eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Order effects when using Hopkinson's multiple criterion scale of discomfort due to glare

Michael G. Kent[a], Steve Fotios[b], Sergio Altomonte[c],*

[a] Department of Architecture and Built Environment, University of Nottingham, Nottingham, United Kingdom
[b] School of Architecture, University of Sheffield, Sheffield, United Kingdom
[c] Architecture et Climat, Université catholique de Louvain, Louvain-la-Neuve, Belgium

ABSTRACT

Hopkinson's multiple criterion scale is widely used to investigate the subjective degree of discomfort due to glare. Using an adjustment procedure, glare source luminance is adjusted to reveal four levels of discomfort, typically: just imperceptible, just acceptable, just uncomfortable, and just intolerable. In many studies, observers are instructed to attend to each level of discomfort in ascending order, from the lowest to the highest criterion. It is likely, however, that any settings made using this approach are influenced by an order bias, and this would affect the reported thresholds of discomfort. To investigate order effects, a Hopkinson-like multiple criterion adjustment experiment was performed, but under three different order sequences: ascending, descending, and randomised. The results revealed substantial bias due to order effects, particularly for lower glare criteria. This demonstrates the need for caution when interpreting subjective evaluations of discomfort due to glare and estimating the robustness of glare indices derived from studies that used Hopkinson's scale and procedure.

1. Introduction

The investigation of glare in buildings has led to the development of several predictive models that have been widely used in studies seeking to evaluate the thresholds of visual discomfort caused by artificial lighting sources [1,2] and by daylight [3–5]. These studies are central to investigating the interactions between building occupants, glare protection strategies (e.g., blind systems), and energy use [6]. Although there have been advancements in the way in which discomfort caused by glare is measured – for example, using photographic techniques and image analysis software in the evaluation of daylit environments [7,8] – much less attention has been placed on experimental procedures.

In this context, Hopkinson's 1940 paper [9] is commonly recognised as the founding study of discomfort due to glare. In this and in many subsequent experiments, Hopkinson used an adjustment procedure whereby the brightness of a source of illumination was adjusted to the point at which observers suggested that the visual scene represented a specified threshold of discomfort glare. Rather than adjusting the source brightness to the borderline between comfort and discomfort (the BCD, as used in some other studies [10]), Hopkinson asked for four settings to be made. These corresponded to four criteria describing different degrees of discomfort due to glare: the multiple criterion scale (MCS) (Table 1).

Using the multiple criterion scale, Petherbridge and Hopkinson [11] later established an empirical relationship between the reported degree of discomfort due to glare and three lighting parameters. This relationship is known as the glare constant.

Actually, Petherbridge and Hopkinson used a variation of the MCS in which the criteria Satisfactory (C) and Just Not Perceptible (D) were replaced by Just Acceptable and Just Imperceptible. Later studies used different labels for their glare response scales, particularly for those corresponding to the two lower degrees of discomfort (Table 2). Note, for example, that the second point of the scale is Just Acceptable in a study [11] but Slightly Uncomfortable in another [14]. Response scale labels are somewhat arbitrary to respondents, since they tend to place evaluated items in relative order of magnitude and may ignore the scale label [16]. If this is not realised by experimenters, different conclusions might be drawn about discomfort for the same response point, despite that not being the intention of the respondent. Yet, variations in response scale labels may affect the subjective evaluations made by respondents, especially when they have difficulty understanding the semantic descriptors used on the scale [17] or the experimenter's interpretation of such definitions. This issue is, however, not addressed in the study described in this article.

The Illuminating Engineering Society Glare Index (IES-GI) [18] was developed from Petherbridge and Hopkinson's glare constant but with

Table 1
The multiple criterion scale and definitions of discomfort glare criteria (from Ref. [9]).

	Discomfort Glare Criteria	Definition
A	Just Intolerable	The change-over point between intolerable and uncomfortable glare. The glare is definitely very uncomfortable, and any increase in the glare would produce intolerable glare.
B	Just Uncomfortable	The point where the glare is definitely distracting but only just uncomfortable. The term "uncomfortable" is associated in some observers with the desire to shade the eyes from the glare sources. The term "distracting" implies that the sources are sufficiently noticeable to attract the observer's attention.
C	Satisfactory	The point where the sources, though still noticeable, are deemed sufficiently unobtrusive to cause no distraction.
D	Just Not Perceptible	The point where the glare is just no longer noticeable. The sources are still visible but merge into the general field of view in such a way as no longer to attract the eyes of the observer.

Table 2
Variations in the labels used for the lower two levels of a 4-point discomfort glare rating scale in selected past studies based on an adjustment procedure.

Rating Point	Hopkinson [9]	Petherbridge & Hopkinson [11]	Stone & Harker [12]	Osterhaus & Bailey [13]	Kim & Koga [14]	Tuaycharoen & Tregenza [15]
C	Satisfactory	Just Acceptable	Just Distracting	Noticeable	Slightly Uncomfortable	Just Noticeable
D	Just Not Perceptible	Just Imperceptible	Just Perceptible	Imperceptible	Just Perceptible	Just Perceptible

Table 3
Discomfort glare criteria of multiple-criterion scales.

Date	Authors	Discomfort Glare Criteria	Derived Indices of Discomfort Glare	
			Glare Model	Formula
1950	Petherbridge & Hopkinson	Just Intolerable Just Uncomfortable Just Acceptable Just Imperceptible	Glare Constant	$G = \dfrac{(B_s^{1.6} \cdot \omega^{0.8})}{B_b}$
1962	IES Panel	Just Intolerable Just Uncomfortable Just Acceptable Just Perceptible	Illuminating Engineering Society Glare Index	$IES - GI = 10\log_{10} \cdot 0.478 \sum_{i=1}^{n} \dfrac{L_s^{1.6} \cdot \omega_s^{0.8}}{L_b \cdot P_s^{1.6}}$
1995	CIE	Not Specified	Unified Glare Rating	$UGR = 8 \log \dfrac{0.25}{L_b} \sum_{i=1}^{n} \left(\dfrac{L_s^2 \cdot \omega_s}{P_s^2} \right)$

