

# Temporal Variables and Personal Factors in Glare Sensation

**MG Kent** MSc<sup>a</sup>, **S Altomonte** PhD<sup>a</sup>, **PR Tregenza** PhD<sup>b</sup> and **R Wilson** PhD<sup>a</sup>

<sup>a</sup> Department of Architecture and Built Environment, University of Nottingham, Nottingham, UK

<sup>b</sup> Sheffield School of Architecture, University of Sheffield, Sheffield, UK

Received 30 December 2014; Revised 22 February 2015; Accepted 25 February 2015

Previous laboratory experiments have provided evidence of an effect of time of day on glare sensation. During the tests, temporal variables and personal factors were measured to analyse their influence on levels of visual discomfort as the day progresses. The results revealed statistically significant and practically relevant tendencies towards greater tolerance to source luminance from artificial lighting at all times of day for earlier chronotypes and for participants not having ingested caffeine. No conclusive evidence was found for the effect of fatigue, sky condition, and prior light exposure on glare sensation throughout the day. These findings suggest that temporal variables and personal factors should be measured in conjunction with visual discomfort levels to explore the causes of the wide individual differences commonly associated with the subjective evaluation of glare sensation.

Address for correspondence: Michael G Kent, Department of Architecture and Built Environment, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

E-mail: [ezxmk4@nottingham.ac.uk](mailto:ezxmk4@nottingham.ac.uk)

Kent, M., Altomonte, S., Tregenza, P., Wilson, R., 2015. Temporal variables and personal factors in glare sensation. *Lighting Research & Technology*. doi: 10.1177/1477153515578310.

## 1. Introduction

A previous series of experiments conducted in a controlled laboratory setting revealed a tendency towards greater tolerance to luminance increases in artificial lighting as the day progresses<sup>1</sup>. This trend was found not to be statistically related to learning, suggesting an effect of time of day on the subjective evaluation of glare sensation. The postulated temporal effect was observed to be directly linked to the independent variable itself (i.e., time of day); that is, as the time gap between sessions of artificial light exposure increased, so did the difference in the source luminance corresponding to reported levels of visual discomfort. In the experiments, all variables known to potentially influence glare (i.e., source and background luminance, position index, size and uniformity of the source, etc.) were controlled. Nevertheless, when luminance values corresponding to votes of glare sensation were regressed against the time of test sessions, a low coefficient of determination showed a large scatter in the results. This suggested that there could be other factors, not controlled through experimental manipulation, that could influence the reported levels of visual discomfort. Even if some indications of confounding variables were detected, the statistical significance and practical relevance of their effect was not sufficient to draw any consistent conclusion, thus requiring further investigation<sup>1</sup>.

Rodriguez and Pattini<sup>2</sup> stated that, although it may not be feasible to completely eliminate the scatter commonly associated with the subjective evaluation of glare sensation, individual differences may be mitigated if these are systematically identified. In this context, various studies in the literature have investigated the role of factors that could be associated with personal variability in reported levels of visual discomfort. In controlled laboratory experiments, Phipps-Nelson *et al.*<sup>3</sup> and Smolders *et al.*<sup>4</sup> found that, when exposing participants to an artificial lighting source producing 1,000 lux at the eye, subjective alertness and psychomotor vigilance tasks – i.e., sustained-attention and reaction-timed tasks measuring the speed of response to a visual stimulus – were increased, whilst sleepiness was reduced. Similarly, Boyce *et al.*<sup>5</sup> detected interactions

between visual sensation, duration of exposure, individual control of lighting, surface reflectance, and task characteristics. A study by Borisuit *et al.*<sup>6</sup> found a potential connection between visual comfort and non-visual responses – that is, the metabolic effects of lighting stimuli affecting circadian processes and hormonal regulation<sup>7</sup> – suggesting a link between visual acceptance and mood, alertness, sleepiness, prior daylight exposure, and time of day. Other studies focusing on the non-visual response system showed that, during the biological night, the sensitivity of photo-biological processes is significantly heightened compared to the daytime<sup>8-10</sup>.

While the temporal and spatial aspects of daylight have been encompassed into climate-based modelling metrics<sup>11</sup>, temporal differences in visual response (e.g., seasonal, daily, and at shorter time scales) are yet to be comprehensively evaluated<sup>12</sup>. Therefore, based on a review of literature and on the data collected in previous experiments, the aim of this study consists in analysing the potential influence of several temporal variables (i.e., subject to changes with time of the day) and personal factors on the scattered results commonly associated with subjective evaluations of glare sensation as the day progresses.

## **2. Method**

### **2.1 Experimental procedure**

Data from a previous series of laboratory tests, conducted with the experimental setup described in Kent *et al.*<sup>1</sup>, were utilised in this study. The experimental procedure required subjects to participate, on the same day, in four test sessions evenly distributed at 3-hour intervals:

- Morning: 09:00 or 09:30
- Afternoon A: 12:00 or 12:30
- Afternoon B: 15:00 or 15:30
- Evening: 18:00 or 18:30

Since discomfort glare is a personal sensation that entails subjective methods of evaluation<sup>13</sup>, participants were asked to make judgments using, as benchmarks, adaptations of Glare Sensation Votes (GSVs)<sup>14-18</sup>: ‘Just Perceptible’, ‘Just Noticeable’, ‘Just Uncomfortable’, and ‘Just Intolerable’. To help subjects giving consistent judgements and avoid individual interpretations due to the abstraction caused by the assessment, time descriptors were used to accompany the GSVs<sup>13, 19</sup>.

At the beginning of the test procedure, subjects were requested to fill in a short questionnaire featuring demographic information (e.g., age, gender, ethnicity) and self-evaluations of personal factors such as photosensitivity and chronotype (chronotype is to be intended as an attribute that reflects individual circadian phases, thus indicating at what time of the day physiological functions – e.g., hormone levels, body temperature, eating and sleeping patterns, etc. – are activated<sup>20</sup>). To assess photosensitivity, participants were asked questions regarding their self-evaluated sensitivity to natural and artificial light, their use of solar protection devices (e.g., sunglasses), their luminous preferences at work (e.g., bright or dim conditions), and their frequency of interaction with environmental controls (e.g., blinds)<sup>21</sup>. In terms of chronotype, since it was considered that this term may not be fully comprehended by all participants, the Munich Chronotype Questionnaire (MCTQ)<sup>22</sup> was used to aid test subjects in providing informed judgements.

During the sessions, participants were asked to direct their gaze towards a visual fixation point located at the centre of a small diffusive screen whose luminance was gradually increased at a constant pace. Subjects were then asked to vocally indicate when their sensation of visual discomfort reached each criterion of GSV due to the brightness of the source, which was then related to corresponding photometric values previously-recorded. After all four GSV judgements had been provided, participants were asked to complete one additional questionnaire reporting their perceived level of fatigue, their caffeine and food ingestion prior to the test, the prevailing sky conditions

and the light exposure (artificial or natural) that they had experienced between test sessions.

## 2.2 Statistical Analysis

A total of 30 participants volunteered to take part to the experiments. Membership to the sample group was tightly controlled, as described in Kent *et al.*<sup>1</sup>.

Consistent with several other discomfort glare studies<sup>2, 23-26</sup>, where feasible the data related to temporal variables and personal factors were reduced to dichotomous categories. This was to avoid making comparisons between large groups of data, whilst also obtaining more robust results by increasing the power of statistical tests and reducing the possibility of occurrence of false discoveries. However, variables that could potentially take on more than two values (e.g., fatigue and chronotype) were measured on ordinal scales, since it was considered that forcing these factors into dichotomous categories may have risked underestimating their effect on glare sensation. For all temporal variables and personal factors, consistent with previous work<sup>1</sup>, the source log luminance was used as the primary evaluation parameter to quantify subjective assessments of visual discomfort at different times of day and explore the bases of individual differences in glare sensation<sup>13, 23-25</sup>. In the analysis of the data, a large scatter appeared when the log luminance was regressed against times of the day<sup>1</sup>. For all variables, graphical and statistical inspection revealed that data were not normally distributed around the mean, thus violating one of the assumptions for a parametric test<sup>27</sup>. Furthermore, the non-parametric Levine's test of homogeneity of variance – a test suitable for non-Gaussian distributions<sup>28</sup> – showed no statistically significant differences ( $p \geq 0.05$ ) between variances. Hence, non-parametric tests, relying on less stringent assumptions than their parametric counterparts, were adopted<sup>29</sup>.

Inferential testing was performed for all temporal variables and personal factors measured during the experimental procedure, analysing the differences in source log luminance corresponding to each criterion of GSV across the four test sessions.

Directionality of the hypothesis was informed by descriptive and visual inspection of central tendencies and graphical displays of the data<sup>30</sup>. However, if no consistent directionality of observed differences could be determined through preliminary exploratory analysis, two-tailed hypothesis testing was applied<sup>31</sup>. Initial inferential tests showed no statistically significant differences between subjective evaluations of glare sensation for all levels of visual discomfort and times of the day based on the gender, ethnicity and photosensitivity of test subjects. Likewise, due to the tightly controlled membership to the test sample, the age of participants was regarded as masked from the experiment, while analysis of the variation of reported food ingestion revealed no statistically significant or practically relevant differences between independent groups. Hence, consideration of these factors was excluded from this investigation.

The statistical significance of the differences in source log luminance considered under dichotomous variables was evaluated using the Mann-Whitney U test. Conversely, statistical significance of *a priori* ordering effects with respect to ordered variables was analysed using the Jonckheere-Terpstra test. The Mann-Whitney U test is appropriate when variables fall under two experimental conditions and different subjects are used in each<sup>32</sup>. The Jonckheere-Terpstra test requires independent samples divided into ranked orders so as to evaluate the strength of trends within the data<sup>33</sup>.

