Lighting Res. Technol. 2021; **53**: 701–725

Comparing perceptions of a dimmable LED lighting system between a real space and a virtual reality display

S Rockcastle PhD[a], **M Danell** MArch[a], **E Calabrese** MArch[a], **G Sollom-Brotherton** BA[a], **A Mahic** MArch[a], **K Van Den Wymelenberg** PhD[a] and **R Davis** PhD[b]
[a]Department of Architecture, University of Oregon, Eugene, OR, USA
[b]Pacific Northwest National Laboratories, Portland, USA

Received 10 January 2020; Revised 17 December 2020; Accepted 4 January 2021

Over the last several decades, designers have used digital screens to view images of real and simulated spaces and make critical design decisions. Screen technology has improved during this time, as technologies like OLED have replaced legacy displays (CRT, plasma, and LCD). These new screens provide a higher pixel resolution, luminous output and contrast ratio. Immersive head-mounted displays now allow designers to view immersive images, and recent developments in real-time rendering have encouraged the uptake of virtual reality (VR) head-mounted displays in mainstream practice and design education. This paper presents an experiment on lighting perception using a series of LED lighting conditions in a real space and a virtual representation of those conditions captured using a 360° high-dynamic-range camera and presented on an HTC Vive Pro HMD. Fifty-three participants were asked to rate each lighting condition by viewing it in a real space ($n = 30$) or via immersive HDR photographs displayed in a VR HMD ($n = 23$). The results show that ratings of *visual comfort*, *pleasantness*, *evenness*, *contrast* and *glare* are similar between the HTC Vive Pro HMD and our real space when evaluating well-lit scenes, but significant differences emerge in dim and highly contrasted scenes for a number of rating scales.

1. Introduction

Over the past several years, new screen displays have been developed to provide visual immersion through virtual reality (VR) head-mounted-displays (HMDs). During this time, VR HMDs have seen increased integration within professional practice, design education and research to offer seemingly more realistic visual and auditory experiences. While architects and engineers have used traditional visualization workflows like 2D rendering and walk-through animations for years, many architects and engineers are starting to use VR for increasingly more realistic and immersive representations of interior space.[1,2] This improved visual experience allows designers to engage clients and inform decisions about aesthetic and perceptual qualities of architectural space at a much lower cost than a full-scale mock-up.[3,4] Visual immersion through a VR HMD can also offer educational opportunities for students to test design concepts within a rendered 3D model and gain immediate feedback from a user perspective.[5,6]

Over the past few years, VR HMDs have rapidly improved in resolution, from 1280 × 800 pixels in the Oculus Rift DK1 to 2880 × 1600 pixels in the HTC Vive Pro. As the quality of these displays improves in

---

Address for correspondence: S Rockcastle, College of Design, School of Architecture & Environment, 5249 University of Oregon, Eugene, OR 97403-5249, USA.
Email: srockcas@uoregon.edu

10.1177/1477153521990039

resolution, brightness, refresh rate and contrast ratio, this technology has and will continue to offer additional value to researchers by providing a highly controlled environment for the reproduction of visual and auditory experiences and the recording of human behaviour such as eye-tracking and EEG brain activity.[7]

While the visual experience provided by current VR HMDs can appear highly realistic, there have only been a handful of studies that have compared the perceptual accuracy of scenes viewed in VR HMDs to real-world lighting scenarios.[8–12] As technological advances in both hardware and software continue, accessibility to these displays will increase – as will the number of lighting design decisions that may be impacted by immersive display technologies.[13] As such, it is timely and relevant to analyse the perceptual accuracy of immersive HMDs so that we may better understand the opportunities and limitations in their application within education, research and design decision-making.

Existing studies that have compared real-world space to their digital representations in VR have generally relied on well-lit daylit or electric lighting scenes and have not considered underlit or intentionally glaring scenes alongside those conditions.[8,10–12] This leaves a gap in our understanding of the perceptual accuracy of VR HMDs for lighting systems that offer broad tunability, such as LEDs and their associated controls. As we see a rise in the use of VR as an educational tool for lighting designers, it becomes critical that we understand the limitations of VR in reproducing inadequate (underlit and glaring) as well as adequate lighting conditions.[14] To be effective, a designer should be able to navigate an underlit or overlit 3D model before lighting systems are deployed and view the environment, as it would appear in the real world. They should also be able to view the broad range of lighting conditions that a LED system may produce to know whether a selected control system is appropriate.

While the above-mentioned studies found that VR offered a reasonable environment for research on visual perception for many, if not all of the conditions they tested, they leave an incomplete understanding of the limitations associated with a VR HMD for interior lighting applications, where critical design decisions may require the assessment of spaces that are underlit or where direct glare from luminaires may cause visual discomfort.

As the use of LED lighting systems continue to dominate the market, it is pertinent that displays being used to make qualitative decisions by a design team must be able to reproduce scene conditions that are created using these technologies. While researchers acknowledge the limited range of luminance that can be produced by these screens, professionals and educators are still using VR HMDs to visualize environments where dim or glaring lighting may occur. While these screens may still be useful in informing design decisions, we need to understand when their use may lead to an over-or-under specification of light or inadequate consideration of glare. By comparing data collected in real-world lighting scenes and virtual representations of those scenes in a VR HMD, we can offer a better understanding of the potentials and limitations associated with VR HMD devices and suggest when they should be used to make informed design decisions.

This paper presents the first phase of a two-part experiment; comparing subjective impressions collected in a physical electric lighting environment to those collected in an immersive representation of that environment using HDR photography and an HTC Vive Pro VR HMD. Lighting conditions for this experiment represent a range of illumination levels, from very dim to adequately bright and highly contrasted using recessed and track LED fixtures in a studio workplace setting, excluding daylight. Our survey asks participants to answer questions about visual comfort, pleasantness, evenness, brightness, contrast and glare perception for a range of lighting conditions. In addition to our subjective ratings, we recorded reading times for a visual acuity test in each scene. The conditions used in this experiment are produced using dimmable LED recessed and track lighting as well as a table-top task lamp with a LED source.

2. Background

There are a number of studies that have compared perceptual responses between real and virtual scenes on a range of digital display devices.[8,15–17] Newsham *et al.*[18] compared the ratings of images for a range of lighting conditions on 2D HDR and LDR screens. The authors found that participants rated scenes on the LDR display as more pleasant, more realistic, more comfortable, less glaring and less hostile than on the HDR display. The authors also compared these ratings to those collected in the real space during a previous study and concluded that despite the difference in luminance range between screens, images displayed on *both* the LDR and HDR screens could provide a reasonable representation of reality under many lighting conditions.

Contrary to the findings in Newsham *et al.*,[18] research completed by Akyuz *et al.*[19] found that participants generally preferred scenes presented on an HDR display to those presented on an LDR display when ranking images according to preference for all but 2 of the 10 scenes. Newsham *et al.*[20] found a similar trend when comparing responses in a real-world scene to images displayed on HDR and LDR displays. Images that contained large areas of high-luminance were rated as significantly more realistic on the HDR display than the LDR display.

The development of brighter displays has, in many ways, emerged as a response to advances in HDR photography or high dynamic range imaging (HDRI). HDRI allows researchers and designers to capture (or render) photometrically accurate scene data and analyse that data for metrics relating to brightness, contrast, or glare. Using software like Photosphere to generate, calibrate and compress HDR images, we are now able to capture lighting scenes within 10% error in luminance as compared to the real space.[21] Using HDRI, HDR displays allow for the reproduction and display of light levels across a much broader range than previously possible, improving perceived reality.[22]

With the development of brighter OLED displays and immersive imaging techniques, VR HMDs have emerged onto the market at a rapid pace. Recent studies have used VR HMDs to determine user preferences within a rendered or photographed space,[14,23–25] as well as to compare responses collected in a real-world space to those collected in digitally rendered or photographed scene.[8,12,26] A study by Abd-Alhamid *et al.*[10] compared ratings of electric lighting scenes captured using a quasi-stereoscopic HDR imaging technique to those in a real-world space. Although a single electric lighting condition was used in this study, there were no significant differences between the real-world and VR groups for ratings of brightness, distribution, interest, excitement and colour variety. Significant differences were noted for ratings of detail, contrast, colourfulness, spaciousness, complexity, colour temperature and pleasantness. The authors also found a significant effect of display device on task completion times, indicating a gap in task performance between real-world and VR scenes. The results from this study suggest that VR display technologies could adequately reproduce some subjective ratings as compared to real-world environments, but not task performance results.

Chen *et al.*[11] compared real-world spaces to those presented in VR using a mobile phone-based HMD, through video footage or through photography to differentiate ratings collected using these media. This experiment found that ratings were significantly correlated ($p < 0.5$) between real-world and VR scenes, but not between real-world and video or photograph. While the authors found significant differences between the real-world and VR for several attributes, VR still resulted in the closest perceptual match to real-world scenes, followed by video footage and then photographs. This order remained consistent for five of the six attributes; presence, reality, field depth and stereo, but not comfort. It should be noted that this study relied on a mobile phone-based HMD and not a tethered VR HMD, limiting the luminance range, pixel resolution and refresh rate of the images, which may contribute to the perceived differences between real-world and VR scenes.

With the rapid trajectory that VR lighting research is taking, a study by Chamilothori *et al.*[8] discusses the precautionary elements to consider when conducting experiments on lighting perception. The authors note that verbal questionnaires are preferred over written questionnaires due to presence of the headset and that symmetrical number scales (1–5) offer a more intuitive range for verbal delivery. Their findings also conclude the importance of tone-mapping and image compression in order to present the most perceptually accurate image based on the selected display device, as has been noted by other key studies.[16,27,28]

In summary, past studies have compared ratings of images presented on a range of different screen types, including LDR, HDR and VR and found mixed-results in terms of perceptual accuracy, preference or performance. Existing research that compared the reproducibility of lighting scenes in real-world and VR environments found that despite the limited luminance range of VR HMDs, VR can offer a reasonable proxy for reality in well-lit scenes. However, there is a lack of studies that have used underlit (dim) or intentionally contrasted (glaring) scene conditions which are easily achievable using LED lighting and continuous dimming controls. This paper explores the capacity of VR to reproduce perceptual ratings of electric lighting conditions as compared to a real space and understand the limits of VR as a design tool and human-factors research apparatus.