Where: B_s = source luminance (fL), ω_s = apparent source size (sr), B_b = background luminance (fL), L_s = source luminance (cd/m^2), L_b = background luminance (cd/m^2), P_s = Guth's Position Index ($-$).

three changes: 1) a logarithmic function was introduced to account for the sensitivity of the visual system [19]; 2) Guth's position index was used to characterise changes in discomfort associated with the position of the source relative to the line of sight [10]; and, 3) a constant (K = 0.478) was used to convert from foot lamberts (used in the glare constant) to the international system of units (candelas). The IES-GI also features the same discomfort glare criteria found on the original multiple criterion scale, apart from Just Imperceptible that was changed to Just Perceptible [19] (Table 3). Based on Hopkinson's MCS, the IES-GI became the first recognised discomfort glare index. However, it is the Unified Glare Rating (UGR) system that is currently recommended in the lighting standards published by the Society of Light and Lighting (SLL) [20], the Illuminating Engineering Society of North America [21] and, originally, by the International Commission on Illumination (CIE) [22]. The UGR was also developed from the glare experiments carried out by Petherbridge and Hopkinson [11] and from the IES-GI, although without further experimental evaluations of discomfort glare sensation. During the development of the UGR, in fact, Sorensen [23] stated that the basic formula of the IES-GI should have been accepted as the "*dogmatic truth*". Therefore, the Petherbridge and Hopkinson's study [11] was adopted as the basis for current glare indices with the assumption that there was no need for further investigation to determine the validity of their findings.

The intended purpose of glare indices is to provide robust predictions of the likely visual discomfort reported by an observer in a given luminous environment. However, when calculated glare indices are compared against subjective evaluations of discomfort, the results often show a large degree of scatter [15]. In this article, we discuss the influence of one aspect of experimental procedure on errors between predicted and reported discomfort.

In his experiments, Hopkinson [9] asked the observers to first adjust the source luminance to mark the lowest level of discomfort (Just Not Perceptible), then to the next higher criterion (Satisfactory), and then to Just Uncomfortable and Just Intolerable. In other words, the four criteria settings were always considered in an ascending order of increasing magnitude of discomfort. A similar procedure was later repeated by following studies [11,24], where observers were instructed to vary the luminance of the glare source to meet the lowest of the four criteria (Just Imperceptible), and then incrementally the others until they had reached a level of discomfort that resembled intolerable glare. Such a procedure is likely to influence the settings made, in other words bringing an experimental bias.

In a repeated-measures design, where observers provide adjustments to more than one level of discomfort, it is to be expected that the current setting might be influenced by the preceding ones (i.e., an order effect). Order effects are linked to a variety of causes in which the experimental outcome is confounded by the sequence of evaluations made by test subjects [25]. Here, we consider two likely explanations for order effects that could influence the subjective evaluation of glare sensation using an adjustment procedure.

First, when adjustments are made to the four separate discomfort criteria in an ascending order, the successive setting will always be to a

Table 4
Mean luminance settings when adjustments were made to the four criteria of discomfort glare in different sequential orders (data from Ref. [29]).

Discomfort Glare Criteria	Luminance (cd/m^2)	
	Hopkinson's Ascending Order (JP, JA, JU, JI)	Secondary Sequence (JU, JI, JA, JP)
Just Perceptible	418	1042
Just Acceptable	1330	2189
Just Uncomfortable	2836	3110
Just Intolerable	4501	5501

Note: JP = Just Perceptible, JA = Just Acceptable, JU = Just Uncomfortable, JI = Just Intolerable.

higher luminance than the previous one, as anything else would be a nonsensical reaction to an increase in magnitude of the glare descriptor. This means that if the first setting was an overestimate, the following settings are also likely to be so. Second, in any successive trials, the preceding settings have an influence regardless of the direction of magnitude change. Settings, in fact, are biased towards the condition prevalent at the outset of the adjustment. This is generally called anchoring. In adjustment procedures, anchoring has been shown to influence settings of preferred light level [26], favoured colour appearance [27], and discomfort due to glare [28].

It is, therefore, plausible that any settings achieved using Hopkinson's multiple criterion scale in a strict ascending sequence might be erroneous due to order effects. The influence of an order effect can be seen in the results of Pulpitlova and Detkova [29] who used a secondary sequence in addition to Hopkinson's ascending-only order in an adjustment procedure. This secondary order was nearly a reversal of that used by Hopkinson, other than for the lowest two levels of discomfort (Table 4). The study showed that the mean luminance settings in the secondary sequence were consistently higher than those in Hopkinson's ascending order for all four discomfort criteria. One approach to off-setting an ordering effect consists in considering the four glare criteria in a randomised order [30]. Unfortunately, Pulpitlova & Detkova [29] did not use this approach, and thus their data do not reveal the likely outcome.

In summary, there is reason to believe that the ascending-only order in which the multiple criterion scale was used by Hopkinson and others has an influence on the settings made. Evidence of an order effect would support the hypothesis that the results from the Petherbridge and Hopkinson's study [11] – and, hence, the subsequent IES-GI and UGR glare indices and thresholds – provide an incorrect estimate of the luminance level (or, more precisely, of the relationship between the background luminance level and the glare source) associated with each discomfort glare criterion. To test this hypothesis, an experiment was carried out investigating whether the order in which the criteria on the multiple criterion scale are evaluated, when using an adjustment procedure, affects the luminance associated with each given magnitude of discomfort due to glare.

2. Method

2.1. Experimental design

The experiment was designed to explore whether order effects could be detected using controlled conditions in a laboratory test. The background luminance, the glare source area, and its position in the field of vision of the observer, were all held constant. The influence of extraneous light sources (e.g., daylight) was masked from the experimental setting.

The experimental layout (Fig. 1) was based on the apparatus used in previous studies of discomfort due to glare [15,28]. The lighting test chamber was semi-hexagonal in plan. The interior surfaces (2.7 m in

height) were matte white, and three 3W LED lamps were mounted above the participant's head to provide background illumination. Luminance measures were taken from the position of the participant, using a Minolta LS-100 mounted on a tripod. The mean background luminance, calculated from 17 individual measurements taken on a regular grid symmetrical about the central fixation point, was held at 65 cd/m^2 throughout the procedure.