As null hypothesis significance testing (NHST) depends both on the size of the sample and on the size of the postulated effect under examination<sup>34</sup>, the emphasis of the inferential tests was placed primarily on the effect size (i.e., a standardised measure of the magnitude of the observed difference between sample groups)<sup>1</sup> and not only on its statistical significance (which, particularly for uneven sample groups, could confound effect size and sample size). The effect size shows if the predictor variable has any practical relevance<sup>35, 36</sup>, and was calculated by making use of equivalence between the standardised measure of the difference detected between sample groups and the Pearson's coefficient  $r$ , according to the following formula<sup>32, 37</sup> :

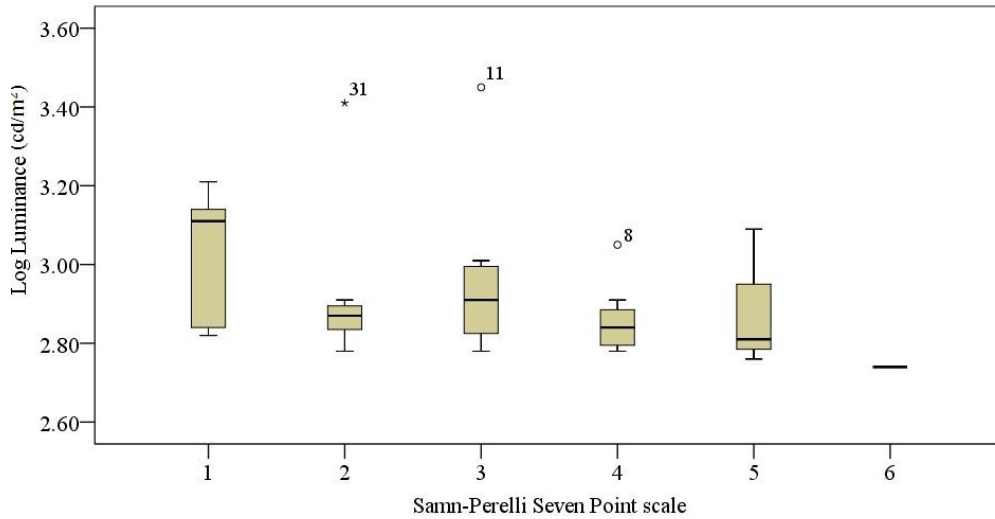
$$\text{Effect Size} = \frac{Z_{\text{score}}}{\sqrt{N}} \quad (1)$$

where N is the number of observations, and the test-statistic ( $Z_{\text{score}}$ ) was extracted from the Mann-Whitney U and Jonckheere-Terpstra inferential tests. The interpretation of the outcome was derived from the tables provided by Ferguson<sup>38</sup>, where benchmark values are given for ‘small’, ‘moderate’, and ‘large’ effect sizes ( $r \geq 0.20, 0.50, \text{ and } 0.80$ , respectively). Values below 0.20 were considered as not substantive (i.e., ‘negligible’), and thus not presenting a practically relevant effect.

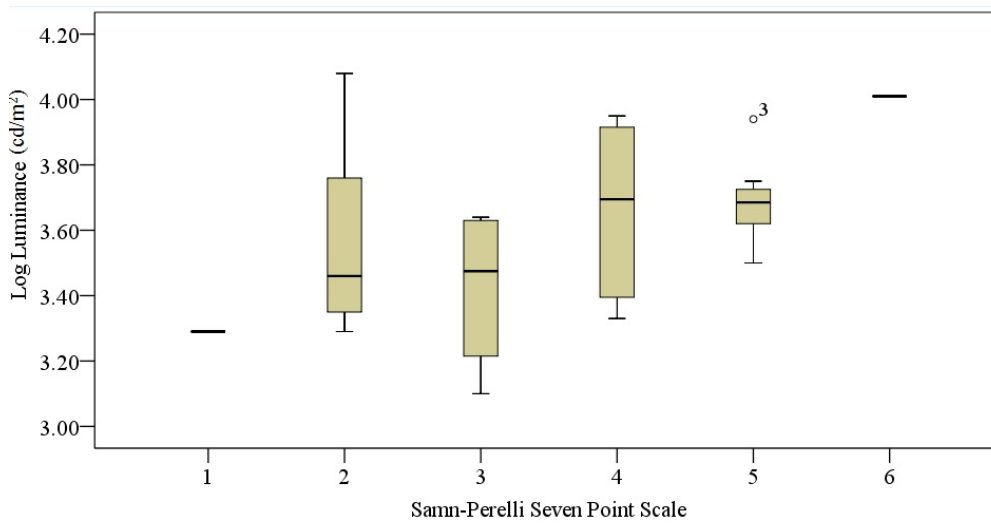
### **3. Results and discussion**

#### **3.1 Fatigue**

Figures 1 and 2 present selected boxplots related to consideration of the temporal variable fatigue for test sessions and GSVs of particular interest. The figures provide the values recorded in the Afternoon B session for the ‘Just Perceptible’ Glare Sensation Vote (Figure 1), and in the Evening for the ‘Just Intolerable’ GSV criterion (Figure 2). The figures plot, on the y-axis, the source log luminance ( $\text{cd/m}^2$ ) and, on the x-axis, the self-reported levels of fatigue at which participants expressed the various GSVs. Fatigue is ordered according to the Samn-Perelli seven-point scale, which divides the dependent variable into categories of increasing magnitude<sup>39</sup>. Table 1 provides a distribution of test subjects based on the feeling of fatigue self-reported at each session according to the Samn-Perelli scale, and gives a synthetic description of each point. The seventh point of the scale has not been included in the figures since none of the test subjects reached that level of fatigue at the time of the testing.



**Figure 1.** Boxplots for the variable Fatigue in the Afternoon B session for ‘Just Perceptible’



**Figure 2.** Boxplots for the variable Fatigue in the Evening session for ‘Just Intolerable’

From graphical inspection of the data it can be observed that, in the Afternoon B session and at low levels of glare sensation (‘Just Perceptible’, Figure 1), it is difficult to make any univocal interpretation of the distributions of minimum, 25<sup>th</sup> percentile, median, 75<sup>th</sup> percentile, and maximum statistical parameters, even if a general trend



seems to emerge for the ‘Just Perceptible’ GSV to be reported at a lower source log luminance with higher subjective fatigue. Conversely, in the Evening and at a greater level of visual discomfort (‘Just Intolerable’, Figure 2), there seems to be a much clearer tendency for this GSV to be reported by test subjects at a higher source log luminance as the magnitude of perceived fatigue increases.

Tables 2-5 report the J-value, the test statistic ( $Z_{\text{score}}$ ), the Monte Carlo simulated lower ( $CI_L$ ) and upper ( $CI_U$ ) 95% confidence intervals for the statistical significance, the statistical significance ( $p$ -value), and the effect size ( $r$ ) for the Jonckheere-Terpstra tests. Initial exploratory analysis revealed no prevailing directionality of central tendencies, hence the alternative hypothesis supported either a direct or an inverse relationship (i.e., as fatigue intensifies the log luminance respectively increases or decreases) between variables. The effect size provides a measure of the magnitude of these relationships, and its sign indicates whether the trend is direct (positive) or inverse (negative).

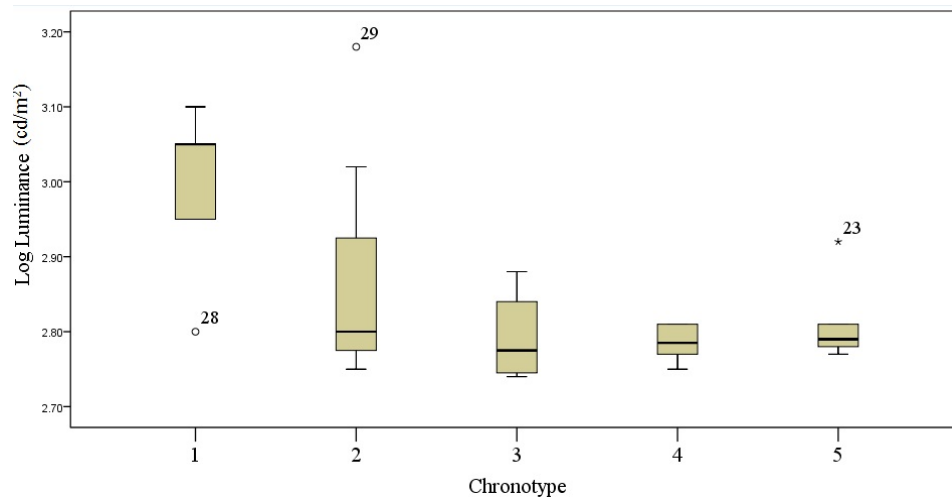
The two-tailed Jonckheere-Terpstra tests show evidence of only two weakly statistically significant differences ( $p \leq 0.05$ ). Specifically, for the GSV ‘Just Perceptible’ during the Afternoon B session ( $J= 136.00$ ,  $p= 0.04$ ,  $r= -0.36$  (small)), the results demonstrate a significant and practically relevant inverse relationship between fatigue and source log luminance. This suggests a higher tolerance to luminance levels for less fatigued subjects. Conversely, for the ‘Just Intolerable’ criterion during the Evening session ( $J= 245.50$ ,  $p= 0.02$ ,  $r= 0.40$  (small)), the findings reveal a substantive direct relationship between variables, indicating that test subjects self-assessing to be more fatigued reported the ‘Just Intolerable’ GSV at higher levels of source log luminance. Interestingly, although results were mostly not statistically significant, some effects of small, yet relevant, size ( $0.20 \leq r < 0.50$ ) could be detected for all GSVs.

