HUMAN, EARTH AND FUTURE
Ital Publication

ISSN: 2785-2997

Available online at www.HEFJournal.org

Journal of
Human, Earth, and Future

Vol. 3, No. 2, June, 2022

Investigation of Colour Naturalness in Lighting: A Comparative Study

Andrew Neil Chalmers [1*]

[1] *Institute of Biomedical Technologies, Auckland University of Technology, Auckland City, New Zealand*

Received 04 March 2022; Revised 14 May 2022; Accepted 19 May 2022; Published 01 June 2022

Abstract

This paper is concerned with improving the acceptability of LED light sources, since their long life and high efficiency have contributed to their widespread adoption in many applications. Concerns remain, however, in relation to their colour properties. The purpose of this paper is therefore to promote discussion of the naturalness concept among the users, specifiers, and manufacturers of lighting sources and systems, in the hope that this may provide a valid pathway to the classification of light-source colour properties. An overview is presented of experimental investigations aimed at establishing visually meaningful metrics for the colour quality of various light sources, predominantly LEDs. All the cases presented here included colour naturalness as at least one of the dimensions studied, and one has to conclude that naturalness is a property of great interest to lighting engineers and scientists. Because the majority of naturalness studies have invoked the use of pre-existing colour quality metrics, the paper also includes an overview of some of the major such metrics and their features. The paper also identifies two important concerns relating to Naturalness: the need to agree on an acceptable definition of colour naturalness in lighting; and how to standardize or compare the results of disparate investigations. Finally, the paper proposes the concept of a colour fidelity continuum, in the ultimate hope of uniting the various approaches to lighting colour quality.

Keywords: Clean Technology; LED Lighting; Colour in Lighting; Colour Rendition; Colour Quality; Colour Naturalness.

1. Introduction

The purpose of this paper is to promote discussion of the naturalness concept among the lighting fraternity. As will become evident, there are two main difficulties in writing about Naturalness in Lighting: First is the problem of a lack of consensus in the definition of lighting naturalness; next (and this follows from the first) is the wide range of different approaches adopted in the attempts to achieve subjective evaluations of Naturalness in experimental conditions.

The concept of Naturalness was thrust to the forefront of the LED lighting community in 2020 when one LED supplier proposed Spectral Similarity [1] (explained later) as a metric for the naturalness of a light source (i.e. to classify the capacity of a light source to illuminate coloured surfaces in such a way as to provide the "*most natural*" view of the colours). To put this into context: the CIE (International Commission on Illumination), and the lighting community generally, have for several decades been grappling with the inadequacies of the colour rendering index (CRI, symbol Ra) [2]; and a number of alternative, or supplementary, metrics have been proposed in attempts to provide lighting users with more useful data.

* Corresponding author: andrew.chalmers@aut.ac.nz

http://dx.doi.org/10.28991/HEF-2022-03-02-04

The widespread availability of LED lighting in the new millennium has been nothing short of revolutionary. The lumen-per-watt efficacy has climbed to unprecedented levels, while the physical size of the sources has shrunk to sizes measured in millimetres, making for significant increases in efficiency in luminaire performance. In the realm of colour, LEDs are providing the source designer with an unparalleled range of narrow-band monochromatic spectra, which can be combined in ways to create white light sources of widely differing SPDs (spectral power distributions) with a correspondingly wide range of colour properties. These can be further expanded by developments in the formulation of phosphor materials which are able to act as wavelength converters, often with wide-band outputs.

Until relatively recently, the only colour specifications available to users were the CCT (correlated colour temperature) which largely determines the atmosphere created by the lighting, plus the CIE colour rendering index, Ra, as a guide to the colour quality of the source. At a time when fluorescent sources first dominated the interior lighting scene, these two metrics were generally sufficient. However, the CRI was found to be lamentably deficient when LEDs came onto the general lighting scene. Many LED lamps with supposedly "*good*" Ra values (and high luminous efficacies) were found to give unacceptable colour performance for users. The CIE and other professional and standardization bodies have been working on this problem for much of the past two decades.

Because a range of metrics already exists for determining the colour properties of sources (e.g. correlated colour temperature, colour rendering, colour fidelity, etc.) one approach has been to assess the degree of correlation of subjective Naturalness with one or more of these existing metrics (or combinations thereof). For this reason, Section 2 of this paper will give an overview of the existing colour indices and their evolution over the past (roughly) six decades. This subject matter has been included for the benefit of readers unfamiliar with the technology, and it can be omitted by those who are already conversant with the field. Following this, in Section 3, the question of Naturalness and some possible naturalness metrics will be explored in detail.

