Taylor & Francis
Taylor & Francis Group

LEUKOS

The journal of the Illuminating Engineering Society of North America

ISSN: 1550-2724 (Print) 1550-2716 (Online) Journal homepage: http://ies.tandfonline.com/loi/ulks20

Evaluating a New Suite of Luminance-Based Design Metrics for Predicting Human Visual Comfort in Offices with Daylight

Kevin Van Den Wymelenberg & Mehlika Inanici

To cite this article: Kevin Van Den Wymelenberg & Mehlika Inanici (2015): Evaluating a New Suite of Luminance-Based Design Metrics for Predicting Human Visual Comfort in Offices with Daylight, LEUKOS, DOI: 10.1080/15502724.2015.1062392

To link to this article: http://dx.doi.org/10.1080/15502724.2015.1062392

LEUKOS, 00:1–26, 2015
Copyright © Illuminating Engineering Society
ISSN: 1550-2724 print / 1550-2716 online
DOI: 10.1080/15502724.2015.1062392

Taylor & Francis
Taylor & Francis Group

Evaluating a New Suite of Luminance-Based Design Metrics for Predicting Human Visual Comfort in Offices with Daylight

Kevin Van Den Wymelenberg[1] and Mehlika Inanici[2]
[1]University of Idaho Integrated Design Lab, Boise, Idaho, USA
[2]Department of Architecture, University of Washington, Seattle, Washington, USA

ABSTRACT A new suite of visual comfort metrics is proposed and evaluated for their ability to explain the variability in subjective human responses in a mock private office environment with daylight. Participants ($n = 48$) rated visual comfort and preference factors, including 1488 discreet appraisals, and these subjective results were correlated against more than 2000 unique luminance-based metrics that were captured using high dynamic range photography techniques. Importantly, luminance-based metrics were more capable than illuminance-based metrics for fitting the range of human subjective responses to data from visual preference questionnaire items. No metrics based upon the entire scene ranked in the top 20 squared correlation coefficients, nor did any based upon illuminance or irradiance data, nor did any of the studied glare indices, luminance ratios, or contrast ratios. The standard deviation of window luminance was the metric that best fit human subjective responses to visual preference on seven of 12 questionnaire items (with $r^2 = 0.43$). Luminance metrics calculated using the horizontal 40° band (a scene-independent mask) and the window area (a scene-dependent mask) represented the majority of the top 20 squared correlation coefficients for almost all subjective visual preference questionnaire items. The strongest multiple regression model was for the semantic differential rating (too dim–too bright) of the window wall ($_{\text{adj}}R^2 = 0.49$) and was built upon three variables; standard deviation of window luminance, the 50th percentile luminance value from the lower view window, and mean luminance of the 40° horizontal band.

KEYWORDS controls, daylighting, discomfort glare, visual perception

Received 15 January 2015; revised 11 June 2015; accepted 11 June 2015.

Address correspondence to Kevin Van Den Wymelenberg, University of Oregon, Eugene, OR 97403, USA.
E-mail: kvanden5@uoregon.edu

Color versions of one or more of the figures and tables in the article can be found online at www.tandfonline.com/ulks.

1. INTRODUCTION

Daylighting is commonly lauded as a high-performance architectural design strategy that can reduce reliance on electric lighting and provide several human benefits, including health, comfort, and productivity. It is also well known that if daylight is not carefully designed and controlled, it can cause discomfort glare, disability glare, and veiling reflections on computer screens used by office workers. Most designers rely on intuition or simulated measures of horizontal illuminance to inform daylighting design choices; rarely are visual comfort or discomfort rigorously evaluated during design. This is in large part due to lack of confidence in the available metrics and the difficulty to calculate them during design stages.

Original research conducted within a private daylit office, including 48 participants and a 6-month repeated-measures experimental design is presented. A previous paper from this research has established that the current illuminance- and luminance-based visual comfort metrics are limited in their ability to reliably predict visual comfort in offices with daylight, and fundamental deficiencies with current metrics have been documented [Van Den Wymelenberg and Inanici 2014]. Because human perception of brightness closely relates to measures of luminance, and because office tasks are dominated by vertical tasks, it is hypothesized that luminance-based measures from the occupants' point of view are more likely than illuminance-based measures to correlate closely with subjective assessments of visual comfort in office environments. A brief literature review of established and recently proposed illuminance- and luminance-based visual comfort metrics is provided. This article details a wide range of newly proposed luminance-based metrics. Finally, it presents results and provides discussion and conclusion of the most promising metrics identified.

1.1. Established Metrics

Horizontal illuminance is the most common lighting design metric, yet human preference of horizontal illuminance has been shown to vary widely under electric light [Boyce and others 2006; Newsham and Veitch 2001; Veitch and Newsham 2000] as well as daylight sources [Laurentin and others 2000; Van Den Wymelenberg and Inanici 2014; Van Den Wymelenberg and others 2010]. Some evidence for an upper threshold of comfortable horizontal illumination was shown between 2000 and 4300 lux; however, some individuals preferred values as high as 5000 lux. Recent research [Van Den Wymelenberg and Inanici 2014] suggested that vertical illuminance (E_v) is more capable than horizontal measures of illuminance at predicting visual comfort.

Beyond measures of illuminance, luminance ratios are likely the next most common lighting design metric used in practice. Van Den Wymelenberg and Inanici [2014] found some evidence for a 22:1 ratio as the borderline between comfort and discomfort (BCD) when measuring between the mean window luminance and a mean task luminance, supporting some preliminary evidence by others [Sutter and others 2006]. However, only the most advanced building design teams routinely evaluate visual comfort associated with daylight by using luminance or glare metrics despite recent software developments ["DIVA for Rhino" n.d.; Fraunhofer n.d.; Kumaragurubaran and Inanici 2013] that have increased access to these data. This is possibly due to limitations in usefulness of results or time and expertise required to generate results. This highlights the ongoing need to improve the metrics used to predict human comfort for design decisions and there are similar needs for improved metrics to support building lighting and shading control.

The IES recently published LM-83 [Heschong and Van Den Wymelenberg 2012; IESNA-Daylight Metrics Committee 2012], which documents the definition and calculation procedures for the first two human factors–based IES-adopted daylighting design metrics, spatial daylight autonomy (examining annual daylight sufficiency), and annual sunlight exposure (examining annual potential risk of excessive sunlight penetration). Though this is important progress, LM-83 stresses the need for additional metrics " . . . to allow a daylighting design or daylit space to be further evaluated relative to other aspects of a daylit space, such as uniformity, contrast, or glare, and eventually human health and building energy impacts [IESNA-Daylight Metrics Committee 2012, p. 2]." Van Den Wymelenberg [2014] stated that luminance-based metrics are likely to prove useful in this endeavour and encourages their development. Zaikina and others [2014a, 2014b] provide further support for the need and usefulness of improved luminance-based metrics to help describe observer visual perception. This article presents results from a broad exploration of both illuminance- and luminance-based lighting design metrics and their ability to predict visual comfort and discomfort, in order to provide guidance to researchers, designers, codes and standards organizations, and the lighting and automated blinds controls industry.

2. METHODOLOGY

2.1. Research Procedures and Setting

The research design was repeated measures and included 48 participants (45 repeated) for daylong (Table 1) experiments in a daylit mock private office in Boise, Idaho (documented in Van Den Wymelenberg and Inanici [2014] in detail). The University of Idaho Internal Review Board (IRB) has approved that this study was in compliance with

TABLE 1 Typical participant-day[a]

Condition order was changed monthly to avoid bias

Time (min)	Activity	Description
Put blinds down and rotated closed and electric lights on at full power to begin each participant-day		
9:50 AM (50)	Conditions 1–3 by participant	C1—Participant directed to create MP daylight environment C2—Participant directed to improve environment by adding electric light C3—Participant directed to worsen environment by adjusting electric light
10:40 AM (10)	Morning break	Put blinds all the way up and turn the electric lights off
10:50 AM (50)	Conditions 4–6 by participant	C4—Participant directed to create JU glare daylight environment C5—Can participant improve environment adding electric light? C6—Participant directed to just correct the glare problem by adjusting blinds
11:40 AM (20)	Condition 7 by participant	C7—Participant directed to create MP integrated lighting environment
12:00 PM (60)	Lunch break	Put blinds all the way up and turn the electric lights off
1:00 PM (50)	Conditions 8–10 by researcher with participant confirmation	C8—Participant directed to create MP daylighting environment C9—Researcher sets an intentionally dark scene (blinds all the way down, no electric lights) C10—Participant directed to create JU glare scene from daylight alone
1:50 PM (20)	Afternoon break	Put blinds all the way up and turn the electric lights off.
2:10 PM (50)	Conditions 11–13 by researcher with participant confirmation	C11—Participant directed to create and maintain the MP integrated lighting environment C12—Leaving electric light as previous, researcher closes blinds all the way C13—Leaving electric light as previous, Participant directed to open blinds just enough to create a JU glare scene
		Put blinds all the way up and turn the electric lights off
3:00 PM (50)	Conditions 14–16 by researcher with participant confirmation	C14—Participant directed to create and maintain the MP integrated lighting environment C15—Leaving blinds as pervious, participant directed to dim electric light until just too dim (or until off) C16—Leaving blinds as previous, participant directed to increase electric lights until just too bright (or until on full)
3:50 PM (10)	Debrief/dismiss	

[a] From Van Den Wymelenberg and Inanici [2014].

all Human Subject guidelines (project #10-187) and the University of Washington has an Authorization Agreement for this project with the University of Idaho IRB (HSD 40217).

Figure 1 (left) demonstrates the setting as seen from a participant's point of view. Participants spent two full working days, one in summer and one in fall, assessing a range of visual conditions under naturally occurring sky conditions while manipulating blind height, blind tilt, and ambient electric lighting levels. In the first round of the study (June 29–September 20), 94% of the study hours were "sunny," 2% had "few" or "scattered" clouds, and 4% were "broken" overcast or fully "overcast." In the second round of the study (September 21–December 19), 71% of the study hours were "sunny," 7% had "few" or "scattered" clouds, and 22% were "broken" overcast or fully "overcast." Extensive illuminance and luminance data were collected in an identical adjacent room (equipment room). The participant room and equipment room were each fitted with a semiperforated daylight guiding a motorized louver blind with manual control (via remote or computer; Fig. 1, right). A single manually dimmable (by remote)

Fig. 1 (left) scene from a participant's point of view; (right) light redirecting blind.

T5HO recessed direct electric light fixture was located near the center of the room.

2.2. Questionnaire Items

Participants independently created 16 unique lighting conditions per instruction (Table 1) and a researcher confirmed that participants had created the intended scene. The scenes created by the participants were monitored by the researcher via remote on-screen display and the researcher was available to answer any questions of the participants either by phone or in person. The scenes created by the researcher were established remotely or in person and were verbally verified by participants. Participants then rated the following items for each condition. For questions one through seven, participants rated the following statements using a 7-point Likert-type scale (7 = very strongly agree, 6 = strongly agree, 5 = agree, 4 = neither agree or disagree, 3 = disagree, 2 = strongly disagree, 1 = very strongly disagree):

1. This is a visually comfortable environment for office work. (QU1)
2. I am pleased with the visual appearance of the office. (QU2)
3. I like the vertical surface brightness. (QU3)
4. I am satisfied with the amount of light for computer work. (QU4)
5. I am satisfied with the amount of light for paper-based reading work. (QU5)
6. The computer screen is legible and does not have reflections. (QU6)
7. The lighting is distributed well. (QU7)

The participants rated the following items using a slider bar semantic differential scale from "too bright" (scored as 100) to "too dim" (scored as zero) with neither too bright nor too dim midway between (scored as 50):

1. When I look up from my desk the scene I see in front of me seems: (front-scene)
2. When I look to my left the scene that I see seems: (left-scene)
3. When I look to my right the scene that I see seems: (right-scene)
4. I find the ceiling to be: (ceiling)