A desk with a diffusive white surface was mounted within the wooden partitions. The participant's head was placed at a height of 1.2 m from the floor, set on a chin rest, facing a small diffusive screen (0.08 m × 0.04 m) made from three layers of translucent paper and positioned in front of a computer projector. The screen subtended an angle at the eye of 0.009 steradians and provided a variable luminance in the range between 200 and 32,000 cd/m^2. The luminance of the screen was increased using the relative brightness function of an image editing software operated by the experimenter. The relative brightness was increased at even intervals and spot-point luminance measurements were collected at each adjustment. Luminance readings at each interval were plot on a polynomial fit line to obtain values between each luminance point. This allowed for precise luminance values in repeated trials.

2.2. Experimental procedure

During the experimental procedure, test participants were asked to make judgements of discomfort due to glare using a Hopkinson-like multiple criterion scale [11,24]. Since variations in the interpretation of the four glare criteria could influence the evaluation provided, definitions of each criterion were linked to time-span descriptors so as to support a more consistent understanding [31,32] (see Appendix).

The discomfort criteria were used in three different order sequences:

- *Ascending*: Just Imperceptible, Just Acceptable, Just Uncomfortable, Just Intolerable
- *Descending*: Just Intolerable, Just Uncomfortable, Just Acceptable, Just Imperceptible
- *Randomised*: the four glare criteria were evaluated by participants in a shuffled order

The glare criteria used in this procedure were exactly the same of the four utilised in Petherbridge and Hopkinson's original study [11] with the ascending-only order sequence from where the glare constant formula was derived. Comparing the luminances set for each criterion in the three different sequences would demonstrate whether or not order had any significant effect. A repeated-measures design was used.

At the outset of the experiment, the glare source (i.e., the diffusive screen) was set to an initial luminance corresponding to an IES-GI of 10 (Just Imperceptible). It has been shown that initial settings (anchors) have an influence on luminance adjustments [28], hence for consistency this experiment used only this one anchor at the start of each block of trials. In fact, this anchor was used only for the first trial, and then the luminances set by participants became the anchor for the subsequent setting. At their arrival, participants were asked to adjust the chair so that their head was placed comfortably on the chin rest at the correct viewing position. A set of instructions was then given, including a definition of discomfort glare, the meaning of the four criteria, and a description of the procedure. For each trial, the experimenter adjusted the luminance of the glare source at a controlled pace according to the participant's instruction. Subjects were asked whether they would like the luminance of the glare source to be increased, decreased, or kept at its current brightness to reach a glare sensation corresponding to each of the four predefined criteria, in the order described in one of the three sequences. When participants vocally indicated that the scene represented the specific criterion of discomfort, the screen luminance was recorded. The test procedure was repeated until the participant had provided all four criteria of glare sensation
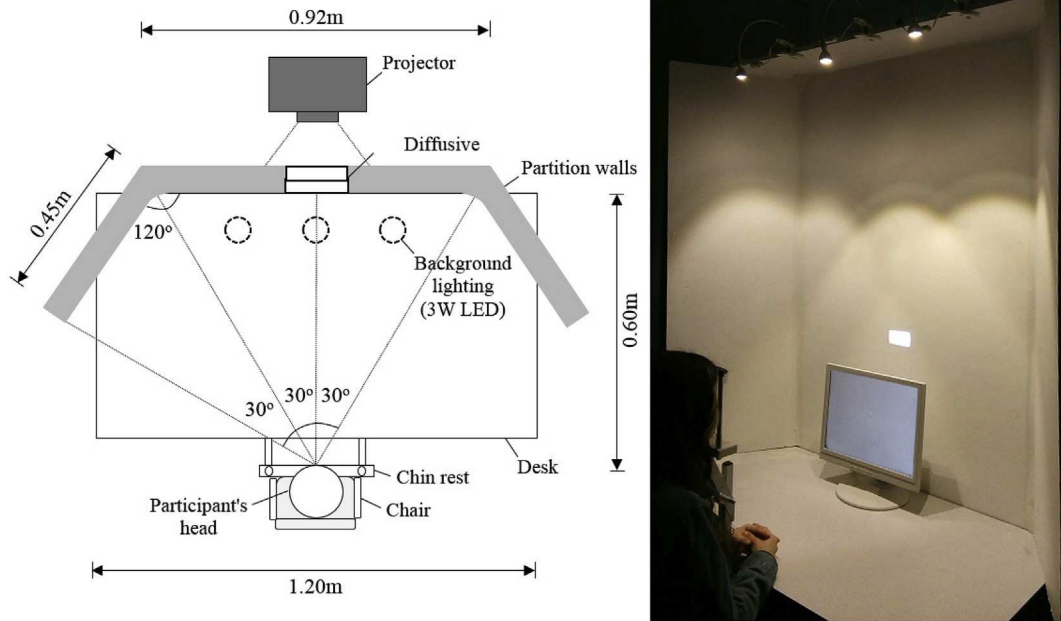
Fig. 1. Plan of the experimental layout and photograph of the lighting test chamber with a participant seated at the viewing position.

under each of the three sequences (i.e., ascending, descending, and randomised), these being administered in a random order. Each test session lasted approximately 30 min.

Twenty participants volunteered to this experiment. They were recruited via an online advertisement addressed to all students at the Department of Architecture and Built Environment, University of Nottingham (UK). The sample included 7 males and 13 females, with a mean age of 24.2 (SD = 5.76), 7 test participants wore corrective lenses during the tests, and all self-certified as having no other health or eye problems.

3. Results

Fig. 2 shows the mean luminances of the glare source, and their associated standard deviations, at the point in which participants reported each criterion of glare sensation under the three test sequences. Visual inspection of the plots suggests that mean source luminances were highest when adjustment settings were made using a descending

sequence for each glare criterion. The lowest mean luminances for the Just Imperceptible and Just Acceptable criteria were found under the ascending sequence, and under the randomised sequence for Just Uncomfortable and Just Intolerable discomfort levels.