### **3.2 Chronotype**

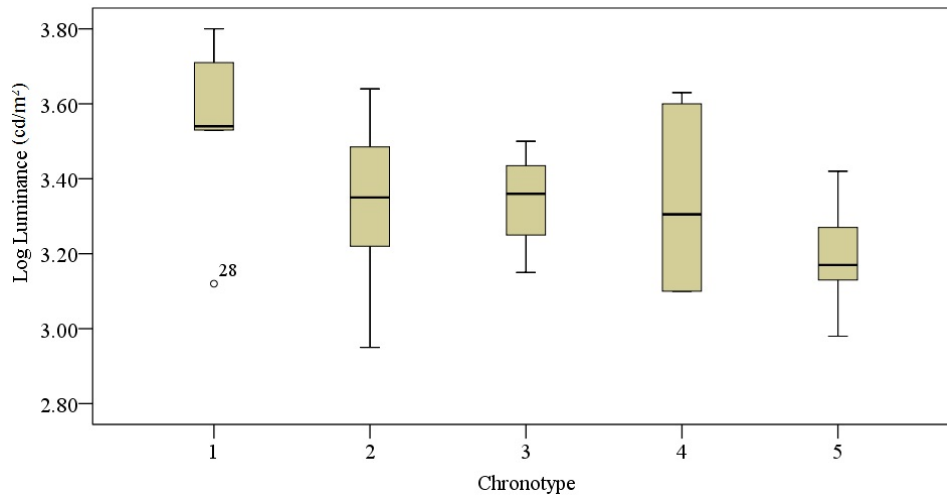
Figures 3 and 4 present selected boxplots referred to consideration of the variable chronotype in the Morning session for the ‘Just Perceptible’ (Figure 3) and the ‘Just

Intolerable' Glare Sensation Votes (Figure 4). Although the characteristics of chronotype depend on sleep phases (i.e., when subjects typically fall asleep and wake up within their circadian sleep cycles), this factor can be considered as an intrinsic personal trait not varying with time of the day. To identify this individual attribute, test subjects were requested to complete the Munich Chronotype Questionnaire (MCTQ)<sup>21</sup>, which categorises chronotypes using an ordinal scale ranging from 'Extremely Early'= Type 0 to 'Extremely Late'= Type 6. The distribution of participants according to their self-reported chronotype was: Type 0= 0, Type 1= 5, Type 2= 11, Type 3= 3, Type 4= 4, Type 5= 7, and Type 6= 0.

Figures 3 and 4 plot, on the y-axis, the source log luminance ( $\text{cd/m}^2$ ) of the small diffusive screen at which subjects provided the Glare Sensation Votes (GSVs) and, on the x-axis, the chronotype reported by participants. Since no subject was self-assessed as extremely early or late chronotype, Types 0 and 6 have not been included in the figures.



**Figure 3.** Boxplots for the variable Chronotype in the Morning session for 'Just Perceptible'



**Figure 4.** Boxplots for the variable Chronotype in the Morning session for ‘Just Intolerable’

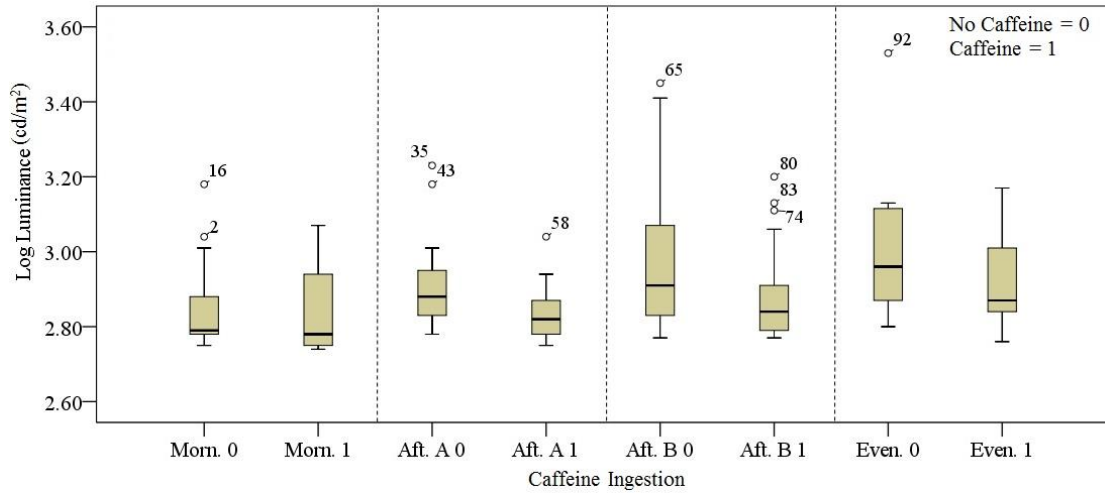
Graphical inspection of statistical parameters suggests a general trend for the criteria of GSV to be reported at lower levels of source log luminance as the chronotype of participants progresses from early to late type. That is, in the morning session, earlier chronotype subjects seem to be associated with a higher tolerance to source luminance. Tables 6-9 report the J-value, the test statistic ( $Z_{score}$ ), the Monte Carlo simulated lower ( $CI_L$ ) and upper ( $CI_U$ ) 95% confidence intervals for the statistical significance, the statistical significance ( $p$ -value), and the effect size ( $r$ ) for the Jonckheere-Terpstra tests. Initial exploratory analysis revealed a directional inverse relationship between variables (i.e., earlier chronotypes tended to report GSVs at higher levels of source luminance). Hence, the alternative hypothesis considered only statistical significance of an inverse effect, and the effect size measured the magnitude of the trend.

The consistently negative sign of the test statistic and of the effect size is coherent with the adoption of a one-tailed test of significance. The inferential analysis provides evidence that the inverse relationship between chronotype and source log luminance is significant ( $p \leq 0.01$ ) in four cases and weakly significant ( $p \leq 0.05$ ) in six cases out of 16.

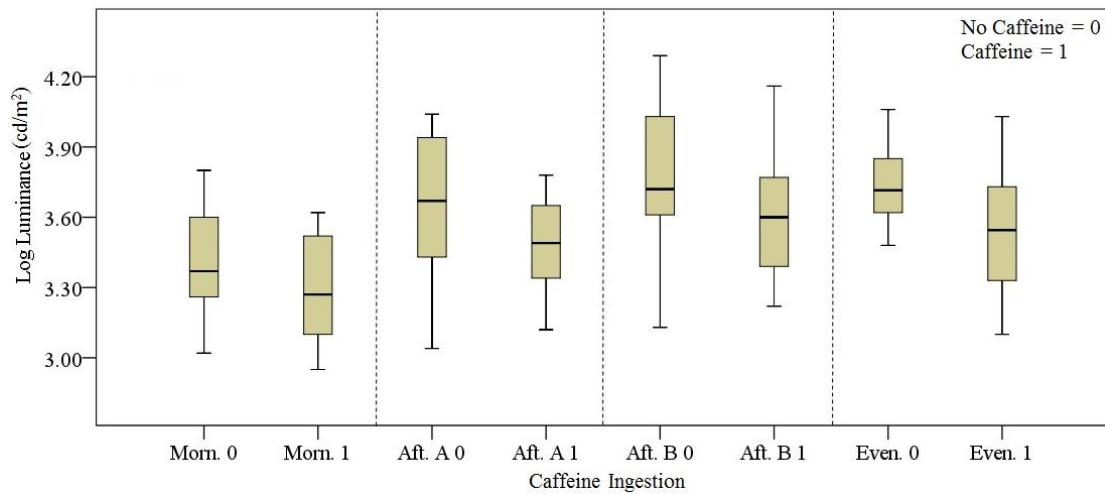
The results also indicate a consistently substantive magnitude of the effect of chronotype for all Glare Sensation Votes and test sessions, showing 14 small but practically relevant effect sizes ( $0.20 \leq r < 0.50$ ). The inferential results therefore seem to confirm the initial exploratory findings, providing statistically and practically significant evidence to the hypothesis that earlier chronotype test subjects were able to tolerate higher levels of source log luminance for all levels of visual discomfort and at all times of the day. Across all GSVs, no statistically significant differences were found for the Afternoon A session, this resulting in the only two negligible ( $r < 0.20$ ) effect sizes detected ( $r = -0.16$  for 'Just Uncomfortable', and  $r = -0.14$  for 'Just Intolerable').

### **3.3 Caffeine ingestion**

Figures 5 and 6 present selected boxplots taking into account the temporal variable caffeine ingestion. Similar to all other dichotomous variables, the figures plot, on the y-axis, the source log luminance ( $\text{cd}/\text{m}^2$ ) at which subjects reported the various GSVs and, on the x-axis, the different test sessions binary coded, in this case, as follows: 'No Caffeine' = 0, and 'Caffeine' = 1. So, for example, the 'Morn. 0' and 'Morn. 1' groups feature the data from the Morning session given by participants who declared to have not ingested (0) or to have ingested (1) caffeine before the test (or between test sessions).



**Figure 5.** Boxplots for the variable Caffeine Ingestion and for ‘Just Perceptible’



**Figure 6.** Boxplots for the variable Caffeine Ingestion and for ‘Just Intolerable’

Graphical inspection of the distribution of data reveals a tendency for the statistical parameters to correspond to higher levels of source log luminance for the ‘No Caffeine’ group (binary coded as 0) at each level of glare sensation and at all times of the day.