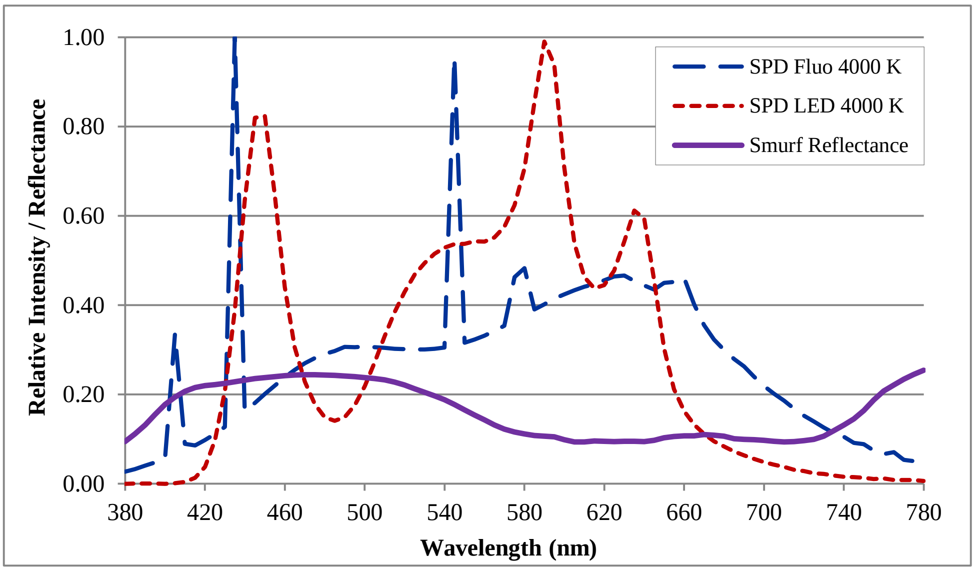
Note that this paper will focus on the performance of light sources in relation to their ability to provide a natural portrayal of the colours of objects in the scenes they illuminate, and it concentrates largely on LED-type sources. It will not consider the specific effects of variations in the CCT of sources (as it is hoped ultimately to discover a Naturalness metric that will apply independently of CCT) and the concept of dynamic lighting * is excluded.

2. Sources, Surfaces, and Colour Metrics

2.1. A Short Case Study

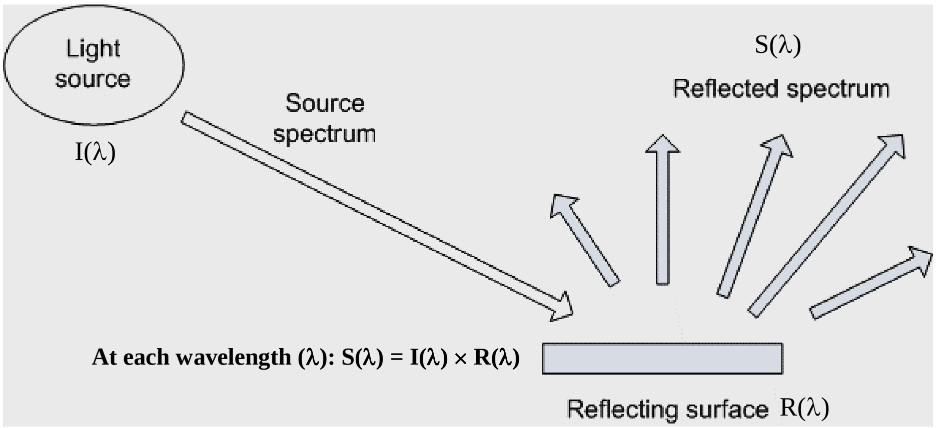
One may question why there is any need for colour rendition metrics – i.e. the quantities like Ra, Qf, Rf, Rg, etc., as devised by CIE, NIST (National Institute of Standards and Technology, USA.), IES (Illuminating Engineering Society of North America), and others, and outlined in the following section. They are necessary because of the human visual attribute of metamerism which, in the lighting context, means that light sources can look the same even though their spectral compositions may be very different. Figure 1(a) shows two sources with the same CCT (4000 K), and hence the same colour appearance, but clearly different SPDs. Figure 1(b) illustrates the general principle that the light reflected off any surface depends on both the source SPD and the spectral reflectance of the surface.

