



**HAL**  
open science

# Quantitative and Qualitative Evaluation of IES TM-30 Indices for the Color Preference of Landscape Oil Paintings

Ching-Wei Lin, Sophie Jost-Boissard, Youri Meuret, Kevin Smet

## ► To cite this version:

Ching-Wei Lin, Sophie Jost-Boissard, Youri Meuret, Kevin Smet. Quantitative and Qualitative Evaluation of IES TM-30 Indices for the Color Preference of Landscape Oil Paintings. LEUKOS, 2026, pp.1-24. <10.1080/15502724.2025.2566689>. <hal-05544970>

**HAL Id: hal-05544970**

**<https://hal.science/hal-05544970v1>**

Submitted on 10 Mar 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Quantitative and Qualitative Evaluation of IES TM-30 Indices for the Color Preference of Landscape Oil Paintings

Ching-wei Lin [ORCID](#)<sup>a\*</sup>, Sophie Jost-Boissard [ORCID](#)<sup>b</sup>, Youri Meuret [ORCID](#)<sup>a</sup>, and Kevin A.G. Smet [ORCID](#)<sup>a</sup>

<sup>a</sup>ESAT-WaveCore/Light & Lighting Laboratory, KU Leuven, Ghent, Belgium; <sup>b</sup>ENTPE, Ecole Centrale de Lyon, Vaulx-en-Velin, France

Ching-wei Lin [jingwei.edu@gmail.com](mailto:jingwei.edu@gmail.com) Address: ESAT/Light & Lighting Laboratory, KU Leuven, Gebroeders De Smetstraat 1, Ghent 9000, Belgium.

Provide short biographical notes on all contributors here if the journal requires them: not required for LEUKOS.

## ORCID

C-W L: <http://orcid.org/0000-0002-9813-3826>

S J-B: <https://orcid.org/0000-0002-2607-6543>

Y M: <http://orcid.org/0000-0002-2815-5915>

KAG S: <http://orcid.org/0000-0003-3825-6274>

**Acknowledgments.** Author C-W Lin would like to thank Taiwan–KU Leuven scholarship to support his PhD study. During the preparation and revision of the manuscript, Perplexity AI, featuring a Deep Research capability launched in February 2025 was used solely for improving English writing.

## Funding

This work was supported by the Ministry of Education, Taiwan under Taiwan–KU Leuven scholarship program of Ching-wei Lin; and KU Leuven under the starting grant of Kevin A.G. Smet [grant number STG/17/013].

# Quantitative and Qualitative Evaluation of IES TM-30 Indices for the Color Preference of Landscape Oil Paintings

ANSI/IES TM-30-20 Annex E provides guidelines for evaluating the color rendition performance of interior lighting, which is incorporated in the latest ANSI/IES RP-30-20 standard as a guide for museum lighting designers. However, one general guideline may not be suitable for all types of museum exhibitions. This study investigates the relationship between TM-30 indices and preference for landscape oil paintings, assesses the applicability of TM-30 color preference criteria, and examines whether the results depend on the painting color composition. Nine 3000 K lighting spectra were systematically generated with varying  $R_f$ ,  $R_g$ , and best-fit ellipse rotation angles  $\psi$  on the color vector graph. Through a two-alternative-forced-choice paired comparison, the preference of 18 participants for six landscape oil paintings with various colors was evaluated. Their choices were converted to quantitative interval scores for each spectrum. Based on the statistical difference between scores, the preference criteria were adjusted by lowering  $R_f$  thresholds and dividing the most preferred level into two levels. The less strict  $R_f$  requirement might be attributed to the lack of memory color for unfamiliar paintings. Repeated measure ANOVAs revealed that the six paintings can be classified into two groups of three paintings: a high- $R_{cs,hl}$ -preferred group and a high- $R_g$ -preferred group. Quantitative answers from participants to questionnaires provided potential methods for categorizing paintings into these groups. This study concludes with three recommendations for narrowing the selection of luminaires for landscape oil paintings: one for each group, and the adjusted preference criteria for paintings that cannot be clearly classified into either group.

Keywords: color preference; TM-30; museum lighting; paintings; color rendering

## 1. Introduction

A good art museum lighting requires careful consideration of multiple aspects. Numerous studies have contributed to the development of standards for lighting designers, including ANSI/IES RP-30-20 Recommended Practice: Lighting Museums

(IES 2020) and ANSI/IES TM-30 IES Method for Evaluating Light Source Color Rendition (IES 2024). However, these standards were not specifically developed for artwork illumination, and their applicability to the preference of oil paintings is evaluated in this study.

### ***1.1 Art museum lighting***

Museum lighting needs for art exhibition can be categorized into two approaches. The first focuses on the visual experience of the visitors. Kesner (1993) identified the visual needs of art museum visitors, including minimizing direct glare, maintaining acceptable gallery brightness, ensuring accurate artwork color, providing clear artwork forms, and showing rich artwork color. Among these, color accuracy and clear forms were the most important (Kesner 1993). The second approach is the intension of the artists, aiming to make the artwork appear closer to what the artist wants to show. For example, Schanda et al. (2016) introduced a new color fidelity calculation to reproduce the painting as seen under daylight while maintaining a warm-white light source for museum lighting.

However, in museum lighting, there is always a challenge to balance preservation and exhibition. Strict preservation reduces light exposure but this compromises exhibition quality, and vice versa (Ajmat et al. 2011). Researchers have studied methods to minimize light damage. For example, Durmus et al. (2020) demonstrated that artwork light absorption can be considerably reduced while maintaining the color appearance. Cuttle (2020) proposed a three-band source that creates an environment with equal visual satisfaction and correlated color temperature (CCT) as tungsten halogen spotlights, but with substantially lower irradiance on the paintings.

## *1.2 Factors influencing visual perception of paintings*

Paintings are among the most colorful objects in an art museum, and their appearance is influenced by various environmental factors, particularly lighting. Lighting conditions are typically classified into four main categories: spatial, temporal, intensity, and spectral (Royer et al. 2022). Among all factors, illuminance (an intensity parameter) and CCT (a spectral parameter) were the most commonly examined in previous studies, and often in conjunction.

Both have been found to impact perception, but when both illuminance and CCT are varied, illuminance tends to dominate. Bellia et al. (2020) investigated painting perception by modifying CCT of the lighting, illuminance on two paintings, and the wall color in the background. They found that illuminance is the main parameter affecting the perception of paintings. Wall color, CCT, and the combination of these parameters impacted the perception depending on the chromatic composition of the painting.

Based on several studies, an illuminance of 200 lux is appropriate for paintings under certain conditions. In a study involving one watercolor painting and four oil paintings, Loe et al. (1982) concluded that “a painting illuminance approaching 200 lux was required to give visual satisfaction for an illuminant with high color rendering index (CRI) and high gamut area.” The 200 lux level has been adopted in several studies (Bellia et al. 2020; Chen et al. 2016; Nascimento and Masuda 2014; Zhai et al. 2015, 2016) and is also specified as the upper limit for oil paintings to minimize damage from too intense illumination according to the latest Illumination Engineering Society (IES) museum lighting standard—ANSI/IES RP-30-20 Recommended Practice: Lighting Museums (IES 2020).

Some studies kept the illuminance fixed and found an effect of CCT on the visual perception and visual appreciation of the illuminated artwork. For example, Feltrin et al. (2020) examined the effects of altering the dominant color of paintings, CCT of lighting, and wall lightness (white, gray, or black), on various visual perceptions including warmth, vividness, and brightness of the lighting, attractiveness of painting colors, the appreciation of the wall color, and the appreciation of the overall arrangement. Among the three altered variables, they found that CCT has the largest effect on the evaluation of each measured perceptual factor.

There is disagreement about the most preferred CCT when viewing paintings. Kruithof (1941) defined a pleasant zone, which became the most famous rule about CCT for interior lighting design. However, Chen et al. (2016) showed that Kruithof's rule only partially agreed with their visual experiment results. Scuello et al. (2004) concluded that a CCT of 3600 K was preferred with an illuminance on the painting of approximately 200 – 250 lux (a result mentioned in IES RP-30-20). Pinto et al. (2008) found that a CCT of approximately 5100 K maximized visual preference; Nascimento and Masuda (2014) reported a most preferred CCT of 5500 K for a real scene.

Although illuminance and CCT are important lighting factors affecting painting perception, they are not within the scope of this study which focuses on the effects of color rendering. Based on the conditions applied in a local fine-art museum and taking into account the aforementioned studies, 3000 K with a vertical illuminance of 150 lux at the painting center were chosen. Paintings are typically created to have a pleasing appearance to viewers. Thus, color preference should be a key consideration in the lighting design process for paintings and is therefore the focus of this article.

### ***1.3 Lighting standards on color rendering: ANSI/IES TM-30***

Color rendering is also important in the context of museum lighting as it directly

influences the color appearance of exhibited objects. The IES RP-30-20 guideline, applicable to various museum types from art to archaeology and natural sciences, recommends that lighting designers utilize color rendering metrics to narrow their light source choices and finalize specifications with mockups (IES 2020). This standard also recommends the ANSI/IES TM-30-18, IES Method for Evaluating Light Source Color Rendition (IES 2024) in favor of the CIE General Color Rendering Index ( $R_a$ ; colloquially, CRI) (CIE 1995).

IES TM-30, first published in 2015 as “TM-30-15” and updated to “TM-30-18,” “TM-30-20,” and “TM-30-24” in 2018, 2020, and 2024, respectively, quantifies a light source’s color rendition with several indices. The fidelity index  $R_f$  and the gamut index  $R_g$  are calculated based on the color appearances of 99 color evaluation samples (CES). In the  $R_g$  calculation process, 16 hue angle bins are established. To illustrate and detail the CES color appearances, a color vector graphic (CVG) and additional indices are defined: local chroma shift  $R_{cs,hj}$ , local hue shift  $R_{hs,hj}$ , and local color fidelity  $R_{f,hj}$  where  $j = 1, 2, \dots, 16$  represent each hue angle bin.

The IES TM-30 Annex E provides guidance for applying the TM-30 indices to a desired lighting design based on three design intents—color preference (P), color vividness (V), and color fidelity (F)—, each subdivided in three priority levels (1, 2, 3) from high to low. For example, an environment illuminated with a lighting spectrum that fits the P1 criteria ( $R_f \geq 78$ ,  $R_g \geq 95$ , and  $-1\% \leq R_{cs,h1} \leq 15\%$ ) is considered more preferred than one illuminated with a spectrum that only fits the P3 criteria ( $R_f \geq 70$ ,  $R_g \geq 89$ , and  $-12\% \leq R_{cs,h1} \leq 23\%$ ). Spectra not fitting the P3 criteria are considered less preferred and are marked as P-. These priority symbols (1, 2, 3, -) are also applied to color vividness and color fidelity intents. The color preference specification criteria in IES TM-30 Annex E are based on five experiments (IES 2024) evaluating daily life

objects (Esposito and Houser 2019; Royer et al. 2017, 2018, 2020; Zhang et al. 2017), assuming a polychromatic environment with average horizontal illuminance between 200 lux to 700 lux and uniform chromaticity.

While IES RP-30-20 is understandably recommending IES TM-30 for museum lighting designers, TM-30 Annex E may not be entirely suitable for paintings. First, museum painting illuminance is often 200 lux or lower (IES 2020), and is at the lower limit of the Annex E illuminance range (IES 2024). Second, TM-30 does not consider the Hunt, Bezold-Brucke, or Stevens effects, limiting its accuracy in describing color perception under varying light intensities. Though research is ongoing (Ohno and Li 2023), current findings have not yet extended Annex E's applicable illuminance range, potentially making it unreliable for low-illuminance museum lighting. Furthermore, TM-30 is based on studies using familiar objects from daily life where memory color influences preference judgement. However, memory colors might be less relevant for unfamiliar paintings due to the lack of a memory reference, potentially affecting color appearance preference and thus deviating from the color preference criteria derived from daily life objects.