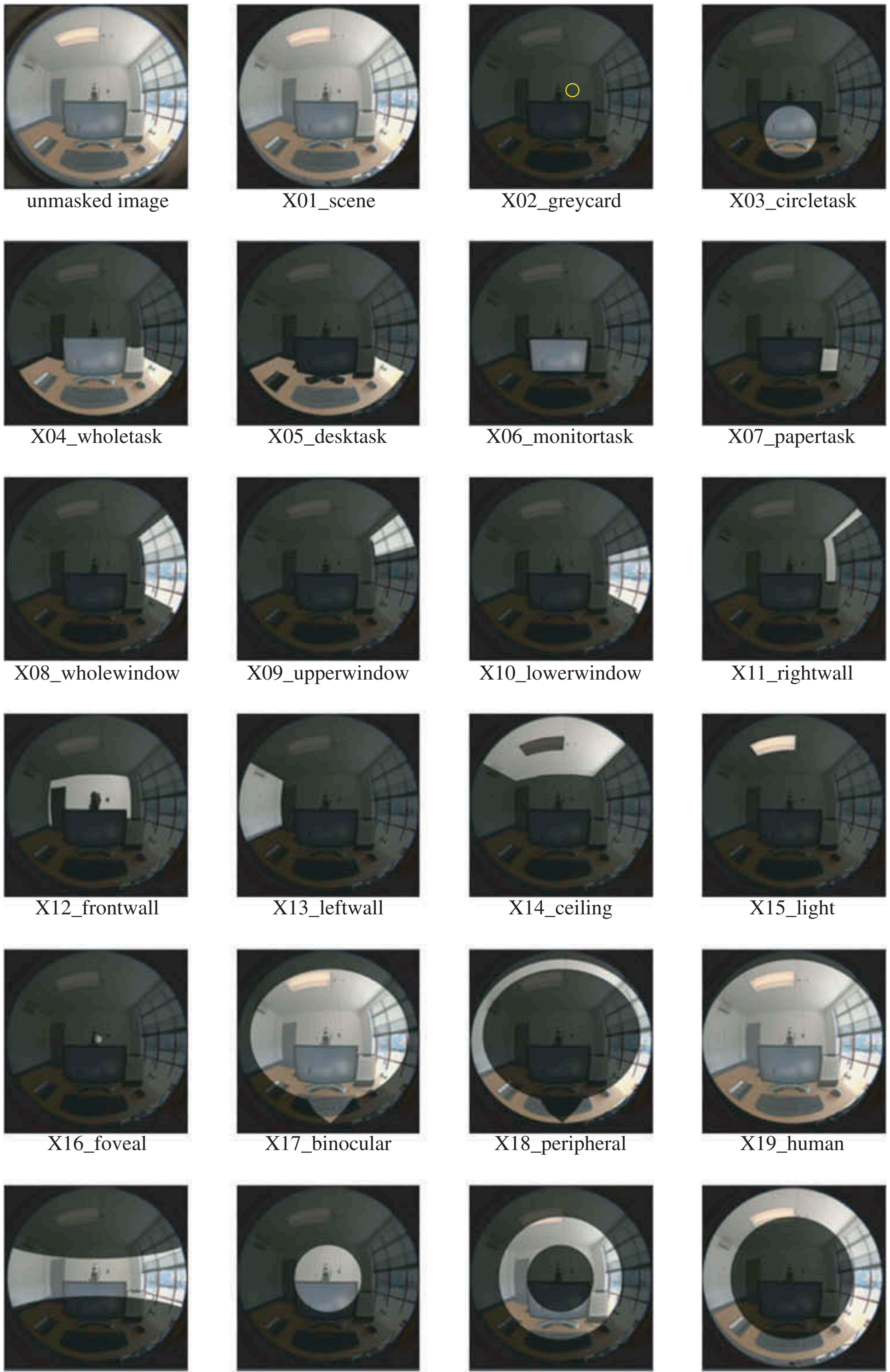
2.3. Analysis methods

High dynamic range (HDR) photography was collected for 93 participant-days and 16 conditions per day, resulting in 1488 individual HDR data sets captured. Selected scenes were removed due to excessive daylight variability [Van Den Wymelenberg 2012] during the HDR capture sequence (100 participant-scenes, 6.7%) or because participants accidentally turned electric lights on when they were supposed to be off (four participant-scenes, 0.27%). Therefore, results from a total of 1379 HDR scenes are reported herein. For data analysis, descriptive statistics and inferential statistics were employed. One-way and two-way, paired and unpaired t tests used a 95% confidence interval, and Pearson and Spearman correlations were both conducted.

2.3.1. Luminance Metrics and Scene Masks

The research reported here is the result of a comprehensive study where over 2000 unique luminance metrics were tested using the equidistant fisheye HDR data sets. The complete list of metrics tested is described briefly below, but only selected results are reported for brevity in this article. In order to better understand specific areas within scenes, 23 masked regions were examined (as shown in Fig. 2) using the 6-month data set. Several masks are scene dependent and others are scene independent. Several metrics, as follows, were calculated for each mask:

- Minimum, maximum, mean ($\bar{x}$), standard deviation (σ), coefficient of variation ($\sigma/\bar{x}$) of mask luminance.
- Several luminance percentiles (2nd, 10th, 50th, 75th, 90th, 98th) and ratios of these (e.g. 2nd, 98th percentile).
- Percentage of mask pixels above or below certain absolute luminance thresholds (below 5, 10, 40, 50, 100, 250, 500, 1000 cd/m^2; above 1500, 2000, 2500, 3000,



unmasked image
X01_scene
X02_greycard
X03_circletask
X04_wholetask
X05_desktask
X06_monitortask
X07_papertask
X08_wholewindow
X09_upperwindow
X10_lowerwindow
X11_rightwall
X12_frontwall
X13_leftwall
X14_ceiling
X15_light
X16_foveal
X17_binocular
X18_peripheral
X19_human
X20_40band (horizontal 40°band)
X21_0_60 (central 60° fov)
X22_60_120 (fov from 60°-120°)
X23_120_180 (fov from 120°-180°)

Fig. 2 **Masks applied to an example scene (X01 is Mask 01).**

4000, 5000 cd/m^2) and ratios of these (for example, percentage below 5 cd/m^2: percentage above 5000 cd/m^2).

The following glare metrics were calculated for the entire scene only (Mask 01):

- Daylight Glare Probability (DGP), Daylight Glare Index (DGI), Visual Comfort Probability (VCP), Uniform Glare Rating (UGR), CIE Glare Index, and the average luminance of the glare sources identified within the entire scene, calculated using Evalglare version 1.11

(Fraunhofer ISE). Evalglare output was generated using two glare source identification methods. One method was based upon mean luminance of scene (Mask 01) and task (Mask 03) multipliers ($3*\bar{x}$, $5*\bar{x}$, $7*\bar{x}$, $10*\bar{x}$) and the second used absolute luminance values (1500, 2000, 2500, 3000, 4000, 5000 cd/m^2) within the scene to identify glare sources. Once the glare sources were identified the glare indices were calculated.

- Radiance *findglare* and *glarendx* programs [Ward 2011] were used to calculate DGI for the entire scene using the default method glare source identification method ($7 * \bar{x}$) and the same six absolute luminance values listed in the previous bullet.

A number of additional metrics were calculated using data from multiple masks detailed elsewhere [Van Den Wymelenberg 2012]. These include basic luminance ratios, contrast ratios, and comparisons of mean and standard deviation values between several masks. The luminance ratio metrics examine simple ratios between mean values of the task (Mask 03 mean L, Mask 21 mean L) and either adaptation background values (mean L of scene, Mask 22 mean L, Mask 23 mean L), background variability (SD L of scene, Mask 23 SD L), or high scene luminance values (90th percentile L value of scene, 98th percentile L value of scene, mean luminance of window). The luminance contrast metrics include two combinations of masks to develop task (Mask 03, Mask 21) to background (Mask 01, Mask 22) luminance contrast ratios. The mean to standard deviation ratios examine the brightness of the central 60° of vision to the variation of luminance in the entire scene (Mask 01) or the noncentral vision (Mask 23). Loe and colleagues [1994] presented luminance analysis of an electrically illuminated scene with metrics extracted from the 40° horizontal band that proved promising and thus (Mask 20) was repeated herein. In addition to these approximately 2000 luminance metrics, illuminance and irradiance data were collected as reported previously [Van Den Wymelenberg and Inanici 2014].

3. RESULTS

The results from the comprehensive study are organized as follows: (1) correlation matrices are presented for the top 20 ranked single regression metrics, then selected existing luminance-based metrics, (2) detailed results for two example top-ranked single regression metrics, and (3) multiple linear regression results for the two top-ranked multiple-metric models.

Several abbreviations are utilized in the remainder of the article. “Most preferred” is written “MP” and “JU” refers to “just uncomfortable” scenes; illuminance is “E” and luminance is “L.” “C8” refers to condition eight (whereby the participant created a MP daylight environment during the afternoon) and “C10” refers to condition 10 (whereby the participant created a JU daylight environment in the afternoon); therefore, “C8C10” refers to data grouped using conditions eight and ten together. “QU1” refers to question 1, et cetera. “Task” refers to Mask 03, X03 as shown in Fig. 2 unless specified.

3.1. Linear and Nonlinear Regression

Pearson pairwise squared correlation coefficients were calculated for the entire set of illuminance- and luminance-based metrics as described in the previous section. Table 2 provides the top 20 ranked metrics and subjective questionnaire items for conditions C1, C2, C4, C6, C7, C8, C10, C11, C13, and C14 (described as the “composite data set”) and selected lighting metrics are given in Table 3. The “filtered” designation was appended to the condition string (for example, composite_data_set_filtered) in cases where uncomfortable data were filtered out of the MP data set and comfortable data were filtered out of the JU data set based upon the responses to QU4. The rationale is that the participant was unable to create the intended setting. Results from seven Likert items (QU1–QU7), four semantic differential items evaluating locational perception of brightness (too dim–too bright for each of “front-scene,” “left-scene,” “right-scene,” and “ceiling” as described in Tables 2 and 3), and the overall scene preference semantic differential (from least preferred–most preferred; coded “light-in-scene”) item are summarized for top ranking and additional selected metrics. These results are presented in ranked order by the item right-scene, with relative ranks listed in the leftmost column of each table and the abbreviated metric names in the next column to the right. Right-scene was selected to rank the metrics in Table 2 for two reasons. One reason is that right-scene represented the highest overall squared correlation coefficients for candidate metrics within the data set. Another reason is that right-scene was the semantic differential item that had the highest correlation with all of the Likert-type items. The results in Table 2 represent r^2 values. These values are generally higher than adjusted-r^2 (${}_{\text{adj}}r^2$) values; however, given the substantial size of this sample data, the r^2 and ${}_{\text{adj}}r^2$ figures are almost identical. For example, for the standard deviation of window luminance relative to right-scene, the $r^2 = 0.4252$ is inconsequently higher than the ${}_{\text{adj}}r^2 =$

TABLE 2 Top 20 ranked r^2 values ordered by right-scene (using composite_data_set_filtered)[a]

Rank	Top 20 metrics (ordered by right-scene)	QU1	QU2	QU3	QU4	QU5	QU6	QU7	Likert_all	Front-scene	Left-scene	Right-scene	Ceiling	Light-in-scene
1	SD L window	**0.298**	**0.254**	**0.163**	**0.302**	**0.149**	0.281	**0.232**	**0.288**	0.091	0.072	**0.425**	0.113	0.065
2	25th percentile L value lower window	0.271	0.224	0.141	0.289	0.133	0.269	0.218	0.264	0.124	0.100	0.389	0.079	0.077
3	50th percentile L value lower window	0.245	0.192	0.126	0.251	0.115	0.224	0.185	0.228	0.145	0.103	0.370	0.062	0.085
4	25th percentile L value window	0.257	0.223	0.159	0.287	0.130	**0.288**	0.216	0.266	0.113	0.113	0.359	0.101	0.042
5	Mean L of window	0.241	0.193	0.128	0.250	0.117	0.237	0.178	0.229	0.113	0.092	0.340	0.077	0.079
6	Mean L of the 40° horizontal band	0.244	0.194	0.130	0.252	0.113	0.236	0.189	0.231	0.125	0.110	0.331	0.067	0.079
7	75th percentile L value left wall	0.242	0.208	0.156	0.263	0.122	0.270	0.207	0.250	0.123	0.172	0.324	0.103	0.093
8	10th percentile L value ceiling	0.226	0.192	0.135	0.249	0.108	0.253	0.185	0.229	0.118	0.144	0.321	0.099	0.105
9	10th percentile L value lower window	0.241	0.207	0.124	0.257	0.136	0.249	0.209	0.243	0.080	0.072	0.320	0.076	0.044
10	Percentage of pixels in 40° horizontal band below 1000 cd/m^2	0.215	0.159	0.113	0.227	0.091	0.203	0.156	0.198	0.134	0.107	0.318	0.052	0.114
11	50th percentile L value in 120°–180° FOV	0.231	0.200	0.143	0.255	0.110	0.268	0.195	0.239	0.105	0.136	0.315	0.101	0.078
12	75th percentile L value in 40° horizontal band	0.230	0.191	0.135	0.254	0.108	0.269	0.190	0.234	0.119	0.140	0.313	0.092	0.081
13	25th percentile L value ceiling	0.219	0.185	0.131	0.242	0.105	0.249	0.180	0.223	0.107	0.140	0.310	0.094	0.085
14	25th percentile L value 40° horizontal band	0.225	0.188	0.130	0.247	0.108	0.267	0.191	0.230	0.102	0.126	0.308	0.106	0.074
15	2nd percentile L value ceiling	0.213	0.181	0.125	0.235	0.100	0.241	0.174	0.216	0.112	0.136	0.307	0.094	0.103