The standard deviations about the mean source luminances become larger when adjustments were made to higher levels of discomfort, and this is consistent across all three order sequences. A similar trend could also be seen in previous glare studies that used the adjustment procedure with an ascending-only sequence and a single anchor (initial luminance setting) [33], and an ascending-only sequence with three different anchors [28]. This is due to the values involved being also larger: normalising the data reveals that the variance is approximately even for all the discomfort criteria.

Null Hypothesis Significance Testing (NHST) was performed to determine if the differences in source luminance for the same reported glare criterion were statistically significant across the three test sequences. However, since NHST is dependent on both the size of the sample and the effect under examination [34], emphasis of the
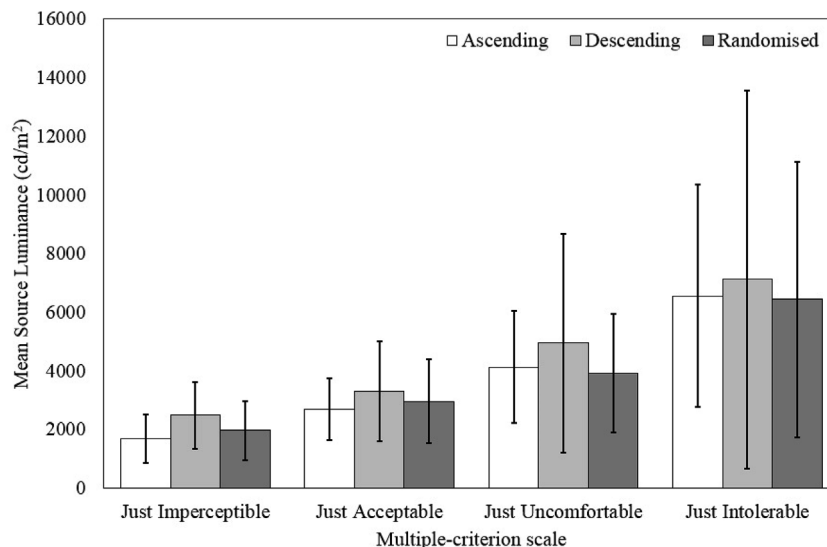

Fig. 2. Mean source luminances for the four discomfort glare criteria under the ascending, descending and randomised test sequences. Error bars show standard deviations.

inferential analysis was placed on the effect size (i.e., a standardised, scale-free, measure of the observed difference across the independent variable) [35] and not only on the *p*-value (which, in cases of small sample sizes, could confound the outcome) [36,37]. Therefore, increasing values of effect size would indicate a larger influence of order bias.

To analyse the data, a repeated-measures Analysis of Variance (RM-ANOVA) was used to compare against each other the source luminance settings for each criterion of reported glare sensation across the three different order sequences. An initial inspection of the data using graphical (Q-Q plots) and statistical (Shapiro-Wilk and Kolmogorov-Smirnov) tests did not suggest the data to be drawn from a normally-distributed population. Therefore, to relax the assumption of normality, a bias-corrected bootstrap was performed on both descriptive (differences in means and their 95% upper and lower confidence intervals) and statistical (standard errors and *p*-values) parameters [38,39]. The Maulchly's test of sphericity was used to test whether the variances of the differences between all paired comparisons of the within-subject variable (i.e., the order sequence) were equal [40,41]. The tests detected statistical significance across the independent groups for the criteria of Just Uncomfortable and Just Intolerable (Table 5). Since the assumption of sphericity had not been met (i.e., the variances of the differences were not all equal), the method of Greenhouse-Geisser was used to adjust the degrees of freedom (df) and calculate a conservative *F*-test ratio protected against Type I errors [42].

Table 6 presents the results of the RM-ANOVA, providing the glare criteria on the multiple criterion scale, the test statistic (F) and the degrees of freedom, the statistical significance (*p*-value), and the effect size (partial eta squared, η_p^2) of the differences detected. The interpretation of the outcome was derived from the tables recommended by Ferguson [43], where benchmarks are given for small, moderate, and strong effect sizes ($\eta_p^2 \geq 0.04$, 0.25, and 0.64, respectively). Values of η_p^2 below 0.04 denote not substantive (i.e., negligible and not practically relevant) differences.

The results of the inferential analysis show that the differences across the independent variable (order sequence) are highly significant in one case (Just Imperceptible), weakly significant in one case (Just Acceptable), and not significant for the remaining two glare criteria. The differences have a substantive effect size ranging from moderate ($0.25 \leq \eta_p^2 < 0.64$ in one case: Just Imperceptible) to small ($0.04 \leq \eta_p^2 < 0.25$ in two cases: Just Acceptable and Just Uncomfortable). Not substantive differences were detected for the Just Intolerable criterion ($\eta_p^2 < 0.04$, negligible). In these data, the magnitude of the effect size decreases when considering a higher level of discomfort. The effect of sequence (order of discomfort glare criteria) on the luminance settings made by test participants appears to be smaller at higher levels of visual discomfort, which confirms the observations drawn from Fig. 2.

Post-hoc testing was performed to compare all combinations of the independent variable for each level of glare sensation. Statistical significance of the differences was calculated using paired *t*-tests to determine the locations of the variations detected in the RM-ANOVA. Directionality of the statistical test was determined through inspection of central tendencies and graphical displays of the data (two-tailed

Table 5
Maulchly's test of sphericity used to compare the variances of the differences between the three order sequences for each discomfort glare criterion.

Discomfort Glare Criteria	Maulchly's W	Approximate χ	df	Statistical Significance (*p*-value)
Just Imperceptible	0.87	2.62	2	0.27 n.s.
Just Acceptable	0.94	1.06	2	0.59 n.s.
Just Uncomfortable	0.38	17.67	2	0.00***
Just Intolerable	0.55	10.63	2	0.01**

Table 6
RM-ANOVA (with Greenhouse-Geisser corrections) across the three order sequences for each discomfort glare criterion.