Four studies have already evaluated painting color preference with TM-30 indices (Bao and Wei 2021; Feltrin et al. 2020; Mundinger and Houser 2023; Wei et al. 2020), but none of their results could confirm TM-30-20 color preference levels. Bao and Wei (2021), Wei et al. (2020), and Mundinger and Houser (2023) observed that increasing  $R_g$ , up to a limit, enhances preference. However, they only considered one or two paintings, raising questions about the general applicability of their results to other paintings and limiting the ability to verify color preference levels. Feltrin et al. (2020) compared five paintings, but set TM-30 indices as control variables ( $R_f = 88$  or  $89$  and  $R_g = 101$  or  $102$ ), which precluded the verification of the levels.

#### ***1.4 Objective of this study***

The aim of this study is to evaluate TM-30 indices on the color preference of several landscape oil paintings. The research questions are:

- Are TM-30 Annex E color preference criteria applicable to landscape oil paintings?
- If not, how can the criteria be adjusted?
- Does color preference depend on the color composition of the painting?

This research serves as an initial step in studying painting preference, focusing on and limiting to landscape oil paintings, a common type in fine art museums. By comparing paintings with similar material (oil paints), theme (landscape), and style (realistic) but varying colors, this study investigates whether the results are consistent or affected by painting color compositions. Consistent results will enable future studies on other paintings to broaden the application. Inconsistent results will necessitate methods to classify paintings to design a preferred lighting spectrum.

Most aforementioned studies on color rendition and on painting color preference primarily consider quantitative evaluation results (scores). While some authors have interpreted participants preference from scores, such as “observers prefer more neutral lighting for museum displays” (Feltrin et al. 2020), qualitative answers are rarely reported. This leaves uncertainty about whether participants would agree with the authors’ interpretation. For instance, even though the preferred lighting in Feltrin's study was “physically” more neutral, did the participants “perceive and notice” that paintings under this preferred lighting appeared more neutral? By providing space for comments, Munding and Houser (2023) collected participants’ reasoning for their evaluations, helping identify under- and over-saturated scene. In this study both

quantitative scores and qualitative responses (answers to open questions) were collected to strengthen and validate the findings.

## **2. Method**

### ***2.1 Apparatus and test space***

Oil paintings were hung in a museum-like environment and illuminated by a 7-channel spectral tunable luminaire (SPECTRA TUNE 7! – LEDMOTIVE). Participants sat in a high chair in front of the painting at a distance of 0.8 m, with their eyes nearly level with the middle of the painting. Mid-gray cloths covered the frames of the paintings and the walls, serving as the background of the visual field of the participants (Fig. 1). The lighting spectra were determined using the measured spectral radiance of a white reference tile located at the center of the painting. The spectral radiance of the tile and that of the background were measured with a JETI Specbos 1211L radiometer, and the reflectance spectrum of the tile was measured with a Hunterlab Ultrascan Pro Spectrophotometer under a specular reflectance included 8/d geometry condition.

[Fig. 1 near here]

### ***2.2 Lighting spectra***

The luminaire generated nine spectra with a fixed background color. The reflected light of the gray background under these spectra and a 3000 K Planckian radiator were colorimetric metamers for the CIE 1964 standard observer, with  $L_{10} = 5.85 \pm 0.09$  cd/m<sup>2</sup> and a maximum  $\Delta u^*v^*_{10}$  of 0.0017 (Table 1). The luminance value corresponded to a vertical illuminance of approximately 150 lux at the painting center, located in the Kruithof's pleasant zone. This is also the value set by a local fine-art museum, where future in-situ experiments will be held. The nine spectra were determined to provide

different  $R_f$  and  $R_g$ . When  $R_f$  is low, two spectra with identical  $R_f$  and  $R_g$  values may have quite different CVG and different rendering on objects (Esposito and Houser 2019). Therefore, an extra parameter is required to characterize such spectra. To prevent bias towards any hue bin, the best-fit ellipse rotation angle  $\psi$ , established by Esposito and Houser (2019), was chosen. Among the 16 TM-30 hue bin boundaries, the angle of  $22.5^\circ$  was selected as it maximizes the number of different  $R_f$  and  $R_g$  combinations. Its orthogonal angle,  $-67.5^\circ$ , was also investigated. Additionally, a high-fidelity spectrum ( $f_{95}g_{100}$ ) with a CVG close to a circle (flattening of the best-fit ellipse equals to 0.03) was tested. Table 1 and Fig. 2 present the properties of these spectra.

[Table 1 and Fig. 2 near here]

### **2.3 Paintings**

The test objects were six oil paintings on canvas by different unknown painters. All paintings had the same theme (landscape) and a similar style (realistic), but their colors varied (Fig. 3). The paintings were named as “one-house,” “red-windmill,” “blue-windmill,” “three-houses,” “black-windmill,” and “forest,” and labeled as W1, W2, W3, M1, M2, and C1, respectively. The first three had warm tones, the last one had a cool tone, and the other two had multi-colors. The dimensions of M2, excluding the frame, were  $78\text{ cm} \times 58\text{ cm}$ , while the other paintings were all  $48\text{ cm} \times 39\text{ cm}$ . The outer region of M2 was covered with mid-gray cloths, so all test paintings had the same visible size. Additionally, one small  $17\text{ cm} \times 27\text{ cm}$  oil painting on wood was used during the practice period. The color distribution of each test painting is shown in Fig. A1 in Appendix A. The average CAM02-UCS colorfulness ( $M'$ ) is 11.6 for all paintings.

[Fig. 3 near here]

## ***2.4 Participants***

Eighteen naïve participants who never saw the paintings before and who wore black clothes participated in the experiment. Their ages ranged from 21 to 38 years, with a median of 28 years. They all had color-normal vision as confirmed by the Ishihara test. They were separated into six groups with a balanced painting order, gender and ethnicity (Appendix B). Four participants had a lighting background, and two of them were color experts. The sample size of 18 was chosen to fulfill a counter-balanced design on spectrum order (multiple of nine), painting order (multiple of six), and fatigue (multiple of two). Although lower than the 30 participants recommended by Royer et al. (2022), the sample size was not increased due to the time duration of the experiment.

## ***2.5 Procedure***

A sequential two-alternative forced choice (2AFC) paired comparison method was used. While categorical judgment or continuous rating scales are common in color rendering studies (Esposito and Houser 2019; Feltrin et al. 2020; Gao et al. 2020; Huang et al. 2018; Jost-Boissard et al. 2016; Mundinger and Houser 2023; Royer et al. 2017, 2018; Wei et al. 2017; Zhai et al. 2015, 2016), they were avoided here as participants, unfamiliar with the paintings, might struggle with the anchoring point of the scale. Some researchers address this by showing extreme lighting spectra initially (Royer et al. 2017, 2018). However, selecting extreme conditions is subjective because the most and least preferred is unknown before their experiment. Thus, following Jost-Boissard et al. (2015), a 2AFC paired comparison was selected.

A 36-pair optimal paired comparison order proposed by Ross (1934) was created based on the nine lighting spectra. The original Ross order was assigned to half of the participants, while the other half received the reversed order counterbalancing the effect

of fatigue. The starting lighting spectrum was also counterbalanced. Consequently, for each painting, each lighting spectrum was assigned to two of the 18 participants: one following the original Ross order and one following the reversed order.

Participants completed the experiment individually on three days within two months. Fig. 4 shows the procedure for one day (upper line) and for one test painting (lower line). Each day began with written instructions under the 3000 K high-fidelity spectrum doubling as the chromatic adaptation light source; the process took ten minutes on the first day and two minutes as a recall on the other days. Then participants practiced the experimental procedure with the small practice painting for at least 10 minutes before judging two test paintings.

Participants evaluated the overall color appearance of each painting in terms of four criteria: preference, saturation, brightness, and normalness (defined in the written instructions) (Appendix C), one criteria at a time. For each lighting pair, each spectrum was applied for two seconds. Participants were then forced to select the most preferred, the most saturated, the brightest, or the most normal appearance. If undecided, they could switch back and forth between the two spectra for an unlimited number of times before making their final judgment. Spectra switched immediately without transition in-between the paired spectra and between pairs. Participants explained their judgment strategies via a questionnaire after every 18 pairs (one session). On average, they took around one hour to complete one test painting. This paper focuses solely on preference; results for the other criteria will be the subject of another paper.

[Fig. 4 near here]

## ***2.6 Questionnaire***

The questionnaire filled out after each preference judgement session is presented in Appendix D. There was no time limit for answering the questionnaire.

## 2.7 Data analysis

Using MATLAB code by Pérez-Ortiz and Mantiuk (2017), the paired comparison results were converted to a one-dimensional interval scale, with values called just-objectionable-differences scores (JODs). Pérez-Ortiz and Mantiuk applied the concept from ISO 20462 and ensured a one-unit distance on the scale yields a 75%:25% proportion in a 2AFC paired comparison (ISO 2005; Keelan and Urabe 2003). The scores are relative and shifted such that the average score across the nine spectra equals zero. When two spectra are equally preferable—each chosen by 50% of participants in a 2AFC paired comparison—their JOD scores are identical, resulting in a distance of zero. As the distance between two spectra increases, a larger proportion of participants prefer the higher-scoring spectrum. For instance, a distance of one unit corresponds to a preference with 75% people for the higher-scoring spectrum. A distance of two units corresponds to a preference with 91% people for the higher-scoring one.

The mathematical expression of the JODs is below. Based on the assumption of a Gaussian distribution for perceived quality of each condition, equation (1) expresses the relation between the proportion  $P$  of two paired conditions and their score distance  $x$ , using a cumulative normal distribution with a scaling factor of 0.6745 for the score distance  $x$ :

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0.6745x} e^{-t^2/2} dt \quad (1)$$

. The difference between scores can derive the proportion  $P$  by (1). For example, if conditions R, S, and T score 0, 0.5, and 1, respectively, the proportion choosing T over R is 75% ( $P = 0.75$  when  $x = 1$ ), while T over S equals S over R at 63% ( $P = 0.63$  when  $x = 0.5$ ). When  $x$  equals two,  $P$  equals 0.91 or 91% as mentioned earlier. The

code uses maximum likelihood estimation to output the estimated JODs of each spectrum that maximizes the probability of observing the collected data.

The statistical differences between each spectrum pair were calculated using the bootstrapping method (embedded in the code) with 3000 simulated samples. This method provided 95% confidence intervals for each score and two-tailed  $p$ -values for each pair. These  $p$ -values determined the statistical differences between paired spectra. For each painting, the JODs of each lighting spectrum were calculated based on the judgements of all participants. Additionally, “the pooled scores” are the JODs derived from the combined judgements across all six test paintings.

Since  $\psi$  is an angle with a period of  $360^\circ$ , sinusoidal regression was used to analyze the relationship between  $\psi_{\text{meas}}$  and the JODs, while Pearson and Spearman correlations examine the relationship between the other color quality indices and the JODs.

A two-way repeated measure ANOVA was implemented by using function *Anova* in the R package “car” and functions *factorial\_design* and *anova\_summary* in the R package “rstatix” (Fox and Weisberg 2018; Kassambara 2023; R Core Team 2023). The test assessed the interaction of lighting spectra and paintings on perceived preference. The ANOVA requires linear preference scores for each spectrum, painting, and participant (9 spectra  $\times$  6 paintings  $\times$  18 participants design); a method by Veitch et al. (2014) derived the linear scores from the paired comparison judgements. The interaction effect investigated whether preferred spectra are painting-dependent. Additionally, Pearson correlations of JOD scores between paintings evaluated the inter-painting preference relationship.

### **3. Quantitate results: JOD scores (JODs)**

Based on the JOD preference scores (see details in subsection 3.2), test paintings were

grouped into two groups: Group 1 (Gp1) comprising W1, W2, and M1, and Group 2 (Gp2) comprising W3, M2, and C1. Group scores were then estimated by pooling the data from paintings within each group.