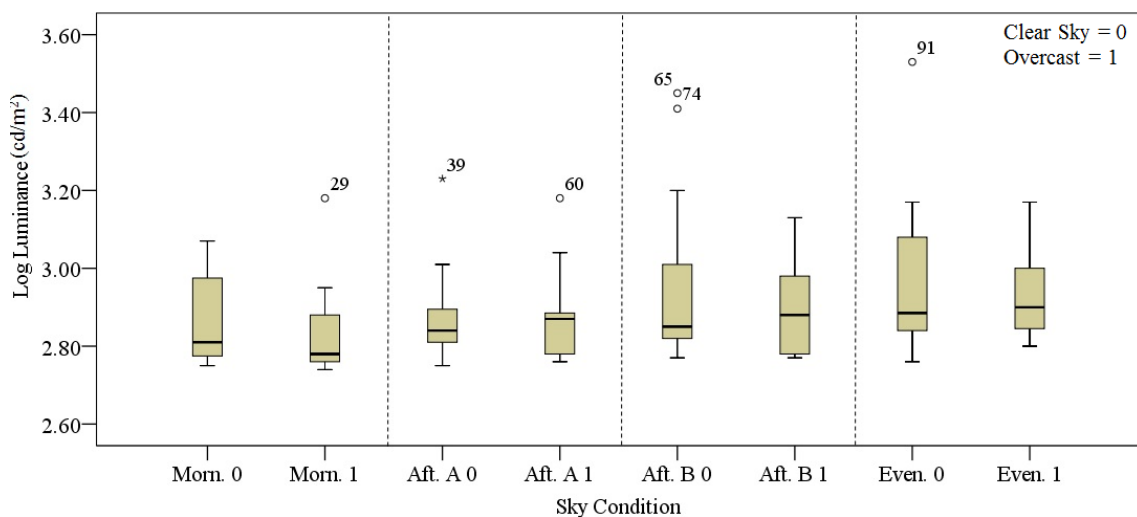
Tables 10-13 provide, for every GSV criterion, the sample size ( $N$ ) of the dichotomous groups ( $x_0$  corresponding to the 'No Caffeine' and  $x_1$  to the 'Caffeine' group), the medians ( $M_{dn}$ ) and interquartile ranges (IQR) of the log luminance at each session, the median difference ( $\Delta M_{dn}$ , 'No Caffeine' vs. 'Caffeine') between the groups and the outcome of its statistical significance (NHST,  $p$ -value calculated with a one-tailed test), the mean rank for the groups, the Mann-Whitney test statistic ( $U$ ), and the effect size ( $r$ ).

Analysis of descriptive statistics ( $\Delta M_{dn}$ ) suggests that test subjects featured in the 'No Caffeine' group show higher tolerance to the source log luminance for all Glare Sensation Votes and at each time of the day. In fact, the difference  $\Delta M_{dn}$  in median log luminance between the 'No Caffeine' and the 'Caffeine' groups is always positive, hence supporting the directionality of the initial alternative hypothesis and justifying the adoption of a one-tailed Mann-Whitney  $U$  test of significance. The results of the inferential tests indicate that differences across the independent variable are significant ( $p \leq 0.01$ ) in one case and weakly significant ( $p \leq 0.05$ ) in eight out of 16 cases. The outcomes also demonstrate that the differences over the temporal variable are substantive, with small but practically relevant effect sizes ( $0.20 \leq r < 0.50$ ) detected in all but one case ( $r = -0.18$  for the 'Just Intolerable' criterion in the Afternoon B session). With respect to the interpretation of the results, contrary to the previous Jonckheere-Terpstra tests, it is important to note that the sign of the effect size is not used here to indicate whether the relationship between variables is inverse or direct. The Mann-Whitney  $U$ , in fact, is a rank-order test used for assessing differences between the distributions of two independent groups when they are combined into a single sample<sup>31</sup>. When ranking the data for the two groups, the test compares the ranks of one group to the average ranks of both groups to determine if the ranks of each of the two samples are significantly different. In so doing, the test takes the lowest sum of ranks as the test statistic  $U$ , which is then converted to a  $Z_{score}$  for the calculation of the effect size. Since the test statistic  $U$  is set to equal the smaller of the two adjusted sums of ranks (and the

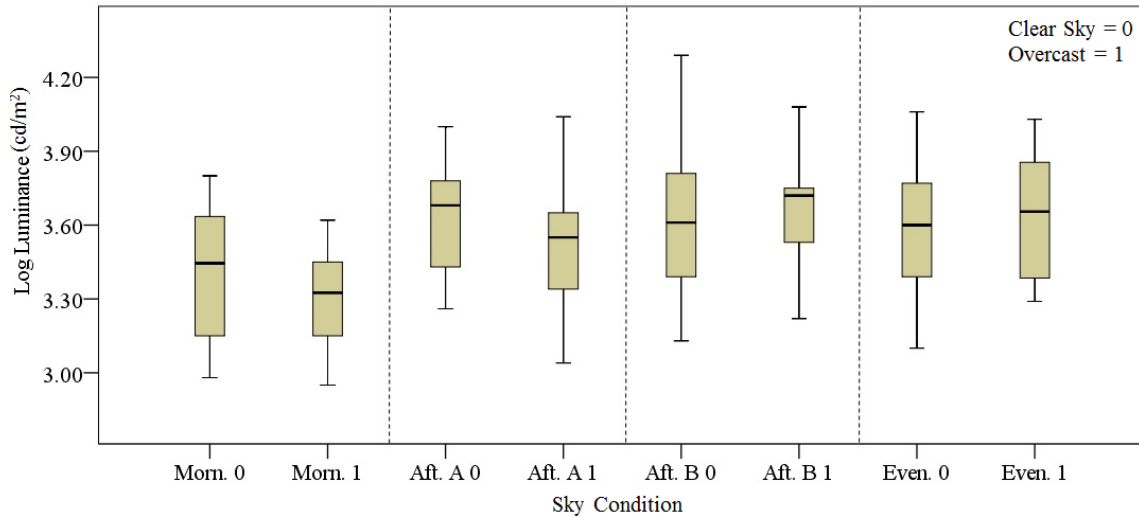
lowest sum of ranks will be evidently lower than the average ranks), in a one-tailed test the sign of the  $Z_{score}$  will always be negative (as will the effect size, which therefore has to be interpreted in its absolute value). In these tests, the directionality of the difference between groups is ascertained by looking at the boxplots of distributions, this supported by inspection of descriptive statistics (e.g.,  $\Delta M_{dn}$ ). As a matter of fact, although means or medians are not part of the procedures for obtaining and testing the test statistic, the Mann-Whitney U is considered as a powerful tool for indirectly measuring differences between these parameters<sup>40</sup>.

### 3.4 Sky condition

Figures 7 and 8 present selected boxplots related to consideration of the variable sky condition. The prevailing sky condition before the start of (or between) test sessions was reported by participants and binary coded as: ‘Clear Sky’= 0 and ‘Overcast’= 1. Test subjects were advised to include partially cloudy conditions under the ‘Clear Sky’ option.



**Figure 7.** Boxplots for the variable Sky Condition and for ‘Just Perceptible’



**Figure 8.** Boxplots for the variable Sky Condition and for ‘Just Intolerable’

Graphical inspection of statistical parameters does not lead to any univocal interpretation of a prevailing tendency for any of the Glare Sensation Votes.

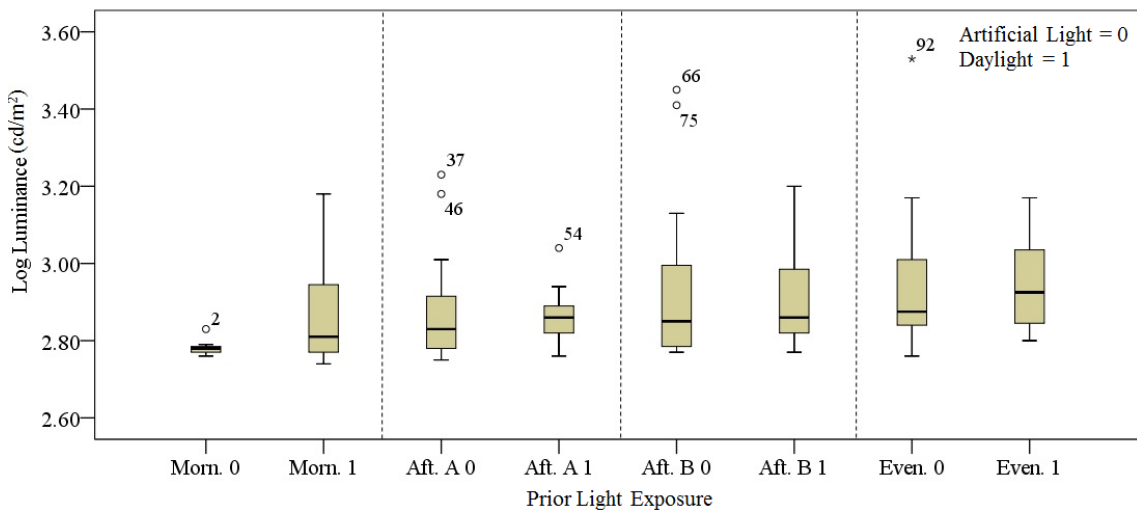
Tables 14-17 report the sample size ( $N$ ) of the dichotomous groups ( $x_0$  corresponding to the ‘Clear Sky’ and  $x_1$  to the ‘Overcast’ group), the medians ( $M_{dn}$ ) and interquartile ranges (IQR) of the log luminance for all times of the day, the median difference ( $\Delta M_{dn}$ , ‘Clear Sky’ vs. ‘Overcast’) between the groups and the outcome of its statistical significance (NHST,  $p$ -value calculated with a two-tailed test), the mean rank for the  $x_0$  and  $x_1$  groups, the test statistic ( $U$ ), and the effect size ( $r$ ).

The two-tailed Mann-Whitney  $U$  tests did not detect any statistically significant differences across the 16 cases considered. However, the effects over the independent variable are practically relevant ( $r \geq 0.20$ ) in five cases. In particular, the Morning session corresponds to substantive effect sizes for differences between independent groups across all GSV criteria ( $p = 0.19$  CI [0.18, 0.20],  $r = 0.25$  for ‘Just Perceptible’;  $p = 0.19$  CI [0.17, 0.19],  $r = 0.24$  for ‘Just Noticeable’;  $p = 0.19$  CI [0.18, 0.20];  $r = 0.24$  for ‘Just Uncomfortable’; and,  $p = 0.23$  CI [0.22, 0.24];  $r = 0.22$  for ‘Just Intolerable’).