Discomfort Glare Criteria	F (df)	Statistical Significance (*p*-value)	Effect Size (η_p^2)
Just Imperceptible	7.96 (1.76, 33.47)	0.00***	0.30
Just Acceptable	3.50 (1.89, 35.95)	0.04*	0.16
Just Uncomfortable	2.14 (1.27, 23.38)	0.15 n.s.	0.10
Just Intolerable	0.59 (1.38, 26.28)	0.50 n.s.	0.03

$\eta_p^2 < 0.04$ = negligible; $0.04 \leq \eta_p^2 < 0.25$ = small; $0.25 \leq \eta_p^2 < 0.64$ = moderate; $\eta_p^2 \geq 0.64$ = strong.

hypothesis) [44,45]. In this analysis, the effect size was calculated by making use of equivalence between the observed differences and the Pearson's coefficient r, according to Equation (1):

$$r = \sqrt{\frac{t^2}{t^2 + df}} \tag{1}$$

where, t is the test statistic extracted from the *t*-test, and df is the degrees of freedom.

The interpretation of the outcome was derived again from Ferguson [43] to infer small, moderate, and strong effect sizes (Pearson's r ≥ 0.20, 0.50, 0.80, respectively). Absolute values below 0.20 were considered negligible and not substantive effects.

Table 7 reports the results of the paired *t*-tests, providing, for each criterion of glare sensation, the comparison under examination (order sequences denoted as group 1 and 2), the mean (M) and standard deviations (SD) for each group corresponding to the luminance of the glare source set by participants using the adjustment, the differences between means (ΔM), their statistical significance (NHST) and the bias-corrected 95% upper (CI_U) and lower (CI_L) confidence intervals, and the effect size (r).

Inspection of descriptive and inferential statistics shows no consistent directionality of the sign for the mean differences and the effect sizes across all comparisons, this being consistent with the adoption of a two-tailed hypothesis. Out of the 12 comparisons presented in Table 7, the differences between mean values of source luminance are highly significant in one case, significant in four cases, weakly significant in three cases, and not significant in the remaining four cases. For all settings made to the highest criterion of discomfort (Just Intolerable), the effect of test order was not suggested to be statistically significant. The differences detected were mostly of substantive magnitude, with effect sizes ranging from moderate (0.50 ≤ r < 0.80 in two cases) to small (0.20 ≤ r < 0.50 in seven cases). Negligible effects were detected for three comparisons (r < 0.20).

4. Discussion and limitations

A controlled experiment was set up to explore whether the order in which the criteria on Hopkinson's MCS are evaluated when using an adjustment procedure affects the luminance associated with each perceived magnitude of discomfort glare. Inferential analysis of the data confirmed that the order of test sequences had significant and substantive influence on the final setting made by test participants for the same perceived level of discomfort due to glare. More specifically, the order effect on glare settings appeared to be larger at lower levels of glare sensation (i.e., Just Imperceptible and Just Acceptable).

Before drawing conclusions on the theoretical and design implications of these results, some methodological limitations need to be acknowledged. Among these, it should be considered that, in this experiment, each block of tests commenced with the glare source set at

Table 7
Comparisons of mean (M) and standard deviation (SD) of the luminances set within the three order sequences using the paired samples t-test and their effect sizes.

Discomfort Glare Criteria	Sequences (1 vs. 2)	M(SD)$_1$	M(SD)$_2$	ΔM^{NHST}	BCa CI$_L$,CI$_U$	Effect Size (r)
Just Imperceptible	Asc. vs. Des.	1676 (829)	2484 (1123)	− 807***	− 1,217, − 398	− 0.69
	Asc. vs. Ran.	1676 (829)	1972 (1005)	− 296*	− 686, 138	− 0.36
	Des. vs. Ran.	2484 (1123)	1972 (1005)	511**	13, 1010	0.44
Just Acceptable	Asc. vs. Des.	2686 (1065)	3317 (1707)	− 631**	− 1,100, − 161	− 0.54
	Asc. vs. Ran.	2686 (1065)	2962 (1419)	− 276*	− 746, 193	− 0.27
	Des. vs. Ran.	3317 (1707)	2962 (1419)	354*	− 203, 911	0.29
Just Uncomfortable	Asc. vs. Des.	4130 (1905)	4044 (3718)	− 815**	− 2,186, 557	− 0.27
	Asc. vs. Ran.	4130 (1905)	3922 (2034)	207 n.s.	− 331, 746	0.18
	Des. vs. Ran.	4044 (3718)	3922 (2034)	1022**	− 165, 2210	0.38
Just Intolerable	Asc. vs. Des.	6562 (3783)	7116 (6459)	− 554 n.s.	− 2,240, 1131	− 0.16
	Asc. vs. Ran.	6562 (3783)	6443 (4702)	120 n.s.	− 708, 947	0.03
	Des. vs. Ran.	7116 (6459)	6443 (4702)	674 n.s.	− 835, 2183	0.21

Note: Asc. = Ascending, Des. = Descending, Ran. = Randomised.
With Bonferroni corrections: ***highly significant; **significant; *weakly significant; n.s. = not significant.
r < 0.20 = negligible; 0.20 ≤ r < 0.50 = small; 0.50 ≤ r < 0.80 = moderate; r ≥ 0.80 = strong.

low luminance (low anchor, corresponding to Just Imperceptible according to the IES-GI).

Under the ascending sequence, the first setting (Just Imperceptible) that participants were asked to assess was effectively the same at which the anchor was set. This might have induced a bias, in that participants may have felt compelled to adjust only in an upwards direction from this low anchor. However, Table 8 suggests that this was not the case, since the numbers of participants making settings either above or below the anchor are approximately equal (respectively, 8 and 9 subjects). Under the descending sequence, the first setting that participants were asked to assess corresponded to Just Intolerable glare. All 20 participants set a luminance higher than the initial anchor, i.e. their response was in the expected direction. Finally, under the randomised sequence, the first setting that participants were asked to assess could be any of the four glare criteria. For the tests where the first setting required was Just Imperceptible, there is again an equal distribution of settings above (2) and below (2) the anchor, as with the ascending sequence. In the tests where the first setting to be assessed was either Just Acceptable, Just Uncomfortable, or Just Intolerable, in all cases the luminance set was above that of the anchor, as with the descending sequence.