3. Methods

The space we selected for our experiment was located on the University of Oregon campus and was renovated in 2017 as part of the Department of Architecture for overflow studio space. At that time, the space was outfitted with new LED fixtures to provide tunable lighting for pin-ups, presentations and meetings. This section will introduce the selection of scene conditions, the methods used to capture each lighting condition, the post-processing workflow, display specifications, measurement capturing process and the design and delivery of the experiment. Statistical methods will be presented in Section 4.

3.1 Selection of scene conditions

To compare perceptual responses across a range of electric LED lighting conditions that include both very dim and highly contrasted scenes, the authors selected eight lighting conditions. Figure 1 shows these conditions and their names: ‘Neutral’, ‘All high’, ‘All low’, ‘Track high’, ‘Track low’, ‘Overhead high’, ‘Overhead low’ and ‘Task on’. These names were not shared during the experiment but are used in this paper to code each condition. Subjects always entered the experiment under the ‘neutral’ condition to allow for adaptation to a mid-level luminance range, while all subsequent conditions were shown to the subjects in random order. For this purpose, the neutral condition will sometimes be referred to as the starting condition.

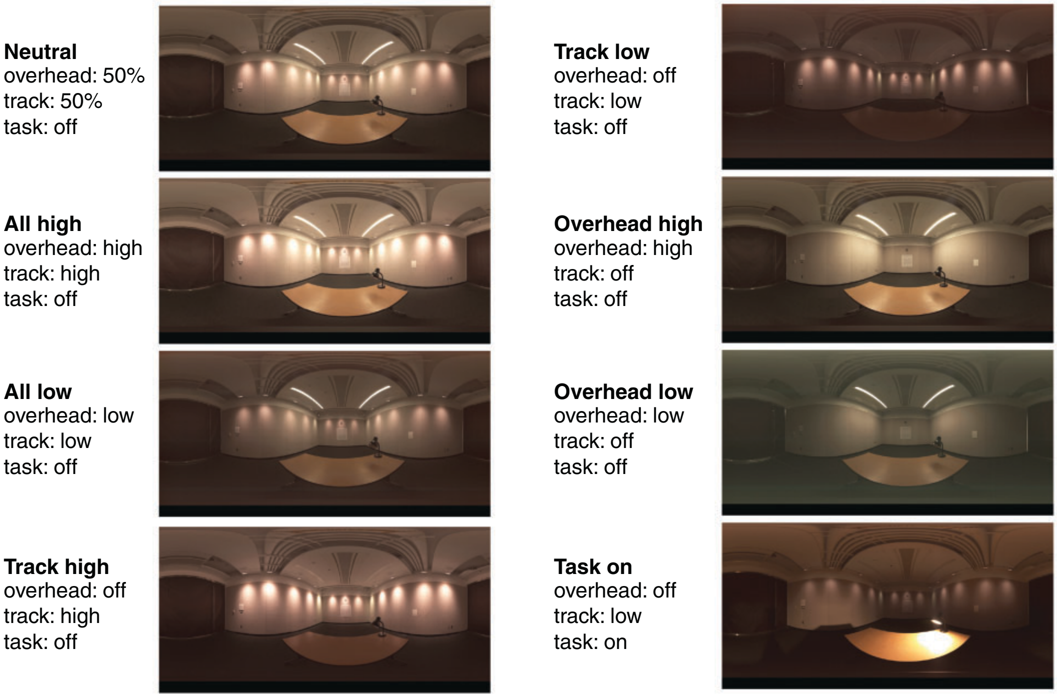


Figure 1. The eight lighting conditions; neutral, all high, all low, track high, track low, overhead high, overhead low and task on

Photosphere was used to measure the average luminance of each condition across the 180° field-of-view facing forward from the chair. Our conditions resulted in the following average luminance values: Neutral (9.31 cd/m^2), All high (14.8 cd/m^2), All low (0.76 cd/m^2), Track high (7.03 cd/m^2), Track low (0.30 cd/m^2), Overhead high (9.16 cd/m^2), Overhead low (0.46 cd/m2) and Task on (3 cd/m^2). While these average luminance values appear low, they include luminance values from across the full 180° field-of-view, including the black rubber flooring. To provide a better sense of the luminous conditions perceived by the participant in the space, a Minolta CL 500A meter was used to capture a 1 m horizontal grid of lux measurements across the real space for a selection of scene conditions. Figure 2(b) shows the location of measurements taken in the room, excluding those directly adjacent to the back wall and outside the field-of-view, each located 76 cm from the ground. For the neutral condition, the average illuminance was 171 lx, with a min of 95 lx and a max of 252 lx. The all high condition resulted in an average illuminance of 287 lx, with a min of 147 lx and a max of 412 lx, as reported in Figure 2(b). The all low condition had an average illuminance of 12 lx, with a min of 6 lx and a max of 17 lx.

3.2 HDR scene capture

To capture images for display on our VR HMD, we used a SPHERON-VR SCENECAM 2.0 camera to generate 360° × 180° HDR images. This camera captures a dynamic-range of 26 f-stops with 96-bit image data per pixel. Using the pre-scan setting to determine an appropriate lens length and

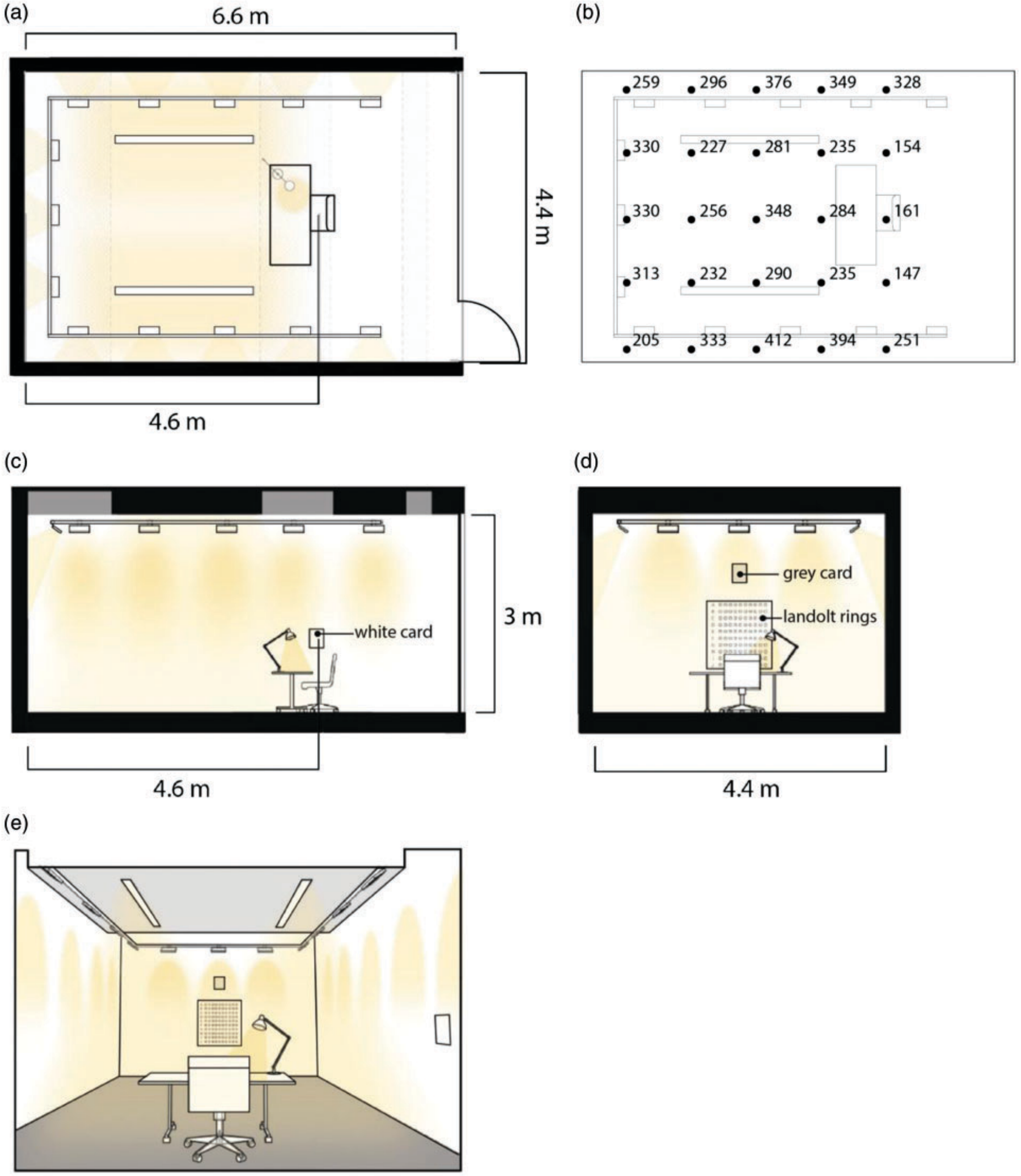


Figure 2. The dimensions and layout for our experimental space in (a) plan, (b) lux measurements for the 'all high' condition, (c) longitudinal section, (d) cross-section and (e) perspective

shutter speed for each lighting condition, we ran the settings in Table 1 to capture each condition using the 26-f-stop setting with lens-length set to infinity.

Using each scene scan took between 31 and 34 min. The use of variable aperture and shutter speeds allowed us to capture a 26-f-stop these settings range that is most

Table 1. SceneCam 2.0 settings for lens aperture and shutter speed

Condition	Lens aperture	Shutter speed
Neutral	4	1/15
All high	5.6	1/15
All low	2.8	1/8
Track high	4	1/8
Track low	2.8	1/8
Over. high	4	1/8
Over. low	2.8	1/8
Task on	2.8	1/8

Figure 3. Our spectrophotometer (Minolta CL 500A) mounted in the HTC Vive Pro HMD. This apparatus was rotated to capture readings in 45° angular increments for each lighting condition

appropriate for each lighting condition. At the time of capture, we used a luminance meter (Minolta LS-160) to take physical luminance measurements of grey and white cards mounted to the front and side walls in the space. The grey card located above the visual acuity test was centred at 2.1 m from the ground and approximately 4.6 m from the right wall. The white card was located 4.6 m from the front wall and approximately 1.1 m from the ground to be directly perpendicular and at sitting head height of participants for image capturing (Figure 2(c) and (d)). We calibrated each HDR image to the grey-card measurements using Photosphere and calculated the error between grey and white-card readings. This error ranged from 1% to 5% across all eight lighting conditions, which is aligned with previously reported accuracies for HDR capture methods.[21]

3.3 Image-processing workflow

There is a lack of tone-mapping operators (TMOs) developed specifically for the representation of photometrically accurate and immersive HDR images displayed in VR. Despite this limitation, the authors considered multiple TMOs for use in this experiment, including Reinhard *et al*.[29] and PCOND via RADIANCE.[30] The Reinhard TMO, used in Abd-Alhamid *et al*.,[10] was found by our authors to produce odd colour artefacts from wall-washing track LEDs in our HDR images. After trying multiple operators and settings, we decided to use the PCOND TMO to compress the luminance values from an HDR range to the range of the HTC Vive Pro. As discussed by Chamilothori *et al*.,[8] future work is needed to understand whether an alternative algorithm is more appropriate for immersive display technologies.