### ***3.1 JODs and statistical significance of the pooled data***

Table 2 presents the pooled preference JOD scores for each spectrum, as well as the statistical significant difference based on the bootstrapping method. It also includes measured TM-30 indices, corresponding color preference levels, and an adjusted level proposed in subsection 5.2. Fig. 5 illustrates the estimated JODs by spectrum, with error bars representing 95% confidence intervals of the pooled scores. It also displays JODs of Gp1 and Gp2. The JODs by painting are shown graphically in Fig. A2 in Appendix A.

[Table 2 and Fig. 5 near here]

### ***3.2 The effect of paintings***

To assess consistency among landscape oil paintings, a two-way repeated measures ANOVA was conducted. The analysis revealed a significant interaction between spectra and paintings ( $p < 0.05$ ), indicating that spectrum preference depends on the painting (Table E1 in Appendix E). The interaction has been visualized in Fig. 5.

Table 3 presents the Pearson correlation coefficients between the JODs of each painting. Based on these correlations and Fig. 5, two painting groups arose: W1, W2, and M1 in Gp1; W3, M2, and C1 in Gp2. Subsequent repeated measures ANOVAs of linear scores within each group revealed no significant painting-spectrum interaction, confirming the validity of these groupings (Table E2 in Appendix E).

[Table 3 near here]

Pooled data from paintings in each group were used to estimate preference JODs of each group. Fig. 5 and Tables 4 and 5 present the scores along with the statistical differences between spectra within each group. Tables 4 and 5 also include the TM-30 preference levels and the adjusted preference levels proposed in Table 2. The proposed levels classify the preferences for each group better than the TM-30 Annex E levels. The classification is generally accurately, with minor exceptions: (1) Gp1: No perceived difference between P'2 and P'3. (2) Gp2: No perceived difference between P'2 and P'1. (Note: P'2 and P'3 were considered as different since P'2 still significantly differed from several P'3 spectra).

[Table 4 and Table 5 near here]

### ***3.3 Correlation between JODs and color quality indices***

Correlations between JODs and lighting indices ( $R_f$ ,  $R_g$ , CRI  $R_a$ ,  $R_{cs,h1}$ , and  $R_{f,h1}$ ), along with the sinusoidal regression to  $\psi$ , are presented in the Appendices (see Fig. A3, Fig. A4, and Appendix F). For Gp1-painting, JODs significantly relate to  $\psi$  and  $R_{cs,h1}$  ( $p < 0.05$ ). For Gp2-painting, JODs significantly correlate to  $R_g$  ( $p < 0.05$ ), while in Gp1, only M1 exhibits the correlation. In summary, Gp1 was named the “high- $R_{cs,h1}$ -preferred group” because for every painting in this group, higher  $R_{cs,h1}$  and  $\psi$  closer to  $5^\circ$  are preferred; Gp2 was named the “high- $R_g$ -preferred group” because for every painting in this group, higher  $R_g$  is preferred.

## **4. Qualitative results: answers to the questionnaires**

Despite the metameric background colors, participants generally described the lighting as “reddish” vs. “greenish,” or as “warm” vs. “cool” in their questionnaire answers, based on the appearance of the paintings. Answers to questions 4 and 5 showed no common trend among participants and are therefore not reported in this paper; the

answers to remaining questions are provided below.

#### ***4.1 Question 1: judgement criteria***

Question 1 asked participants to explain how they judged and what influenced their choices. Table G1 in Appendix G lists the frequency of commonly reported terms in the 108 (18 × 6) answers. As shown in the last row, participants often prefer a colorful appearance over high fidelity. The most frequent reasons included: “(more) red/orange (but not too much)” or “saturated/vivid/colorful/... (but not too much).” However, the reasons varied by the painting. “Red,” “orange,” “warm,” or “red/orange but not too much” were predominant for Gp1-paintings. Preference for a greener appearance was reported only for W3 and C1.

#### ***4.2 Question 2: area(s) of greatest impact***

In Question 2, participants marked the area(s) that had the greatest impact on their preference evaluations. Table G2 in Appendix G presents the top six selected objects for each painting, their colors, and the selection frequency. Notably, M1 yielded interesting results: Despite predominant green areas (Fig. 3 and A1), the three red houses in the center had the greatest impact on the evaluation for all participants and were the only objects unanimously marked by all participants.

#### ***4.3 Question 3: contradictions between different parts of the painting***

In Question 3 of Session 2, participants reported any contradictions in preference evaluation between different parts of the painting and their decision-making process if contradictions occurred. Their answers indicated that contradictions were uncommon, with only 35 out of the 108 (18 × 6) answers reporting contradictions. The answers revealed that some participants noticed orange and green were often oppositely

enhanced when comparing two spectra: orange objects (including red and brown) and green objects were most frequently reported. One participant noted: “flashy orange leaves vs. flashy green grass were both nice”.

## **5. Discussion**

The quantitative preference JOD scores and their correlation with the lighting color quality indices suggest the six test paintings can be divided into two groups with three paintings in each group: the high- $R_{cs,h1}$ -preferred group Gp1 (W1, W2, M1) and the high- $R_g$ -preferred group Gp2 (W3, M2, C1). Based on the statistical significance of the differences between spectra, adjusted preference levels derived from TM-30 Annex E were proposed. This section explores the properties and possible reasons for the two groups in subsection 5.1, investigates the adjusted preference levels in subsection 5.2, and inspects the limitation and future directions of this study in subsection 5.3.

### ***5.1 Properties and possible reasons for the two preference groups***

Bellia et al. (2020) found that the chromatic composition of a painting influences the effect of CCT on its perception. Results in subsection 3.2 similarly indicates that preferred spectra depend on the painting, allowing the six test paintings to be grouped in two: the high- $R_{cs,h1}$ -preferred group Gp1 and the high- $R_g$ -preferred group Gp2. Within each group, JODs are strongly correlated and not significantly affected by specific paintings, while rankings differ between groups.

Identifying a method to classify paintings properly into these groups could enable better preference prediction through two distinct guidelines, rather than one general guideline. The approach of adapting lighting to painting content aligns with the proposal by Okuda et al. (2023): “paintings of landscapes in evening and autumn should be illuminated with low CCT, and the opposite being true for landscape in summer or

winter.” Based on the result of this study, high  $R_{cs,h1}$  and high  $R_g$  can serve as a guideline for Gp1 and Gp2, respectively.

The key content factors differentiating the two groups remain to be determined. Several studies have examined lighting color quality with “multiple” paintings (three or more) (Chen et al. 2016; Feltrin et al. 2020; Nascimento and Masuda 2014; Nascimento et al. 2021; Okuda et al. 2023; Zhai et al. 2015, 2016). However, only two compared results across different paintings. Feltrin et al. (2020) found that paintings with different predominant colors significantly impacted background appreciation, warmth, and vividness. Okuda et al. (2023) observed that lighting with higher CCTs was favorable, particularly for paintings containing much green. Possible options to classify the paintings in this study are discussed below.

The overall painting tone does not provide clear hints, as warm-tone paintings (W1, W2, W3) are distributed across groups. Similarly, the quantitative amount of colors in each hue bin does not appear to be a determining factor, because paintings with more green colors (bins 5 and 6) than yellow (bin 4) (M1 and C1), and paintings with some cyan colors (bins 8 to 10) (M1 and M2) are also separated between groups.

Qualitative answers from participants suggest potential factors distinguishing the two groups (Table G2):

- (1) Attention drawn by reddish and orange objects: In M1, the houses, particularly their red roofs, had the most dominant impact on preference evaluation, despite larger green and blue areas in this painting. Similarly, in W1 and W2 (also high- $R_{cs,h1}$ -preferred paintings), orange and brown objects attracted the most attention. The focus on red and orange objects may explain the preference for red-enhanced color. However, for W3, a high- $R_g$ -preferred painting, a

substantial attention to orange and brown objects was also noted by the participants.

- (2) Attention to green objects: All Gp2-paintings received equal or more than half of the attention on green objects, while all Gp1-paintings did not. As described in subsection 4.3, several participants experienced contradictory preferences between red and green objects when comparing some lighting spectra. Consequently, in Gp2-paintings, where participants' preference were split between more saturated red and green, the global  $R_g$  might better explain the preference, leading to a preference on high- $R_g$  spectra.
- (3) Presence and attention to strictly red objects: Red objects (distinct from orange or brown) were present and received attention in every Gp1-painting but not in Gp2-painting. They are the windmill top (W2) and roofs (W1 and M1). The presence of red may have led to a preference for red-enhanced high- $R_{cs,h1}$  spectra.

Based on the results, three suggestions for landscape oil paintings are proposed to lighting designers:

- For paintings that have eye-catching red objects or dominant orange objects in the scene, high- $R_{cs,h1}$  luminaires are suggested to give an appreciated color appearance.
- For paintings that have eye-catching green objects in the scene, high- $R_g$  luminaires are suggested to give an appreciated color appearance.
- General guidelines are suggested for paintings that either meet both conditions or neither condition.

## 5.2 Comparison and adjustment to IES TM-30-20 Annex E

Table 2 lists the color preference levels of lighting spectra as defined in IES TM-30-20. While these levels broadly align with the pooled data from the six test paintings, some discrepancies exist. One P- spectrum ( $f_{65g90}^+$ ) achieves comparable favorability to a P3 spectrum ( $f_{85g100}^-$ ), and a P2 spectrum ( $f_{75g110}^+$ ) is preferred over P1 spectra ( $f_{85g100}^+$  and  $f_{85g110}^+$ ). Referring to the methods applied by Royer et al. (2020), statistically significant differences between lighting spectra were used to refine the criteria. The adjusted levels (P'1, P'2, P'3, and P'-) are presented in Table 2.

The adjusted level aligned with the result better, justifying a definition of the criteria. The TM-30 criteria for  $R_f$  were lowered to adjust the levels of  $f_{65g90}^+$  and  $f_{75g110}^+$ . Two potential approaches to distinguish P'2 and P'1 emerged:

- (1) Setting an  $R_g$  lower threshold at 105
- (2) Setting an  $R_{cs,h1}$  threshold as follows:  $5\% \leq R_{cs,h1} \leq 11\%$ , extrapolated from the original TM-30 Annex E threshold.

The current data cannot conclusively determine which approach is superior, as each index appears to be a suitable criterion for its respective painting group. Given that the adjusted  $R_{cs,h1}$  criterion is an extrapolation while  $R_g$  is not, the second approach was selected. Consequently, a set of new criteria with minimal deviation from TM-30-20 is proposed:

- P'1:  $R_f \geq 77$ ,  $R_g \geq 95$ ,  $5\% \leq R_{cs,h1} \leq 11\%$  (new level)
- P'2:  $R_f \geq 77$ ,  $R_g \geq 95$ ,  $-1\% \leq R_{cs,h1} \leq 15\%$  (original P1 with an adjusted lower  $R_f$ )
- P'3:  $R_f \geq 65$ ,  $R_g \geq 89$ ,  $-12\% \leq R_{cs,h1} \leq 23\%$  (original P3 with an adjusted lower  $R_f$ )

It should be noted that the two adjusted  $R_f$  thresholds could be further lowered while maintaining coherence with the results. The values 65 and 77 were chosen to minimize the difference with the original TM-30-20 criteria. Future studies incorporating a boarder range of lighting spectra could determine more optimal values.

The original TM-30 P2 and P3 levels are combined as one P'3 level, possibly due to the limited number of participants failing to establish a statistically significant difference between P2 and P3. Within the P'3 group, Spectra  $f_{75}g_{90}^+$  and  $f_{75}g_{100}^+$  differentiate P2 and P3, with a JOD gap of 0.14—the largest among all neighboring gaps ( $f_{65}g_{90}^+/f_{85}g_{100}^-$ ,  $f_{85}g_{100}^-/f_{75}g_{90}^+$ , etc.). If the study had larger sample size, this gap might have become statistically significant, potentially establishing a new P'4 level separated from P'3 with the TM-30 P2 criteria. In other words, a new P'3 level (using TM-30 P2 criterion:  $R_f \geq 75$ ,  $R_g \geq 92$ ,  $-7\% \leq R_{cs,h1} \leq 19\%$ ) and a new P'4 level (using the P'3 criterion above:  $R_f \geq 65$ ,  $R_g \geq 89$ ,  $-12\% \leq R_{cs,h1} \leq 23\%$ ) could be proposed.