NB: This pair of SPDs has been chosen randomly from an available set of real source spectra † with CCT of 4000 K (which is a widely-used CCT in commercial and educational lighting installations).



(a)

* Defined by Philips as lighting that mimics the natural rhythm of night and day that the human body responds to by positively influencing the "body clock" [3].
† Included in the MCRI Calculator spreadsheet published by Smet *et al.* [4] where they are designated as "Fluorescent4k" and "WAR4k".



(b)

Figure 1. (a) Two example metameric spectra, Fluorescent source and LED combination, together with the surface reflectance spectrum of a Smurf's body [4]; (b) Illustrating that the reflected light depends on both the surface and source spectra

Figure 2(a) illustrates the appearance of a typical Smurf manikin. The spectral reflectance curve plotted in Figure 1(a) was obtained from measurements of Smurf "*Skin*" by Smet et al. [4]. We note that different spectral compositions of sources inevitably lead to differences in appearance of the coloured surfaces being illuminated, as in Figure 2 (b) which shows the two curves for the Smurf's reflected-light SPDs, respectively under the two 4000 K sources defined in Figure 1(a). When evaluated in the CIELAB colour space, these two results have a colour difference ΔE_{ab} of 4.5 units which is generally regarded as noticeable.

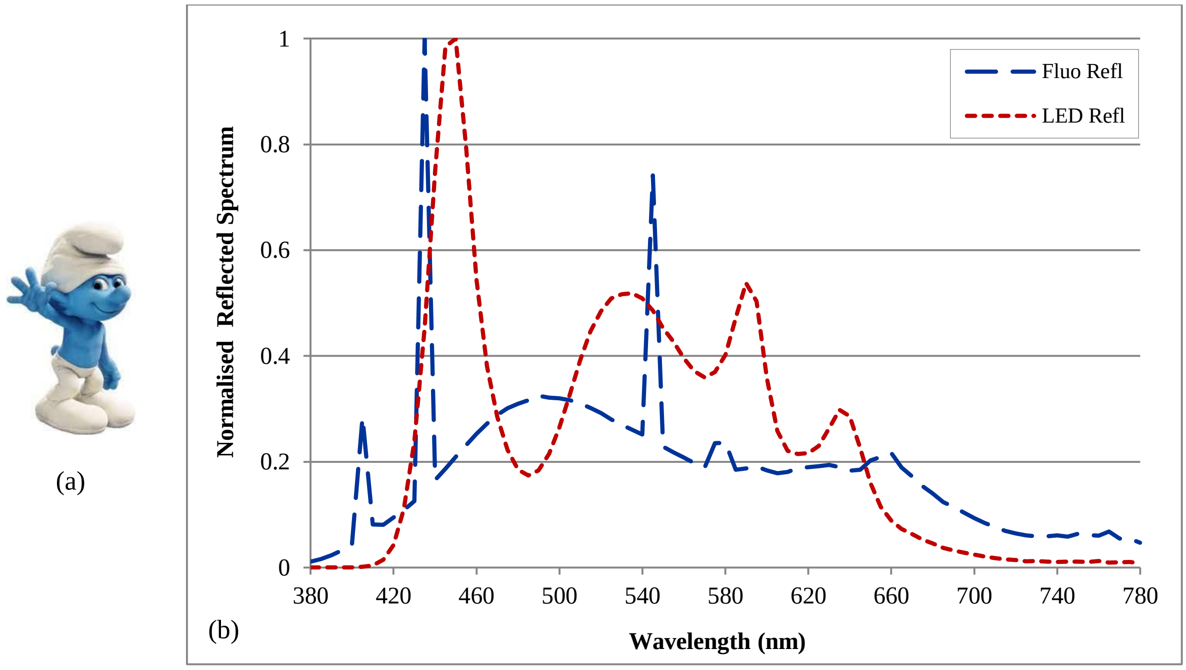


Figure 2. (a) Image of a typical Smurf figurine; (b) Spectra of the reflected light from a Smurf body illuminated, respectively, by each of the two 4000-K sources: Fluorescent: long dash; LED combination: short dash

Note that a similar procedure can be applied to any number of different surface colours, leading to an array of ΔE_{ab} results. The questions that may then arise are along the lines of: Which source gives the more accurate colours?; or the more pleasing colours ?; or the more natural colours.? It is the attempts to answer these types of questions that have led to the development of the various systems of colour rendition evaluation reviewed below.