(*Continued*)

TABLE 2 (Continued)

Rank	Top 20 metrics (ordered by right-scene)	QU1	QU2	QU3	QU4	QU5	QU6	QU7	Likert_all	Front-scene	Left-scene	Right-scene	Ceiling	Light-in-scene
16	90th percentile L value 40° horizontal band	0.212	0.159	0.108	0.217	0.090	0.203	0.153	0.194	**0.152**	0.112	0.305	0.052	0.070
17	50th percentile L value peripheral FOV (Mask 18)	0.230	0.202	0.145	0.251	0.111	0.271	0.198	0.240	0.094	0.122	0.303	0.097	0.059
18	50th percentile L value 40° horizontal band	0.211	0.179	0.129	0.235	0.099	0.247	0.176	0.217	0.111	0.141	0.300	0.101	0.091
19	50th percentile L value window	0.204	0.161	0.124	0.221	0.092	0.215	0.155	0.199	0.137	0.108	0.299	0.072	0.046
20	25th percentile human FOV (Mask 19)	0.216	0.189	0.133	0.242	0.101	0.270	0.187	0.227	0.096	0.121	0.298	0.112	0.077

[a] QU1—This is a visually comfortable environment for office work.
QU2—I am pleased with the visual appearance of the office.
QU3—I like the vertical surface brightness.
QU4—I am satisfied with the amount of light for computer work.
QU5—I am satisfied with the amount of light for paper-based reading work.
QU6—The computer screen is legible and does not have reflections.
QU7—The lighting is distributed well.
Front-scene—When I look up from my desk does the scene I see in front of me seems (too dim–too bright).
Left-scene—When I look to my left the scene that I see seems (too dim–too bright).
Right-scene—When I look to my right the scene that I see seems (too dim–too bright); this direct included the window.
Ceiling—I find the ceiling to be (too dim–too bright).
Light-in-scene—I find this lighting condition to be: (least preferred–most preferred) [C9, C10, C12, C13, C15, C16 only].

Bolded numbers indicate that the metric's r^2 value was the highest ranked overall for a specific item. Pink fill indicates that the metric's r^2 value was greater than or equal to 0.20 and yellow fill indicates a value greater than or equal to 0.10 but less than 0.20.

TABLE 3 **Selected r^2 values ordered by "right-scene" (using composite_data_set_filtered); selected* results reported below for reference were originally published in Van Den Wymelenberg and Inanici [2014]; however, in this article, Evalglare version 1.11 was used for DGP and DGI results**[a]

Rank	Other metrics of interest (by right-scene)	QU1	QU2	QU3	QU4	QU5	QU6	QU7	Likert_all	Front-scene	Left-scene	Right-scene	Ceiling	Light-in-scene
30	*E vertical top of monitor ($E_{v\text{-monitor}}$)	**0.239**	**0.200**	**0.150**	**0.260**	0.118	**0.283**	**0.213**	**0.250**	0.104	0.131	**0.298**	0.091	0.049
53	SD L 40° horizontal band	0.211	0.171	0.115	0.209	0.098	0.185	0.169	0.198	0.076	0.058	0.280	0.046	0.016
59	*Mean L scene	0.200	0.172	0.115	0.223	0.097	0.243	0.168	0.207	0.085	0.104	0.278	0.095	0.065
78	*SD L scene	0.202	0.179	0.114	0.219	0.104	0.223	0.174	0.207	0.059	0.060	0.276	0.089	0.063
141	*E vertical at camera ($E_{v\text{-eye}}$)	0.207	0.170	0.121	0.235	0.097	0.263	0.181	0.216	0.085	0.103	0.267	**0.118**	0.010
150	Mean L glare sources (5* mean L scene)	0.195	0.169	0.106	0.210	0.099	0.222	0.162	0.198	0.067	0.074	0.260	0.096	0.053
152	Percentage of pixels in 40° horizontal band above 2000 cd/m^2	0.164	0.120	0.082	0.162	0.065	0.146	0.103	0.143	**0.126**	0.097	0.257	0.045	**0.115**
182	Mean L brightest 10% scene pixels	0.175	0.150	0.096	0.193	0.087	0.209	0.145	0.179	0.068	0.076	0.241	0.078	0.049
237	98th percentile L value scene	0.164	0.138	0.086	0.178	0.079	0.182	0.131	0.163	0.082	0.080	0.221	0.063	0.057
254	*Percentage of pixels in scene above 2000 cd/m^2	0.152	0.120	0.082	0.162	0.060	0.189	0.113	0.148	0.090	0.088	0.214	0.076	0.055
261	*DGP (5* mean L task)	0.148	0.122	0.087	0.175	0.063	0.199	0.124	0.154	0.071	0.078	0.212	0.060	0.049
272	*DGP (5* mean L scene)	0.138	0.114	0.082	0.164	0.059	0.188	0.117	0.145	0.069	0.078	0.210	0.062	0.064
313	DGI (>500 cd/m^2)	0.126	0.104	0.072	0.133	0.062	0.137	0.099	0.125	0.058	0.039	0.197	0.048	0.078
326	Percentage of pixels in window above 2000 cd/m^2	0.124	0.084	0.061	0.130	0.043	0.123	0.073	0.107	0.101	0.072	0.192	0.037	0.068
459	*Irradiance vertical at SW exterior (adj.)	0.107	0.100	0.067	0.132	0.061	0.145	0.106	0.122	0.038	0.070	0.149	0.051	0.008
437	*Mean L window: Mean L task	0.091	0.058	0.040	0.096	0.032	0.068	0.050	0.074	0.061	0.028	0.145	0.016	0.005
484	*E horizontal desktop	0.107	0.113	0.086	0.118	**0.128**	0.114	0.135	0.137	0.010	0.020	0.113	0.018	0.001

(*Continued*)

TABLE 3 (Continued)

Rank	Other metrics of interest (by right-scene)	QU1	QU2	QU3	QU4	QU5	QU6	QU7	Likert_all	Front-scene	Left-scene	Right-scene	Ceiling	Light-in-scene
503	25th:75th percentile L value in scene	0.054	0.035	0.027	0.068	0.020	0.071	0.032	0.051	0.052	0.067	0.099	0.026	0.012
509	Mean L central 60° FOV: Mean L scene	0.090	0.075	0.056	0.093	0.030	0.100	0.079	0.088	0.033	0.030	0.097	0.041	0.009
510	*E horizontal top of monitor	0.079	0.076	0.049	0.087	0.043	0.119	0.075	0.089	0.046	0.052	0.096	0.074	0.049
528	Mean L central 60° FOV: Mean L 120°–180° FOV	0.088	0.073	0.056	0.087	0.030	0.091	0.077	0.085	0.029	0.025	0.090	0.037	0.009
531	Percentage of pixels in scene below 30 cd/m^2	0.051	0.042	0.035	0.078	0.017	0.106	0.044	0.061	0.042	0.052	0.088	0.035	0.017
536	Mean L front wall: Mean L task	0.051	0.033	0.029	0.068	0.011	0.064	0.031	0.047	0.056	0.058	0.086	0.025	0.000
552	*Mean L task: Mean L scene	0.058	0.042	0.039	0.076	0.014	0.085	0.045	0.059	0.047	0.41	0.081	0.028	0.000
582	Mean L central 60° FOV: SD L 120°–180° FOV	0.074	0.064	0.051	0.061	0.032	0.055	0.061	0.068	0.008	0.004	0.064	0.024	0.023
584	Mean L central 60° FOV: SD L scene	0.074	0.063	0.052	0.063	0.033	0.056	0.060	0.069	0.008	0.003	0.063	0.022	0.022
629	*Coefficient of variation L scene	0.051	0.050	0.030	0.038	0.032	0.031	0.050	0.048	0.000	0.000	0.043	0.017	0.004

[a] QU1—This is a visually comfortable environment for office work.
QU2—I am pleased with the visual appearance of the office.
QU3—I like the vertical surface brightness.
QU4—I am satisfied with the amount of light for computer work.
QU5—I am satisfied with the amount of light for paper-based reading work.
QU6—The computer screen is legible and does not have reflections.
QU7—The lighting is distributed well.
Front-scene—When I look up from my desk does the scene I see in front of me seems (too dim–too bright).
Left-scene—When I look to my left the scene that I see seems (too dim–too bright).
Right-scene—When I look to my right the scene that I see seems (too dim–too bright); this direct included the window.
Ceiling—I find the ceiling to be (too dim–too bright).
Light-in-scene—I find this lighting condition to be: (least preferred–most preferred) [C9, C10, C12, C13, C15, C16 only].
Bolded numbers indicate that the metric's r^2 value was the highest ranked overall for a specific item. Pink fill indicates that the metric's r^2 value was greater than or equal to 0.20 and yellow fill indicates a value greater than or equal to 0.10 but less than 0.20.

0.4245. This example is based upon a single regression with 690 degrees of freedom (F_{1690}), where $F_{1690} = 510.4$, and has significance at $P < 0.00001$ (written ${}_{\text{adj}}r^2 = 0.43$, $F_{1690} = 510.4$, $P < 0.01$).

Luminance metrics based upon Mask 08, Mask 10, and Mask 20 are the most common among the top 20 metrics for right-scene. Standard deviation of window luminance has the highest squared correlation coefficient for six of the seven Likert items as well as right-scene. There are several metrics based upon the horizontal 40° band within the field of view (FOV; Mask 20) in the top 20, including mean luminance of the 40° horizontal band, percentage of pixels in the 40° horizontal band below 1000 cd/m^2, and the 75th percentile luminance value within the 40° horizontal band. No metrics based upon the entire scene (Mask 01) ranked in the top 20, nor did any based upon illuminance or irradiance data, DGI, DGP (or any other glare indices), luminance ratios, or contrast ratios. As noted above, right-scene is the item with the overall highest squared correlation coefficient. However, QU1 ("This is a visually comfortable environment for office work"), QU4 ("I am satisfied with the amount of light for computer work"), and QU6 ("The computer screen is legible and does not have reflections") are clustered together as the next highest and address a different construct than right-scene, namely, a more holistic assessment of human visual preference and acceptance. Of these, QU1 represents the most general characterization of visual comfort and is therefore reported in addition to right-scene in the detailed examples below.

The following section presents detailed results for two top-rated metrics based on the composite data set. Table 4 and Table 6 below include results for both filtered and unfiltered data for the composite data set as well as for a data set using just C8 with C10 (C8C10). Data for C8C10 were also reported because these conditions represent MP (C8) and JU (C10) daylight conditions (only adjusting blinds, electric lights off) in close time step (<30 min) and in the afternoon when sun penetration was possible. Scatter plots are shown with a first-degree line of fit as well as a *loess* (locally weighted polynomial regression smoothing) polynomial line for each metric relative to both QU1 and right-scene. The correlation coefficients shown are adjusted r^2 values (${}_{\text{adj}}r^2$) and represent the first-degree line of fit. Loess methods are suggested for nonparametric and exploratory analyses [Cleveland and Devlin 1988] and are well-suited to the nature of this research. However, the adjusted correlation coefficients are reported for the first-degree line of fit so as not to be overstated due to the potential "overfitting" of the loess curve. Each metric was also plotted using C8C10 data, ordered by the metric result, and data points were color-coded by the subjective responses to QU1. These plots are useful in discerning the most preferred and least preferred ranges of the metric as well as the typical changeover range, described hereafter as the bounded-borderline between comfort and discomfort, or bounded-BCD, following [Van Den Wymelenberg and Inanici 2014]. They are therefore the most useful for indicating recommended performance criteria; however, these must be considered preliminary in nature.