The adjustment to lower  $R_f$  can be explained by the influence of memory color. Among the three indices involved in the preference judgment,  $R_f$  measures how close the color under the test source matches the color under a reference illuminant. Even though preferred object color may differ from memory color (Bartleson and Bray 1962), considerable deviations from memory color in familiar objects often leads to a dislike. All five experiments acting as the primary data for TM-30-20 Annex E included familiar objects (Esposito and Houser 2019; IES 2024; Royer et al. 2017, 2018, 2020; Zhang et al. 2017), suggesting potential influence from memory. In contrast, participants in this study were unfamiliar with the test paintings and lacked a reference for the paintings' appearance under a reference illuminant. As a result, memory color likely does not play a role in this experiment. In other words, for art, especially lesser-

known paintings, memory colors are likely less relevant for preference. Therefore, the effect of  $R_f$  is reduced, resulting in less strict criterion and a lower threshold.

### ***5.3 Limitations of this study and future research directions***

The study has limitations related to lighting conditions, painting types, and wall color, which points towards future research. The suggestions given in subsections 5.1 and 5.2 are based on the tested conditions and may not apply when conditions deviate considerably from the tested range.

An illuminance of 150 lux and a CCT at 3000 K were selected to meet the requirements of the collaborating fine art museum. Loe et al. (1982) found that 200 lux is required for visual satisfaction, a value also used in other studies (Bellia et al. 2020; Chen et al. 2016; Nascimento and Masuda 2014; Zhai et al. 2015, 2016) and this value is recommended as the standard for oil paintings (IES 2020). For paintings made with high or medium light sensitive materials (*e.g.*, watercolors, pastels), illuminance is limited to 50 lux (IES 2020). Therefore, we encourage future research to investigate other illuminance levels commonly used in museums. The choice of CCT often depends on lighting design experts and may differ from 3000 K. Future studies could explore a broader range of CCTs. The test conditions are in the range of  $R_f \geq 75$  (with one exception  $f_{65g90^+}$ ),  $90 \leq R_g \leq 110$ , and  $-19\% \leq R_{cs,h1} \leq 9\%$ . If luminaire spectra differ substantially from this range, the suggested high- $R_g$  and high- $R_{cs,h1}$  preference for each group may not hold due to oversaturation effect observed in other studies (Mundinger and Houser 2023; Wei et al. 2017).

As mentioned in subsection 5.2, another future research direction involves applying more lighting spectra to refine the criteria for the adjusted preference levels. However, due to limitations of the luminaire, the range of  $R_f$  and  $R_g$  could not be expanded at the 150 lux and 3000 K lighting condition. When conducting color

rendering experiments, real luminaires are subject to spectral limitations, which restricts the test range. If the reflectance of paintings is accurately obtained (for example with an hyperspectral camera), it is possible to simulate the color appearance of paintings under virtual illuminants. This simulation could enable visual experiments on a calibrated display, provided that the simulated colors remain within the display's gamut and luminance range. This method broadens the test range. Once the most preferred condition is identified through display experiments, results could be validated in a real scene by redesigning the luminaires. Although this idea requires advanced techniques (*e.g.*, hyperspectral imaging, spectrum simulation, display calibration, luminaire design), it presents a promising avenue for future research on optimal spectra for paintings.

The findings are limited to oil paintings with a mid-gray background color. The landscape oil painting was chosen as an initial step to study how painting color composition affects preference. However, these results may not apply to other painting types. Future research could explore other common subjects (*e.g.*, portraits) and materials (*e.g.*, watercolors) to broaden applicability. Since background color influences the painting perception (Bellia et al. 2020; Feltrin et al. 2020), these results may not hold if the frame color or the wall color differs seriously from mid-gray.

Even when applying these findings to similar conditions (lighting, painting type, and background color), caution is advised against treating these suggestions as golden rules. Royer et al. (2022) emphasize that “any one set of objects is insufficient to make generalized color rendition recommendations for a specific application.” Further research with varied setups and paintings is necessary to verify whether the suggestions in subsection 5.1 and the adjusted levels in subsection 5.2 can be generalized for landscape oil paintings.

Moreover, even if these suggestions and adjusted levels prove broadly applicable, the museum lighting standard IES RP-30-20 recommends using them to narrow the selection of light sources rather than for making definitive choice (IES 2020). The standard recommends that final choice should be based on effectively illuminating the target paintings, a process where the lighting design expertise plays a crucial role (IES 2020).

## 6. Conclusion

This study aimed to verify whether the TM-30 Annex E color preference criteria apply to landscape oil paintings and whether the results depend on painting color compositions. To achieve this, six landscape oil paintings with varying color composition were illuminated under nine 3000 K metameric spectra, systematically constructed with different  $R_f$ ,  $R_g$ , and  $\psi$  values. Using a 2AFC paired comparison method, 18 participants evaluated the color appearance of the paintings in terms of preference. The evaluation results were converted into quantitative interval JOD scores, while qualitative reasons for participants' judgments were collected through questionnaires.

Based on the statistically significant difference in pooled JOD scores across spectra, the TM-30 Annex E criteria (P1, P2, and P3) were adjusted to P'1, P'2, and P'3 as follows:

- P'1 criterion:  $R_f \geq 77$ ,  $R_g \geq 95$ ,  $5\% \leq R_{cs,h1} \leq 11\%$
- P'2 criterion:  $R_f \geq 77$ ,  $R_g \geq 95$ ,  $-1\% \leq R_{cs,h1} \leq 15\%$
- P'3 criterion:  $R_f \geq 65$ ,  $R_g \geq 89$ ,  $-12\% \leq R_{cs,h1} \leq 23\%$

If this study had included a larger number of participants, some JOD scores differences might have reached statistical significance. In that case, the TM-30 P2

criterion ( $R_f \geq 75$ ,  $R_g \geq 92$ ,  $-7\% \leq R_{cs,h1} \leq 19\%$ ) could have been used to divide the P'3 level into separate P'3 and P'4 levels.

The adjustment of criteria P'2 and P'3 from P1 and P3—particularly the lowering of  $R_f$  threshold—suggests that a less strict fidelity requirement is suitable when evaluating lighting preference of paintings. This implies that fidelity is less critical in painting lighting design compared to its importance in TM-30-20, which is intended for general interior applications. This phenomenon may be attributed to the absence of memory color for unfamiliar paintings. Expanding the range of lighting spectra in future studies could further refine the  $R_f$  criteria.

The JOD scores for each painting under each spectrum illustrate inconsistencies among paintings, revealing two painting categories: a “high- $R_{cs,h1}$ -preferred group” and a “high- $R_g$ -preferred group.” While using general guidelines for all paintings is acceptable, adapting guidelines to each group yields better results.  $R_{cs,h1}$  is a reliable predictor of preference in the first group, while  $R_g$  serves as a strong predictor in the second group. Based on qualitative answers from the questionnaire, the following suggestions help narrow the choice of light sources for landscape oil paintings:

- For paintings that have eye-catching red objects or dominant orange objects in the scene, high- $R_{cs,h1}$  luminaires are suggested to give an appreciated color appearance.
- For paintings that have eye-catching green objects in the scene, high- $R_g$  luminaires are suggested to give an appreciated color appearance.
- General guidelines such as the adjusted criteria (P'1, P'2, and P'3) are suggested for paintings that either meet both conditions or neither condition.

The first and second suggestions are based on the evidence within the tested range (CCT  $\approx$  3000 K, illuminance  $\approx$  150 lux, mid-gray background,  $R_f \geq 75$ ,  $90 \leq R_g \leq 110$ ,  $-19\% \leq R_{cs,h1} \leq 9\%$ ) and may not hold for a lighting spectrum outside this range, where oversaturation effects may need to be considered. Moreover, as stated in the museum lighting standard IES RP-30-20 (IES 2020), after narrowing options through these suggestions, it is recommended to finalize the choice by effectively illuminating the target painting and allowing lighting design expertise to play a significant role.

Future research can focus on broadening both the lighting range and painting types to provide more comprehensive guidance for museum lighting designers. To extend beyond the spectral limitations of real luminaires, advanced techniques such as hyperspectral imaging, spectrum simulation, and display calibration techniques could be employed to conduct experiments on displays in future studies.

### **Ethics Declarations**

This work complies with the principles and regulations of the European Union (GDPR) and the national legislation regarding data protection of personal data. The project is approved by the Social and Societal Ethics Committee of KU Leuven with reference number G-2021-4302-R2. All participants read and signed an informed consent before doing the experiment. (During the peer review period, part of the information is removed from the anonymous manuscript.)

### **Funding**

This work was supported by the scholarship of the primary author and the starting grant of the last author. (During the peer review period, detailed information of the funding is on the title page of the manuscript with author details.)

### **Disclosure statement**

The authors report there are no competing interests to declare.

## Appendix A. Supplementary figures

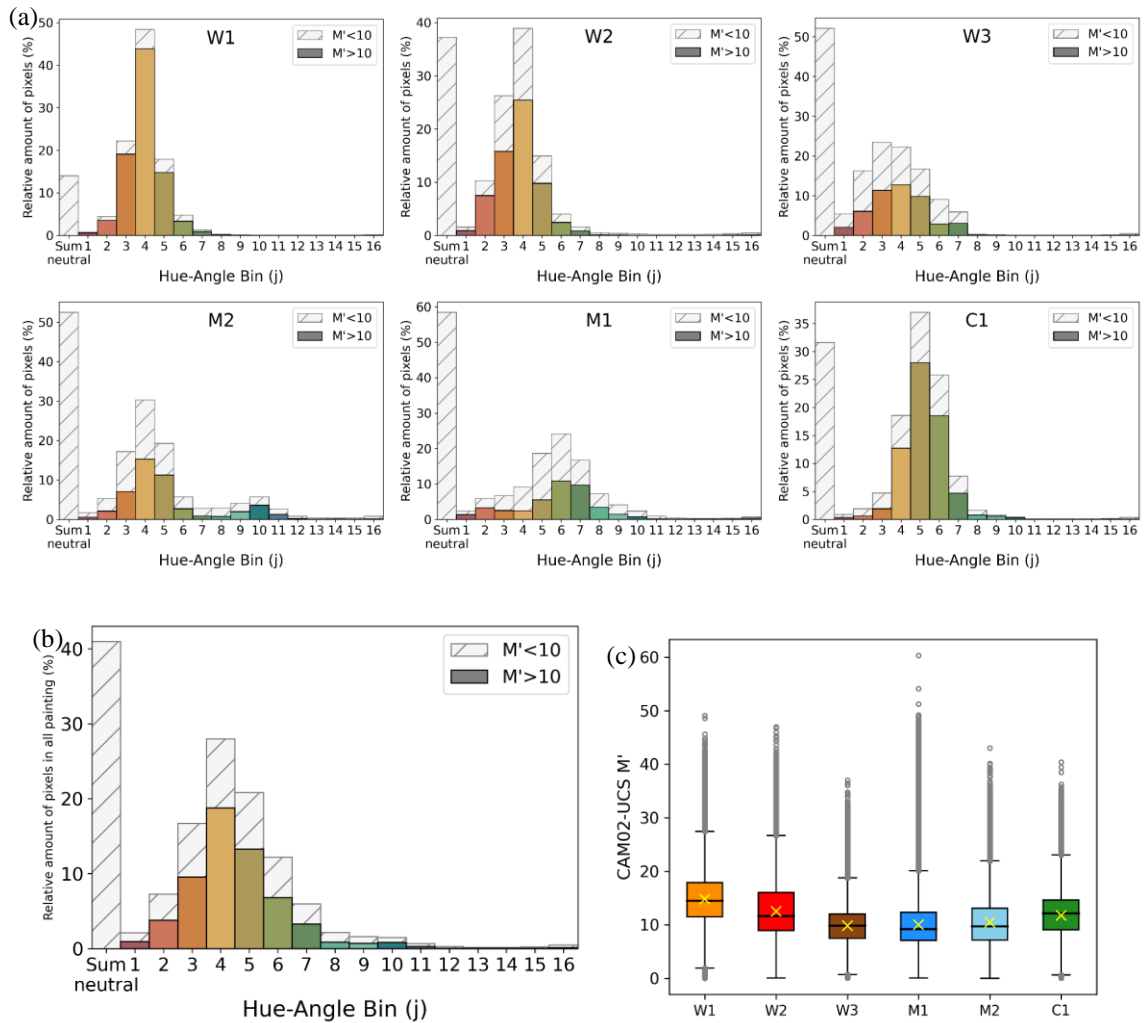


Fig. A1. Color distribution of the test oil paintings under the high-fidelity spectrum ( $f_{95g100}$ ). The histograms (a) and (b) show the percentage of pixels in each hue bin from tristimulus images captured by a colorimetric camera (LMK 6-5 color – TechnoTeam). (a): each painting; (b): sum of all paintings. The boxplot (c) shows the CAM02-UCS colorfulness ( $M'$ ) distribution of each painting, and yellow markers represent the average  $M'$  of each painting. The average  $M'$  for the six paintings is 11.6. When converting CIE 1931  $2^\circ$  XYZ values to CAM02-UCS space, the reference white point was defined by the XYZ values of the white patch on a Macbeth ColorChecker under  $f_{95g100}$ . Pixels with  $M' < 10$  in the CAM02-UCS a-b plane were considered near-neutral and are presented separately in (a) and (b). The sum of near-neutral pixels across all 16 hue bins is shown in the first column. The threshold value 10 was chosen subjectively based on visual inspection of neutral pixels in the paintings.