While PCOND offers a number of options, the authors used a *linear* (-l) rather than *human* (-s) compression when the human contrast sensitivity setting (-s) was found to create noticeable artefacts on the wall and desk from both the track and task sources. The max luminance (-u) was set to the recorded range of the HTC Vive Pro as reported in Section 3.4 Given the generally low dynamic range of our scenes (i.e. no daylight), the authors determined that a linear compression was acceptable for our images.

3.4 Scenes in virtual reality

The authors selected the HTC Vive Pro HMD for use in this experiment based on advances in pixel resolution (1440 × 1600 pixels per eye) over its predecessor the HTC Vive and competitors like the Oculus Rift CV1. The HTC Vive Pro uses a 3.5″ diagonal AMOLED display with a 90 Hz refresh rate, 150° horizontal and 110° vertical field-of-view. Using a Minolta CL 500 A spectrophotometer facing into the device within a dark room (Figure 3), we measured a max luminance (with a white image: RGB 255, 255, 255) of 135.5 cd/m^2 with a control minimum of 0.0 cd/m^2 (with a black image: RGB 0, 0, 0). These images were displayed using the UNITY gaming engine, and the contrast ratio was calculated at 2711.

To project scenes within our VR HMD, we saved calibrated and tone-mapped HDR images as .bmp files with a gamma correction of 2.2. The authors used the process outlined in Abd-Alhamid *et al.*[10] to compute the appropriate gamma curve for our device and found that like the Oculus Rift CV1 and the HTC Vive, the HTC Vive Pro also uses a 2.2 correction value. These images were then loaded into the UNITY gaming engine and assigned to a spherical skybox with keypress controls for each scene so that individual scenes could be selected and controlled in our VR HMD. This method is adopted from Rockcastle *et al.*[24] and Chamilothori *et al.*[8] but uses the Vive rather than the Oculus platform. It should be noted that our scenes were monoscopic rather than stereoscopic, meaning that both eyes saw the same image. At this time, there are no HDR cameras that are capable of capturing a continuous 26 f-stop 360° HDR scan for two eyes simultaneously. Workflows like those introduced in Abd-Alhamid *et al.*[10] have used an SLR camera to stitch together images and offset the camera tripod to capture two eyes, but these scenes are only stereoscopic in one direction, as the tripod does not rotate with an offset. Workflows used by Chamilothori *et al.*[8] do produce a stereoscopic projection, but use renderings via scripts developed by McNeil[31] for Radiance.

3.5 Comparison of spectral intensity

While a rigorous method was used to calibrate, tone-map and compress each HDR image to the HTC Vive Pro HMD, we wanted to understand the VR headset's capacity to reproduce measurable lighting levels within the scene. To compare the intensity and spectrum of light received at eye level in both the real and VR scenes, a spectrophotometer (Minolta CL 500 A) was used to take measurements across 45° radial increments. Using a tri-pod in the real space, we mounted this spectrophotometer at seated eye-level (1.07 m) and took eight radial measurements in each lighting condition. We then captured the same directional readings inside the left eye of the HTC Vive Pro HMD for each lighting condition in our experimental space with all exterior lights turned off and our laptop located behind a blackout curtain (Figure 3). These results will be presented and discussed in Section 4.

3.6 Design and delivery of experiment

The design of this experiment was built using a between-groups comparison to reduce the impact of order bias and allow for greater flexibility in recruitment by limiting an experimental session to one 25-min period. Subjects were recruited via email, social media (Facebook, Instagram), flyer and digital advertisements displayed on screens around campus at the University of Oregon between 5 August and 20 August 2019. Participants were recruited using a 25-dollar incentive as compensation for their time and were asked to schedule a preferred time slot via an online scheduling poll.

Individuals that signed up on the same day were then assigned to either the real space or VR group. We did not randomize their assignment to a group, but instead ran the real space group first due to room availability (5 August–9 August 2019). Subjects were not aware of which group they would be assigned to, as this information was suppressed in the online poll. Subjects were then notified via email of their confirmed appointment time and were asked to review the consent forms in advance of their arrival and contact the researcher with any questions. We advised any participants with a history of psychiatric disorders, heart conditions, or any other serious medical conditions to contact us so that we may evaluate the safety of their participation. Any acknowledged risks were associated with the VR group and participants that contacted us with a psychiatric disorder, heart condition or any other serious medical condition were moved into a different group. Our project received an exempt determination via internal IRB review on 9 May 2019 (protocol: 03292019.025).

At the time of their appointment, participants were asked to review and sign a consent form in a waiting space adjacent to the real space or the VR lab. Both spaces were climate-controlled and contained overhead lighting with limited access to exterior windows. Once the consent form was signed, the researcher provided the participant with a tablet to answer a series of demographic questions related to age, gender and use of corrective lenses. At the end of this short survey, participants were asked to hand the tablet back to the researcher and follow them into the real room or VR lab.

The real space in our study is 6.6 m × 4.6 m in plan with 3 m from the floor to the underside of the suspended ceiling as shown in Figure 2. Walls, pin-up surfaces and ceiling are painted white and the floor is black rubber. Two recessed LED linear fixtures were located in the suspended ceiling, and a series of adjustable LED track luminaires were mounted along the wall and oriented to wash light down onto the wall panels.

A single desk was positioned in the centre of the room facing the back wall with a desk lamp in the right-hand corner. The desk lamp contained an A19 LED (3200 CCT) lamp. All participants in the real group entered the space under a neutral lighting condition which consisted of track and overhead lights on at 50% intensity (step 10 of a 20-step dimming control) and the desk lamp turned off. On the wall facing the participant, a poster was mounted with 10 rows of Landolt rings, as shown in Figure 4. Our rings were spaced 24 mm apart and sized to be legible by someone with a moderately low vision capacity of 5/32.[32–34]

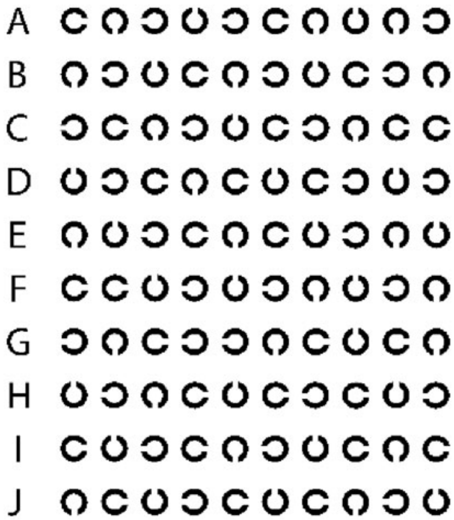


Figure 4. Landolt Rings spaced 24-mm apart and sized relative to a moderately low vision of 5/32[34]

Upon entering the real space, participants were asked to sit down in the chair under the neutral or starting lighting conditions while the researcher explained how to read the Landolt rings from left to right, indicating the side the missing segment was located (in other words, which direction the opening of the ‘C’ was facing). The researcher then initiated the survey on the tablet and asked participants to verbally respond to questions while presenting each of the lighting conditions in a random order. While the neutral condition was always shown first upon entering the room to allow for adaptation to scene lighting, all subsequent lighting conditions were randomized using survey logic. Based on the random order of lighting conditions presented in the survey, the researcher was instructed to change the lighting condition. Between conditions, participants were asked to close their eyes until indicated by the researcher.

Upon opening their eyes in each new lighting scene, participants were read a script instructing them on how to read the Landolt rings on the paper mounted on the back wall. While repetitive, this gave participants around 20 s to adapt to the lighting condition before responding to the first question. The first and last questions in each lighting condition were not randomized. The first question consisted of the Landolt ring test, where participants were instructed to read a designated line from left to right as fast as possible out loud without straining themselves. The line was determined by the lighting condition so that each line was read only once. Researchers recorded the time it took for participants to read the row from start to finish (regardless of errors) as a means to test the visual acuity of the back wall in each lighting condition using a digital stopwatch. While the number of errors were recorded, very few were reported within our data set and the authors decided to use reading time rather than error count as a measure of task performance. The researchers noted difficulty in accurately noting errors while recording reading time and as a result, decided not to analyse error count to avoid conclusions based on potential human error. The last question then asked participants, ‘Would you describe the light in this scene as glaring?’ If they answered yes, an additional piped question asked: ‘On a scale from 1 to 5, would you rate this glare as’:

1) imperceptible
2) perceptible
3) unacceptable
4) uncomfortable
5) intolerable

This was based on Hopkinson’s 9-level glare-rating scale but reduced down to 5-levels for ease of comparison to other questions using a 5-point scale and to reduce verbal fatigue in delivery.[35] The remainder of the questions were delivered in random order and asked participants:

(a) How visually comfortable is the light in this space, with 1 being uncomfortable and 5 being comfortable?
(b) How evenly distributed is the light in this space, with 1 being uneven and 5 being even?
(c) How bright is the light in this space, with 1 being dim and 5 being bright?
(d) How contrasted is the light in this space, with 1 being low contrast and 5 being high contrast?
(e) How pleasant is the light in the space, with 1 being unpleasant and 5 being pleasant?

While the protocol was similar for both real space and VR groups, there were a few key differences which are outlined below. Participants in the VR group were brought into the test space and asked to sit in a chair in the middle of a room surrounded by black-out curtains. The overhead lights were initially left on. Participants were then handed the VR HMD and the researcher helped to adjust the head-straps for the participant to ensure a comfortable fit. Participants were then advised on how to adjust the space between the lenses until the image was clear. Upon wearing the HMD, participants viewed the neutral starting condition, same as in the real group. Overhead lights were then turned off, and instructions on how to complete the Landolt ring test were read to the participant, same as in the real group. Rather than closing their eyes between conditions (as they were asked to do in the real space), participants were placed in a fully black scene while the researcher advanced the survey to the next page. Instructions for the Landolt ring test were always repeated at the start of each condition (in both the real and the VR groups) to allow the participant 20 s to adapt to the scene. At the end of each session, participants were given their compensation in an envelope and thanked for their time.