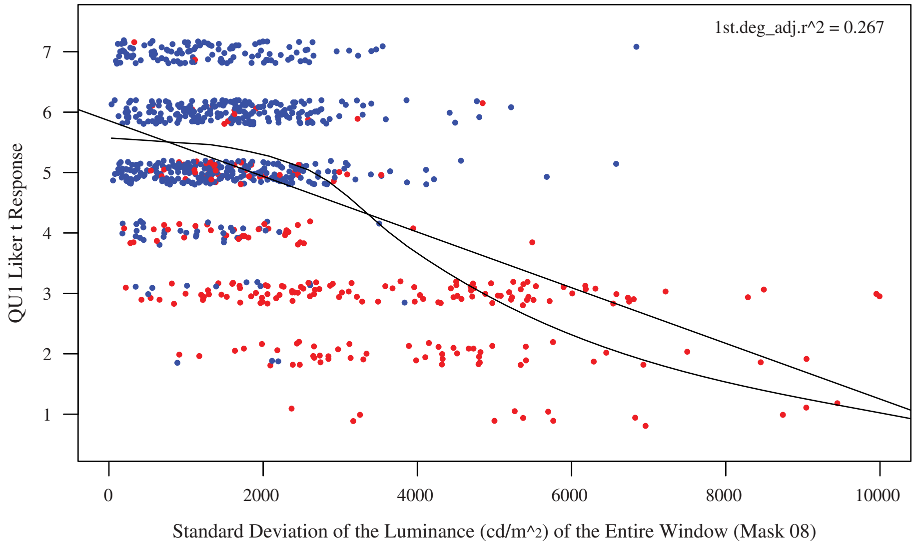
3.1.1. Standard Deviation of Window Luminance (Mask 08)

The standard deviation of the luminance values within the entire window (Mask 08) represents the highest squared correlation coefficient for six of the seven Likert items (all except QU6) as well as right-scene for the composite data set. It is also one of the 10 highest metrics for QU6 and the rating of "ceiling" brightness. Figure 3 represents the ability of the metric to explain the variance in QU1 and right-scene for the composite data set (JU data are red, MP data are blue). The single regression statistics can be seen in Table 4. Finally, Fig. 4 takes the C8C10 data, organizes it by the metric result, and color-codes it by the response to QU1 (where 7 = very strongly agree is dark blue and 1 = very strongly disagree is dark red). This graphic reveals three preliminary thresholds for criteria development as described in Table 5.

3.1.2. Mean Luminance of 40° Horizontal Band (Mask 20)

The mean of the luminance values within the 40° horizontal band (Mask 20 shown in Fig. 2) represents the highest squared correlation coefficient for any metric based upon a scene-independent mask (whereas some masks are space specific; for example, Mask 08). It is one of the 10 highest squared correlation coefficients for QU1, QU2, front-scene, and right-scene and is in the top 20 for QU4, QU5, and QU7. Figure 5 represents the ability of the metric to explain the variance in QU1 and right-scene for the composite data set (JU data are red, MP data are blue). The single regression statistics can be seen in Table 6. Finally, Fig. 6 takes the C8C10 data, organizes it by the metric result, and color-codes it by the response to QU1 (where 7 = very strongly agree is dark blue and 1 = very strongly disagree is dark red). This graphic reveals three preliminary thresholds for criteria development as described in Table 7.

(a) QU1 vs. Standard Deviation of The Luminance of the Entire Window (Mask 08) Using Composite Data Set



(b) Right_Scene vs. Standard Deviation of the Luminance of the Entire Window (Mask 08) Using Composite Data Set

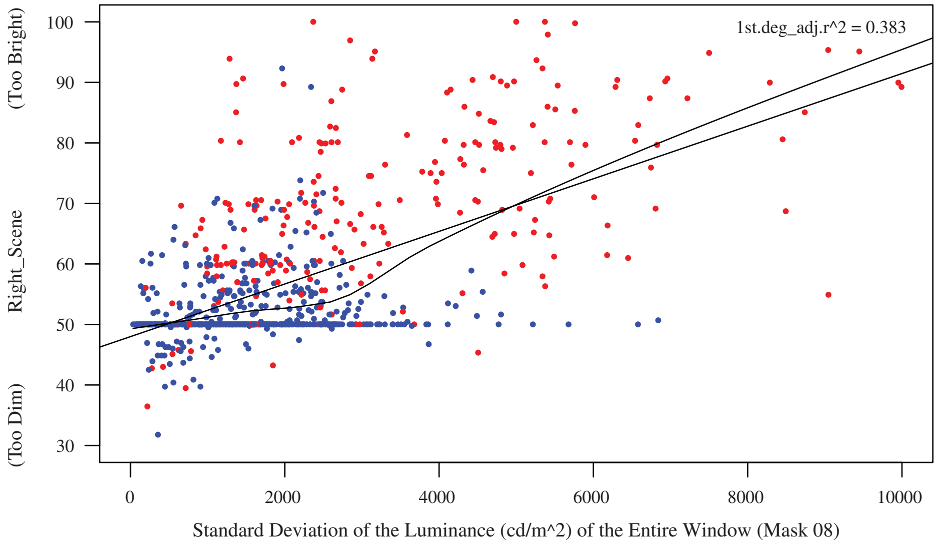


Fig. 3 **Standard deviation L window (Mask 08) versus subjective ratings of QU1 (top) and right-scene (bottom) for the composite data set (JU data are red, MP data are blue).**

3.2. Multiple Regressions and Subjective Responses

Though this research focused primarily on single regression analysis aimed at describing the strengths and limitations of individual metrics, the pursuit of logical multiple-metric models was pursued through multiple linear regression methods. Metrics with the highest squared correlation coefficients were organized in a correlation matrix along with other selected metrics of interest in order to determine which metrics were not highly correlated with one another as a starting point for assembling multiple-metric models. This is an important step in order to be assured that the metrics chosen for the model are addressing separate phenomena. Only logical metric combinations were explored, and a limit of three metrics was applied. The metrics were selected based upon a hypothesis that each might describe one or more important lighting characteristics, such as access to view, scene or surface luminance variability, proportion of the scene that is too bright or dim, and luminance contrast or glare.

TABLE 4 Standard deviation L window, single regression results

C8C10: Standard deviation L window (cd/m^2)				
DV	$_{\text{adj}}r^2$	F-statistic	DF	*P* value
C8C10				
QU1	0.2880	70.98	172	1.39E-14
right-scene	0.3553	96.32	172	2.20E-16
Composite_data_set				
QU1	0.2667	314.10	860	2.20E-16
right-scene	0.3834	536.40	860	2.20E-16
C8C10 Q4-filtered				
QU1	0.3108	59.18	128	3.36E-12
right-scene	0.3526	71.26	128	5.81E-14
Composite_data_set Q4-filtered				
QU1	0.2973	293.40	690	2.20E-16
right-scene	0.4245	510.40	690	2.20E-16

The best model identified for its ability to fit the results of right-scene produced an $_{\text{adj}}R^2 = 0.49$, $F_{3688} = 221.5$, P value < 0.01. It is detailed in Table 8 and was built using the following:

1. Standard deviation of window luminance
2. 50th percentile luminance value from the lower window
3. Mean luminance of the 40° horizontal band

One additional model is reported due to its overall strength and logic. It produced an $_{\text{adj}}R^2 = 0.32$, $F_{3688} = 107.9$, P value < 0.01 for QU1, and as shown in Table 9, an $_{\text{adj}}R^2 = 0.45$, $F_{3688} = 190.7$, P value < 0.01 for right-scene. It was built using the following:

TABLE 5 Standard deviation L window, range and preliminary criteria

C8C10: Standard deviation L window (cd/m^2) range						
Min.	First quartile	Median	Mean	Third quartile	Maximum	Σ
175	1386	2503	2842	3928	9952	1892

Preliminary criteria:

$x < 2500$	Likely to be comfortable
$2500 > x < 4000$	Bounded-BCD
$x > 4000$	Likely to be uncomfortable

1. Standard deviation L window luminance
2. Percentage of pixels in the 40° horizontal band above 2000 cd/m^2
3. Percentage of pixels in the 40° horizontal band below 1000 cd/m^2

4. DISCUSSION

4.1. Illuminance-Based Metrics

Illuminance-based metrics did not rank highly for any subjective items. As previously reported [Van Den Wymelenberg and Inanici 2014], the highest overall squared correlation coefficient for an illuminance-based metric (using the composite data set) was for right-scene and E_v at the top of the monitor measured in the participants' viewing direction, producing $r^2 = 0.298$, whereas the highest luminance-based metric was with

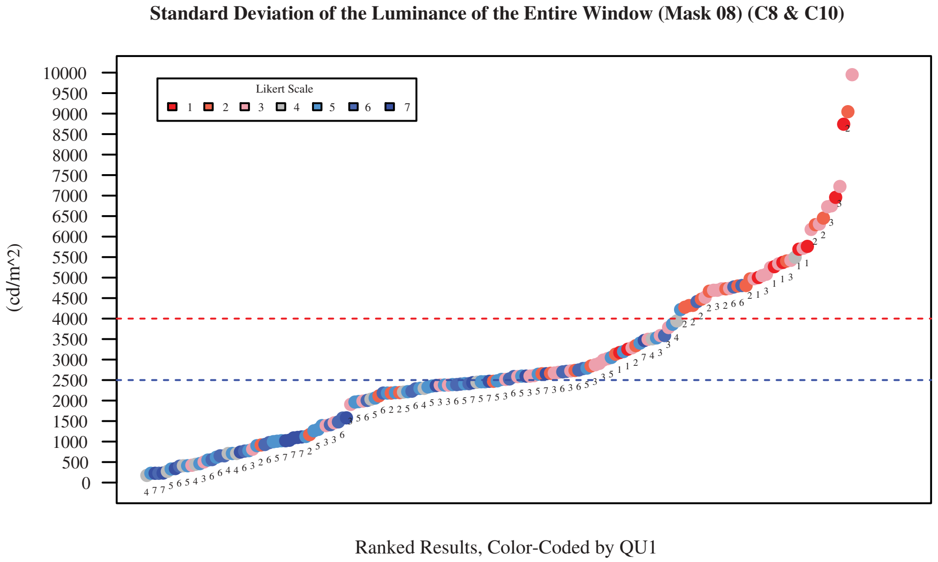
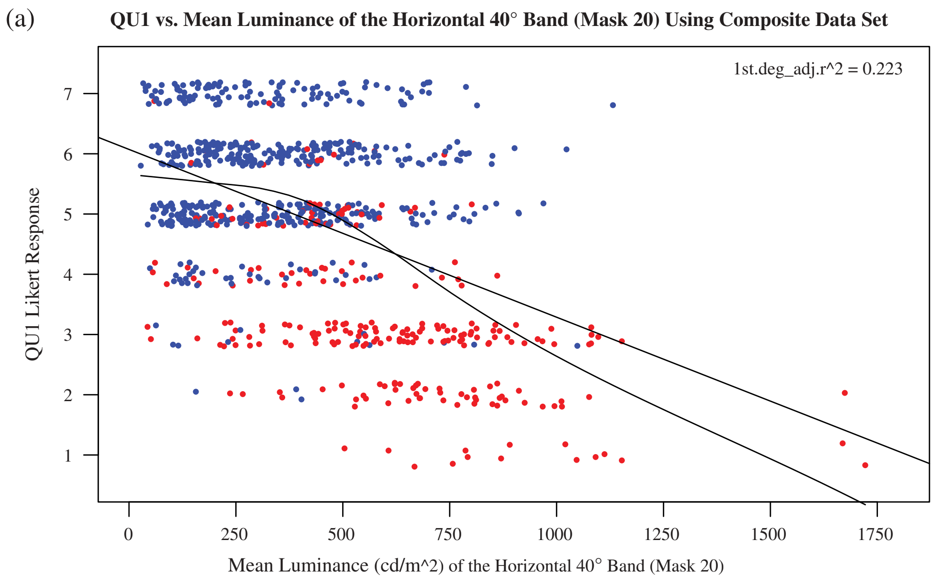


Fig. 4 Standard deviation L window (Mask 08) for C8 and C10. Results ordered by metric and color-coded by response to QU1 (7 = very strongly agree is dark blue; 1 = very strongly disagree is dark red).