This suggests that test participants were not compelled to make their adjustment setting in a certain direction, although this does not provide any indication in terms of the influence that different sequences might have had on the luminances set (i.e., comparing the three orders).

Fig. 3 plots, along the x-axis, the absolute values of the effect sizes (r) estimated from the paired t-tests presented in Table 7. On the y-axis, the graph displays the four glare criteria for the paired comparisons between the three order sequences. For all three contrasts, the effect size is largest for the lowest glare criterion (Just Imperceptible) and tends to become progressively smaller at higher levels of discomfort. For low degrees of visual discomfort, the comparison between the ascending and descending sequences corresponds to the greatest effect

sizes. It must be taken into account, however, that these trends may have been induced by the use of a low initial anchor. Glare evaluations are dependent on the anchor (luminance setting used before the adjustment) [28], and the trends observed here may change when using a higher initial luminance setting.

In interpreting these results, it must be considered that this study used the ascending sequence to replicate Hopkinson's experiments [9,11], the descending as its reverse order, and the randomised sequence as a potential good practice to overcome an order effect [30].

The Ascending vs. Randomised comparison, therefore, reveals the likely differences in results between a study that uses Hopkinson's approach in terms of scale and procedure and one that follows good experimental practice. This comparison suggests that the magnitude of the order effect was significant and substantive (non-negligible effect size, r > 0.20) for the criteria of Just Imperceptible and Just Acceptable, but not for the other two.

One might question whether combining the data obtained under an ascending and a descending order, and using the mean as best estimate, might lead to results that are in accordance with those achieved under a randomised sequence of discomfort glare criteria.

Randomised orders are, in fact, generally considered the most robust experimental approach. Where this is not possible, then taking the mean of results gained using lower and upper anchors may provide the best estimates [46]. To offer an initial exploration of such hypothesis, the mean source luminances of the glare source corresponding to the adjustment settings made for the four discomfort glare criteria under the ascending and descending orders were combined and then compared to the mean source luminance settings made by participants under the randomised sequence. Fig. 4 illustrates the results of the comparison in terms of mean source luminances, standard deviations, and mean differences.

At the lowest two criteria of discomfort glare (Just Imperceptible and Just Acceptable), the plots show a relatively small difference in mean source luminance between the combined and the randomised sequences (respectively, ΔM = 68.11 and 38.93 cd/m^2). At higher glare criteria (Just Uncomfortable and Just Intolerable), the mean luminance values obtained from the combined data are much larger than the adjustment settings made under the randomised sequence (with differences, respectively, of ΔM = 615.08 and 436.24 cd/m^2). This was to be anticipated considering that, as shown in Fig. 2, at lower levels of visual discomfort, the mean source luminance values under the randomised sequence fell between the mean values for the ascending and descending orders. Conversely, for higher discomfort glare criteria, the adjustment settings made under the ascending and descending orders were both performed at higher luminances than the randomised
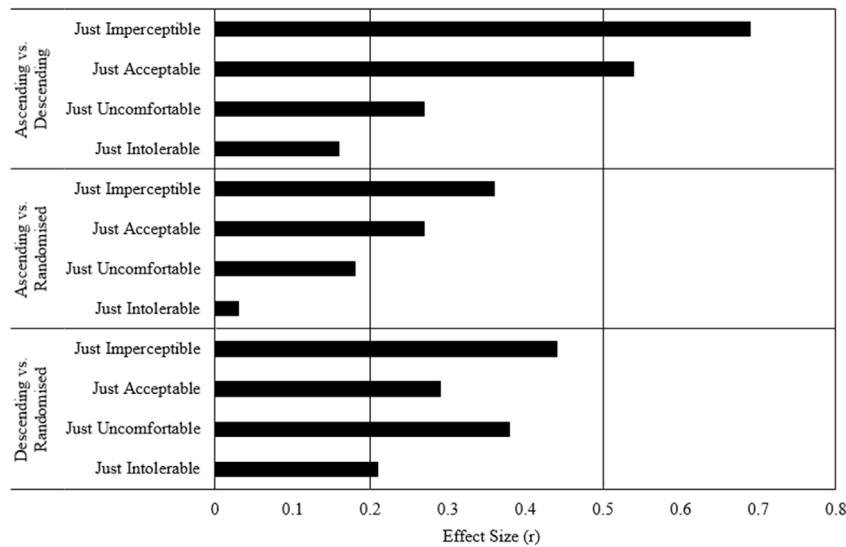
Table 8
Position of the first luminance setting relative to the initial anchor.

Sequence	Initial anchor setting	Below the Anchor	Equal to the Anchor	Above the Anchor
Ascending	Just Imperceptible	8	3	9
Randomised	Just Imperceptible	2	2	2
	Just Acceptable/ Uncomfortable/ Intolerable	0	0	14
Descending	Just Intolerable	0	0	20

Fig. 3. Effect sizes for each comparison and discomfort glare criterion. The vertical lines represent the effect size thresholds suggested by Ferguson: $r < 0.20$ = negligible; $0.20 \le r < 0.50$ = small; $0.50 \le r < 0.80$ = moderate [43].

sequence.

Further testing could not be performed to investigate the statistical and practical significance of the differences detected since, due to the methods used for the collection of our data, the assumption of independence could not be met. In fact, the statistical significance of the differences cannot be calculated when the luminance settings given by the same test participant in separate conditions (e.g., ascending and descending orders) are combined. However, these initial observations can be useful for future experimental designs, particularly in the presence of constraints in terms of time and resources, although they would need to be supported by further experimental research and, possibly, by meta-analysis of data from previous studies.

5. Conclusions

In many studies of glare, the thresholds derived to limit the levels of visual discomfort in buildings have been heavily influenced by prediction models that can be traced back to the original Hopkinson's [9] and Petherbridge and Hopkinson's studies [11]. While it is not common

in discomfort glare research to question the procedures used to derive the data in fundamental studies, previous work demonstrated a need to identify key sources of method bias [28]. The implications of these errors, in fact, can play a significant role in the way in which buildings are designed to protect their occupants from sources of visual discomfort.