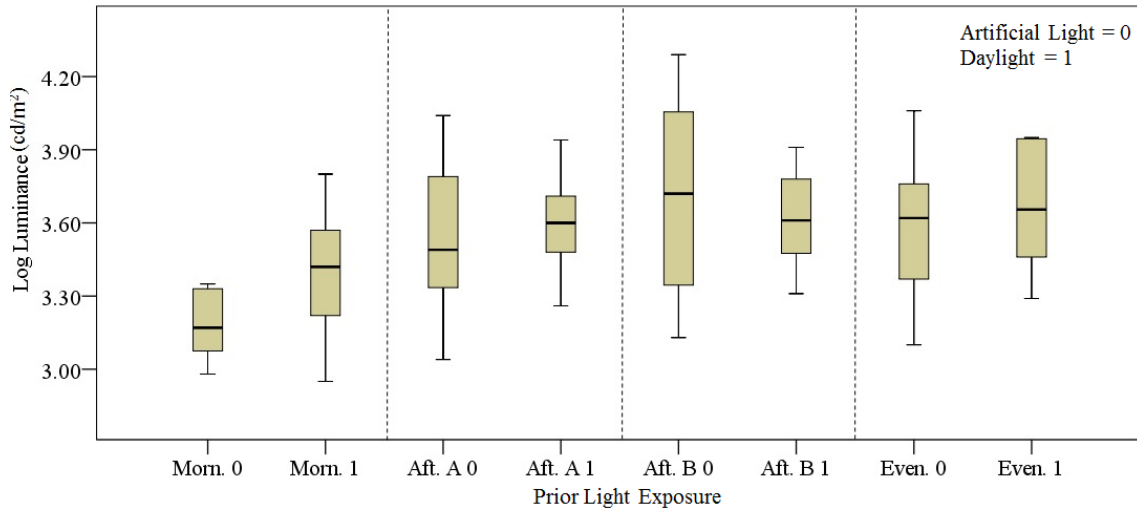


### 3.5 Prior light exposure

Figures 9 and 10 present selected boxplots referred to consideration of prior light exposure. To account for the influence of this variable, subjects were asked questions related to the lighting conditions they had experienced before or between test sessions, for example investigating whether their luminous environment was predominantly artificial or daylight, or whether they had sat next to a window with access to daylight. The dichotomous groups have been coded as: ‘Artificial Light’= 0 and ‘Daylight’= 1.



**Figure 9.** Boxplots for the variable Prior Light Exposure and for ‘Just Perceptible’



**Figure 10.** Boxplots for the variable Prior Light Exposure and for ‘Just Intolerable’

Graphical inspection of statistical parameters does not lead to the identification of a univocal and consistent prevailing tendency, although the distribution of data seems to suggest that, for some test sessions, subjects who had been previously exposed to a daylight environment reported votes of glare sensation at higher levels of log luminance.

Tables 18-21 provide the sample size (N) of groups ( $x_0$  corresponding to the ‘Artificial Light’ and  $x_1$  to the ‘Daylight’ group), the medians ( $M_{dn}$ ) and interquartile ranges (IQR) of the log luminance at each session, the median difference ( $\Delta M_{dn}$ , ‘Artificial Light’ vs. ‘Daylight’) between groups and the outcome of its statistical significance (NHST, p-value calculated with a two-tailed Mann-Whitney test), the mean rank for the  $x_0$  and  $x_1$  groups, the test statistics (U), and the effect size (r).

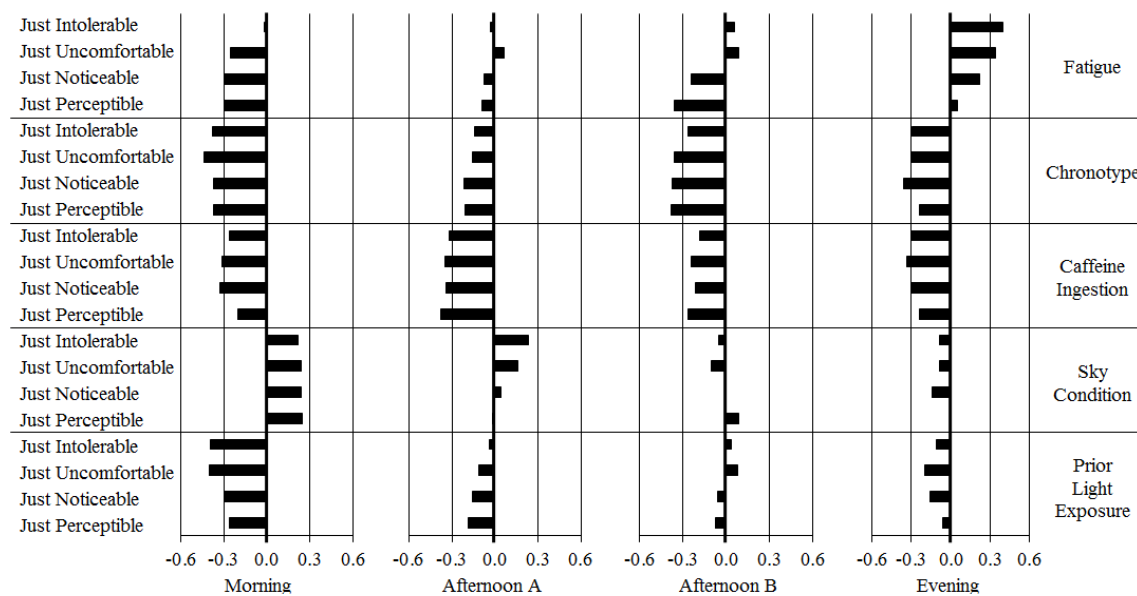
Analysis of the median differences ( $\Delta M_{dn}$ , ‘Artificial Light’ vs. ‘Daylight’) shows that subjects exposed to daylight before the tests reported the GSVs at a higher source log luminance across the various times of the day, with two exceptions. In fact,  $\Delta M_{dn}$  are always negative but in the Afternoon B session for the ‘Just Uncomfortable’ ( $\Delta M_{dn} = 0.21$ ) and ‘Just Intolerable’ ( $\Delta M_{dn} = 0.11$ ) Glare Sensation Votes. However, inferential statistics show detected differences being weakly significant ( $p \leq 0.05$ ) only

in two out of 16 cases, both corresponding to the Morning session for the highest levels of visual discomfort. In these cases, statistically significant differences between the independent groups indicate a higher tolerance to source log luminance for participants having been previously exposed to daylight with a substantive effect size ( $p=0.02$  CI [0.02, 0.03],  $r=-0.40$  for ‘Just Uncomfortable’; and  $p=0.03$  CI [0.03, 0.04],  $r=-0.39$  for ‘Just Intolerable’). Similar to the results obtained for the sky condition, the Morning session consistently corresponds to practically relevant effect sizes, although the detected differences related to the GSV criteria of ‘Just Perceptible’ and ‘Just Noticeable’ are not statistically significant (respectively,  $p=0.23$  CI [0.22, 0.24], and  $p=0.08$  CI [0.07, 0.09]).

#### **4. Discussion**

In interpreting the results of the graphical, descriptive, and inferential analyses, it should be considered that between-subject tests (e.g., Jonckheere-Terpstra and Mann-Whitney U) are less likely to detect statistical significance than repeated-measures tests (e.g., Wilcoxon-Matched Pairs). This is generally due to the individual differences between subjects that can be controlled (albeit not totally eliminated) by homogeneous samples<sup>41</sup>. Conversely, practical relevance (effect size) is independent of the size of sample groups, although it is not always a direct reflection of statistical significance<sup>42</sup>. It is indeed possible for a difference to be statistically significant (that is, unlikely to happen by chance) but trivial in its magnitude ( $r < 0.20$ ), whilst a result may be substantive ( $r \geq 0.20$ ) even if outside the conventional boundaries of statistical significance<sup>34, 43</sup>. Since the effect size provides a standardised measure of the magnitude of the differences between sample groups, it can be considered as a credible estimator of the scatter that, in previous analysis<sup>1</sup>, was observed when regressing the source log luminance at which test subjects provided their subjective vote of visual discomfort against the independent variable (time of the day).

Figure 11 graphically summarizes the effect sizes detected for each GSV criterion reported by participants (y-axis, on the left) related to consideration of temporal variables and personal factors (y-axis, on the right), and times of the day (x-axis).



$r < 0.20$  = negligible;  $0.20 \leq r < 0.50$  = small;  $0.50 \leq r < 0.80$  = moderate;  $r \geq 0.80$  = large

**Figure 11.** Effect sizes for all GSV criteria, variables, and test sessions

Figure 11 can help in the visualisation of the influences of temporal variables and personal factors on the subjective evaluation of glare sensation at different times of day. It should be noted that the sign of the effect size is retained in this figure so as to account for the two-tailed directionality of the effect in the cases of the variables fatigue, sky condition, and prior light exposure. Conversely, for the factors chronotype and caffeine ingestion, the consistently negative sign of the effect sizes substantiates the application of one-tailed alternative hypotheses, resulting from the initial graphical inspection of statistical parameters and supported by the analysis of descriptive statistics.

In Figure 11, the temporal trends for both the fatigue and chronotype variables follow a pattern similar to that found for other proxies such as cognitive performance and body temperature. This is typically exemplified by the ‘post-lunch dip’, apparent from 12:00 to 14:00 hours and creating a drop in individuals’ productivity<sup>44, 45</sup>. This characterisation can be observed by the evident decrease of the effect sizes at the Afternoon A sessions (12:00 or 12:30) with respect to the Morning (9:00 or 9:30). More specifically in terms of chronotype, the inferential analysis provided significant and substantive evidence that earlier chronotypes were able to tolerate higher levels of source log luminance at all times of the day. Consideration of the variation of effect sizes at different test sessions leads to hypothesise that the influence of chronotype on the subjective evaluation of glare sensation reduces around midday before increasing again in the early afternoon with the same directionality of the effect (that is, earlier chronotypes seem to tolerate higher levels of source luminance for the same reported level of visual discomfort).