(a) QU1 vs. Mean Luminance of the Horizontal 40° Band (Mask 20) Using Composite Data Set

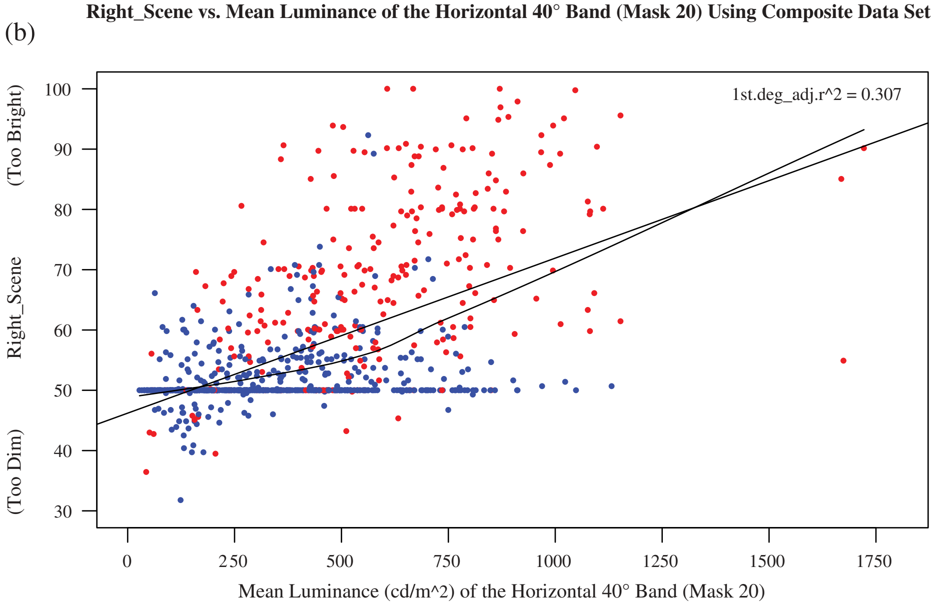

(b) Right_Scene vs. Mean Luminance of the Horizontal 40° Band (Mask 20) Using Composite Data Set

Fig. 5 **Mean L of 40° horizontal band (MASK 20) versus subjective ratings of QU1 (top) and right-scene (bottom) for the composite data set (JU data are red, MP data are blue).**

standard deviation of the window luminance producing a much higher squared correlation coefficient ($r^2 = 0.425$). The next highest squared correlation coefficient for an illuminance-based metric was for QU6 and E_v on the northeast wall ($r^2 = 0.284$) and once again, E_v at the top of monitor in the participants' viewing direction ($r^2 = 0.283$), and the best luminance-based metric for QU6 produced $r^2 = 0.288$ (25th percentile L of the window), a negligible increase.

The finding that luminance-based measures outperformed illuminance-based measures is somewhat contrary to Newsham and others [2008], who noted that E_{desktop} outperformed the best luminance-based measure (luminance ratio 75%:25% pixel value, 0.36 versus 0.31 as reported by Newsham and others 2008). However, their study did not use subjective ratings of human visual preference and acceptance directly as the variable of comparison; rather, they used the participants' electric lighting dimmer choice while performing typical office activities, including paper-based tasks. It could be that the differing result is partly due to the variable used for comparison (dimmer choice rather than subjective responses to comfort questions), it could be explained by differences in the amount of paper-based tasks in the two studies, and it could also be explained by the fact that fewer luminance-based metrics were tested by Newsham and others [2008]. Interestingly,

TABLE 6 Mean L of the 40° horizontal band, single regression results

C8C10: Mean L of the 40° horizontal band (cd/m^2)				
DV	$_{adj}r^2$	F-statistic	DF	P value
C8C10				
QU1	0.2889	71.27	172	1.25E-14
right-scene	0.3230	83.54	172	2.20E-16
Composite_data_set				
QU1	0.2234	248.7	860	2.20E-16
right-scene	0.3075	383.3	860	2.20E-16
C8C10 Q4-filtered				
QU1	0.3615	74.02	128	2.38E-14
right-scene	0.3601	73.6	128	2.72E-14
Composite_data_ Q4-filtered				
QU1	0.2425	222.2	690	2.20E-16
right-scene	0.3305	341.7	690	2.20E-16

desktop illuminance ranked higher using QU5 (paper-based tasks) than it did for all other subjective items examined.

4.2. Luminance-Based Metrics

Luminance-based metrics had higher squared correlation coefficients than illuminance-based metrics for all subjective questionnaire items. Luminance metrics based upon the horizontal 40° band within the FOV (Mask 20) and window masks (Mask 08 and Mask 10) are the most common among the top 20 metrics for right-scene.

TABLE 7 Mean L of the 40° horizontal band, range and preliminary criteria

C8C10: Mean L of the 40° horizontal band (cd/m^2) range						
Min.	First quartile	Median	Mean	Third quartile	Maximum	Σ
51	278	509	533	750	1674	311

Preliminary criteria:

$x < 500$	Likely to be comfortable
$500 > x < 700$	Bounded-BCD
$x > 700$	Likely to be uncomfortable

Results of several previously reported promising metrics are not reported in detail herein (see summary in Table 3). Specifically, the mean luminance of the scene, the percentage of pixels in scene exceeding 2000 cd/m^2 [Van Den Wymelenberg and others 2010], the coefficient of variation of the entire scene [DiLaura and others 2011; Howlett and others 2007], and the ratio of the 75th:25th luminance value in the entire scene [Newsham and others 2008], are not detailed because their r^2 results did not rank among the highest metrics investigated. As indicated previously [Van Den Wymelenberg and Inanici 2014], this finding underscores potential challenges to generalizability for luminance-based metrics.

4.2.1. Standard Deviation L Window

Standard deviation of the window luminance (Mask 08) was the highest overall lighting metric for nearly

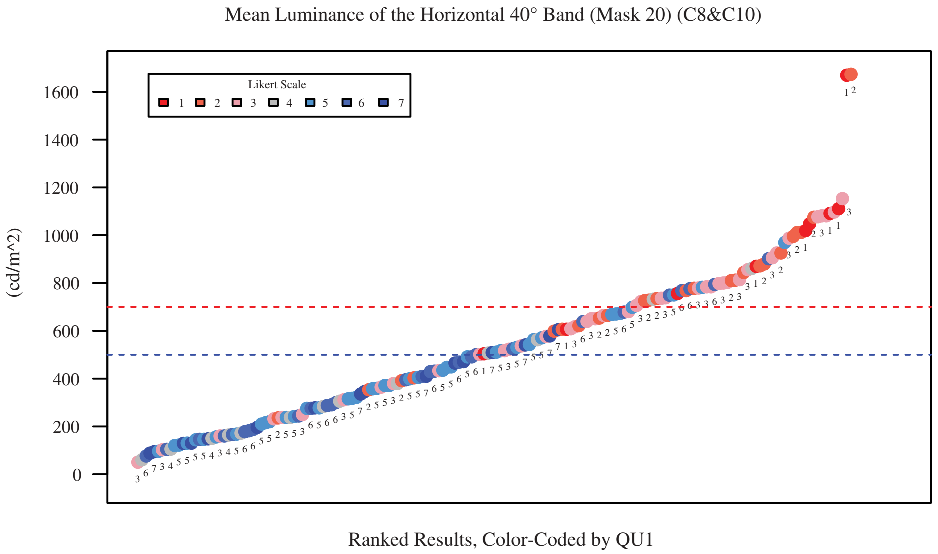


Fig. 6 Mean of L within the 40° horizontal band (C20) for C8 and C10. Results ordered by metric and color-coded by response to QU1 (7 = very strongly agree is dark blue; 1 = very strongly disagree is dark red).

TABLE 8 Multiple regression: right-scene versus standard deviation L window + 50th percentile L the lower window + Mean L of the 40° horizontal band[a]

			Single regression			
DV	Metric	$_{\mathrm{adj}}r^2$	F-statistic	Number of variables	DF	*P* value
right-scene	Standard deviation L window	0.4245	510.4	1	690	2.20E-16
right-scene	50th percentile L value in the lower window	0.3688	404.8	1	690	2.20E-16
right-scene	Mean L of the 40° horizontal band	0.3305	341.7	1	690	2.20E-16
			Multiple regression summary			
			Estimate	SE	*t* value	Pr(>|*t*|)
	(Intercept)		46.0289947	0.5951	77.352	2.00E-16
	Standard deviation L window		0.0040222	0.0003	12.326	2.00E-16
	50th percentile L value in the lower window		0.0153788	0.0017	9.022	2.00E-16
	Mean L of the 40° horizontal band		−0.0147608	0.0031	−4.726	2.78E-06
		$_{\mathrm{adj}}R^2$	F-statistic	Number of variables	DF	*P* value
right-scene	multiple-model	0.4891	221.5	3	688	2.20E-16
			Multiple regression ANOVA table			
		DF	Sum of squares	Mean square	F value	Pr(>F)
	Standard deviation L window	1	41,110	41,110	5.75E+02	2.20E-16
	50th percentile L value in the lower window	1	4793	4793	6.70E+01	1.29E-15
	Mean L of the 40° horizontal band	1	1597	1597	2.23E+01	2.78E-06
	Residuals	688	49,188	71		

[a] Using: Composite_data_set Q4-filtered.

all subjective items. This is an encouraging finding for several reasons. Given this data set, it outperforms all current best practice lighting design metrics with regard to its ability to describe the variance in a range of subjective ratings of visual preference and acceptance in an office with daylight. Again, given this data set, the results of this metric appear to separate into three categories of subjective response (Fig. 4): scenes likely to be comfortable, scenes likely to be uncomfortable, and scenes that fall within a bounded-BCD. Standard deviation is a logical and commonly understood description of variability. Because the window region is often perceived as the brightest light source in spaces with daylight, focusing on the variability of this region is an intuitive approach to support luminance-based design analysis as well as for automated blinds control. In most office applications, defining the window area is straightforward because it is typically defined by clear architectural boundaries, thus supporting both field research and simulation-based design analysis using this metric. It is relatively easy to calculate and is computationally inexpensive. Spreadsheets or available software applications can compute the metric given a defined set of luminance values; thus, it does not require specialized software or scripting. The metric is simple and firmly defined and thus will inherently resist subtle manipulation aimed at improving the fit of the metric to a given sample. Often, studies aimed at improving the fit lead to overfitting the algorithm to the specific sample rather than improving the metric's ability to describe the variability of the population.

Though standard deviation of window luminance has many positive attributes, there are also some drawbacks. This metric requires masking a specific region of an HDR for analysis, necessitating an intermediary step.

TABLE 9 Multiple regression: right-scene versus standard deviation L window + percentage of pixels in the 40° horizontal band above 2000 cd/m^2 + percentage of pixels in the 40° horizontal band below 1000 cd/m^2[a]

Single regression						
DV	Metric	$_{adj}r^2$	F-statistic	Number of variables	DF	*P* value
right-scene	Standard deviation L window	0.4245	510.4	1	690	2.20E-16
right-scene	Percentage of pixels in the 40° horizontal band above 2000 cd/m^2	0.2563	239.1	1	690	2.20E-16
right-scene	Percentage of pixels in the 40° horizontal band below 1000 cd/m^2	0.3176	322.5	1	690	2.20E-16
Multiple regression summary						
			Estimate	SE	*t* value	Pr(>$\lvert t \rvert$)
	(Intercept)		142.1	22.29	6.376	3.34E-10
	Standard deviation L window		0.003446	0.0002702	12.752	2.00E-16
	Percentage of pixels in the 40° horizontal band above 2000 cd/m^2		−68.47	32.48	−2.108	3.54E-02
	Percentage of pixels in the 40° horizontal band below 1000 cd/m^2		−96.83	22.52	−4.3	1.96E-05
		$_{adj}R^2$	F-statistic	Number of variables	DF	*P* value
right-scene	multiple-model	0.4516	190.7	3	688	2.20E-16
Multiple regression ANOVA table						
		DF	Sum of squares	Mean square	F value	Pr(>F)
	Standard deviation L window	1	41,110	41,110	535.768	2.20E-16
	Percentage of pixels in the 40° horizontal band above 2000 cd/m^2	1	1368	1368	17.825	2.75E-05
	Percentage of pixels in the 40° horizontal band below 1000 cd/m^2	1	1419	1419	18.487	1.96E-05
	Residuals	688	52,792	77		

[a] Using: Composite_data_set Q4-filtered.

However, this is now easily accomplished using software such as hdrscope [Kumaragurubaran and Inanici 2013]. Still, because of the demands of the required mask, the metric is highly specific to space and position. That is, every space and every workstation position within a space requires a unique mask to be defined for analysis due to changes in window patterns from space to space or due to the changes in proximity to the window within a given space. This means that either field research or simulation analysis will likely require creating a unique mask for each position and view direction of interest. Future research can examine what resolution of analysis points is required to adequately characterize a given space using this metric. Because of these limitations, it is also useful to examine metrics that are based upon position-independent masks (for example, Mask 01, Mask 19, Mask 20). Using these masks will reduce analysis time because they can be consistently applied across a wide range of spaces and positions.