While it is now considered good practice of experimental research to evaluate different degrees of glare sensation in a randomised order, Hopkinson used an adjustment procedure with a strict ascending order to establish the luminances associated with four degrees of discomfort due to glare. An experiment was carried out to investigate whether the order in which these four levels are set has an effect on the resultant outcome. Compared with the ascending sequence, a randomised order led to significantly and substantively different luminances for the criteria of Just Imperceptible and Just Acceptable glare but did not suggest statistically and practically relevant differences for higher degrees of glare sensation. These results suggest a need to review the thresholds for lower levels of discomfort established in glare studies that have used an adjustment procedure with Hopkinson's multiple criterion scale in an
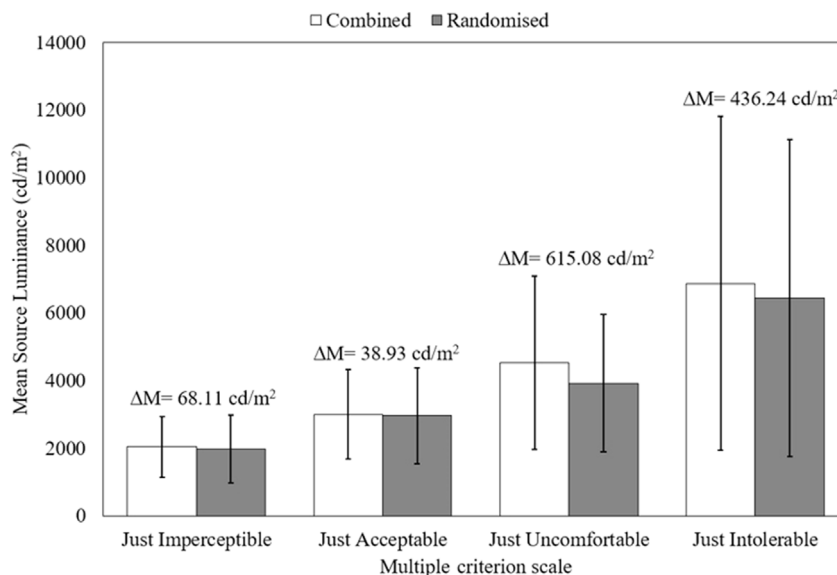

Fig. 4. Mean source luminances for the four discomfort glare criteria under the combined and the randomised test sequences. Error bars show standard deviations.

ascending-only sequence. In addition, since the adjustment procedure has been shown to be dependent on the initial luminance setting (anchor) used [28], on-going research is investigating potential interacting effects of these biases also at higher levels of discomfort (that is, whether using a higher initial anchor would lead to larger differences also at higher glare criteria).

Considering that the Petherbridge and Hopkinson's study [11] has been used as a fundamental reference for the development of most modern glare models, the results of this study raise questions related to the alleged precision of the glare indices (e.g., UGR, IES-GI, etc.) generally used to describe and predict the levels of discomfort experienced by observers from luminous sources. The findings also emphasise the need to derive an experimental procedure that considers the influence of different sources of methodological bias on the thresholds of visual discomfort, which could ultimately be used to propose a new glare prediction model.

Acknowledgements

Appendix. Definitions of discomfort glare as given to test participants

In this experiment, you will be asked to express your own perceived level of visual discomfort when presented with a small diffusive screen, using four thresholds of glare sensation votes (GSVs): Just Imperceptible, Just Acceptable, Just Uncomfortable, and Just Intolerable. These are described below:

- *Just Imperceptible*: when the source of light becomes quite bright without necessarily giving a sensation of glare. As the light source is being adjusted, for a moment, if you were performing a visual task, the source would be something that attracts your attention.
- *Just Acceptable*: this corresponds to a glare sensation that could be tolerated for approximately one day when working at this workstation. If you had to work under this lighting condition at your own work place, you may want to use blinds or other measures to decrease the perceived discomfort.
- *Just Uncomfortable*: this corresponds to a glare sensation that could be tolerated for approximately 15–30 min, for example if finishing a certain task would take this amount of time. After this, adjustments to the lighting conditions would be made, if the same degree of discomfort was present over time.
- *Just Intolerable*: this corresponds to the point where you would no longer be able to work under these lighting conditions for any amount of time and would immediately intervene to change them.

References

[1] S. Altomonte, M.G. Kent, P.R. Tregenza, R. Wilson, Visual task difficulty and temporal influences in glare response, Build. Environ. 95 (2016) 209–226.

[2] G.H. Scheir, P. Hanselaer, P. Bracke, G. Deconinck, W.R. Ryckaert, Calculation of the Unified Glare Rating based on luminance maps for uniform and non-uniform light sources, Build. Environ. 84 (2015) 60–67.

[3] M.G. Kent, S. Altomonte, R. Wilson, P.R. Tregenza, Temporal effects on glare response from daylight, Build. Environ. 113 (2017) 49–64.

[4] J.Y. Suk, M. Schiler, K. Kensek, Investigation of existing discomfort glare indices using human subject study data, Build. Environ. 113 (2017) 121–130.

[5] I. Konstantzos, A. Tzempelikos, Daylight glare evaluation with the sun in the field of view through window shades, Build. Environ. 113 (2017) 65–77.

[6] J. Wienold, Dynamic Simulation of Blind Control Strategies for Visual Comfort and Energy Balance Analysis, International Building Performance Simulation Association, Beijing, China, 2007, pp. 1197–1204.

[7] M. Hirning, S. Coyne, I. Cowling, The use of luminance mapping in developing discomfort glare research, J. Light Vis. Environ. 34 (2) (2010) 101–104.

[8] M.B. Hirning, G.L. Isoardi, V.R. Garcia-Hansen, Prediction of discomfort glare from windows under tropical skies, Build. Environ. 113 (2017) 107–120.

[9] R.G. Hopkinson, Discomfort glare in lighted streets, Trans. Illum. Eng. Soc. 5 (1–9) (1940) 1–32.

[10] M. Luckiesh, S.K. Guth, Brightnesses in visual field at borderline between comfort and discomfort (BCD), Illum. Eng. 44 (11) (1949) 650–670.