2.2. The Colour Fidelity Principle

Before studying the development of different colour rendition systems, it may be beneficial first to extend the preceding case study by looking into the general principle of colour fidelity. Let us assume that the body colour of the Smurf is a colour of interest in terms of a colour fidelity investigation related to the two sources defined in the previous section. To establish the fidelity of each source, one undertakes the two comparisons illustrated in Figure 3. In each instance one finds the reflected SPDs for the test source (fluorescent and LED respectively) for comparison with the reflected SPD obtained when using a reference source (defined here as a Planckian radiator having the same CCT as each test source in turn) –the results of which are as shown in Figure 3.

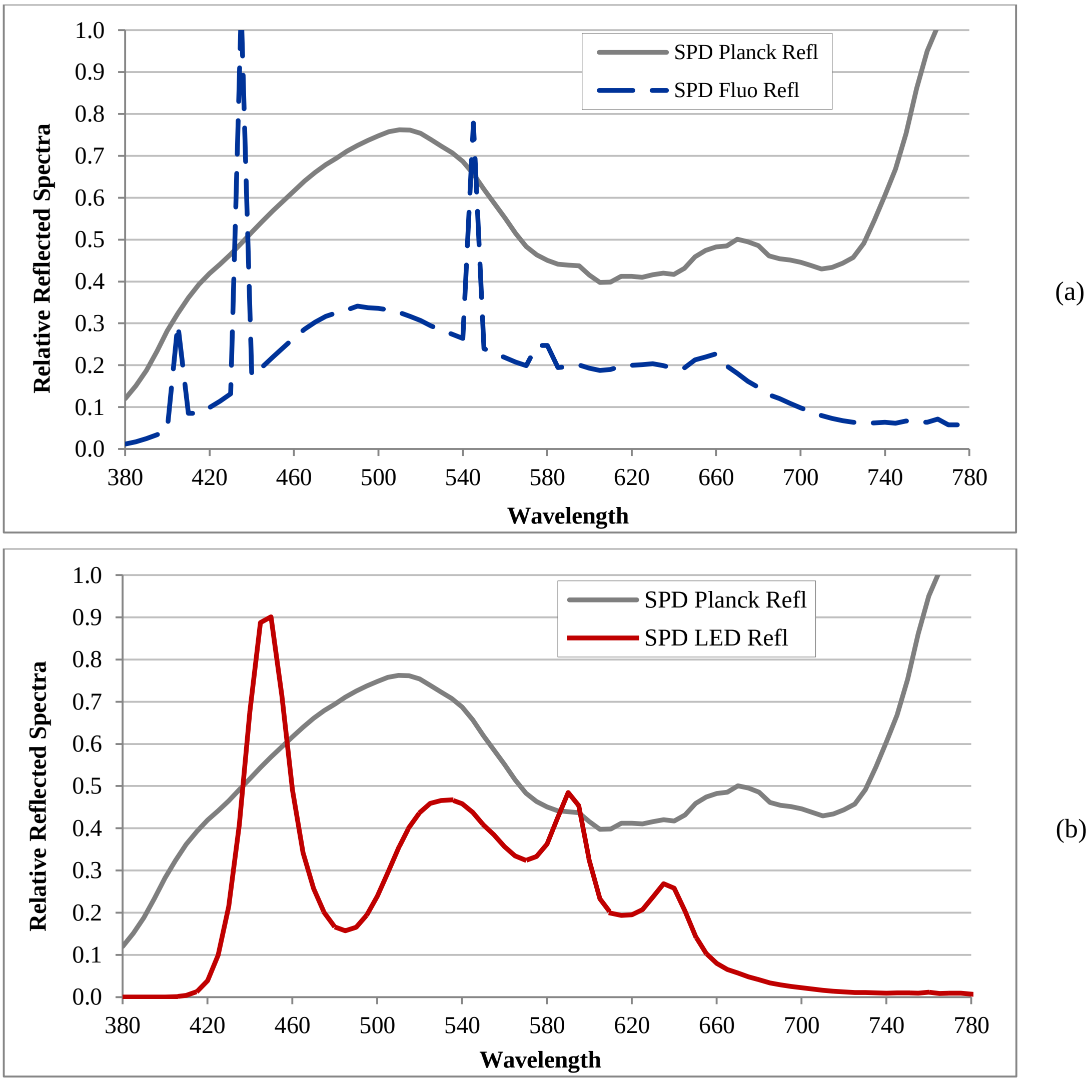


Figure 3. SPDs of reflected light from the body of the Smurf figurine: (a) Comparing the results of the 4000 K Fluorescent source and 4000 K Planckian radiator; (b) Comparing the results of the 4000 K LED-combination source and 4000 K Planckian radiator.

One would generally wish to quantify the results by use of an accepted form of colour difference calculation; as an example, the colour difference in the CIELAB colour space. For the comparisons illustrated here: (a) for the fluorescent source: $\Delta E_{ab(Fluo)} = 1.363$; and (b) for the LED source $\Delta E_{ab(LED)} = 5.783$. The conclusion is that, for this particular surface colour, the fluorescent source gives the higher fidelity (based on the smaller colour difference). In terms of a ***general*** colour fidelity index or ***general*** colour rendering index, it is considered necessary to incorporate as many test colours as is conveniently possible, and to find ways of averaging the results. The colour difference calculations should be carried out using the most widely accepted algorithms available at the time.

2.3. Colour Rendition and Colour Quality

We come now to an examination of major systems of colour evaluation currently in existence; and begin with the CIE Colour Rendering Index. We note that the causes and effects of the problems with the CIE Ra have been discussed in a paper by Van Trigt [5]. In spite of the criticisms it has received, Ra is still the internationally accepted metric for colour rendering [2], and it is included here for completeness' sake. Briefly, the computation method for Ra involves: (a) the identification of the test source; (b) selection of the appropriate reference illuminant* of the same CCT; (c) calculating the colour of a selected test colour sample under each of the sources, using a chromatic adaptation transform to correct for differences between the test source and reference source; (d) calculating the difference between the two resulting sample colours; (e) repeating steps (c) and (d) for the 14 defined CIE test colours; (f) calculating the mean colour difference for the first 8 CIE test colours†; (g) scaling the result (f) by a set constant, and then subtracting from 100 to yield the value of Ra.

