environ. 150 (2019) 144–155, https://doi.org/10.1016/j.buildenv.2019.01.009.
[56] K. Chamilothori, J. Wienold, M. Andersen, Adequacy of immersive virtual reality for the perception of daylit spaces: comparison of real and virtual environments, Leukos 15 (2019) 203–226, https://doi.org/10.1080/15502724.2017.1404918.
[57] 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. ConStruct. 54 (2015) 116–126, https://doi.org/10.1016/j.autcon.2015.03.020.
[58] A. Heydarian, J.P. Carneiro, D. Gerber, B. Becerik-Gerber, Immersive virtual environments, understanding the impact of design features and occupant choice upon lighting for building performance, Build. Environ. 89 (2015) 217–228, https://doi.org/10.1016/j.buildenv.2015.02.038.
[59] Yvonne de Kort, Wijnand IJsselsteijn, Jolien Kooijman, Yvon Schuurmans, Virtual Laboratories: Comparability of Real and Virtual Environments for Environmental Psychology, PRESENCE: Virtual and Augmented Reality 12 (4) (2003) 360–373, https://doi.org/10.1162/105474603322391604.