4. Results

4.1 Population statistics

There were 30 participants in the real group and 23 in the virtual group. In the real group, the population was 11 male, 18 female and 1 other, while the VR group was 8 male and 15 female (as reported by the participants). Age was evenly distributed between groups, with ~80% between 18 and 33 and ~20% between 34 and 58, which reflects the demographics of our university setting. As reported by the participants at the time of the experiment, 13 participants in the real group were wearing corrective lenses (glasses or contacts) and 17 were not. In the VR group, 8 participants were wearing corrective lenses and 15 were not. Participants in the VR group were able to and encouraged to wear their glasses inside the VR HMD. Table 2 shows *t*-test results for gender and corrective lens use for each rating scale (comfort, evenness, brightness, contrast, pleasantness and glare rating) and display group (real or VR).

While the glare rating-scale in our experiment was between 1-imperceptible and 5-intolerable, not all participants rated glare for each image because this rating was only available for those who indicated ‘yes’ to the question: ‘Would you describe the light in this scene as glaring?’ For those participants who indicated ‘no’ to the screener question, we assigned a rank of 0, which explains why some mean ratings are <1 in the analysis in Tables 4 and 5. This hybrid scale was used to maintain the sample size in our analysis.

These *t*-tests cluster the ratings for all lighting conditions together and it should be mentioned that values are not corrected for multiple comparisons so that each null-hypothesis can be evaluated in relative terms. While we do not see a significant ($p < 0.05$) effect of gender on the means of any rating scales for the real or VR groups in Tables 2 and 3, there is a small ($0.2 <$ cohen_d < 0.5) but significant ($p < 0.05$) effect of corrective lens use for the mean values of multiple rating scales in both the real and the VR groups in Tables 4 and 5.[36]

For those wearing corrective lenses, the real group resulted in small, but significantly lower mean rating of evenness and a higher mean rating of glare (Table 5). The VR group also resulted in small, but significant lower mean ratings of comfort and evenness and a higher mean rating of glare Table 4). The authors did not ask for clarification on whether the corrective lenses were glasses, contact lenses, or what the strength of such correction was, and so we cannot determine whether this effect was due to the type of lens a participant was wearing.

Despite these observations, there are some obvious limitations to this analysis. Given the

Table 2. *t*-test results for the effect of gender on mean ratings in the real group

Rating	Comfort		Evenness		Bright		Contrast		Pleasant		Glare rating	
Gender	M	F	M	F	M	F	M	F	M	F	M	F
Mean	2.82	2.90	3.02	2.99	2.52	2.67	2.60	2.79	2.69	2.76	0.92	0.66
Variance	1.60	1.87	1.63	2.03	1.42	1.56	1.41	1.87	1.43	1.67	2.26	2.03
df	230		230		230		230		230		230	
t-stat	0.43		0.20		0.91		1.07		0.37		1.32	
p-value	0.67		0.84		0.37		0.28		0.71		0.19	
cohen_d	0.06		0.03		0.12		0.15		0.05		0.18	

Note: These results assume a two-tail analysis with equal variance and collapse all lighting conditions into one. The sample is male = 11, female = 18.

sample sizes for gender and lens use in our participant population, the power of our *t*-tests is less than 0.35 for Tables 2 and 5 and does not reach the desired power of 0.80 for small effect sizes. We would need a population of at least 50 for each gender and lens use type to reach of power of 0.80 with a small effect size ($0.2 < \text{cohen\_d} < 0.5$). Furthermore, as the authors did not correct individual p-values for multiple comparisons, these findings may over-estimate the type-1 error for the family of comparisons. As such, we

Table 3. *t*-test results for the effect of gender on mean ratings in the VR group

Rating	Comfort		Evenness		Bright		Contrast		Pleasant		Glare rating	
Gender	M	F	M	F	M	F	M	F	M	F	M	F
Mean	3.05	2.78	3.03	3.13	2.81	3.00	2.81	2.88	3.06	2.74	0.84	1.05
Variance	1.44	1.75	1.55	1.79	1.27	1.50	1.17	1.47	1.30	1.81	1.75	2.25
df	182		182		182		182		182		182	
t-stat	1.328		0.463		1.018		0.345		1.623		0.924	
p-value	0.186		0.644		0.310		0.730		0.106		0.356	
cohen_d	**0.206***		0.072		0.158		0.053		**0.251***		0.143	

Note: These results assume a two-tail analysis with equal variance and collapse all lighting conditions into one. The sample is male = 8, female = 15. *Effect sizes (cohen_d) greater than 0.2.

Table 4. *t*-test results for the effect of corrective lens use on mean ratings in the real group

Rating	Comfort		Evenness		Bright		Contrast		Pleasant		Glare rating	
Lens use	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
Mean	2.70	3.01	2.77	3.18	2.46	2.75	2.66	2.79	2.59	2.88	1.04	0.58
Variance	1.63	1.82	1.87	1.75	1.53	1.52	1.82	1.59	1.49	1.61	2.66	1.79
df	238		238		238		238		238		238	
t-stat	1.821		2.371		1.792		0.772		1.776		2.388	
p-value	0.070		**0.019***		0.074		0.441		0.077		**0.018***	
cohen_d	**0.237***		**0.309***		**0.233***		0.101		**0.231***		**0.311***	

Note: These results assume a two-tail analysis with equal variance and collapse all lighting conditions into one. The sample is corrective lens = 13, no corrective lens = 17. **p*-values less than 0.05 and effect sizes (cohen_d) greater than 0.2.

Table 5. *t*-test results for the effect of corrective lens use on mean ratings in the VR group

Rating	Comfort		Evenness		Bright		Contrast		Pleasant		Glare rating	
Lens use	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
Mean	2.59	3.03	2.80	3.25	2.91	2.95	2.78	2.89	2.69	2.94	1.42	0.78
Variance	1.67	1.59	1.56	1.72	1.67	1.29	1.60	1.24	1.65	1.64	2.88	1.59
df	182		182		182		182		182		182	
t-stat	2.191		2.269		0.237		0.610		1.283		2.929	
p-value	**0.030***		**0.024***		0.813		0.542		0.201		**0.004***	
cohen_d	**0.339***		**0.351***		0.037		0.094		0.199		**0.453***	

Note: These results assume a two-tail analysis with equal variance and collapse all lighting conditions into one. The sample is corrective lens=8, no corrective lens = 15).
*p-values less than 0.05 and effect sizes (cohen_d) greater than 0.2.

cannot say whether the effect of lens use was significant overall, despite the strong suggestion that it may impact glare ratings for the real and VR display groups, in particular. In either case, it would be interesting to explore the impact of corrective lens use in more depth for future studies on lighting perception.

4.2 Mean rating and distribution

Figure 5(a) shows a box plot for each rating, lighting condition and display type with responses in the real space shown in solid grey and the VR screen shown in dashed black. Figure 5(b) shows the mean ratings for each display type to illustrate rating trends for each question and lighting condition. While there are differences between mean ratings for some lighting conditions in each question, the overall response trends are similar independently of the display type. In other words, the scenes that are rated as relatively high or low for any single question are similar between real space and VR groups. Systematic group differences are observed in the question about brightness, where the average across conditions resulted in a higher mean brightness perception rating (2.93) in the VR lab group than in the real space group (2.62). It is useful to note those conditions and questions which produced a larger spread in subjective perceptual ratings. Both brightness and contrast show a broad distribution of ratings for the task on condition, where the task light is turned on within the field-of-view. While people consistently rated that condition as more contrasted in the real space, there is also more spread in the distribution of responses for the VR group.

Box plots show higher comfortable ratings for the bright scene conditions (all high and track high in the real space as compared to the VR display. The distribution of ratings for pleasantness are also more spread for the all high and task on condition in VR as compared to the real space. Regardless of lighting condition, the biggest mean differences were observed for the question regarding brightness, where five of eight conditions (including the neutral) were rated as brighter in the VR lab group than the real space. Responses were not normally distributed in our analyses, as shown in Figure 6 and as indicated by a Shapiro-Wilk test. This matrix of distribution plots reveals the difference in distribution between real space and VR lab groups, with normal distribution indicated by a dashed (rather than dotted) line. Some of these differences show a clear shift up or down in the frequency of ratings and others show a bimodal relationship.

For the neutral, overhead high and overhead low conditions, we can see a shift in distribution for the frequency of brightness ratings. A higher frequency of subjects rated these conditions as brighter in the VR lab group than in the real space group. In contrast, if we look at the track low and all low conditions for ratings of pleasantness, the responses to the VR display produce a bimodal distribution. In this case, we can see that subjects rating the track low condition in VR found the condition to be either low (2) or high (4) in pleasantness, whereas subjects viewing the same condition in the real space found the space to be in the middle (3) of the five-point scale. This non-normally distributed data determined our use of a non-parametric test for our statistical analysis in Section 4.3.