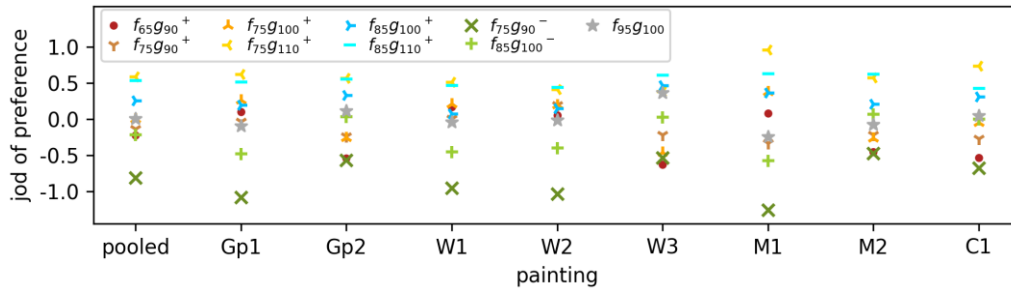


Fig. A2. Estimated preference JODs of each lighting spectrum over different paintings.

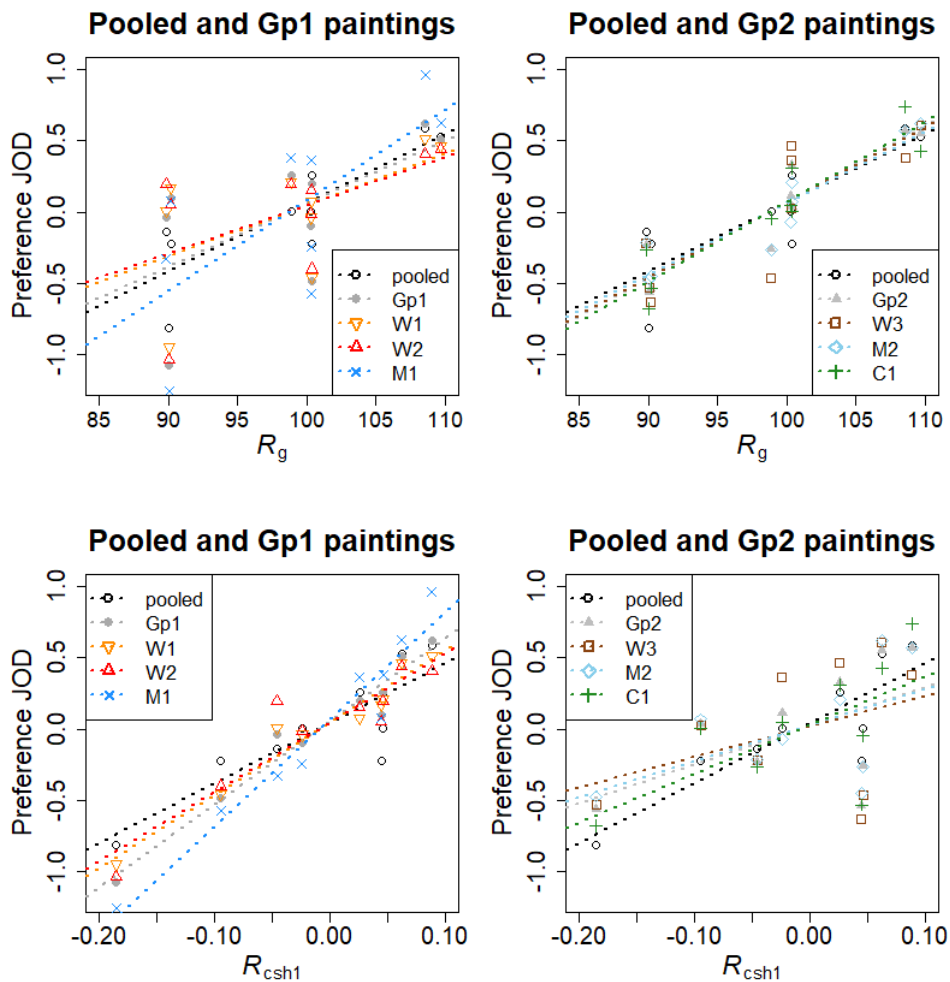


Fig. A3. JODs and their linear regression results with  $R_g$  and  $R_{cs,h1}$ . The markers represent JODs of individual painting, grouped painting, and pooled scores. The lines indicate linear regression results.

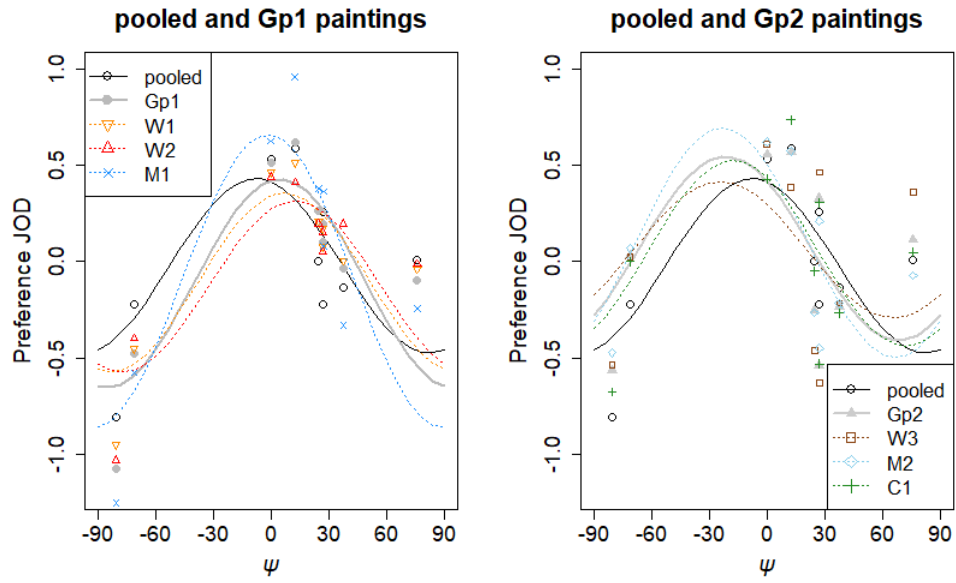


Fig. A4. JODs and the sinusoidal regression results. The markers represent JODs of individual painting, grouped painting, and pooled scores. The lines indicate sinusoidal regression results.

## Appendix B. Participants information and group

Table 2. Counter-balanced Latin square design of the painting order, with the participants' information

Group	Painting presented						Age mean (year)	Participants distribution			
	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	6 <sup>th</sup>		Gender		Ethnicity	
								Male	Female	Asian	European
A	W1	C1	M1	M2	W2	W3	26.7	2	1	1	2
B	M1	W1	W2	C1	W3	M2	27.3	1	2	1	2
C	W2	M1	W3	W1	M2	C1	28.0	2	1	1	2
D	W3	W2	M2	M1	C1	W1	29.3	2	1	2	1
E	M2	W3	C1	W2	W1	M1	30.0	1	2	2	1
F	C1	M2	W1	W3	M1	W2	29.7	2	1	2	1

## Appendix C. Definition of the criteria

1. **Preference:** you like something more than another thing
2. **Saturation:** the colors appear to be more pure or vivid
3. **Brightness:** a lot of light
4. **Normalness:** ordinary or usual; as you would expect an oil painting to look like  
(To help you understand saturation and brightness, please see a printed example)

The example is a paper containing five Munsell charts (5R, 5Y, 5G, 5B, and 5P).

## Appendix D. Questionnaire

Questionnaire after Session 1 (after the 18<sup>th</sup> pair of preference)

1	Explain how you judged in the session you just finished? What made you prefer one appearance to the other? (The answer could be different by the pairs)
2 <sup>a</sup>	On the image, mark the area(s) that had the greatest impact on your preference evaluation in the session. Could you explain how these areas affect you? Why do they have the greatest impact?
3	Do you have further information to say? For example, was the task easy or difficult?

<sup>a</sup> A grayscale image of the painting was shown.

Questionnaire after Session 2 (after the 36<sup>th</sup> pair of preference)

1	Explain how you judged in the session you just finished? What made you prefer one appearance to the other? (The answer could be different by the pairs)
2 <sup>a</sup>	On the image, mark the area(s) that had the greatest impact on your preference evaluation in the session. Could you explain how these areas affect you? Why do they have the greatest impact?
3	In all sessions of this painting, have you ever felt contradictions between different parts of the painting for some pairs? In other words, you like one part of the painting more under X condition, and you like other parts more under Y condition. If yes, what's your strategy to make a judgement for the whole painting?
4	Do you feel that your strategy for making a judgement changes from time to time, from session to session? If yes, explain the changes.
5	Do you have further information to say? For example, was the task easy or difficult?

<sup>a</sup> A grayscale image of the painting was shown.