[11] P. Petherbridge, R.G. Hopkinson, Discomfort glare and the lighting of buildings, Trans. Illum. Eng. Soc. XV (2) (1950) 39–79.

[12] Stone Pt, S.D.P. Harker, Individual and group differences in discomfort glare responses, Light. Res. Technol. 5 (1) (1973) 41–49.

[13] W.K.E. Osterhaus, I.L. Bailey, Large area glare sources and their effect on discomfort and visual performance at computer workstations, Proceedings of the IEEE Industry Applications Society Annual Meeting, 1992 Houston.

[14] W. Kim, Y. Koga, Glare constant G_w for the evaluation of discomfort glare from windows, Sol. Energy 78 (2005) 105–111.

[15] N. Tuaycharoen, P.R. Tregenza, Discomfort glare from interesting images, Light. Res. Technol. 37 (4) (2005) 329–341.

[16] E.C. Poulton, Bias in Quantifying Judgements, Taylor & Francis, 1989.

[17] A.W. Gellatly, D.J. Weintraub, User Reconfigurations of the de Boer Rating Scale for Discomfort Glare, The University of Michigan Transportation Research Institute, 1990.

[18] W. Robinson, H.E. Bellchambers, J.T. Grundy, J. Longmore, W. Burt, H. Hewitt, P. Petherbridge, D. Frith, R.G. Hopkinson, E. Rowlands, The development of the IES glare index system, Trans. Illum. Eng. Soc. 27 (1) (1962) 9–26.

[19] R.G. Hopkinson, A note on the use of indices of glare discomfort for a code of lighting, Trans. Illum. Eng. Soc. 25 (3) (1960) 135–138.

[20] SLL, The SLL Code for Lighting: the Society of Light and Lighting, (2012).

[21] IESNA, Lighting Handbook: Illuminating Engineering Society of North America, tenth ed., IESNA, New York, 2011.

[22] CIE 177-1995, Discomfort Glare in Interior Lighting, CIE, 1995.

[23] K. Sorensen, A modern glare index method, CIE Proceedings of 21st Session, 1987, pp. 108–111 Venice.

[24] R.G. Hopkinson, R.C. Bradley, A study of glare from very large sources, Illum. Eng. 55 (5) (1960) 288–294.

[25] B. Day, I.J. Bateman, R.T. Carson, D. Dupont, J.J. Louviere, S. Morimoto, R. Scarpa, P. Wang, Ordering effects and choice set awareness in repeat-response stated preference studies, J. Environ. Econ. Manag. 63 (2012) 73–91.

[26] Á. Logadóttir, J. Christoffersen, S.A. Fotios, Investigating the use of an adjustment task to set preferred illuminance in a workplace environment, Light. Res. Technol. 43 (4) (2011) 403–422.

[27] Á. Logadóttir, S.A. Fotios, J. Christoffersen, S.S. Hansen, D.D. Corell, C. Dam Hansen, Investigating the use of an adjustment task to set preferred colour of ambient illumination, Colour Research & Application 38 (1) (2013) 46–57.

[28] M.G. Kent, S. Fotios, S. Altomonte, Discomfort Glare Evaluations: the influence of anchor bias in luminance adjustments, Light. Res. Technol. (2017), http://dx.doi.org/10.1177/1477153517734280.

[29] J. Pulpitlova, P. Detkova, Impact of the cultural and social background on the visual perception in living and working perception, Proceedings of the International Symposium: Design of Amenity, 5-8 October, Fukuoka, Japan, 1993, pp. 93–95.

[30] A. Field, G. Hole, How to Design and Report Experiments, Sage, 2013.

[31] D. MacGowan, Correspondence, Light. Res. Technol. 42 (2010) 121–122.

[32] M. Velds, User acceptance studies to evaluate discomfort glare in daylit rooms, Sol. Energy 73 (2) (2002) 95–103.

[33] M.G. Kent, S. Altomonte, P.R. Tregenza, R. Wilson, Discomfort glare and time of day, Light. Res. Technol. 47 (6) (2015) 641–657.

[34] P.D. Ellis, The Essential Guide to Effect Sizes: Statistical Power, Meta-analysis, and the Interpretation of Research Results, Cambridge University Press, Cambridge, 2010.

[35] S. Altomonte, S. Schiavon, M.G. Kent, G. Brager, Indoor environmental quality and occupant satisfaction in green-certified buildings, Build. Res. Inf. (2017), http://dx.doi.org/10.1080/09613218.2018.1383715.

[36] J. Cohen, Things I have learned (so far), Am. Psychol. 45 (12) (1990) 1304–1312.

[37] J. Cohen, The earth is round ($p < .05$), Am. Psychol. 49 (12) (1994) 997–1003.

[38] T.J. DiCiccio, B. Efron, Bootstrap confidence intervals, Stat. Sci. 11 (3) (1996) 189–228.

[39] B. Efron, Better bootstrap confidence intervals, J. Am. Stat. Assoc. 82 (397) (1987) 171–185.

[40] J.W. Maulchy, Significance test for sphericity of a normal n-variate distribution, Ann. Math. Stat. 11 (2) (1940) 204–209.

[41] R.G. O'Brien, M.K. Kaiser, MANOVA method for analyzing repeated measures designs: an extensive primer, Psychol. Bull. 97 (2) (1985) 316–333.

[42] S.W. Greenhouse, S. Geisser, On methods in the analysis of profile data, Psychometrika 24 (2) (1959) 95–112.

[43] C.J. Ferguson, An effect size primer: a guide for clinicians and researchers, Professional Psychology: Research and Practise 40 (5) (2009) 532–538.

[44] D. Hauschke, V.W. Steinijans, Directionality decision for a two-tailed alternative, J. Biopharm. Stat. 6 (2) (1996) 211–218.

[45] G.D. Ruxton, M. Neuhauser, When should we use one-tailed hypothesis testing? Methods in Ecological and Evolution 1 (2) (2010) 114–117.

[46] G.A. Gescheider, Psychophysics: the Fundamentals, Lawrence Erlbaum Associates, 1997.