The effect of caffeine ingestion on glare sensation is relatively consistent across the day (i.e., test subjects featured in the ‘No Caffeine’ group appear to be more tolerant to source luminance at all levels of visual discomfort and at all times of the day), although this influence is less substantive during the Afternoon B session (corresponding at a time after lunch, 15:00 or 15:30). This signals that the practical relevance of the inverse effect of caffeine ingestion over tolerance to source luminance for all criteria of glare sensation reduces in the early afternoon with respect to other times of day. However, it must be remembered that all the detected differences related to consideration of this temporal variable across all GSV criteria were not statistically significant in the Afternoon B session.

Lastly, both sky condition and prior light exposure show practically relevant effect sizes ( $r \geq 0.20$ ) almost exclusively in the Morning session. From the literature, there is evidence to suggest that early morning daylight exposure from 7am to 8am is important to induce alerting effects<sup>46</sup>. In addition, bright light exposure in the first hours of the day

is known to contribute to melatonin (the 'sleep' hormone) suppression<sup>47</sup>. The results obtained seem to confirm that, given the ranges of the independent variable (time of the day) considered in this study, the early morning would likely be the period where access to natural light (which is clearly related to prevailing sky conditions) would be most influential on the reported levels of visual discomfort.

## **5. Conclusions, limitations and future work**

This study, performed under controlled laboratory conditions, has investigated the influence of several temporal variables and personal factors on the subjective evaluation of glare sensation from an artificial lighting source at different times of the day. The main conclusions that can be drawn from the analysis are:

- Statistically significant and practically relevant evidence suggests that earlier chronotype test subjects were able to tolerate higher levels of source luminance for the same reported criteria of visual discomfort at all times of the day. This influence was found to decrease in magnitude in the middle of the day before increasing again in the early afternoon.
- Consideration of caffeine ingestion identified some statistically significant and practically relevant effects leading to an hypothesis that there is higher tolerance to source luminance across all criteria of glare sensation throughout the day for subjects not having ingested caffeine. This influence appeared to be less substantive in the early afternoon, although inferential analysis of the data did not lead to any statistically significant evidence at this time of day.
- Analysis of the results based on age, gender, ethnicity, food ingestion, and self-assessed photosensitivity of participants did not show any statistically significant difference between subjective evaluations of glare sensation at all levels of visual discomfort and for all the test sessions.

- Although some statistically and practically significant differences were detected, the results did not bring any conclusive and consistent evidence of the influence of fatigue, sky conditions, and prior light exposure on individual glare sensation at different levels of visual discomfort and times of the day. However, consideration of variation of effect sizes led to an hypothesis that early morning exposure to daylight would be most influential on the subjective evaluation of glare sensation.

In contextualising the influences detected, some methodological and experimental limitations should be considered. First of all, most variables were measured as dichotomous rather on a continuous, ordinal or interval scale, thus possibly masking some characteristics of their true effects. Also, daylight hours varied across subjects due to the tests having been performed between February and March. Therefore, caution should be used in inspecting the data, particularly at the Afternoon B and Evening sessions for the prior light exposure and sky condition variables.

In interpreting the results, it must also be taken into account that some (or all) of the temporal variables considered may have a high correlation with each other; for example, caffeine ingestion may influence the self-assessment of fatigue, or prior light exposure may be linked to sky condition. Therefore, it may be challenging to systematically isolate from experimental and statistical examination the direct influence of individual variables on the subjective evaluation of discomfort glare as the day progresses.

Nevertheless, the findings of this study clearly indicate that temporal variables and personal factors should be taken into consideration in the investigation of the potential effect of time of the day on visual discomfort, and measured in conjunction with votes of glare sensation so as to increase the likelihood of detecting the effect of interest. This will also contribute to explore the potential causes of scatter previously found when regressing reported levels of visual discomfort against times of the day<sup>1</sup>.

Finally, it must be acknowledged that all the findings presented and discussed in this study were derived from a laboratory setting under artificial lighting conditions. On-going investigations, with larger sample sizes and in a test room with direct access to daylight, are furthering the exploration of the complexity of the influences detected.

## **Funding**

This study is part of on-going PhD research conducted by the first author at the University of Nottingham, and funded by a EPSRC Doctoral Training Award.

## **References**

1. Kent MG, Altomonte S, Tregenza PR, Wilson R. Discomfort glare and time of day. *Lighting Research and Technology*, First published 21 August 2014. doi 1477153514547291.
2. Rodriguez RG, Pattini A. Tolerance of discomfort glare from a large area source for work on a visual display. *Lighting Research and Technology*, 2014; 46(2): 157-170.
3. Phipps-Nelson J, Redman JR, Dijk DJ, Rajaratman SMW. Daytime exposure to bright light, as compared to dim light, decreases sleepiness and improves psychomotor vigilance performance. *Sleep*, 2003; 26 (6): 695-700.
4. Smolders KCHJ, de Kort YAW, Cluitmans PJM. A higher illuminance induces alertness even during office hours: Findings on subjective measures, task performance and heart rate measures. *Physiology and Behavior*, 2012; 107: 7-16.
5. Boyce PR, Veitch JA, Newsham GR, Jones CC, Heerwagen J, Myer M, Hunter CM. Lighting quality and office work: two field simulation experiments. *Lighting Research and Technology*, 2006; 38(3): 191-223.
6. Borisuit A, Linhart F, Scartezzini JL, Munch M. Effects of realistic office daylighting and electric lighting conditions on visual comfort, alertness and mood.



*Lighting Research and Technology*, First published 22 April 2014, doi  
1477153514531518

7. Brainard GC, Hanifin JP, Greeson JM, et al. Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. *Journal of Neuroscience* 2001; 21(16): 6405-6412.
8. Cajochen C, Zeitzer JM, Czeisler CA, et al. Dose-response relationship for light intensity and ocular and electroencephalographic correlates of human alertness. *Behavioural Brain Research* 2000; 115: 75-83.
9. Zeitzer JM, Dijk DJ, Kronauer RE, et al. Sensitivity of the human circadian pacemaker to nocturnal light: melatonin phase resetting and suppression. *Journal of Physiology* 2000; 526(3): 695-702.
10. Zeitzer JM, Khalsa SBS, Boivin DB, et al. Temporal dynamics of late-night photic stimulation of the human circadian timing system. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* 2005; 289: 839-844.
11. Mardaljevic J, Heschong L, Lee E. Daylight metrics and energy savings. *Lighting Research and Technology*, 2009; 41(3): 261-283.
12. Rockcastle S, Andersen M. *Annual Dynamics of Daylight Variability and Contrast. A Simulation-based Approach to Quantifying Visual Effects in Architecture*. London: Springer, 2013.
13. Velds M. User acceptance studies to evaluate discomfort glare in daylight rooms. *Solar Energy*, 2002; 73(2): 95-103.
14. Iwata T, Kimura K. Discomfort caused by wide-source glare. *Energy and Buildings*, 1991; 15(3-4): 391-398.
15. Iwata T, Shukuya M, Somekawa N. Experimental study of discomfort glare caused by windows: Subjective responses to glare from simulated windows. *Journal of Architecture, Planning and Environmental Engineering*, 1992a; 432: 21-30.

16. Iwata T, Shukuya M, Somekawa N, et al. Experimental study of discomfort glare caused by windows: Subjective response to glare from actual windows. *Journal of Architecture, Planning and Environmental Engineering*, 1992b; 439: 19-31.
17. Iwata T, Tokura M. Examination of the limitations of predicted glare sensation vote (PSGV) as a glare index for a large source: Towards a comprehensive development of discomfort glare evaluation. *Lighting Research and Technology*, 1998; 30(2): 81-88.
18. Mochizuki E, Itoh D, Iwata T. Effects on discomfort glare of luminance distribution with a window. *Journal of Environmental Engineering*, 2009; 74(637): 277-282.
19. MacGowan D. Correspondence. *Lighting Research and Technology*, 2010; 42: 121-122.
20. Levandovski R, Sasso E, Hidalgo M.P. Chronotype: a review of the advances, limits and applicability of the main instruments used in the literature to assess human phenotype. *Trends in Psychiatry and Psychotherapy*, 2013; 35(1): 3-11.
21. Aries M. *Human Lighting Demands. Healthy Lighting in an Office Environment*. PhD thesis. Eindhoven: Technische Universiteit Eindhoven, 2005.
22. Roenneberg T, Wirz-Justice A, Meroo M. Life between clocks: Daily temporal patterns of human chronotypes. *Journal of Biological Rhythms*, 2003; 18(1): 80-90.
23. Tuaycharoen N, Tregenza PR. Discomfort glare from interesting images. *Lighting Research and Technology*, 2005; 37(4): 329-341.
24. Tuaycharoen N, Tregenza PR. View and discomfort glare from windows. *Lighting Research and Technology*, 2007; 39(2): 185-200.
25. Wienold J, Christoffersen J. Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras. *Energy and Buildings*, 2006; 38(7): 743-757.

26. Yun GY, Shin JU, Kim JT. Influence of window views on the subjective evaluation of discomfort glare. *Indoor and Built Environment*, 2011; 20(1): 65-74.
27. Heiman GW. *Basic Statistics for the Behavioural Sciences* 7<sup>th</sup> Edition. Stamford: Cengage, 2013.
28. Nordstokke DW, Zumbo BD. A new nonparametric Levene test for equal variances. *Psicologica*, 2010; 31: 401-430.
29. Siegel S. Nonparametric statistics. *The American Statistician*, 1957; 11(3): 13-19.
30. Hauschke D, Steinijans VW. Directionality decision for a two-tailed alternative. *Journal of Biopharmaceutical Statistics*, 1996; 6(2): 211-218.
31. Ruxton GD, Neuhauser M. When should we use one-tailed hypothesis testing? *Methods in Ecology and Evolution*, 2010; 1(2): 114-117.
32. Field A. *Discovering Statistics Using IBM SPSS Statistics*. 4<sup>th</sup> edition. London: Sage, 2013.
33. Jonckheere AR. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 1954; 41(1-2): 133-145.
34. Ellis PD. *The Essential Guide to Effect Sizes: Statistic Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press, 2010.
35. Cohen J. Some statistical issues in psychological research. In Wolman BB, editor. *Handbook of Clinical Psychology*. New York: McGraw-Hill, 1965, pp95-121.
36. Schiavon S, Altomonte S. Influence of factors unrelated to environmental quality on occupant satisfaction in LEED and non-LEED certified buildings. *Building and Environment*, 2014; 77: 148-159.
37. Rosenthal R, DiMatteo MR. Meta-Analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 2001; 52: 59-82.
38. Ferguson CJ. An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practise*, 2009; 40(5): 532-538.