## Appendix E. Repeated measure ANOVA results

Table E1. Summary of the two-way repeated measures ANOVA between spectra and paintings

Effect	DFn	DFd	F	<i>p</i>
spectrum	8	136	11.6	< 0.001
painting	5	85	0	1
spectrum : painting	40	680	1.57	0.015

Mauchly's Test for Sphericity			Sphericity Corrections						
Effect	W	<i>p</i>	Effect	Greenhouse-Geisser			Huynh-Feldt		
				GGe	DF	<i>p</i>	HFe	DF	<i>p</i>
spectrum	1.74e-05	< 0.001	spectrum	0.222	1.78	< 0.001	0.246	1.97	< 0.001

DF: Degrees of freedom DFn: Degrees of freedom for the numerator DFd: Degree of freedom for the denominator  
W: Mauchly's W GGe: Greenhouse-Geisser estimate of sphericity HFe: Huynh-Feldt estimate of sphericity

Table E2. Summary of the two-way repeated measures ANOVA between spectra and paintings in each group

Gp1					Gp2				
Effect	DFn	DFd	F	<i>p</i>	Effect	DFn	DFd	F	<i>p</i>
spectrum	8	136	10.2	< 0.001	spectrum	8	136	7.07	< 0.001
painting	2	34	0	1	painting	2	34	0	1
spectrum : painting	16	272	0.69	0.804	spectrum : painting	16	272	0.55	0.916

Mauchly's Test for Sphericity (Gp1)			Sphericity Corrections (Gp1)						
Effect	W	<i>p</i>	Effect	Greenhouse-Geisser			Huynh-Feldt		
				GGe	DF	<i>p</i>	HFe	DF	<i>p</i>
spectrum	2.40e-05	< 0.001	spectrum	0.227	1.82	< 0.001	0.253	2.02	< 0.001
spectrum : painting	1.30e-11	< 0.001	spectrum : painting	0.198	3.17	0.571	0.249	3.99	0.602

Mauchly's Test for Sphericity (Gp2)			Sphericity Corrections (Gp2)						
Effect	W	<i>p</i>	Effect	Greenhouse-Geisser			Huynh-Feldt		
				GGe	DF	<i>p</i>	HFe	DF	<i>p</i>
spectrum	5.04e-05	< 0.001	spectrum	0.218	1.74	0.004	0.240	1.92	0.003
spectrum : painting	9.51e-13	< 0.001	spectrum : painting	0.206	3.30	0.664	0.262	4.19	0.705

DF: Degrees of freedom DFn: Degrees of freedom for the numerator DFd: Degree of freedom for the denominator  
W: Mauchly's W GGe: Greenhouse-Geisser estimate of sphericity HFe: Huynh-Feldt estimate of sphericity

## Appendix F. Relationship between JODs and lighting indices

Table F1. The correlation coefficient between the JODs and the measured color quality indices

Index	correlation	Pooled	Gp1	Gp2	W1	W2	M1	W3	M2	C1
$R_a$	Pearson	0.16	-0.09	0.42	-0.11	0.01	-0.15	<b>0.62</b>	0.28	0.29
	Spearman	0.22	-0.30	0.35	-0.40	-0.20	-0.23	0.53	0.40	0.33
$R_f$	Pearson	0.32	0.03	<b>0.61</b>	-0.01	0.08	0.01	<b>0.78</b>	0.49	0.48
	Spearman	<b>0.60</b>	0.05	<b>0.73</b>	-0.05	0.10	0.13	<b>0.85</b>	<b>0.73</b>	<b>0.70</b>
$R_g$	Pearson	<b>0.85</b>	<b>0.64</b>	<b>0.93</b>	<b>0.59</b>	0.55	<b>0.71</b>	<b>0.83</b>	<b>0.93</b>	<b>0.93</b>
	Spearman	<b>0.73</b>	0.55	<b>0.83</b>	0.50	0.45	<b>0.62</b>	<b>0.80</b>	<b>0.88</b>	<b>0.88</b>
$R_{cs,h1}$	Pearson	<b>0.86</b>	<b>0.98</b>	0.55	<b>0.98</b>	<b>0.93</b>	<b>0.97</b>	0.40	0.55	<b>0.66</b>
	Spearman	<b>0.75</b>	<b>0.97</b>	0.58	<b>0.98</b>	<b>0.87</b>	<b>0.98</b>	0.42	0.53	<b>0.65</b>
$R_{f,h1}$	Pearson	0.58	<b>0.58</b>	0.46	<b>0.58</b>	<b>0.68</b>	0.49	0.53	0.31	0.46
	Spearman	0.42	0.15	0.32	0.07	0.20	0.22	0.45	0.30	0.35

**bold and underline:** significance level  $p < 0.05$ , **bold:**  $0.05 < p < 0.1$

Table F2. Sinusoidal regression  $R^2$  between  $\psi_{\text{meas}}$  and JODs of pooled, grouped, and individual painting

Index	Pooled	Gp1	Gp2	W1	W2	M1	W3	M2	C1
$\psi_{\text{meas}}$	<b>0.55</b>	<b>0.77</b>	0.29	<b>0.75</b>	<b>0.71</b>	<b>0.76</b>	0.12	0.48	0.34

**bold and underline:** significance level  $p < 0.05$ , **bold:**  $0.05 < p < 0.1$

## Appendix G. Answers to the questionnaires

Table G1. Frequency of terms reported as preferred over 18 participants (question 1)

Group	1A: colorful (general)			1B: colorful (specific color)					2: fidelity					3: others			
	saturated, vivid, colorful, intense, vibrant, pure <sup>d</sup>	colorful/saturated but not too much	distinct, more color difference, high contrast	warm	(more) red/orange	red/orange but not too much	(more) green	(more) blue	real, realistic	normal	natural	neutral	X stays X <sup>e</sup>	balanced, not too pale nor too vivid	prettier, aesthetically pleasing	bright, more light, not too dark	soft, mild, gentle
W1	1			1	4	3			2		1		1	1	2	2	1
W2	1		1	1	7	4						2	1		2		1
M1	5	2	1	2	5	1		3							1	1	1
W3	1		2	1	3	1	2	1	2	1		1	2	1	2	2	1
M2	5	1	1	1	2	2		2	3	1	1		3		1	3	1
C1	5		3	1	2	2	3		1		1	1	1		1	1	1
Sum <sup>a</sup>	17	4	8	7	23	13	5	6	8	2	3	4	8	2	9	9	6
Freq. <sup>b</sup>	26/108			54/108					24/108					NA			
Freq. <sup>c</sup>	72/108																

<sup>a</sup> Sum of all paintings.

<sup>b</sup> Frequency of each sub-group, except for group 3. If a participant reported different terms within the same group for the same painting, these terms were counted as one. Therefore, the number may be less than the sum of values in the row immediately above. The sum of group 3 is meaningless as there is no shared meaning among the terms.

<sup>c</sup> Frequency of each group. If a participant reported different terms for a painting in the same sub-group, these terms were counted as one. Thus, the number 72 is less than the sum of 26 and 54.

<sup>d</sup> Refers to general colors. Terms indicating 'more', such as 'saturated orange' and 'vibrant blue', are excluded but counted in group 1B: colorful (specific color).

<sup>e</sup> X refers to a specific color. Frequency of each color: white (3), brown (1), yellow (1), green (3), blue (1), gray (1). Two participants reported two colors for one painting, so the total frequency across all colors in this note is 10, which is larger than the number 8 shown in the table.

Table G2. Most frequently marked objects as having the greatest impact on the preference evaluation. Their color and frequency are listed, with participant counts out of 18, and session counts in brackets (out of 36). <sup>a</sup>

		1 <sup>st</sup> often marked		2 <sup>nd</sup> often marked		3 <sup>rd</sup> often marked		4 <sup>th</sup> often marked		5 <sup>th</sup> often marked		6 <sup>th</sup> often marked	
		color	freq.	color	freq.	color	freq.	color	freq.	Color	freq.	color	freq.
high- $R_{\text{c,hd}}$ -preferred group (Gp1)	W1	sunset glow		tree on the left of the central trees		big trees at the center		wall of the house		roof of the house		grass	
		orange	14/18 (28/36)	vivid orange	14/18 (27/36)	orange, brown	9/18 (17/36)	white	4/18 (8/36)	red	4/18 (7/36)	green	3/18 (6/36)
	W2	leaves of the central tree		sunset glow		trees on the right		yellow field		windmill		green grass	
		brown	11/18 (19/36)	orange	10/18 (18/36)	orange	10/18 (16/36)	yellow	8/18 (16/36)	red, white	7/18 (11/36)	green	6/18 (11/36)
	M1	houses (roofs) <sup>b</sup>		lake		swan		soil field		cloud		trees	
		red	18/18 (35/36)	cyan	9/18 (15/36)	white	7/18 (12/36)	orange	7/18 (12/36)	white	6/18 (10/36)	green	6/18 <sup>c</sup> (9/36)
high- $R_{\text{g}}$ -preferred group (Gp2)	W3	roof of the house		<u>trees next to the house</u>		<u>light leaves of the trees by the road</u>		blue sky		roof of the windmill		grass	
		brown	12/18 (22/36)	light brown	12/18 (21/36)	orange	12/18 (21/36)	cyan	10/18 (18/36)	blue	9/18 (18/36)	green	9/18 (17/36)
	M2	grass		cloud		<u>roof of houses</u>		<u>body of the windmill</u>		donkey, cart, and the man nearby <sup>d</sup>		blue sky	
		green	11/18 (20/36)	white, gray	10/18 (19/36)	orange	9/18 (18/36)	white, gray	9/18 (18/36)	white, gray, cyan, brown	8/18 (14/36)	cyan	4/18 (8/36)
	C1	crepuscular ray		leaves of central trees		blue waters		yellow ground		grass		mountains behind trees	
		yellow	14/18 (26/36)	green, a bit brown	13/18 (22/36)	cyan	7/18 (12/36)	yellow	6/18 (9/36)	green	4/18 (6/36)	cyan	3/18 (4/36)

<sup>a</sup> Underline: Objects with identical frequencies of participants and sessions should be considered as the same rank. In the table, these objects are ordered based on their position in the scene (left-most object first).

<sup>b</sup> Thirteen participants specifically marked the roofs, while five participants circled the entire houses.

<sup>c</sup> The two frequency numbers remain identical even when counting all green objects: the left tree, two large trees in the foreground, trees on the far right, the green door, green windows, and the grass. It is because at least one tree was marked in the sessions when the door, windows, or grass were also marked.

<sup>d</sup> Seven participants marked these objects within a single circle despite their distinct colors, making individual identification impossible. Only one participant marked these objects separately, specifically marked the jacket and the brown wheel.

## Data availability statement

Data underlying the results presented in this paper will be open to public after its acceptance.

The pre-reserved DOI is: 10.5281/zenodo.14875771

## Data deposition

pre-reserved DOI: 10.5281/zenodo.14875771

## References

- Ajmat R, Sandoval J, Arana Sema F, O'donnell B, Gor S, Alonso H. 2011. Lighting design in museums: exhibition vs. preservation. *WIT Transactions on The Built Environment*, 118:195–206.
- Bao W, Wei M. 2021. Change of Gamut Size for Producing Preferred Color Appearance from 20 to 15000 lux. *LEUKOS*. doi:10.1080/15502724.2019.1587621.
- Bartleson CJ, Bray CP. 1962. On the preferred reproduction of flesh, blue-sky, and green-grass colors. *Photogr Sci Eng*. 6(1):19–25.
- Bellia L, Fragliasso F, Stefanizzi E. 2020. Effects of light source spectrum and background colour on the perception of paintings. *Light Res Technol*. 52(1):36–63. doi: 10.1177/1477153519847254
- Chen H-S, Chou C-J, Luo H-W, Luo M. 2016. Museum lighting environment: Designing a perception zone map and emotional response models. *Lighting Res Technol*. 48(5):589–607. doi:10.1177/1477153515596456
- [CIE] International Commission on Illumination. 1995. CIE 13.3:1995, Method of measuring and specifying colour rendering properties of light sources. Vienna (Austria): CIE.
- Cuttle C. 2000. A proposal to reduce the exposure to light of museum objects without reducing illuminance or the level of visual satisfaction of museum visitors. *J Am Inst Conserv*. 39(2):229–244. doi:10.1179/019713600806082720
- Durmus D, Abdalla D, Duis A, Davis W. 2020. Spectral Optimization to Minimize Light Absorbed by Artwork. *LEUKOS*. 16(1):45–54. doi:10.1080/15502724.2018.1533852
- Esposito T, Houser K. 2019. Models of colour quality over a wide range of spectral power distributions. *Light Res Technol*. 51(3):331–352. doi:10.1177/1477153518765953.
- Feltrin F, Leccese F, Hanselaer P, Smet KAG. 2020. Impact of Illumination Correlated Color Temperature, Background Lightness, and Painting Color Content on Color Appearance and Appreciation of Paintings. *LEUKOS*. 16(1):25–44. doi:10.1080/15502724.2018.1522261.
- Fox J, Weisberg S. 2018. *An R Companion to Applied Regression*. 3rd ed. Thousand Oaks (CA): SAGE. <https://www.john-fox.ca/Companion/index.html>.