4.3 Mann-Whitney test

Table 6 shows Mann Whitney Test results for each lighting scene with ratings from the real and VR groups, treated as two independent samples. The table includes associated *U*-values, median values, p-values (for a two-tail analysis), effect size (effect_r) and *z*-scores for each scene and rating. Scenes and ratings with significant *p*-values are reported in bold $*p < 0.05$ or $**p < 0.01$. Effect sizes are interpreted based on Cohen, 1988.[37] While it is

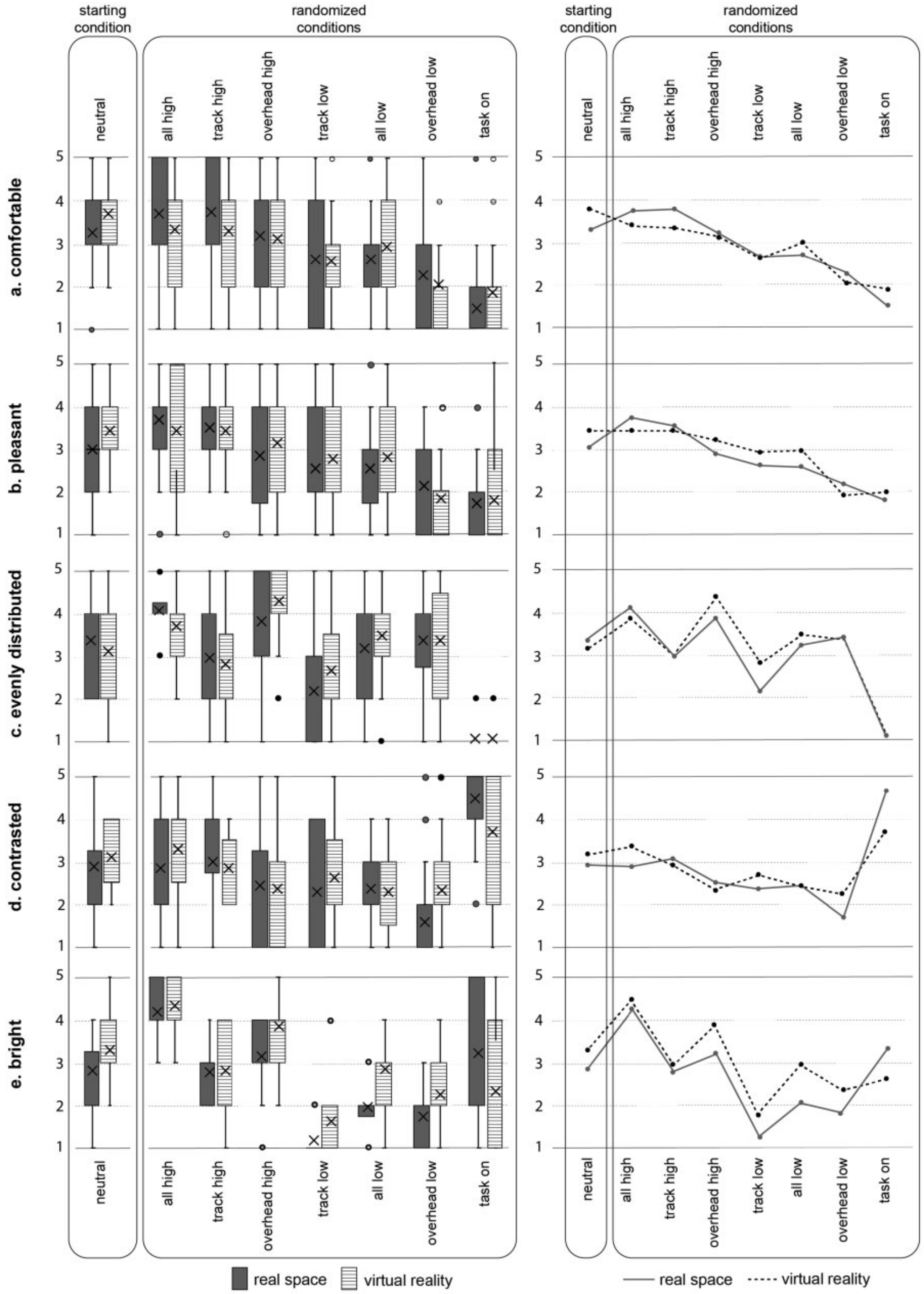


Figure 5. Box plots showing the distribution of subject responses for each question on the left and mean scores for each condition and question on the right. Real spaces are in dark grey and VR scenes are in hatch and dashed line. (a) Comfortable, (b) pleasant, (c) evenly distributed, (d) contrasted and (e) bright

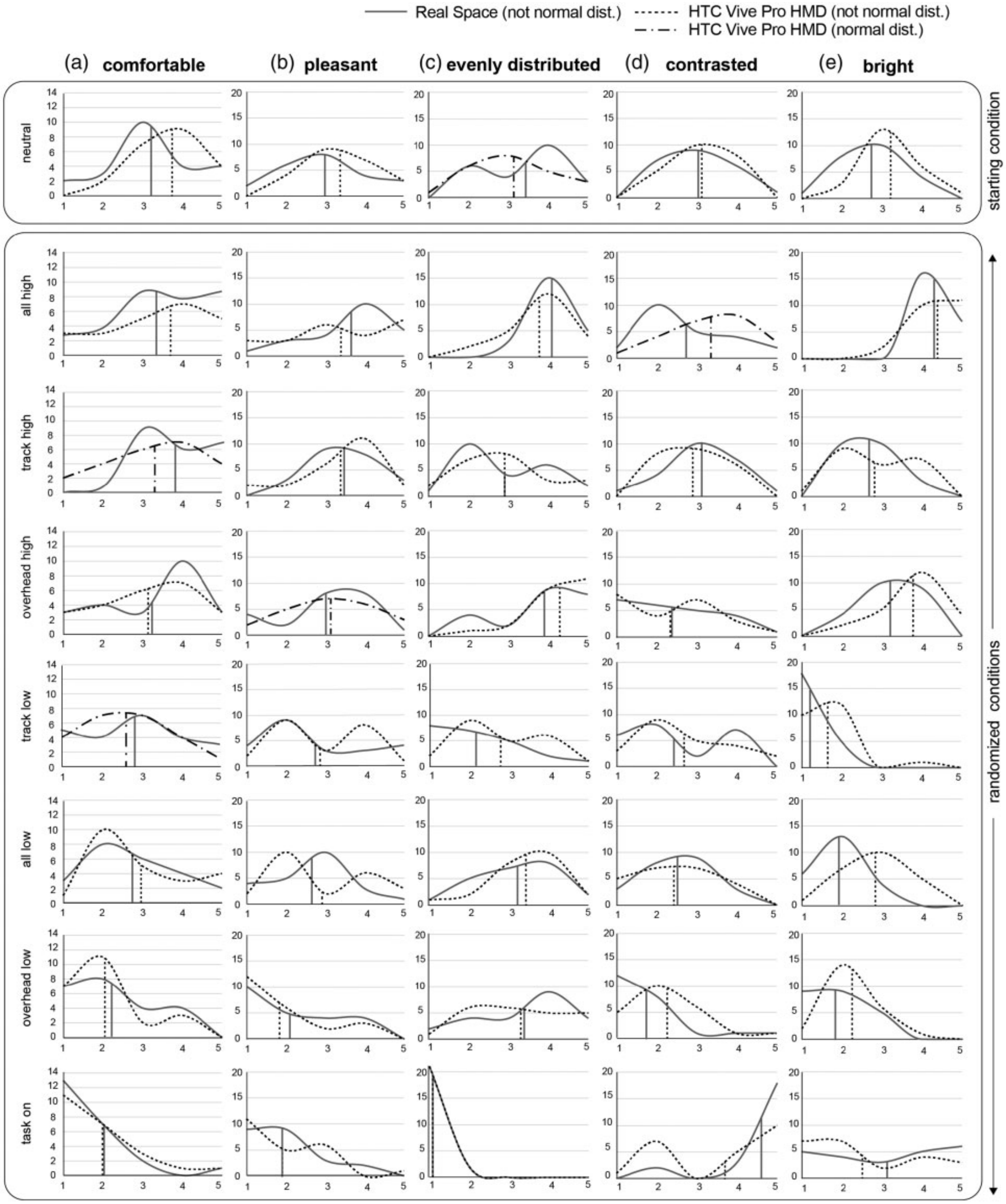


Figure 6. Frequency distribution plots for ratings in real (solid grey line) and virtual (dotted and dashed black lines) groups. Mean values are indicated by a vertical line. Normal distributions are indicated by a line with alternating dash and dot (as calculated using the Shapiro-Wilk Test, with values <0.05). Plots are ordered by question with (a) comfortable, (b) pleasant, (c) evenly distributed, (d) contrasted and (e) bright. The distribution lines for both real and virtual scenes in the 'task on' condition for (c) evenly distributed were indistinguishable

Table 6. Mann-Whitney non-parametric test for two independent samples: real ($n=30$) and VR ($n=23$) groups. Significant *p*-values are indicated in bold (with $p<0.05$* and $p<0.01$**). *p*-values are not corrected for multiple comparisons

Ratings	Mann-Whitney results	Neutral	All high	Track high	Overhead high	Track low	All low	Overhead low	Task on
Comfortable	*U*-value	233	294	273	330	344.5	309	314.5	273
	median (real/VR)	3/4	4/4	4/3	4/3	3/3	3/3	2/2	1/2
	p-value (2-tails)	0.12	0.36	0.20	0.79	1	0.52	0.59	0.20
	effect_r	0.21	0.12	0.18	0.04	0.00	0.08	0.07	0.18
	z-score	1.53	0.91	1.28	0.26	0.00	0.64	0.54	1.28
Pleasant	*U*-value	276	308	340.5	306	290	292.5	291.5	318.5
	median (real/VR)	3/3	4/3	4/4	3/3	2/3	3/2	2/1	1.5/2
	p-value (2-tails)	0.22	0.51	0.94	0.49	0.33	0.35	0.34	0.64
	effect_r	0.17	0.09	0.01	0.09	0.13	0.13	0.13	0.06
	z-score	1.23	0.66	0.07	0.69	0.98	0.93	0.95	0.47
Even	*U*-value	306	286.5	331.5	276	233.5	302	329.5	338
	median (real/VR)	4/3	4/4	3/3	4/4	2/3	3/4	4/3	1/1
	p-value (2-tails)	0.49	0.29	0.82	0.22	**0.05***	0.44	0.79	0.91
	effect_r	0.09	0.14	0.03	0.17	0.27	0.1	0.04	0.02
	z-score	0.69	1.04	0.23	1.23	1.99	0.76	0.27	0.12
Contrasted	*U*-value	282	252.5	319.5	333.5	278	333.5	197	230.5
	median (real/VR)	3/3	3/3	3/3	2/2	2/2	2/2	1/2	5/4
	p-value (2-tails)	0.26	0.10	0.65	0.84	0.23	0.84	**0.01****	**0.04***
	effect_r	0.15	0.23	0.06	0.03	0.16	0.03	0.37	0.28
	z-score	1.21	1.65	0.45	0.20	1.19	0.20	2.65	2.05
Bright	*U*-value	254	282	332.5	21	205	158	218	255
	median (real/VR)	3/3	4/4	3/3	3/4	1/2	2/3	2/2	3.5/2
	p-value (2-tails)	0.10	0.26	0.83	**0.02***	**0.01****	**0.00****	**0.02***	0.11
	effect_r	0.22	0.15	0.03	0.32	0.35	0.46	0.32	0.22
	z-score	1.62	1.12	0.22	2.34	2.5	3.35	2.27	1.60
Glare rating	*U*-value	344.5	267.5	291	301	322	317	228.5	208
	median (real/VR)	1/1	1/1	1/1	1/2	1/1	1/1	1/1	4/3
	p-value (2-tails)	1	0.17	0.33	0.44	0.69	0.62	**0.04***	**0.01****
	effect_r	0.00	0.19	0.13	0.11	0.05	0.07	0.29	0.34
	z-score	0.00	1.38	0.96	0.78	0.40	0.49	2.08	2.45
Reading time	*U*-value	229	230	205.5	212	177	197	191	151.5
	median (real/VR)	5.9/7.5	5.9/6.6	5.7/7.3	5.8/6.8	6.1/7.5	5.9/7.4	6.0/8.0	6.1/7.5
	p-value (2-tails)	**0.04***	**0.04***	**0.01****	**0.02***	**0.00****	**0.01****	**0.01****	**0.00****
	effect_r	0.28	0.28	0.34	0.33	0.41	0.36	0.38	0.38
	z-score	2.07	2.05	2.49	2.34	3.01	2.64	2.75	3.46

**p*-values of ≤ 0.05, ***p*-values of ≤ 0.01.

good practice to correct p-values when multiple comparisons are conducted on the same sample, we chose not to apply the Bonferroni adjustment in order to reveal the relative effect of display type on each rating scale for *each* lighting condition. A universal null-hypothesis is not considered in this paper.