39. Samn SW, Perelli LP, *Estimating aircrew fatigue: A technique with implications to airlift operations*, Report No. SAM-TR-82, Brooks AFB, Texas: USAF School of Aerospace Medicine, 1982.
40. Sawilowsky SS. Mann-Whitney U Test (Wilcoxon Rank-Sum Test). In Salkind NJ and Rasmussen K. *Encyclopaedia of Measurement and Statistics*. London: Sage, 2007, pp565-567.
41. Field A, Hole G. *How to Design and Report Experiments*. London: Sage, 2013.
42. Rosenthal R, Rosnow RL, Rubin DB. *Contrasts and Effect Sizes in Behavioural Research: A Correlation Approach*. Cambridge: Cambridge University Press, 2000.
43. Cohen J. The earth is round ( $p < .05$ ). *American Psychologist*, 1994; 49(12), 997-1003.
44. Monk TH. The post-lunch dip in performance. *Clinics in Sports Medicine*, 2005; 24: 15-23.
45. van Bommel WJM, van den Beld GJ. Lighting for work: a review of visual and biological effects. *Lighting Research and Technology*, 2004; 36(4): 255-266.
46. Okamoto Y, Rea MS, Figueiro MG. Temporal dynamics of EEG activity during short- and long-wavelength light exposures in the early morning. *BMC Research Notes*, 2014; 7(113): 1-6.
47. van Bommel WJM. Non-visual biological effect of lighting and the practical meaning for lighting for work. *Applied Ergonomics*, 2006; 37(4): 461-466.

## **Figure captions**

**Figure 1.** Boxplots for the variable Fatigue in the Afternoon B session for ‘Just Perceptible’

**Figure 2.** Boxplots for the variable Fatigue in the Evening session for ‘Just Intolerable’

**Figure 3.** Boxplots for the variable Chronotype in the Morning session for ‘Just Perceptible’

**Figure 4.** Boxplots for the variable Chronotype in the Morning session for ‘Just Intolerable’

**Figure 5.** Boxplots for the variable Caffeine Ingestion for ‘Just Perceptible’

**Figure 6.** Boxplots for the variable Caffeine Ingestion and for ‘Just Intolerable’

**Figure 7.** Boxplots for the variable Sky Condition and for ‘Just Perceptible’

**Figure 8.** Boxplots for the variable Sky Condition and for ‘Just Intolerable’

**Figure 9.** Boxplots for the variable Prior Light Exposure and for ‘Just Perceptible’

**Figure 10.** Boxplots for the variable Prior Light Exposure and for ‘Just Intolerable’

**Figure 11.** Effect sizes for all GSV criteria, variables, and test sessions

**Table 1.** Samn-Perelli distribution of test subjects

<b>Time of the Day</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
	Fully alert, wide awake	Very lively, responsive, but not at peak	Okay, somewhat fresh	A little tired, less than fresh	Moderately tired, let down	Extremely tired, very difficult to concentrate	Completely exhausted, unable to function effectively
<b>Morning</b>	1	5	11	8	5	0	0
<b>Afternoon A</b>	8	10	7	5	0	0	0
<b>Afternoon B</b>	5	6	7	7	4	1	0
<b>Evening</b>	1	9	4	7	8	1	0

**Table 2.** Jonckheere-Terpstra test for ‘Just Perceptible’

<b>Time of the Day</b>	<b>J-value</b>	<b>Test Statistic</b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>p-value (two-tailed)</b>	<b>Effect Size (r)</b>
<b>Morning</b>	131.00	-1.65	0.09	0.10	0.10 n.s.	-0.30
<b>Afternoon A</b>	162.00	-0.48	0.62	0.64	0.63 n.s.	-0.09
<b>Afternoon B</b>	136.00	-1.99	0.03	0.04	0.04*	-0.36
<b>Evening</b>	192.00	0.30	0.75	0.78	0.76 n.s.	0.05

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 3.** Jonckheere-Terpstra test for ‘Just Noticeable’

<b>Time of the Day</b>	<b>J-value</b>	<b>Test Statistic</b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>p-value (two-tailed)</b>	<b>Effect Size (r)</b>
<b>Morning</b>	131.50	-1.62	0.09	0.10	0.10 n.s.	-0.30
<b>Afternoon A</b>	164.00	-0.48	0.62	0.65	0.63 n.s.	-0.08
<b>Afternoon B</b>	154.50	-1.34	0.17	0.19	0.18 n.s.	-0.24
<b>Evening</b>	217.50	1.19	0.22	0.24	0.23 n.s.	0.22

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large



**Table 4.** Jonckheere-Terpstra test for ‘Just Uncomfortable’

<b>Time of the Day</b>	<b>J-value</b>	<b>Test Statistic</b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>p-value (two-tailed)</b>	<b>Effect Size (r)</b>
<b>Morning</b>	138.50	-1.37	0.17	0.18	0.17 n.s.	-0.25
<b>Afternoon A</b>	166.00	-0.34	0.72	0.74	0.73 n.s.	-0.06
<b>Afternoon B</b>	179.50	-0.47	0.62	0.64	0.63 n.s.	-0.09
<b>Evening</b>	236.00	1.85	0.05	0.06	0.06 n.s.	0.34

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 5.** Jonckheere-Terpstra test for ‘Just Intolerable’

<b>Time of the Day</b>	<b>J-value</b>	<b>Test Statistic</b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>p-value (two-tailed)</b>	<b>Effect Size (r)</b>
<b>Morning</b>	173.50	-0.12	0.90	0.92	0.90 n.s.	-0.02
<b>Afternoon A</b>	171.00	-0.16	0.86	0.88	0.87 n.s.	-0.03
<b>Afternoon B</b>	197.00	0.33	0.74	0.76	0.75 n.s.	0.06
<b>Evening</b>	245.50	2.18	0.02	0.03	0.02*	0.40

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 6.** Jonckheere-Terpstra test for ‘Just Perceptible’

<b>Time of the Day</b>	<b>J-value</b>	<b>Test Statistic</b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>p-value (one-tailed)</b>	<b>Effect Size (r)</b>
<b>Morning</b>	127.00	-2.03	0.01	0.03	0.02*	-0.37
<b>Afternoon A</b>	151.50	-1.17	0.11	0.13	0.12 n.s.	-0.21
<b>Afternoon B</b>	125.50	-2.08	0.01	0.02	0.01**	-0.38
<b>Evening</b>	147.50	-1.30	0.08	0.10	0.09 n.s.	-0.24

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 7.** Jonckheere-Terpstra test for ‘Just Noticeable’

<b>Time of the Day</b>	<b>J-value</b>	<b>Test Statistic</b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>p-value (one-tailed)</b>	<b>Effect Size (r)</b>
<b>Morning</b>	127.00	-2.03	0.02	0.03	0.02*	-0.37
<b>Afternoon A</b>	150.00	-1.22	0.10	0.12	0.11 n.s.	-0.22
<b>Afternoon B</b>	126.00	-2.06	0.01	0.02	0.01**	-0.37
<b>Evening</b>	128.00	-1.99	0.01	0.02	0.02*	-0.36

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 8.** Jonckheere-Terpstra test for ‘Just Uncomfortable’

<b>Time of the Day</b>	<b>J-value</b>	<b>Test Statistic</b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>p-value (one-tailed)</b>	<b>Effect Size (r)</b>
<b>Morning</b>	115.50	-2.43	0.00	0.01	0.01**	-0.44
<b>Afternoon A</b>	160.50	-0.85	0.18	0.20	0.20 n.s.	-0.16
<b>Afternoon B</b>	128.00	-1.99	0.02	0.03	0.02*	-0.36
<b>Evening</b>	138.00	-1.64	0.04	0.06	0.05*	-0.30

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s. = not significant ( $p > 0.05$ )

$r < 0.20$  = negligible;  $0.20 \leq r < 0.50$  = small;  $0.50 \leq r < 0.80$  = moderate;  $r \geq 0.80$  = large

**Table 9.** Jonckheere-Terpstra test for ‘Just Intolerable’

<b>Time of the Day</b>	<b>J-value</b>	<b>Test Statistic</b>	<b>CI<sub>L</sub></b>	<b>CI<sub>U</sub></b>	<b>p-value (one-tailed)</b>	<b>Effect Size (r)</b>
<b>Morning</b>	125.00	-2.10	0.01	0.02	0.01**	-0.38
<b>Afternoon A</b>	163.00	-0.76	0.09	0.12	0.11 n.s.	-0.14
<b>Afternoon B</b>	144.00	-1.43	0.06	0.08	0.07 n.s.	-0.26
<b>Evening</b>	139.00	-1.60	0.04	0.06	0.05*	-0.29

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 10.** Mann-Whitney U test for ‘Just Perceptible’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>No Caffeine M<sub>dn</sub> (IQR)</b>	<b>Caffeine M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(17, 13)	2.79 (0.11)	2.76 (0.19)	0.02 n.s.	17.06	13.46	82.00	-0.20
<b>Afternoon A</b>	(13, 17)	2.88 (0.13)	2.82 (0.09)	0.06**	19.46	12.47	59.00	-0.38
<b>Afternoon B</b>	(9, 21)	2.91 (0.24)	2.84 (0.11)	0.07 n.s.	18.89	14.05	60.50	-0.26
<b>Evening</b>	(8, 22)	2.96 (0.23)	2.87 (0.15)	0.09 n.s.	19.25	14.14	62.00	-0.24