In practice, standard deviation of window luminance is likely to be particularly useful as a metric for blind control purposes rather than for simulation-based daylighting design analysis purposes. Firstly, it is likely that the metric's sensitivity is predicated on some type of blind manipulation or level of scene detail out of a window that is not commonly found during schematic design simulation practice. Secondly, the metric is insensitive to a host of architectural factors that impact daylighting performance inside the envelope. For example, this metric is not likely to be capable of evaluating basic architectural aspects such as room depth, interior finishes, or furnishings. Therefore, it is advisable to calculate this metric as one of several useful inputs for design analysis purposes and consider it as having great potential for inclusion in algorithms controlling automated blind position in real spaces aiming to optimize visual comfort.

Figure 7 summarizes the range of results for standard deviation of window luminance as found in C8C10 and reports the minimum, first quartile, median, mean, third quartile, and maximum results in numerical and graphical manner. It also summarizes the results of each of the subjective responses that correspond to the presented luminance results (see Fig. 8 for a reference scale). This graphic is provided to give the reader a more intuitive understanding of how the metric reacts across a wide range of visual conditions for a single space across time. It is interesting to note that extremely low standard deviation was rated as uncomfortable and the semantic differential results note that the space was too dim overall. This could be the result of a participant who felt that he had to close the blinds to avoid glare (note the small sun spots peeking through blind cord holes) but, in so doing, felt that the space was too dim and rated it as uncomfortable. The cluster of "too dim" ratings in Fig. 3 (bottom) provides further evidence of this finding. This seems to indicate that people may feel a need for a certain amount of window luminance variability, and it is possible that future research will identify a lower "sufficiency" threshold for this or other metrics, thus suggesting that it has use as more than just a glare metric.

4.2.2. Mean L of 40° Horizontal Band (Mask 20)

The mean luminance of the 40° horizontal band produced one of the highest squared correlation coefficients with right-scene ($r^2 = 0.33$ for the composite data set). It is another example of a very simple metric and therefore shares many of the attributes with the standard deviation of window luminance. However, it also has the benefit of being a scene-independent metric. That is, it can be applied directly to any space or position within a space without modification. This metric appears to be robust across time within the space studied herein; however, it may prove too simplistic because it is applied to a broad range of designs. Metrics calculated using Mask 20 hold more promise than metrics calculated using just the window area for evaluating interior architectural daylighting design considerations due to the broader FOV. It is advisable to use this metric in combination with other metrics that describe variability, such as the standard deviation of the window luminance or standard deviation of the same 40° horizontal band. This metric has squared correlation coefficients with E_v as follows; $r^2 = 0.66$ with E_v at the top of the camera and $r^2 = 0.76$ with E_v the top of the monitor.

Figure 9 summarizes the range of results for mean luminance of 40° horizontal band as found in C8C10 as well as corresponding subjective responses. This graphic provides a visual representation of a wide range of visual conditions for a single space across time. Similar to the scenario described in the previous section, it is interesting to note that extremely low mean luminance of 40° horizontal band was rated as uncomfortable because it was too dim. In this case, it seems to be due to very dark outdoor conditions rather than extreme sunlight forcing blinds closed as noted in the previous section for standard deviation of window luminance. The clusters of "too dim" ratings in Fig. 5 (bottom) provide further evidence of this finding.

4.2.3. Mean Window L

Sutter and others [2006] reported that only 25% of people accepted mean window luminance greater than 3200 cd/m^2, and Lee and others [2007] used a proxy of 2000 cd/m^2 of mean window luminance as a "sky brightness" control signal for roller blinds at the *New York Times* headquarters. This study found that only 25% of participants who elected to leave blinds open during MP conditions accepted mean window luminance above approximately 2250 cd/m^2 and that the typical participant who left blinds open for MP conditions accepted approximately 1750 cd/m^2. Participants who lowered blinds between 25% and 75% of the way down typically accepted mean window luminance between 1100 and1500 cd/m^2, and 25% accepted a range between 1250 and 2000 cd/m^2. The bounded-BCD for this metric is between 2000 and 2500 cd/m^2 as seen in Fig. 10, and its ability to predict

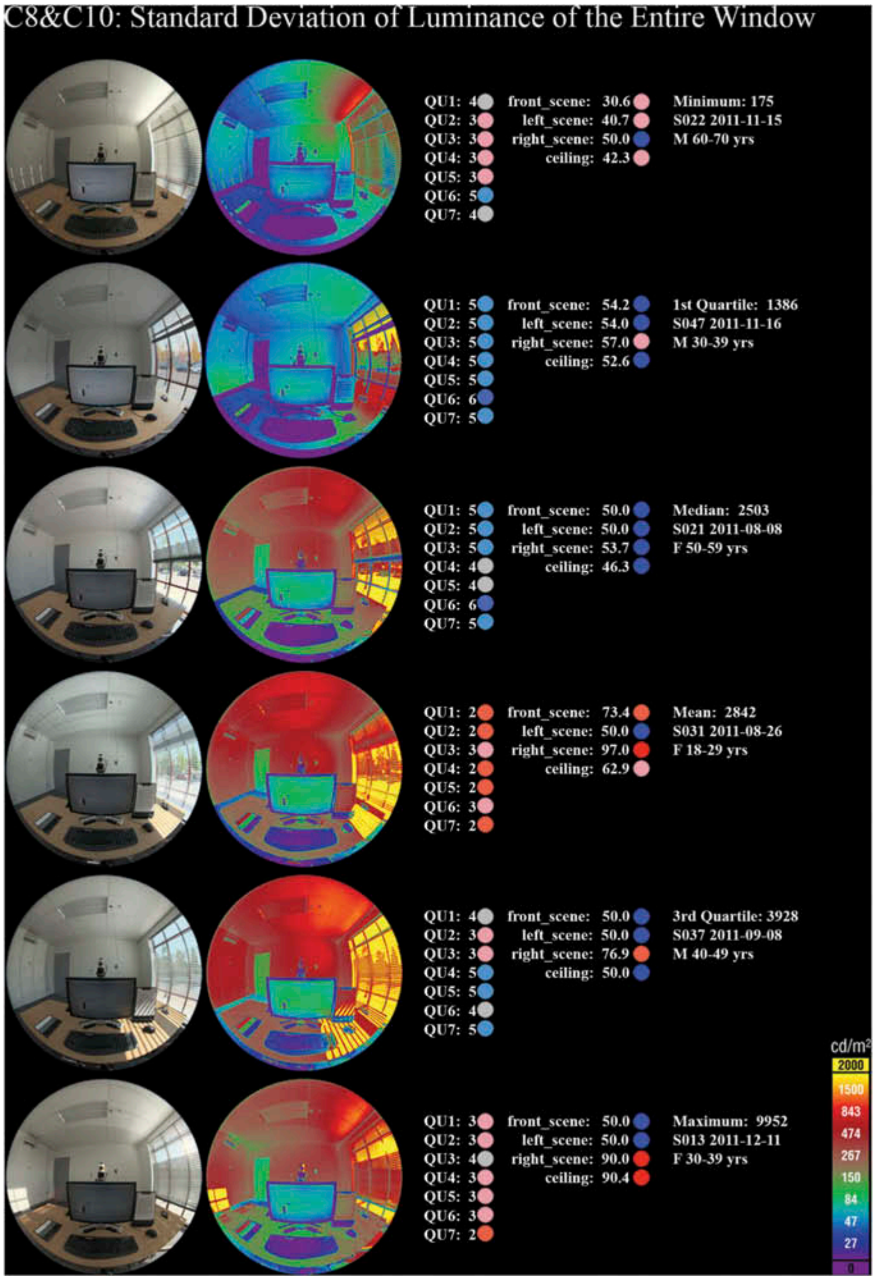


Fig. 7 Summary range of the results for standard deviation L window including tone-mapped image, false color L plot, and subjective response data (minimum result at top, maximum result at bottom); color scales per Fig. 8.

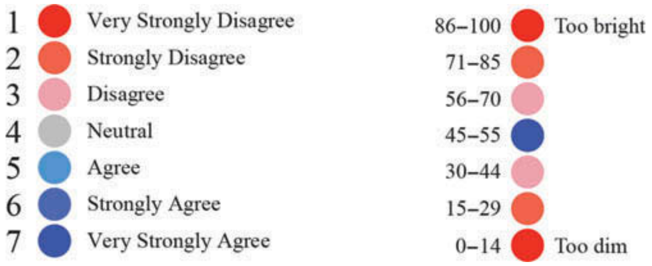


Fig. 8 Scale for use with summary range of the metrics figures.

visual comfort was $_{\mathrm{adj}}r^2 = 0.23$ and to predict whether the right-scene was too dim or too bright was $_{\mathrm{adj}}r^2 = 0.33$. Generally lower mean window luminance values were found than reported by Sutter and others [2006] and Lee and others [2007]. One possible explanation for the difference is that the *New York Times* headquarters uses shade fabric that tends to occlude a greater amount of view than the blinds used in this study; thus, occupants may accept brighter window luminance to preserve

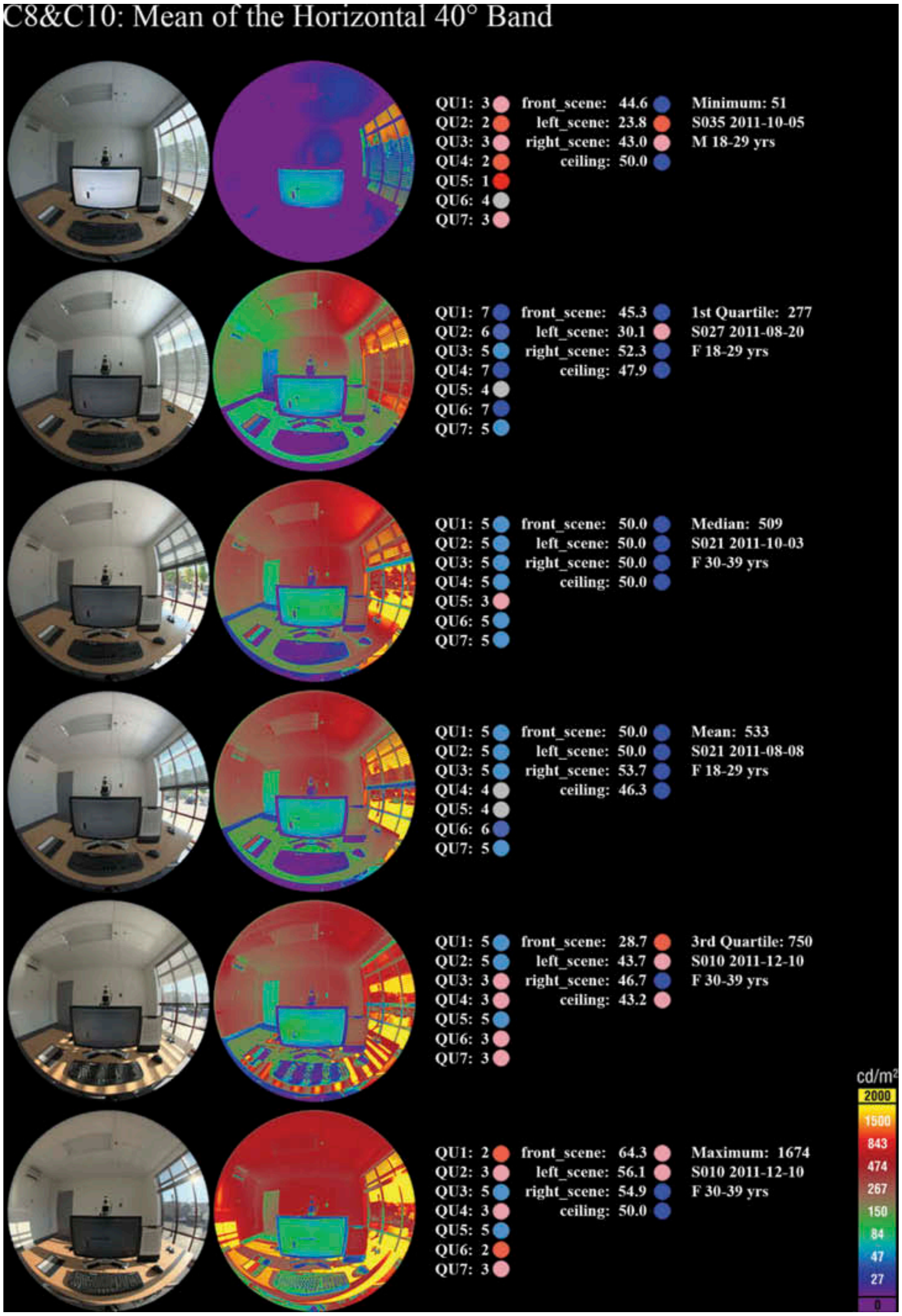


Fig. 9 Summary range of the results for mean L of the 40° horizontal band including tone-mapped image, false color L plot, and subjective response data (minimum result at top, maximum result at bottom); color scale per Fig. 8.

views. Other research supports this notion [Boubekri and Boyer 1992; Chauvel and others 1982; Tuaycharoen 2011; Tuaycharoen and Tregenza 2007].