It should be noted that not all participants rated glare for each condition, because those who indicated that there was no glare in the first question skipped the ranking question. We therefore produced a hybrid mean rating for the purpose of this analysis and that presented in Section 4.5. Those participants who indicated that there was no glare based on the first glare question, we assigned a rank of 1. Participants who indicated that there was glare but then went on to rank that glare as 1-imperceptible were also assigned a rank of 1. All other rankings between 2-perceptible and 5-intolerable were included without adjustment. This hybrid glare rating scale is slightly different than the one presented in Table 2, where participants who answered ‘no’ to the screener question were assigned a rank of 0.

The results reveal that for questions regarding comfort and pleasantness, there are no significant differences between the real space and VR lab groups for any of the lighting conditions under consideration. The question about evenness resulted in one significant difference ($p = 0.05$) with a small effect ($r = 0.27$) between groups for the track low condition, while ratings for contrast resulted in two significant differences between groups for the overhead low ($p = 0.01$, effect_r = 0.37) and task on ($p = 0.04$, effect_r = 0.28) conditions. The brightness question resulted in four significant differences for the overhead high ($p = 0.02$, effect_r = 0.32), track low ($p = 0.01$, effect_r = ,0.35) all low ($p = 0.00$, effect_r = 0.46) and overhead low ($p = 0.02$, effect_r = 0.32) conditions. The *glare rating* resulted in two significant differences between groups for the overhead low ($p = 0.04$, effect_r = 0.29) and task on ($p = 0.01$, effect_r = 0.34) conditions. Results for reading time in the visual acuity test will be discussed in the Section 4.4.

The limited horizontal (150°) and vertical (110°) view angle of the HTC Vive Pro is thought to contribute to the difference in ratings for contrasted and glare rating questions because subjects who adjusted their view away from the lamp on the right side of the desk could eliminate the source entirely from their field-of-view in the task on condition. This resulted in the most significant differences in ratings between the real and VR groups. The LED source on the desk was much brighter and persistent in the peripheral field-of-view when observed in the real space and may reveal a limitation of the limited field-of-view in current HMD headsets like the HTC Vive Pro. New wide-angle headsets like the StarVR One XT, with a 210° horizontal field-of-view, may result in better glare perception even without higher luminance range (https://www.starvr.com/).

4.4 Visual acuity

As for visual acuity, reading times for the Landolt ring test in the VR lab group took an average of 1.1 s longer than in the real space group across all scenes. The results of the Mann-Whitney non-parametric test for two independent samples in Table 6 show this difference to be statistically significant for all eight scenes, including the starting neutral condition. This difference resulted in a medium effect size ($0.3 < \text{effect\_r} > 0.5$) and $p < 0.01$ for the track high, track low, all low, overhead low and task on conditions and a small effect ($\text{effect\_r} < 0.03$) and $0.01 < p < 0.05$ for neutral, all high and overhead high conditions. The systematic difference between real space and VR groups may be due to the pixel resolution of the VR HMD. While the resolution of HTC Vive Pro was state of the art (1140 × 1600 pixels per eye) at the time of this experiment, the researchers noted that participants moved their head slightly up and down to read the rings in response to haloing effects of the screen.

That being said, the conditions that achieved the most significant difference in reading times with the largest effect size (track high, track low, all low, overhead low and task on) were also those with the lowest average scene luminance (reported in Section 3.1). Further research is needed to determine whether VR can reproduce an acceptable task environment for reading small symbols and text, but these results suggest that reading speed is impacted more in VR than in a real space under dim lighting conditions. It should be noted that Abd-Alhamid *et al.*[10] also found that participants took longer on visual tasks in VR than in the real space, although their task was not located in the same orientation. The purpose of including this test in our study was to expose a participant’s ability to decipher patterns and objects in VR at the same speed as in a real space, although recent work by Radianti *et al.*[38] discusses the potential use of VR as a surrogate educational environment. For VR to offer an equal learning environment to a real classroom, the results of our study suggest that tasks should be located on adequately bright surfaces and that additional work may still be needed to improve the pixel resolution of those tasks within the scene.

4.5 Glare count and glare rank

The results for glare count (Q: Is the light in the scene glaring? ‘Yes/No’) and glare rank (If ‘Yes’ then rate this glare from imperceptible to intolerable on a five-point scale) reveal notable differences. Figure 7(a) shows the percentage of people who responded ‘yes’ to the screener question about whether the scene was glaring. Our results show that seven out of eight scenes resulted in a higher percentage of participants who said the scene was glaring in the VR group than in the real group, which is particularly interesting because intensity levels of the light sources in the real space were found to be brighter than the light measured from the HTC Vive Pro HMD (discussed in Section 4.5). The task on condition was the only condition where fewer people rated the light as glaring in the VR group (91%) than in the real space group (100%).

As mentioned in Section 4.3, the modified ranking used in Table 3 and Figure 7(b) allows us to compare responses in the real space and the VR display using the modified Hopkinson scale for all scene conditions and participants. For seven out of the eight conditions, the difference in mean rank was less than 1 point on the five-point scale. If we look at glare count alone, we might conclude that there are obvious differences between the real space and VR lab groups, but when we ask people to rate the glare in each scene, the mean rating and Mann-Whitney test for that rating does not show significant difference between the real space and VR lab groups for six out of eight of the lighting conditions. According to the Mann-Whitney Test in Table 3, there were no significant differences between real and VR groups for all but the dimmest (overhead low) and the most contrasted (task on) lighting condition. To understand this phenomenon in more depth, a broader range of glare-inducing lighting conditions, including variations in source position would need to be studied.

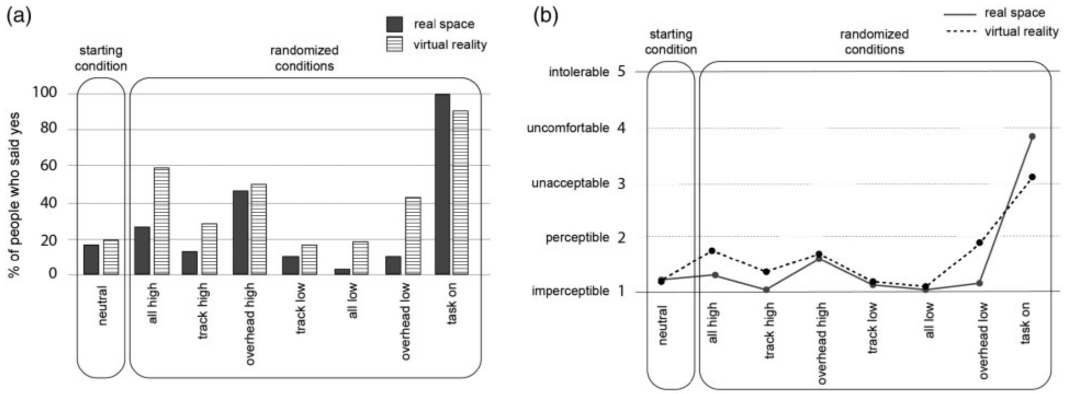


Figure 7. (a) The percentage of people who said the lighting condition was glaring between real and virtual reality groups and (b) the average glare rating (from 1 to 3) for those who responded that glare was present

4.6 Intensity and spectral power distribution

Figure 8 shows a comparison of vertical illuminance (lux) across a (a) 180° vertical and horizontal viewing angle and (b) a single steradian in both the real space and the VR HMD for a selection of lighting conditions. Figure 8(c) shows the spectral power distribution for the front-facing view direction (also across a 180° vertical and horizontal view angle). We chose to show lux values for both a 180° vertical and horizontal view angle (Figure 8(a)) and for a single steradian (Figure 8(b)) to compare the overall difference in brightness in the real space and VR HMD to more localized differences that can vary greatly depending on the source and view direction.