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 11.** Mann-Whitney U test for ‘Just Noticeable’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>No Caffeine M<sub>dn</sub> (IQR)</b>	<b>Caffeine M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(17, 13)	2.93 (0.15)	2.83 (0.36)	0.10*	18.06	12.15	67.00	-0.33
<b>Afternoon A</b>	(13, 17)	3.11 (0.23)	2.99 (0.09)	0.12*	18.92	12.88	66.00	-0.34
<b>Afternoon B</b>	(9, 21)	3.26 (0.30)	3.01 (0.35)	0.25 n.s.	18.33	14.29	69.00	-0.21
<b>Evening</b>	(8, 22)	3.24 (0.25)	3.07 (0.17)	0.17*	19.88	13.91	53.00	-0.30

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large



**Table 12.** Mann-Whitney U test for ‘Just Uncomfortable’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>No Caffeine M<sub>dn</sub> (IQR)</b>	<b>Caffeine M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U-value</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(17, 13)	3.17 (0.25)	2.98 (0.37)	0.19*	17.85	12.42	70.50	-0.31
<b>Afternoon A</b>	(13, 17)	3.40 (0.36)	3.20 (0.23)	0.21*	19.04	12.79	64.50	-0.35
<b>Afternoon B</b>	(9, 21)	3.51 (0.48)	3.30 (0.42)	0.21 n.s.	18.61	14.17	67.50	-0.24
<b>Evening</b>	(8, 22)	3.48 (0.24)	3.28 (0.40)	0.20*	20.38	13.73	49.00	-0.33

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 13.** Mann-Whitney U test for ‘Just Intolerable’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>No Caffeine M<sub>dn</sub> (IQR)</b>	<b>Caffeine M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U-value</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(17, 13)	3.37 (0.34)	3.27 (0.43)	0.10 n.s.	17.50	12.88	76.50	-0.26
<b>Afternoon A</b>	(13, 17)	3.67 (0.51)	3.49 (0.31)	0.18*	18.73	13.03	68.50	-0.32
<b>Afternoon B</b>	(9, 21)	3.73 (0.42)	3.60 (0.38)	0.13 n.s.	17.89	14.48	73.00	-0.18
<b>Evening</b>	(8, 22)	3.72 (0.18)	3.55 (0.39)	0.17*	19.94	13.89	52.50	-0.30

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 14.** Mann-Whitney U test for ‘Just Perceptible’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>Clear Sky M<sub>dn</sub> (IQR)</b>	<b>Overcast M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U-value</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(12, 18)	2.81 (0.18)	2.78 (0.10)	0.03 n.s.	17.31	13.43	83.00	0.25
<b>Afternoon A</b>	(19, 11)	2.84 (0.08)	2.87 (0.10)	-0.03 n.s.	15.37	15.73	102.00	-0.02
<b>Afternoon B</b>	(21, 9)	2.85 (0.19)	2.88 (0.20)	-0.03 n.s.	16.00	14.33	84.00	0.09
<b>Evening</b>	(18, 12)	2.89 (0.23)	2.90 (0.15)	-0.01 n.s.	15.50	15.50	108.00	0.00

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 15.** Mann-Whitney U test for ‘Just Noticeable’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>Clear Sky M<sub>dn</sub> (IQR)</b>	<b>Overcast M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U-value</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(12, 18)	2.95 (0.27)	2.85 (0.14)	0.09 n.s.	17.96	13.86	78.50	0.24
<b>Afternoon A</b>	(19, 11)	3.02 (0.23)	3.01 (0.23)	0.01 n.s.	15.79	15.00	99.00	0.04
<b>Afternoon B</b>	(21, 9)	3.01 (0.33)	3.08 (0.21)	-0.08 n.s.	15.43	15.67	93.00	-0.01
<b>Evening</b>	(18, 12)	3.08 (0.24)	3.10 (0.24)	-0.03 n.s.	14.53	16.96	90.50	-0.14

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 16.** Mann-Whitney U test for ‘Just Uncomfortable’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>Clear Sky M<sub>dn</sub> (IQR)</b>	<b>Overcast M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U-value</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(12, 18)	3.18 (0.25)	3.01 (0.14)	0.18 n.s.	18.00	13.83	73.00	0.24
<b>Afternoon A</b>	(19, 11)	3.34 (0.28)	3.21 (0.35)	0.13 n.s.	17.12	14.26	89.50	0.16
<b>Afternoon B</b>	(21, 9)	3.27 (0.43)	3.48 (0.30)	-0.22 n.s.	14.90	16.89	82.00	-0.10
<b>Evening</b>	(18, 12)	3.33 (0.29)	3.40 (0.35)	-0.07 n.s.	14.94	16.33	98.00	-0.08

\*p $\leq$ 0.05; \*\*p $\leq$ 0.01; \*\*\*p $\leq$ 0.001; n.s.= not significant

r $<$ 0.20= negligible; 0.20 $\leq$ r $<$ 0.50= small; 0.50 $\leq$ r $<$ 0.80= moderate; r $\geq$ 0.80= large

**Table 17.** Mann-Whitney U test for ‘Just Intolerable’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>Clear Sky M<sub>dn</sub> (IQR)</b>	<b>Overcast M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U-value</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(12, 18)	3.45 (0.39)	3.29 (0.28)	0.16 n.s.	17.79	13.97	80.50	0.22
<b>Afternoon A</b>	(19, 11)	3.59 (0.32)	3.49 (0.33)	0.09 n.s.	17.77	13.76	81.00	0.23
<b>Afternoon B</b>	(21, 9)	3.61 (0.42)	3.72 (0.22)	-0.11 n.s.	15.24	16.11	89.00	-0.05
<b>Evening</b>	(18, 12)	3.60 (0.35)	3.66 (0.37)	-0.06 n.s.	14.94	16.33	98.00	-0.08

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 18.** Mann-Whitney U test for ‘Just Perceptible’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>Artificial M<sub>dn</sub> (IQR)</b>	<b>Daylight M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U-value</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(7, 23)	2.78 (0.01)	2.81 (0.18)	-0.03 n.s.	11.71	16.65	54.00	-0.26
<b>Afternoon A</b>	(16, 14)	2.83 (0.12)	2.86 (0.07)	-0.03 n.s.	14.03	17.18	88.50	-0.19
<b>Afternoon B</b>	(15, 15)	2.85 (0.20)	2.86 (0.16)	-0.01 n.s.	14.73	16.27	101.00	-0.07
<b>Evening</b>	(22, 8)	2.88 (0.17)	2.92 (0.17)	-0.05 n.s.	14.95	17.00	76.00	-0.06

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 19.** Mann-Whitney U test for ‘Just Noticeable’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>Artificial M<sub>dn</sub> (IQR)</b>	<b>Daylight M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U-value</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(7, 23)	2.86 (0.07)	2.97 (0.34)	-0.11 n.s.	10.29	17.09	44.00	-0.30
<b>Afternoon A</b>	(16, 14)	2.99 (0.27)	3.05 (0.21)	-0.06 n.s.	13.91	17.32	86.50	-0.16
<b>Afternoon B</b>	(15, 15)	3.06 (0.37)	3.08 (0.28)	-0.03 n.s.	14.93	16.07	104.00	-0.06
<b>Evening</b>	(22, 8)	3.08 (0.23)	3.13 (0.23)	-0.06 n.s.	14.66	17.81	69.50	-0.16

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large



**Table 20.** Mann-Whitney U test for ‘Just Uncomfortable’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>Artificial M<sub>dn</sub> (IQR)</b>	<b>Daylight M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U-value</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(7, 23)	3.02 (0.11)	3.18 (0.34)	-0.17*	9.07	17.46	35.50	-0.40
<b>Afternoon A</b>	(16, 14)	3.24 (0.37)	3.34 (0.20)	-0.10 n.s.	14.69	16.43	99.00	-0.11
<b>Afternoon B</b>	(15, 15)	3.51 (0.54)	3.30 (0.35)	0.21 n.s.	16.20	14.80	102.00	0.08
<b>Evening</b>	(22, 8)	3.34 (0.37)	3.41 (0.37)	-0.07 n.s.	14.43	18.44	64.50	-0.20

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large

**Table 21.** Mann-Whitney U test for ‘Just Intolerable’

<b>Time of the Day</b>	<b>N (x<sub>0</sub>, x<sub>1</sub>)</b>	<b>Artificial M<sub>dn</sub> (IQR)</b>	<b>Daylight M<sub>dn</sub> (IQR)</b>	<b><math>\Delta M_{dn}^{NHST}</math></b>	<b>Mean Rank<sub>(x<sub>0</sub>)</sub></b>	<b>Mean Rank<sub>(x<sub>1</sub>)</sub></b>	<b>U-value</b>	<b>Effect Size (r)</b>
<b>Morning</b>	(7, 23)	3.17 (0.26)	3.42 (0.35)	-0.25*	9.29	17.39	37.00	-0.39
<b>Afternoon A</b>	(16, 14)	3.49 (0.45)	3.60 (0.21)	-0.10 n.s.	15.19	15.86	107.00	-0.04
<b>Afternoon B</b>	(15, 15)	3.72 (0.71)	3.61 (0.30)	0.11 n.s.	15.83	15.17	107.50	0.04
<b>Evening</b>	(22, 8)	3.62 (0.37)	3.66 (0.44)	-0.04 n.s.	14.93	17.06	75.50	-0.11

\*\*\* $p \leq 0.001$ ; \*\* $p \leq 0.01$ ; \* $p \leq 0.05$ ; n.s.= not significant ( $p > 0.05$ )

$r < 0.20$ = negligible;  $0.20 \leq r < 0.50$ = small;  $0.50 \leq r < 0.80$ = moderate;  $r \geq 0.80$ = large