4.2.4. Multiple Regressions

Multiple regression models were studied to determine whether a select group of no more than three variables could be combined to improve the predictive capabilities. Multiple regression models are inherently more complex than single regressions, are difficult to visualize, and therefore are more difficult to put into useful terms for practitioners with regard to recommended design criteria. However, because multiple regression models are also inherently stronger at predicting the variation in a subjective response than single regressions, and because this improved strength could be useful for both computational design analysis techniques and environmental control

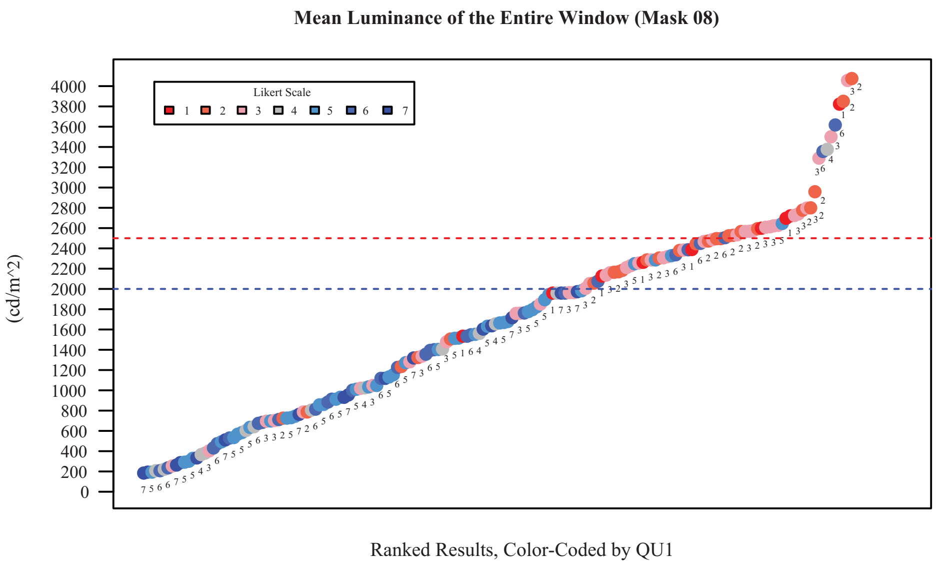


Fig. 10 **Mean of window L for C8 and C10, results ordered by metric and color-coded by response to QU1 (7 = very strongly agree is dark blue; 1 = very strongly disagree is dark red).**

purposes, a few of the stronger models are worth calculating to support design decisions and reporting in future research. The strongest multiple regression model was for the semantic differential rating (too dim–too bright) of the window wall ($_{\text{adj}}R^2 = 0.49$) and was built upon three variables, standard deviation of window luminance, the 50th percentile luminance value from the lower window, and the mean luminance of the 40° horizontal band. This model could be argued to address glare, access to view, and overall room brightness in the three variables respectively. One additional model is discussed due to its overall strength and sound logic. It produced an $_{\text{adj}}R^2 = 0.32$ for QU1 and $_{\text{adj}}R^2 = 0.45$ for right-scene, not the highest for either item individually but nearly the strongest when considering both. It was built using standard deviation of window luminance, percentage of pixels in the 40° horizontal band above 2000 cd/m^2, and percentage of pixels in the 40° horizontal band below 1000 cd/m^2 and can be argued to address luminance variability within the window, the amount of scene above an excessive threshold, and the amount of the scene below a lower threshold.

4.3. Usefulness and Limitations of Absolute Thresholds

Absolute luminance thresholds have several strengths and limitations. One of their greatest strengths is their simplicity, but this is also their downfall. It is well established that wide variability in human preference and acceptance of luminous conditions exists between individuals. Wide variability in luminous conditions also exists between different spaces and for a space with daylight over time. There is also evidence that human sensitivity and expectations for light change with time of day [Newsham and others 2008]. Therefore, it is not likely that any absolute threshold, whether it is illuminance or luminance based, will decisively differentiate between comfortable and uncomfortable luminous conditions in all cases. However, absolute thresholds appear to be useful at establishing extreme upper as well as extreme lower thresholds that are highly likely to be considered uncomfortable. This research design was better suited at addressing the very bright extremes, but some evidence of very dim extremes was identified and discussed above.

The use of absolute thresholds is likely to be more generalizable between individuals, between spaces, and across time if it is translated from a singular threshold value into a bounded-BCD criteria. This section also suggests that absolute thresholds can be more successful in fitting participants' subjective responses to luminous conditions with measured data if metrics based upon fixed thresholds (for example, percentage of pixels in the 40° horizontal band above 2000 cd/m^2, mean luminance of the 40° horizontal band, mean luminance of window) are considered together with metrics describing a scene's luminance variability (for example, standard deviation of window luminance, standard deviation of luminance in the 40° horizontal band).

4.4. Generalizability and the Preliminary Nature of Recommended Design Criteria

As with any research, these findings must be understood within the context in which they were developed. This investigation was conducted in a highly controlled daylighting laboratory, with daylight from one side and with participants in three age groups from 18 to 70 years of age. In these ways, it can be said to be reasonably similar to some other laboratory studies [Newsham and others 2008; Wienold and Christoffersen 2006]. There was one solar orientation (southwest), one view direction (parallel to the window looking southeast), and a single participant occupied the study space at a given time. Therefore, these findings apply most directly to private offices and are more challenging to apply to open-plan office environments due to several effects unique to this space type. The most notable of these are the multiple concurrent viewpoints of occupants in open-plan office environments, the diversity in occupants' visual preference and acceptance levels, and the social psychological aspects that may influence occupant behavior with regard to environmental control.

The recommended criteria are likely to be strongly influenced by the space configuration, the view direction relative to daylight sources, and, possibly, by subtle differences in position relative to daylight sources while maintaining similar view direction. Some of these effects (view direction, space configuration) can be illustrated by comparing findings from a two-day pilot study [Van Den Wymelenberg and others 2010] with the 6-month study. The two study spaces are relatively similar; however, in the pilot study the participants faced northwest and in the 6-month study the participants faced southeast, and the blind type and configuration were different. It appears that these relatively subtle architectural differences produce substantially different output with regard to some of the metrics tested, whereas other metrics perform similarly in both spaces.

There are several confounding differences between the pilot and the 6-month study, including widely varying sun angles, sky conditions, and blind types; the differences between the results of some metrics appear to persist even when isolating many of these factors. For example, the percentage of scene pixels above 2000 cd/m^2 was calculated for both the pilot study and the 6-month study. Figure 11 shows two comparisons between the pilot study and the 6-month study. In one example (Fig. 11, top), data from the pilot study captured on December 16 at 1:56 PM (Fig. 11, top left) are compared to data from the 6-month study captured on December 19 at 2:16 PM (Fig. 11, top right), both on relatively clear sunny days with partially deployed blinds. In the pilot study, the percentage of the scene above 2000 cd/m^2 is 16%, whereas the 6-month study shows only 4.5%. In a second example (Fig. 11, bottom), data from the pilot study captured on December 16 at 2:55 PM (Fig. 11, bottom left) are compared to data from the 6-month study captured on December 10 at 2:19 PM (Fig. 11, bottom right), both on relatively clear sunny days with blinds mostly open. In the pilot study the percentage of the scene above 2000 cd/m^2 is 19%, whereas the 6-month study shows only 9%. According to the preliminary criteria suggested by the 6-month study [Van Den Wymelenberg 2012, p 185], all four of these scenes are either "likely to be uncomfortable" (>6.5%) or in the bounded-BCD region (3.5% > x < 6.5%). This indicates some robustness of the metric, but these preliminary criteria are likely to overestimate the number of scenes within the pilot study space designated as "likely to be uncomfortable." This is likely due to the different viewing direction and the subsequent relationship of the sun patch within the FOV (that is, in the pilot study space the sun hits the wall in the FOV), along with the variability in occupants' responses. Furthermore, the results for percentage of the scene above 2000 cd/m^2 did not rank in the top 20 metrics for any subjective questionnaire items in the 6-month study but it was a strong metric in the pilot study.

Fortunately, some metrics appear to be robust against this phenomenon. For example, standard deviation of window luminance performs similarly between the pilot study and the 6-month study. This metric was not originally tested during the pilot phase and was calculated using the pilot data after it emerged as a strong metric from the 6-month study. This provides some initial corroboration. Figure 12 represents the ability of the metric to explain the variance in QU1 (visually_comfortable) and shows an $_{\text{adj}}r^2 = 0.28$ for the pilot study sample. Finally, Fig. 13 takes these data, organizes them by the metric result, and color-codes them by the response to QU1. The preliminary thresholds established from the 6-month study are plotted for the bounded-BCD using the pilot study sample. The results from the pilot study for this metric are very similar to the findings from the 6-month study, with the exception of a few extreme cases that have higher standard deviation in window luminance than was found in the 6-month study. This suggests that the potential for glare in the pilot study space is indeed higher than that of the 6-month study space. Most important, the preference (MP)

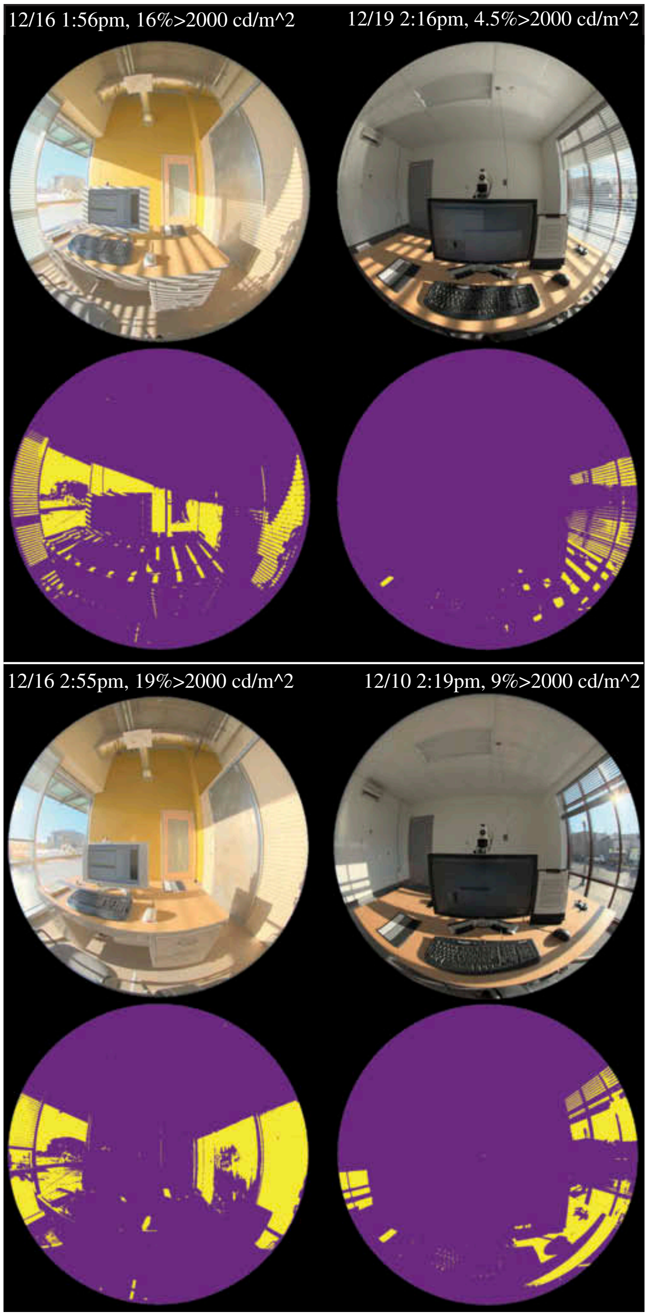


Fig. 11 Similar date- and time-stamped JU scenes; pilot (left), 6-month study (right); yellow > 2000 cd/m^2.

data follow the bounded-BCD criteria established in the 6-month study. These findings are a first, but important, step toward establishing the reliability of this metric and the corresponding design criteria.