- Gao H, Wang Z, Zhu D, Zhang C, Zou N. 2020. Research on the influence of lighting mode and CCT on the lighting design of art museum based on subjective experiment. *AIP Adv.* 10(12):125024. doi:10.1063/5.0022997.
- Huang Z, Liu Q, Westland S, Pointer MR, Luo MR, Xiao K. 2018. Light dominates colour preference when correlated colour temperature differs. *Light Res Technol.* 50(7):995–1012. doi:10.1177/1477153517713542.
- [IES] Illuminating Engineering Society. 2020. ANSI/IES RP-30-20, Recommended Practice: Lighting Museums. New York (NY): IES.
- [IES] Illuminating Engineering Society. 2024. ANSI/IES TM-30-24 Technical Memorandum: IES Method for Evaluating Light Source Color Rendition. New York (NY): IES.
- [ISO] International Organization for Standardization. 2005. ISO 20462-1:2005 Photography — Psychophysical experimental methods for estimating image quality — Part 1: Overview of psychophysical elements. Geneva (Switzerland): ISO.
- Jost-Boissard S, Avouac P, Fontoynt M. 2015. Assessing the colour quality of LED sources: Naturalness, attractiveness, colourfulness and colour difference. *Light Res Technol.* 47(7):769–794. doi:10.1177/1477153514555882.
- Jost-Boissard S, Avouac P, Fontoynt M. 2016. Preferred Color Rendition of Skin under LED Sources. *LEUKOS.* 12(1–2):79–93. doi:10.1080/15502724.2015.1060499.
- Kassambara A. 2023. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. Ver. 0.7.2 [software]. <https://CRAN.R-project.org/package=rstatix>.
- Keelan BW, Urabe H. 2003. ISO 20462: a psychophysical image quality measurement standard. In: Miyake Y, Rasmussen DR, eds. *Image Quality and System Performance*. Proceedings of SPIE 5294; 2004 Jan 18–22; San Jose, CA. Bellingham (WA): SPIE. 181–189. doi:10.1117/12.532064.
- Kesner CW. 1993. Museum Exhibition Lighting: Visitor Needs and Perceptions of Quality. *J Illum Eng Soc.* 22(1):45–54. doi: 10.1080/00994480.1993.10748017
- Kruithof AA. 1941. Tubular luminescence lamps for general illumination. *Philips Tech Rev.* 3:65–96.
- Loe DL, Rowlands E, Watson NF. 1982. Preferred lighting conditions for the display of oil and watercolour paintings. *Light Res Technol.* 14(4):173–192. doi:10.1177/096032718201400401
- Munding J, Houser K. 2023. Fine art under

- low illuminance: Gamut and tint. *Light Res Technol.* 56(6): 570–592.  
doi:10.1177/14771535231172100.
- Nascimento SMC, Herdeiro CFM, Gomes AE, Linhares JMM, Kondo T, Nakauchi S. 2021. The Best CCT for Appreciation of Paintings under Daylight Illuminants is Different for Occidental and Oriental Viewers. *LEUKOS.* 17(3):310–318.  
doi:10.1080/15502724.2020.1761828.
- Nascimento SMC, Masuda O. 2014. Best lighting for visual appreciation of artistic paintings—experiments with real paintings and real illumination. *JOSA A.* 31(4):A214–A219. doi:10.1364/JOSAA.31.00A214.
- Ohno Y, Li J. 2023. A colour fidelity model based on Hunt effect. In: *Proceedings of the 30th CIE Quadrennial Session.* Ljubljana, Slovenia: CIE, 2023.
- Okuda S, Yagi S, Taguchi S, Okajima K, Melgosa M. 2023. Influence of lighting colour on visual evaluation of landscape paintings - Focus on some Claude Monet's artworks. In: De Sainz Molestina D, Galluzzo L, Rizzo F, Spallazzo D, eds. *Life-Changing Design. Proceedings of IASDR Congress 2023; 2023 Oct 9–13; Milan, Italy.* London (UK): Design Research Society. doi:10.21606/iasdr.2023.625.
- Pérez-Ortiz M, Mantiuk RK. 2017. A practical guide and software for analysing pairwise comparison experiments. preprint arXiv.  
doi:10.48550/arxiv.1712.03686.
- Pinto PD, Linhares JM, Nascimento SM. 2008. Correlated color temperature preferred by observers for illumination of artistic paintings. *JOSA A.* 25:623–630.  
doi:10.1364/JOSAA.25.000623
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ross RT. 1934. Optimum orders for the presentation of pairs in the method of paired comparisons. *J Educ Psychol.* 25(5):375–382.
- Royer MP, Houser K, Durmus D, Esposito T, Wei M. 2022. Recommended methods for conducting human factors experiments on the subjective evaluation of colour rendition. *Light Res Technol.* 54(3):199–236. doi:10.1177/147715352110198.
- Royer MP, Wilkerson A, Wei M, Houser K, Davis R. 2017. Human perceptions of colour rendition vary with average fidelity, average gamut, and gamut shape. *Light Res Technol.* 49(8):966–991. doi:10.1177/1477153516663615.

- Royer MP, Wilkerson A, Wei M. 2018. Human perceptions of colour rendition at different chromaticities. *Light Res Technol.* 50(7):965–994. doi:10.1177/1477153517725974.
- Royer MP, Wei M, Wilkerson A, Safranek S. 2020. Experimental validation of colour rendition specification criteria based on ANSI/IES TM-30-18. *Light Res Technol.* 52(3):323–349. doi:10.1177/1477153519857625.
- Scuello M, Abramov I, Gordon J, Weintraub, S. 2004. Museum lighting: Optimizing the illuminant. *Color Res Appl.* 29:121–127. doi: 10.1002/col.10231
- Schanda J. Csuti P. Szabó F. 2016. A New Concept of Color Fidelity for Museum Lighting: Based on an Experiment in the Sistine Chapel. *LEUKOS.* 12(1-2):71–77. doi:10.1080/15502724.2014.978503
- Veitch JA, Whitehead LA, Mossman M, Pilditch TD. 2014. Chromaticity-matched but spectrally different light source effects on simple and complex color judgments. *Color Res Appl.* 39(3):263–274. doi:10.1002/COL.21811.
- Wei M, Bao W, Huang H-P. 2020. Consideration of Light Level in Specifying Light Source Color Rendition. *LEUKOS.* 16(1):55–65. doi:10.1080/15502724.2018.1448992.
- Wei M, Houser KW, David A, Krames MR. 2017. Colour gamut size and shape influence colour preference. *Light Res Technol.* 49(8):992–1014. doi:10.1177/1477153516651472.
- Zhai Q, Luo MR, Liu X. 2015. The impact of illuminance and colour temperature on viewing fine art paintings under LED lighting. *Light Res Technol.* 47(7):795–809. doi:10.1177/1477153514541832.
- Zhai Q, Luo MR, Liu X. 2016. The impact of LED lighting parameters on viewing fine art paintings. *Light Res Technol.* 48(6):711–725. doi:10.1177/1477153515578468.
- Zhang F, Xu H, Feng H. 2017. Toward a unified model for predicting color quality of light sources. *Appl Opt.* 56(29):8186–8195. doi:10.1364/ao.56.008186.

Table 1. Target (gray shaded) and measured characteristics of the 9 lighting spectra.

ID	Background			$E_{\perp}$ (lux)	Target and measured criteria						CRI		
	$Y_{10}$	$u'_{10}$	$v'_{10}$		$R_{f \text{ tar}}$	$R_{f \text{ meas}}$	$R_{g \text{ tar}}$	$R_{g \text{ meas}}$	$\psi_{\text{tar}}$	$\psi_{\text{meas}}$	$R_a$	$R_{\text{cs,h1}}$	$R_{\text{f,h1}}$
	(cd/m <sup>2</sup> )								(°)	(°)			
$f_{65}g_{90}^+$	5.88	0.2622	0.5199	142.6	65	65	90	90	22.5	26.9	61	4%	82
$f_{75}g_{90}^+$	5.91	0.2634	0.5196	148.0	75	76	90	90	22.5	37.4	77	-5%	85
$f_{75}g_{100}^+$	5.89	0.2632	0.5202	145.2	75	75	100	99	22.5	24.6	70	5%	87
$f_{75}g_{110}^+$	5.77	0.2621	0.5189	144.2	75	77	110	109	22.5	12.4	68	9%	80
$f_{85}g_{100}^+$	5.85	0.2630	0.5194	149.1	85	85	100	100	22.5	26.9	83	3%	90
$f_{85}g_{110}^+$	5.89	0.2630	0.5188	150.2	85	86	110	110	22.5	-0.2	80	6%	86
$f_{75}g_{90}^-$	5.86	0.2628	0.5190	159.8	75	76	90	90	-67.5	-80.7	70	-19%	67
$f_{85}g_{100}^-$	5.94	0.2631	0.5197	158.9	85	84	100	100	-67.5	-71.4	85	-9%	83
$f_{95}g_{100}$	5.85	0.2627	0.5199	154.1	95	94	100	100	NA	75.9	97	-2%	94

Table 2. Pooled preference JOD scores for each spectrum, ordered based on the JOD scores. Spectra sharing a bar are not significantly different based on bootstrapping results: thick bar  $p > 0.1$ , thin bar  $0.05 < p < 0.1$  <sup>a</sup>

Spectrum	$f_{75}g_{90}^-$	$f_{65}g_{90}^+$	$f_{85}g_{100}^-$	$f_{75}g_{90}^+$	$f_{75}g_{100}^+$	$f_{95}g_{100}$	$f_{85}g_{100}^+$	$f_{85}g_{110}^+$	$f_{75}g_{110}^+$	
$R_f$ meas	76	65	84	76	75	94	85	86	77	
$R_g$ meas	90	90	100	90	99	100	100	110	109	
$R_{cs,h1}$ meas	-19%	4%	-9%	-5%	5%	-2%	3%	6%	9%	
$\psi$ meas (°)	-80.7	26.9	-71.4	37.4	24.6	75.9	26.9	-0.2	12.4	
TM-30 Annex E preference level	P-	P-	P3	P3	P2	P2	P1	P1	P2	
Preference JOD	-0.81	-0.22	-0.22	-0.14	0.00	0.01	0.26	0.53	0.59	
groups based on bootstrapping		—————						—————	—————	
Adjusted preference level <sup>b</sup>	P'-	P'3	P'3	P'3	P'3	P'3	P'2	P'1	P'1	

<sup>a</sup>  $f_{85}g_{100}^+$  differs from  $f_{75}g_{100}^+$  with  $p = 0.06$ ;  $f_{85}g_{100}^+$  differs from the rest significantly ( $p < 0.05$ )  
<sup>b</sup> Details are described in subsection 5.2

Table 3. Pearson correlation coefficients  $R^2$  of the preference JOD scores between paintings

	Pooled	Gp1			Gp2		
		W1	W2	M1	W3	M2	C1
pooled		<b><u>.800</u></b>	<b><u>.792</u></b>	<b><u>.867</u></b>	<b><u>.605</u></b>	<b><u>.769</u></b>	<b><u>.866</u></b>
W1			<b><u>.955</u></b>	<b><u>.913</u></b>	.208	.358	<b><u>.475</u></b>
W2				<b><u>.828</u></b>	.247	.353	<b><u>.475</u></b>
M1					.268	<b><u>.481</u></b>	<b><u>.634</u></b>
W3						<b><u>.800</u></b>	<b><u>.740</u></b>
M2							<b><u>.886</u></b>
C1							

**bold:**  $\geq 0.4$ , **bold and underline:**  $\geq 0.7$

Table 4. Pooled preference JOD scores of each spectrum for Gp1-paintings. Spectra sharing a bar are not significantly different based on bootstrapping results: thick bar  $p > 0.1$ , thin dashed bar (see remark <sup>a</sup>)

ID	$f_{75}g_{90}^-$	$f_{85}g_{100}^-$	$f_{95}g_{100}$	$f_{75}g_{90}^+$	$f_{65}g_{90}^+$	$f_{85}g_{100}^+$	$f_{75}g_{100}^+$	$f_{85}g_{110}^+$	$f_{75}g_{110}^+$
Preference JOD	-1.08	-0.48	-0.10	-0.04	0.10	0.20	0.26	0.51	0.62
groups based on bootstrapping <sup>b</sup>		---	---	---	---	---	---	---	---
TM-30 Annex E preference level	P-	P3	P2	P3	P-	P1	P2	P1	P2
Adjusted level in Table 2	P'	P'3	P'3	P'3	P'3	P'2	P'3	P'1	P'1

<sup>a</sup> **thin dashed bar:**  $f_{85}g_{100}^-$  significantly differs from the others ( $p < 0.05$ ) except for  $f_{75}g_{90}^+$  ( $p = 0.07$ ) and  $f_{65}g_{90}^+$  ( $p = 0.14$ )

<sup>b</sup> **Exceptional pairs** (no significant difference between bars):  $f_{65}g_{90}^+|f_{85}g_{110}^+$  ( $p = 0.08$ ) and  $f_{75}g_{100}^+|f_{85}g_{110}^+$  ( $p = 0.15$ ).