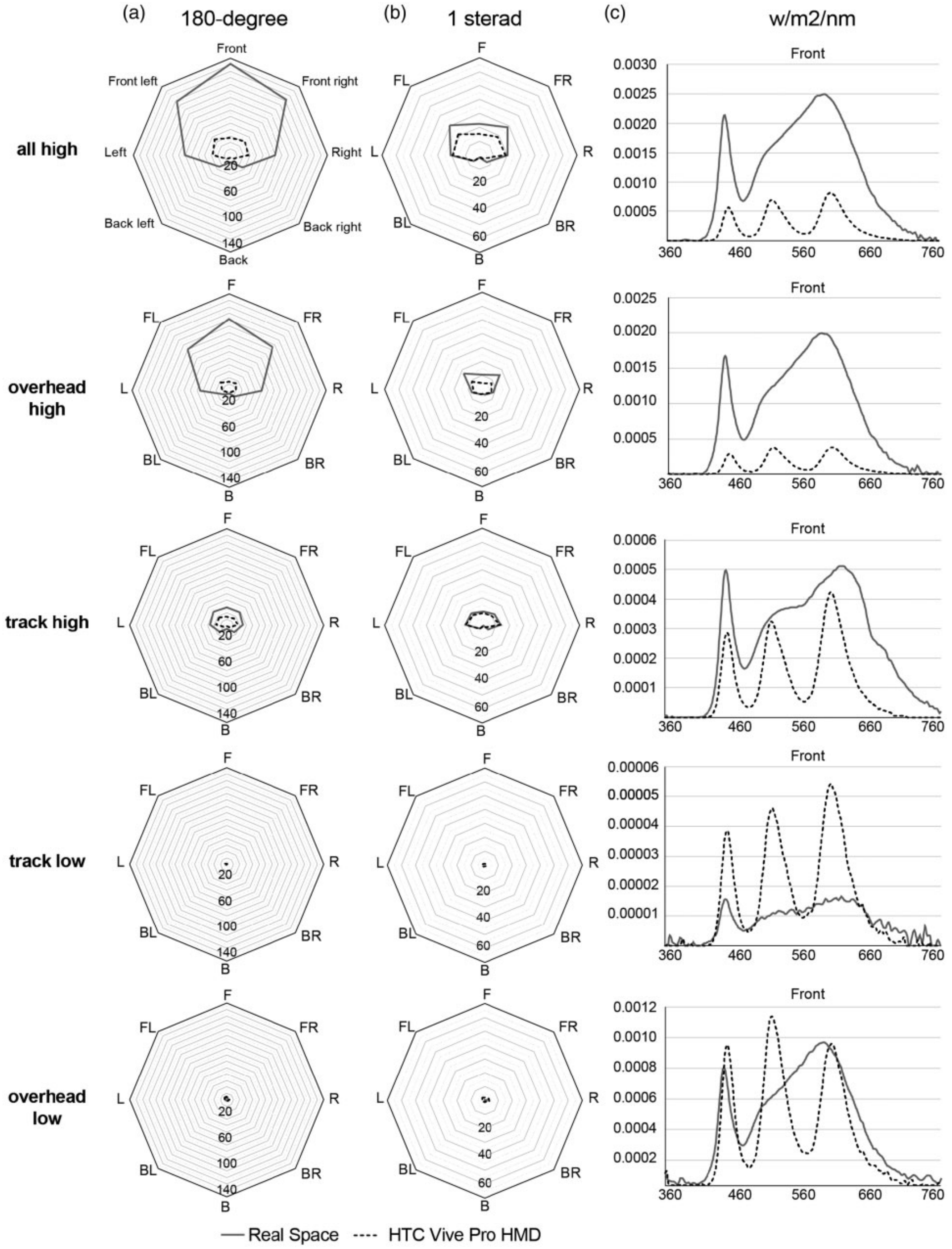
Figure 8(a) reveals a measurable difference between the real space and the VR HMD for the all high and overhead high groups when you consider the full 180° viewing angle. This is due to the contribution of light from overhead luminaires. When you focus instead on a single steradian for those conditions, as shown in Figure 8(b), there is less variation in lux as this narrow viewing angle cuts off the overhead light sources as a direct contribution and measures the light that falls instead on the walls.

On the contrary, if we compare the 180° vertical and horizontal view angle to a single steradian for the task on condition, we can see lux levels in the front right (FR) view direction are much higher in the real space than they are in the VR HMD. This difference, along with the limited 110° view field of the HTC Vive Pro HMD, may contribute to the underestimation of glare and brightness in the VR group for this particular condition (as seen in Figures 5(e) and 7(b). Finally, the spectral power distribution of lighting conditions across the front-facing view-direction (Figure 8(c)) reveal a difference in spectral intensity between wavelengths of light. The OLED screen on the HTC Vive Pro HMD produces consistent peaks at 455 nm, 520 nm, and 615 nm, regardless of the lighting condition.

These graphs also reveal the difference in spectral power between real space and VR HMD. Most lighting conditions produce higher spectral power (and more evenly distributed power across wavelengths) in the real space than in the VR HMD, except for the track low condition, where the VR HMD produces greater spectral power than in the real space. It should be noted that the x-axis on these graphs is scaled to the data, so peaks shown in the track low condition are relatively small in comparison to those in the other conditions. Nonetheless, the researchers noted that it was difficult to produce depth in the darkest regions of an image using the HTC Vive Pro HMD. While the AMOLED display of the device can turn-off to produce truly black pixels, the tone-mapping process compresses the overall luminance range and can minimize the depth of variation in pixels in both the very dark and the very bright range. This may be contributing to differences in brightness and contrast perception in the dim lighting conditions.

5. Discussion

For questions related to visual comfort and pleasantness, the results of this paper reveal that there were no significant differences between real and virtual groups for any of the lighting conditions presented to our participants. Questions related to evenness resulted in one significant difference and questions related to brightness, contrast and glare rating resulted in significant differences between two or more lighting conditions in the real and VR lab groups. In general, these differences were more pronounced in the dimmer and more highly contrasted scenes. The brightest and most uniformly lit neutral, all high and track high scene conditions resulted in no significant differences between


(a)
180-degree
(b)
1 sterad
(c)
w/m2/nm
all high
overhead high
track high
track low
overhead low
Front
Front left
Front right
Left
Right
Back left
Back right
Back
F
FL
FR
L
R
BL
BR
B
20
60
100
140
20
40
60
Front
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0.00006
0.00005
0.00004
0.00003
0.00002
0.00001
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
360
460
560
660
760
Real Space
HTC Vive Pro HMD

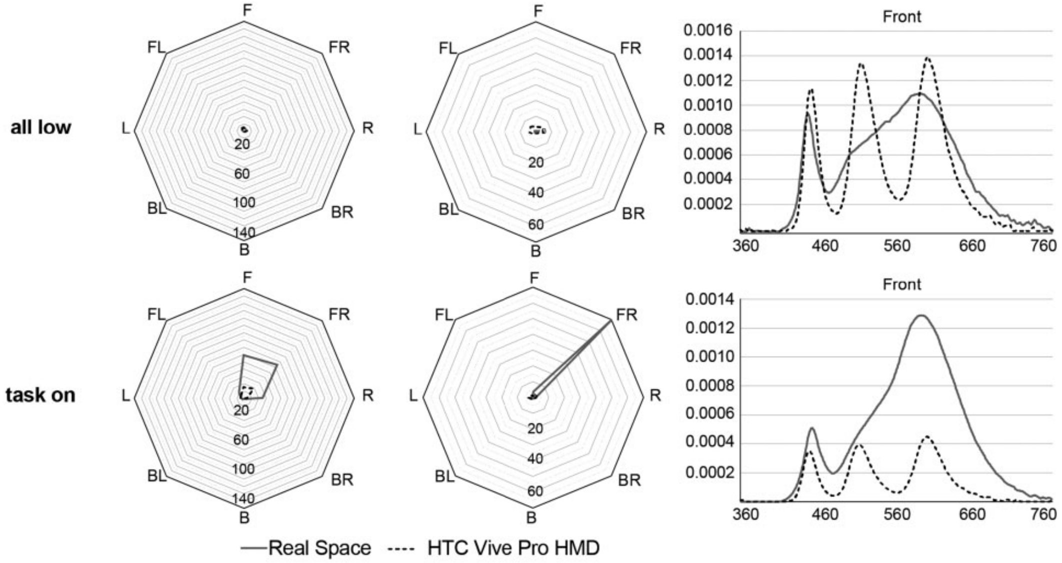


Figure 8. Measurements captured in both the real space and VR HMD for (a) lux across a 180° vertical and horizontal viewing angle in 45° increments, (b) lux across a single steradian in 45° increments and (c) spectral power distribution, all using a Minolta CL 500A spectrophotometer

real and VR for any of the subjective rating scales.

Despite the dimmer display range of the HTC Vive Pro HMD (as compared to light levels in the real space), participants systematically rated lighting conditions in the VR HMD as brighter and more glaring than those viewed in the real space (for all but the task on condition). Participants in the VR lab group also took significantly longer to complete the Landolt ring visual acuity test than those viewing the lighting conditions in the real space. This was true for all seven lighting conditions and the neutral starting condition. This finding is in agreement with Heydarian *et al.*[23] who also found a significantly lower reading speed in immersive virtual environments than their real space comparison. This could be a useful consideration for researchers wanting to use VR as an experimental environment to conduct research on visual performance or behaviour.

While the HTC Vive Pro HMD could not reproduce the same lux or spectral distribution of the real space, this study reveals the potential need for development in tone-mapping and image compression methods to improve the perceptual accuracy of lighting scenes presented in VR HMDs like the HTC Vive Pro. While Chamilothori[39] found no significant differences in brightness or uniformity ratings in VR (using the pcond TMO) when compared to real scenes, this may be due in part to differences in scene conditions and VR HMD manufacturers between the two studies. The current study uses an HTC Vive Pro instead of the Oculus Rift (used in Chamilothori[39]) and considered low-light and highly contrasted LED lighting scenes as opposed to daylit environments. Given the measured differences in spectral power between the lighting conditions captured in the real and virtual groups, it would also be valuable to integrate a visual acuity test related to colour in future work, such as proposed by Abd-Alhamid *et al.*[10] This, alongside subjective questions related to colour saturation and/or vividness could help expose any limitations of perceived colour between the real and VR groups.

The population statistics (Section 4.1) raise an interesting set of questions related to the impact of corrective lenses on glare perception in virtual environments. It appears that participants wearing corrective lenses had a small, but significantly lower mean rating for questions related to comfort and evenness in the VR group, but also a higher mean rating of glare. Future work should explore this issue with a larger sample to confirm these potential effects as the limited sample size in our analysis limits our ability to produce definitive conclusions. In our current analysis, the Mann-Whitney results include participants both with and without corrective lens use. Due to the potential impact of lens use on more than one rating scale in our population statistics analysis, future work may consider a larger and balanced population of lens-wearing and non-lens-wearing participants in order to analyse each sample population separately. Furthermore, as the authors did not include a follow-up question to specify the type of lens correction resulting in these effects (glasses/contacts or strength of correction), it would be interesting to know whether glasses that filter blue-light are behind this effect or if it is a matter of lens thickness (related to prescription and specific visual impairment).

The researchers acknowledge that any of the specific image capture and processing techniques used in this study could be contributing to systematic error in the final displayed images. As such, future work should compare the impact of different tone-mapping algorithms and VR HMDs to understand the interaction between software and hardware variables. In the absence of a dynamic tone-mapping algorithm developed specifically for the compression of HDR images in immersive HMDs, this work may provide motivation for additional development in this area.

The researchers also acknowledge that the HDR images captured in this study were monoscopic (both eyes saw the same image) and that stereoscopic scenes may improve the feeling of presence in our VR scenes. Other studies that have used stereoscopic images in perceptual research with VR have used renderings[8] through a method of generating a rotational offset camera in Radiance,[31] or quasi-stereo HDR photographs,[10] where the camera tripod was offset to capture a front-facing stereoscopic scene. Given the nature of our HDR camera and its rotational capture method, we could not offset the scene to create a stereoscopic image without manipulating our tripod and developing a new piece of software to run the camera.