Despite this initial corroboration for standard deviation of window luminance, these findings suggest that the performance criteria must be interpreted as preliminary in nature. The results for the pilot from the metric percentage of the scene pixels above 2000 cd/m^2 show that it can adequately and reliably differentiate MP and JU scenes within subjects, but it, like other metrics, will likely need to be fine-tuned to space- and position-specific considerations across buildings or different facade orientations within a building. This may limit some of the new metric's usefulness in the near-term as generic design performance criteria and suggests that further research from a wide range of space types, orientations, positions, and viewing directions within spaces is necessary to confirm, refine, or reshape the criteria. However, some metrics, such as standard deviation of window luminance, appear to be

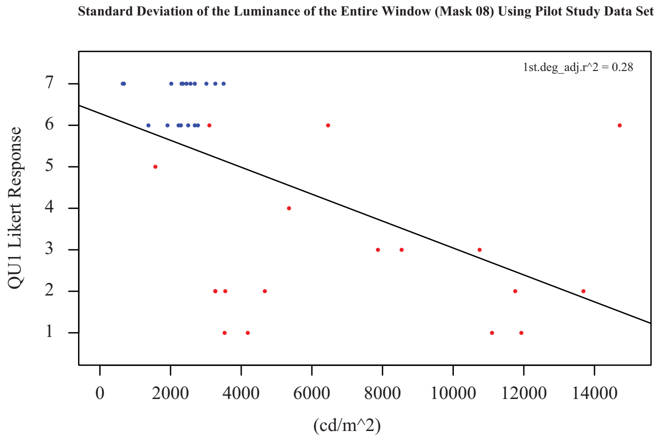


Fig. 12 **Standard deviation L window (Mask 08-pilot) by QU1 (JU data are red, MP data are blue).**

robust across at least two similar spaces, even with different view directions. Future research will reveal which metrics are highly dependent upon various factors (space, view direction, latitude, et cetera). Finally, the potential differences that may emerge for some metrics from future research do not necessarily present the same limitation for applications of environmental controls as they do for recommended design criteria. This is because environmental control applications will require some level of site-specific calibration and commissioning, and this can allow for fine-tuning of the metric's recommended criteria to individual spaces.

5. CONCLUSION

This article reports original results of a 6-month visual comfort human factors investigation in a single occupancy office fitted with exhaustive lighting data collection equipment and aims to establish a new suite of luminance-based lighting quality metrics that, with additional research, can support lighting design guidelines for improved visual comfort in spaces with daylight. Data from a sample including 48 participants were collected using repeated-measures design in an office space under natural and systematically categorized daylight conditions. Key conclusions include the following:

- Luminance-based metrics were more capable than illuminance-based metrics for fitting the range of subjective responses to questionnaire items pertaining to visual preference. Therefore, establishing reliable luminance-based metrics and design criteria that can guide designers in early design stages and be referenced by authors of design guidelines and standards should lead to increased occupant satisfaction in spaces achieving these criteria.
- The standard deviation of window luminance was the metric that best fit human subjective responses to visual preference on 7 of 12 questionnaire items. This metric consistently differentiated between MP and JU scenes and the recommended bounded-BCD was robust for two unique view directions and room configurations (pilot study room and 6-month study room). This

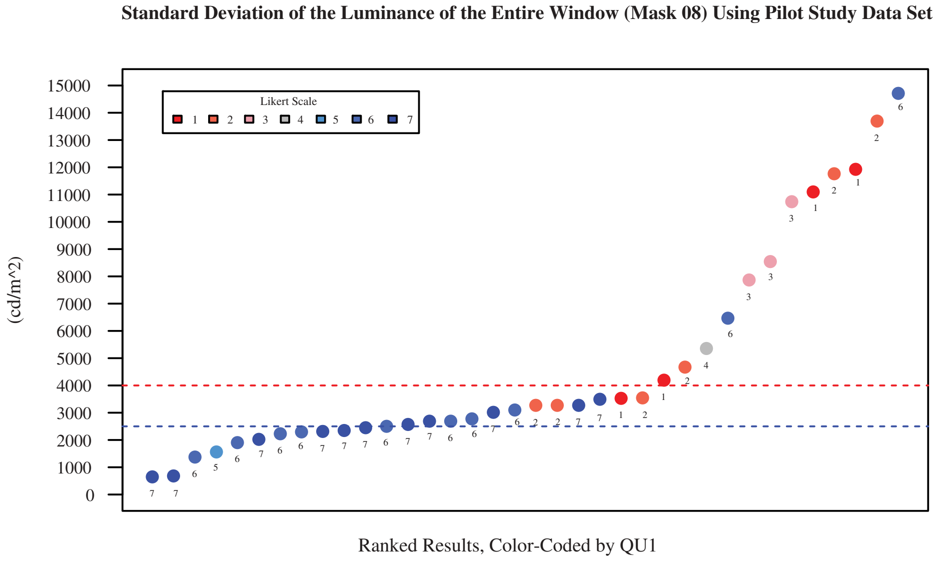


Fig. 13 **Standard deviation L window (Mask 08-pilot) for pilot-preferred (MP) and pilot-just disturbing (JU); results ordered by metric and color-coded by response to QU1, showing the bounded-BCD of 6-month study (7 = very strongly agree is dark blue; 1 = very strongly disagree is dark red).**

metric is easily understood; however, it requires a space- and position-specific mask to be created for each new space or position for which it is calculated.

- Luminance metrics calculated within the 40° horizontal band (Mask 20) ranked in the top 20 squared correlation coefficients for almost all subjective visual preference questionnaire items. Metrics calculated within this masked region were highly ranked more frequently and more consistently than metrics derived from any other region of analysis, including the masked region representing the entire 180° by 180° FOV (Mask 01). This is fortunate because this mask can be calculated with cameras incapable of capturing a full 180° by 180° FOV, thus reducing the cost of equipment for either purposes of research data collection or luminous environmental control.
- The strongest multiple regression model was for the semantic differential rating (too dim–too bright) of the window wall ($_{\mathrm{adj}}R^2 = 0.49$) and was built upon three variables below. These variables should be viewed as some of the most meaningful metrics to support future research and luminance-based environmental control. Because of the statistical significance of this multiple regression, it can be said that these metrics represent unique constructs of the luminous environment. Given the improved predictive ability of the multiple regression models as compared to any single regression model, they will be beneficial to future research and luminance-based environmental control:
  - Standard deviation of window luminance
  - 50th percentile luminance value from the lower window
  - Mean luminance of the 40° horizontal band

The human factors research-based results presented here are useful to evaluate and support recommendations for improved integrated lighting design strategies, computational analysis methods, and lighting and blind control technologies and to guide future visual comfort field research.

ACKNOWLEDGMENTS

Special thanks are due to Dr. Judith Heerwagen, Dr. Peter Johnson, and Professor Joel Loveland. We express our gratitude to University of Idaho Integrated Design Lab staff Brad Acker, Julia Day, Alen Mahic, and Dr. Ery Djunaedy for their assistance.

FUNDING

This work was conducted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Built Environment at the University of Washington and as a research project at the University of Idaho Integrated Design Lab. In these capacities the following organizations provided financial support for aspects of this project. The University of Washington PhD in the Built Environment Program through fellowships, the Nuckolls Fund for Educational Lighting through the Edison Price Fellowship, the Illuminating Engineering Society of North America, specifically the New York City Section's support through the Richard Kelly Grant and the Golden Gate Section's support through the Robert E. Thunen Memorial Scholarship, the International Association of Lighting Designers through the Lighting Design Alliance Scholarship, the Northwest Energy Efficiency Alliance and Idaho Power Company through support of technology demonstrations, Construction Specialties, and Warema International for partial product donations.

REFERENCES

Boubekri M, Boyer LL. 1992. Effect of window size and sunlight presence on glare. Lighting Res Technol. 24(2):69-74.

Boyce PR, Veitch JA, Newsham GR, Jones CC, Heerwagen J, Myer M, Hunter CM. 2006. Occupant use of switching and dimming controls in offices. Lighting Res Technol. 38(4):358-376.

Chauvel P, Collins JB, Dogniaux R, Longmore J. 1982. Glare from windows: current views of the problem. Lighting Res Technol. 14(1):31-46.

Cleveland WS, Devlin SJ. 1988. Locally weighted regression: an approach to regression analysis by local fitting. J Am Stat Assoc. 83(403):596-610.

DiLaura DL, Houser KW, Mistrick RG, Steffy GR. 2011. The lighting handbook. 10th ed. New York (NY): Illuminating Engineering Society of North America. 1328 p.

DIVA for Rhino. (n.d.). DIVA. Solemma LLC. <http://diva4rhino.com/> Accessed 2014 Apr 9.

Fraunhofer. (n.d.). evalglare-en.tar.gz. <http://www.ise.fraunhofer.de/en/areas-of-business-and-market-areas/applied-optics-and-functional-surfaces/downloads-englisch/software/evalglare-en.tar.gz> Accessed 2014 Apr 9.

Heschong L, Van Den Wymelenberg KG. 2012. Annual daylight performance metrics. In: Building synapses: connections in lighting. Minneapolis (MN); 2012 November 11–13; New York (NY): IESNA. p. 1–13.

Howlett O, Heschong L, McHugh J. 2007. Scoping study for daylight metrics from luminance maps. LEUKOS. 3(3):201-215.

IESNA-Daylight Metrics Committee. 2012. Lighting measurement #83, spatial daylight autonomy (sDA) and annual sunlight exposure (ASE). Illuminating Engineering Society of North America. <https://www.ies.org/store/product/approved-method-iesspatial-daylight-autonomy-sda-and-annual-sunlight-exposurease-1287.cfm> Accessed Aug 2015.

Kumaragurubaran V, Inanici M. 2013. hdrscope: high dynamic range image processing toolkit for lighting simulations and analysis. In: International Building Performance Simulation Association (IBPSA) 2013 Conference; Chambery, France; 2013 Aug 25–28.

Laurentin C, Bermtto V, Fontoynont M. 2000. Effect of thermal conditions and light source type on visual comfort appraisal. Lighting Res Technol. 32(4):223-233.

Lee ES, Clear RD, Ward G, Fernandes L. 2007. Commissioning and verification procedures for the automated roller shade system at the *New York Times* headquarters, New York, New York. Berkeley (CA): Lawrence Berkeley National Laboratory. <http://windows.lbl.gov/comm_perf/nyt_pubs.html> Accessed July 2009.

Loe DL, Mansfield KP, Rowlands E. 1994. Appearance of lit environment and its relevance in lighting design: experimental study. Lighting Res Technol. 26(3):119-133.

Newsham GR, Aries MBC, Mancini S, Faye G. 2008. Individual control of electric lighting in a daylit space. Lighting Res Technol. 40(1):25-41.

Newsham GR, Veitch JA. 2001. Lighting quality recommendations for VDT offices: a new method of derivation. Lighting Res Technol. 33(2):97-113.

Sutter Y, Dumortier D, Fontoynont M. 2006. The use of shading systems in VDU task offices: a pilot study. Energ Buildings. 38(7):780-789.

Tuaycharoen N. 2011. Windows are less glaring when there is a preferred view. Built Environment Sri Lanka. 9-10(1-2):45-55.

Tuaycharoen N, Tregenza PR. 2007. View and discomfort glare from windows. Lighting Res Technol. 39(2):185-200.

Van Den Wymelenberg KG. 2012. Evaluating human visual preference and performance in an office environment using luminance-based metrics [Dissertation]. Seattle (WA): University of Washington. <http://www.idlboise.com/papers/KevinVanDenWymelenberg-phd> Accessed Aug 2012.

Van Den Wymelenberg KG. 2014. Visual comfort, discomfort glare, and occupant fenestration control: developing a research agenda, LEUKOS. 10(4):207-221.

Van Den Wymelenberg KG, Inanici M. 2014. A critical investigation of common lighting design metrics for predicting human visual comfort in offices with daylight. LEUKOS. 10(3):145-164.

Van Den Wymelenberg KG, Inanici MN, Johnson PW. 2010. The effect of luminance distribution patterns on occupant preference in a daylit office environment. LEUKOS. 7(2):103-122.

Veitch JA, Newsham GR. 2000. Preferred luminous conditions in open-plan offices: research and practice recommendations. Lighting Res Technol. 32(4):199-212.

Ward G. 2011. The RADIANCE 4.0 synthetic imaging system. Version 4.0. <http://www.radiance-online.org/> Accessed June 2011.

Wienold J, Christoffersen J. 2006. Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras. Energ Buildings. 38(7):743-757.

Zaikina V, Matusiak BS, Klöckner CA. 2014a. Luminance-based measures of contour distinctness of 3D objects as a component of light modeling. LEUKOS. 11(1):31-45.

Zaikina V, Matusiak BS, Klöckner CA. 2014b. Luminance-based measures of shape and detail distinctness of 3D objects as important predictors of light modeling concept. Results of a full-scale study pairing proposed measures with subjective responses. LEUKOS. 11(4): 193-207.