Table 5. Pooled preference JOD scores of each spectrum for Gp2-paintings. Spectra sharing a thin bar are not significantly different based on bootstrapping results ( $p > 0.05$ )

ID	$f_{75}g_{90}^-$	$f_{65}g_{90}^+$	$f_{75}g_{100}^+$	$f_{75}g_{90}^+$	$f_{85}g_{100}^-$	$f_{95}g_{100}$	$f_{85}g_{100}^+$	$f_{85}g_{110}^+$	$f_{75}g_{110}^+$	
Preference JOD	-0.56	-0.54	-0.26	-0.23	0.03	0.11	0.33	0.55	0.57	
groups based on bootstrapping *										
TM-30 Annex E preference level	P-	P-	P2	P3	P3	P2	P1	P1	P2	
Adjusted level in Table 2	P-	P3	P3	P3	P3	P3	P2	P1	P1	

\***Exceptional pairs** (significant difference within a bar):  $f_{65}g_{90}^+/f_{75}g_{100}^+$  ( $p = 0.03$ ),  $f_{85}g_{100}^-/f_{85}g_{110}^+$  ( $p = 0.01$ ),  $f_{95}g_{100}/f_{85}g_{110}^+$  ( $p = 0.001$ )

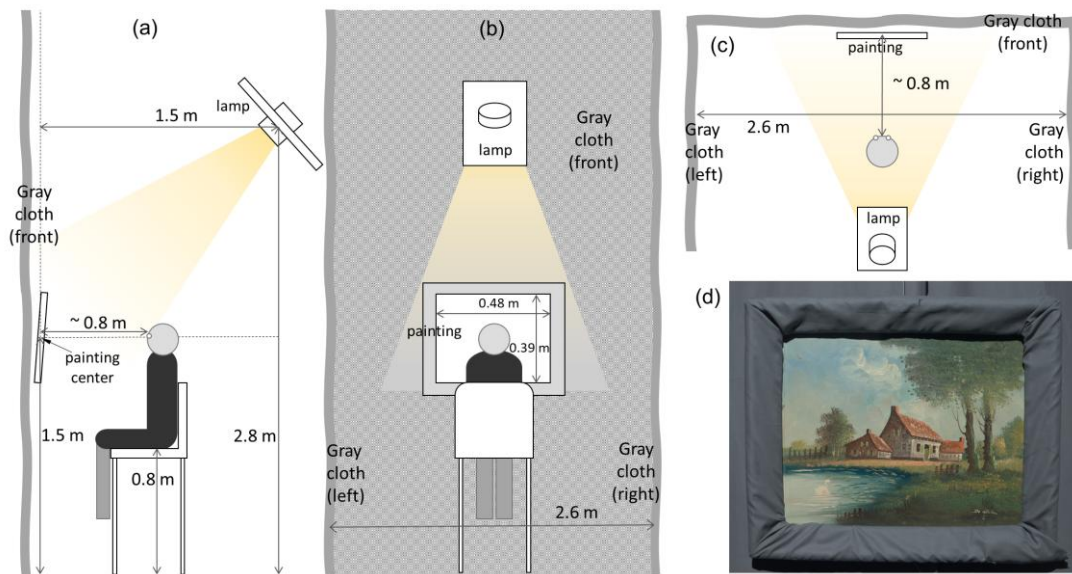


Fig. 1. Experimental layout. Participants evaluated the color appearance of an oil painting in a room with three walls covered by mid-gray cloths. A spectrally tunable luminaire illuminated the painting. (a) Side view. (b) Back view. (c) Top view. (d) View from participants: an oil painting with its frame covered by gray cloth.

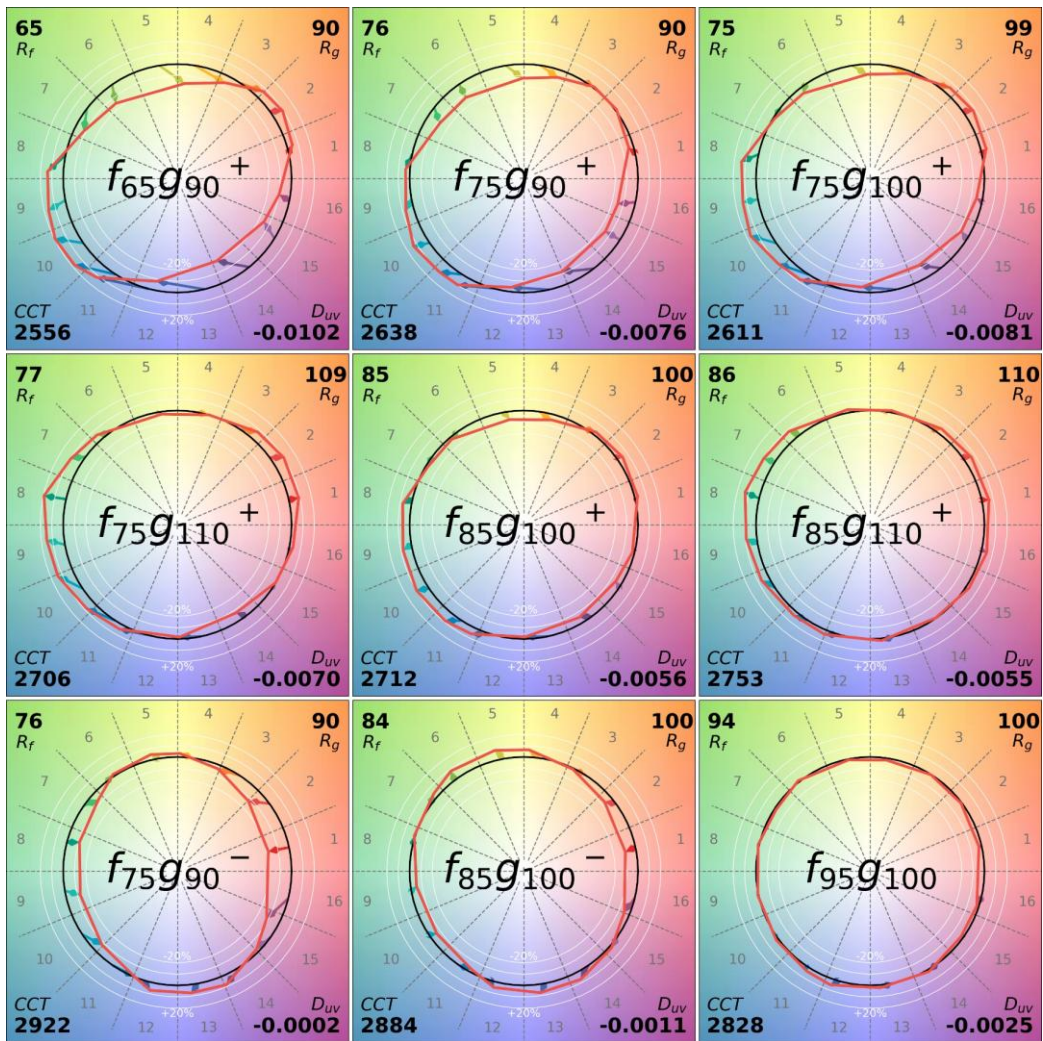


Fig. 2. TM-30 Color Vector Graphic and indices of each spectrum. Under these spectra and a 3000 K Planckian radiation, the reflected light from the gray background were colorimetric metamers for the CIE 1964 standard observer. In TM-30, the CCT and  $D_{uv}$  values are calculated using the CIE 1931 2° standard observer; therefore, these CCT values are lower than the 3000 K value obtained with the 10° 1964 standard observer.



Fig. 3. Six test oil paintings on canvas evaluated during the experiment under the high-fidelity spectrum  $f_{95g_{100}}$ . Images are transformed to a daylight appearance by applying a von Kries chromatic adaptation transform from the 3000 K white point under illumination  $f_{95g_{100}}$  (as measured on a white patch of a Macbeth ColorChecker) to CIE D65. Regions outside the fixed visible area of 48 cm  $\times$  39 cm were covered with mid-gray cloths to serve as background.

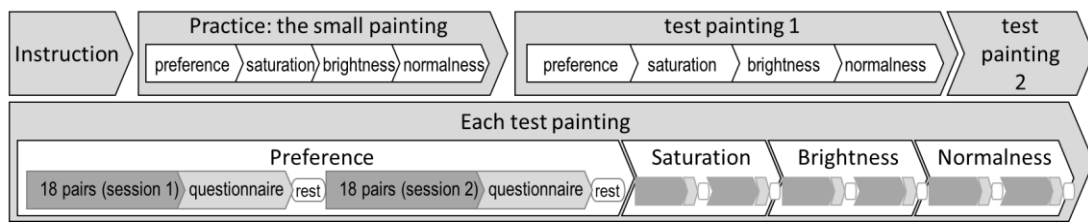


Fig. 4. Upper line: One-day experimental procedure. Lower line: Procedure for each test painting.

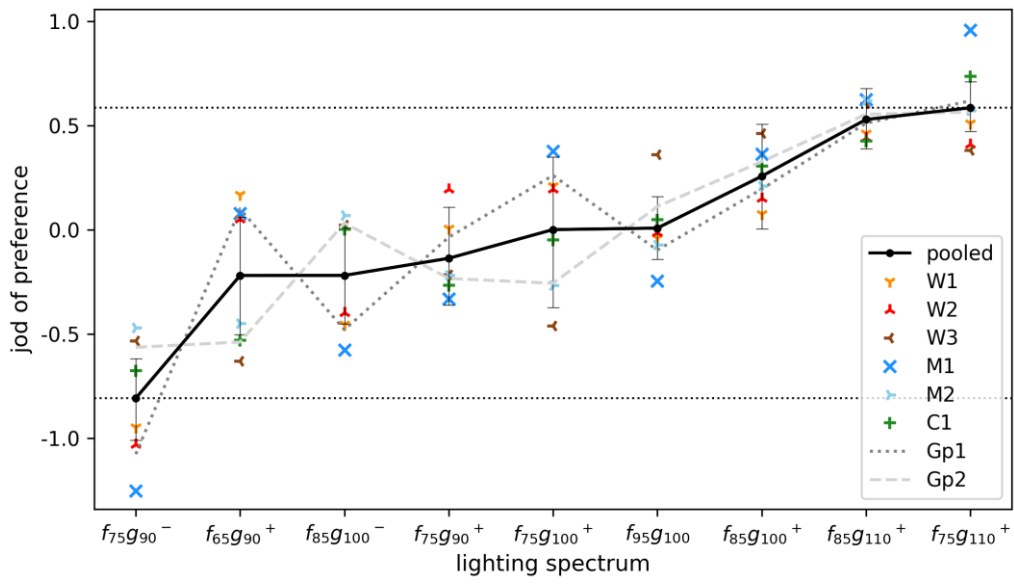


Fig. 5. Estimated preference JODs of each painting over different lighting spectra. The lighting spectra are ordered in ascending of pooled JOD scores. Horizontal dotted lines represent the minimum and maximum pooled scores. Thin error bars indicate the 95% confidence intervals of the pooled score, estimated from the bootstrapping results.