6. Conclusion

Overall, the findings from this experiment reveal that current VR HMDs can be a reasonable surrogate for real-world lighting environments when evaluating well-lit (neither too dim nor too contrasted) electric lighting scenes for subjective ratings about visual comfort, pleasantness, evenness, contrast, and glare rating. For lighting scenes that are dim or high in contrast and potential glare, VR may not provide an adequate medium for reproducing accurate lighting perception. For questions related to brightness, the HTC Vive Pro HMD may produce scenes that appear systematically brighter in VR than they would in a real space, but this may be improved with a TMO optimized for VR HMDs.

These findings build upon those by Abd-Alhamid *et al.*,[10] who found no significant difference in rated brightness between real and VR groups. The authors of that study relied on objectively well-lit spaces (average of 498 lx) and used an HTC Vive rather than an HTC Vive Pro HMD. Despite the difference in HMD, our neutral and all high lighting conditions also showed no significant differences in ratings of brightness. Despite this, our results did find significant differences in ratings of brightness between the real and VR groups for spaces that were dim or highly contrasted. For designers and students seeking to evaluate the quality of a lighting design or for researchers looking to use VR as a surrogate for full-scale mock-ups, it is important to note these limitations and use caution when using VR to determine whether a space appears adequately bright or if it poses a glare risk. Future work should seek to include an even broader range of lighting conditions alongside those considered in this study, such as LED lighting for high acuity tasks (lx > 800) as well as daylight. While this study compares lighting in the dim, moderately well-lit and highly contrasted ranges, additional conditions would need to be included to test the upper range of brightness in the HTC Vive Pro.

While many of our subjective rating scales performed well in our VR HMD, the differences in brightness and contrast reveal a potential disconnect between the virtual reality in which designers are making qualitative design decisions and the physical reality those virtual images are meant to represent. The impacts of this disconnect could result in the under or over-specification of electric light sources in a VR HMD that produce an undesirable result in the physical reality. Given the fast speed of development in VR HMDS and the rate at which designers are integrating these displays into their decision-making and presentation arsenal, this study is one of many more that are needed to fully understand the capacity of VR HMDs to produce lighting scenes that are perceptually accurate when compared to the real-world.

Acknowledgements

The team acknowledges contributions Dale Northcutt for assistance in the collection of data specific to the HTC Vive Pro HMD.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research in this paper was funded by the Pacific Northwest National Laboratory.

ORCID iDs

Siobhan Rockcastle https://orcid.org/0000-0002-2514-281X
R Davis https://orcid.org/0000-0003-2460-6771

References

1 Jevard P, Mullins M. *The use of VR technologies to enhance methods for lighting design practice*: *7th eCAADe Regional International Symposium,* Aalborg, Denmark, May 2–3: 2019: 97–102.
2 Paes D, Arantes E, Irazzy J. Immersive environment for improving the understanding of architectural 3D models: comparing user spatial perception between immersive and traditional virtual reality systems. *Autom Constr* 2017; 84: 292–303.
3 Klerk RD, Duarte AM, Medeiros DP, Duarte JP, Joaquim J, Lopes DS. Usability studies on building early stage architectural models in virtual reality. *Autom Constr* 2019; 103: 104–116.
4 Panya D, Seo HJ, Park HJ, Lee WJ, Choo S. *A study of kinetic façade modelling performance using virtual reality*: *36th International Symposium on Automation and Robotics in Construction (ISARC),* Banff, Canada, May 21–24: 2019: 356–359.

5 Ozhen DS, Afacan Y, Surer E. Usability of virtual reality for basic design education: a comparative study with paper-based design. *Int J Technol Design Educ* 2019: 1–21.

6 Kreutzberg A. *Visualising architectural lighting concept with 360°panoramas*: *Proceedings of the 36th eCAADe Conference,* Lodz, Poland, 19–21 September 2018, pp. 745–752.

7 Marin-Morales J, Higuera-Trujillo JL, Greco A, Guixeras J, Llinares C, Gentili C, Scilinog EP, Alcaniz M, Valenza G. Real vs. *immersive-virtual emotional experience: analysis of psycho-physiological patterns in a free exploration of an art museum*. Plos One 2019; 14: e0223881.

8 Chamilothori K, Wienold J, Andersen M. Adequacy of immersive virtual reality for the perception of daylit spaces: comparison of real and virtual environments. *Leukos* 2018; 15: 203–226.

9 Higuera-Trujillo JL, Maldonado JLT, Millán CL. Psychological and physiological human responses to simulated and real environments: a comparison between photographs, 360 panoramas, and virtual reality. *Appl Ergon* 2017; 65: 398–409.

10 Abd-Alhamid F, Kent M, Bennetta C, Calautita J, Wu Y. Developing an innovative method for visual perception evaluation in a physical-based virtual environment. *Build Environ* 2019; 162: 106278.

11 Chen Y, Cui Z, Hao L. Virtual reality in lighting research: comparing physical and virtual lighting environments. *Light Res Technol* 2019; 51: 820–837.

12 Kus B. *A comparative study on spatial perception in real and virtual office environments under different lighting conditions*. Master's Thesis. Ankara: Bilkent University, 2019.

13 El Beheiry M, Doutreligne S, Caporal C, Ostertag C, Dahan C, Masson J. Virtual reality: beyond visualization. *J Mol Biol* 2019; 431: 1315–1321.

14 Wong M, Du J, Zhang Z, Liu Y, Chen S, Lee S. An experience-based interactive lighting design approach using BIM and VR: a case study. *IOP Conf Series: Earth Environ Sci* 2019; 238: 012006.

15 Cauwerts C. *Influence of presentation modes on visual perceptions of daylit spaces*. PhD thesis. Louvain-la-Neuve: Universite catholique de Louvain (UCL), 2003.

16 Murdoch M, Stokkermans M, Lambooij M. Towards perceptual accuracy in 3D visualizations of illuminated indoor environments. *J Solid State Light* 2015; 2: 1–19.

17 Chessa M, Maiello G, Borsari A, Bex P. The perceptual quality of the oculus rift for immersive virtual reality. *Human-Computer Interact* 2019; 34: 51–82.

18 Newsham G, Seetzen H, Veitch J, Chaudhuri A, Whitehead L. *Lighting quality evaluations using images on a high-dynamic range display*: *Proceedings of the ARCC/EAAE Conference on Architectural Research*. Montreal, Quebec, 22–25 May 2002, pp. 1–9.

19 Akyuz AO, Fleming RW, Riecke BE, Reinhard E. Do HDR displays support LDR content?: A psychophysical evaluation. *ACM Trans Graphics* 2007; 26: 1–7.

20 Newsham G, Cetegen D, Veitch JA, Whitehead L. Comparing lighting quality evaluations of real scenes with those from high dynamic range and conventional images. *ACM Trans Appl Res* 2010; 7: 1–25.

21 Inanici M. Evaluation of high dynamic range photography as a luminance data acquisition system. *Light Res Technol* 2006; 38: 123–136.

22 Cetegen D. Veitch JA, Newsham GR. *View size and office luminance effects on employee satisfaction*: *Proceedings of Balkan Light,* Ljubljana, Solvenia, Oct 7: 2008: 243–252.

23 Heydarian A, Pantazis E, Carneiro J, Gerber D. *Towards understanding end-user lighting preferences in office spaces by using immersive virtual environments*: *ASCE International Workshop of Computing in Civil Engineering*, Austin, TX, 21–23 June 2015, pp.1–8.

24 Rockcastle S, Chamilothori K, Andersen M. *An experiment in virtual reality to measure daylight-driven interest in rendered architectural scenes*: *Proceedings of Building Simulation*, San Francisco, USA, 2–4 September 2017, pp. 2797–2806.

25 Sawyer A, Chamilothori K. *Influence of subjective impressions of a space on brightness satisfaction: an experimental study in virtual reality*: *Proceedings of SimAUD*, Atlanta, USA, 7–9 April 2019, pp.1–8.

26 Heydarian A, Carneiro J, Gerber D, Becerik-Gerber B, Hayes T, Wood W. Immersive virtual environments versus physical built environments: a benchmarking study for building design and user-built environment explorations. *Autom Constr* 2015; 54: 116–126.

27 Melo M, Bessa M, Debattista K, Chalmers A. Evaluation of tone-mapping operators for HDR video under different ambient luminance levels. *Comput Graphics* 2015: 38–49.

28 Raffin M, Guarnieri G. *Tone mapping and enhancement of high dynamic range images based on a model of visual perception*: *International Conference on Computer Graphics and Imaging*, Innsbruck, Austria, Feb 13–15: 2008: 190–195.

29 Reinhard C, Stark M, Shirley P, Ferwerda J. *Photographic tone reproduction for digital images: Proceedings to SIGGRAPH '02*, San Antonio, USA, Jul 21–26: 2002: 267–276.

30 Ward G. *The RADIANCE lighting simulation and rendering system*: *Proceedings of SIGGRAPH Conference*, Orlando, USA, Jul 24–29: 459–472.

31 McNeil A. *Rendering omni-directional stereo images with radiance*: *Radiance Workshop*. Padova, Italy, 29 August 2016, pp. 1–58.

32 Evans J. *Standards for Visual Acuity*. USANational Institute for Standard and Technology, 2006.

33 Vrabel P, Bernecker C, Mistrick R. Visual performance and visual clarity under electric light sources: part 1 – visual perfirmance. *J Illumin Eng Soc* 1995; 27: 69–80.

34 Council of Opthamology International. Visual acuity measurement standard. *Italian J Opthalmol* 1988; 2: 1–15.

35 Hopkinson R. Glare from windows. *Constr Res Dev* 1970; 2: 98–105.

36 Cohen J. Statistical power analysis. *Curr Direct Psychol Sci* 1992; 1: 98–101.

37 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. London: Routledge, 1988.

38 Radianti J, Majchrzak T, Fromm J, Wohlgenannt I. A systematic review of immersive virtual reality applications for higher education: design elements lessons learned, and research agenda. *Comput Educ* 2020; 147: 103778.

39 Chamilothori K. *Perceptual effects of daylight patterns in architecture*. PhD thesis. Lausanne: Ecole Polytechnique Federale de Lausanne, 2019.