---

* In the CIE system, the reference is the Planckian radiator for CCTs below 5000 K, and for CCTs ≥ 5000 K it is to be one of the CIE daylight illuminants.

† The remaining six test colours provide additional information; e.g. test colour 9 (a red) gives a rating R9 that has been widely used as an indicator of a source's red rendition performance.

An updated method for estimating the colour quality of a source was published by NIST [6] and is known as the CQS (colour quality scale). It was also based on a colour difference approach, using the CIE definition of reference illuminant, but with 15 newly-defined test colours, and using more up-to-date colour difference and adaptation calculations, as well as more sophisticated mathematical procedures to produce the final results, which are given in three main forms: Qa (general quality index), Qf (fidelity index) and Qp (preference index, which is not included in version 9 of the CQS). A colour gamut index Qg is also defined. Another more modern approach, known as TM-30-15, was published by the IES [7], again using a colour difference method but with a modified definition of the CIE reference illuminant*, and using a new set of 99 test colours. It included new advances in colour difference, chromatic adaptation, and averaging computations, to yield two main outputs: Rf (colour fidelity index) and Rg (gamut area index).

In 2017 the CIE adopted their own colour fidelity index Rf [8], a modified version of the IES Rf, but using the same basic calculations that were built into the IES system. The IES has more recently published TM-30-18, a revised version of the 2015 method, harmonizing the Rf calculation with the CIE. A useful set of slides outlining the merits of the TM-30 methods, and providing comparisons with the CIE Ra method, is available online [9]. Meanwhile, there have been many other efforts over the past two decades to find or refine these and other metrics that may have the potential to serve as measures of colour quality (also including Naturalness) which are summarized in [4]. One of these, often referred to as the MCRI (memory colour rendering index) is singled out for mention here, and explained in detail by Smet & Hanselaer [10]. The metric was based on the assumption that the colour rendition or colour quality of a light source improves when the colours of familiar objects are rendered more closely to what is expected or recalled. The method made use of ten test colours, being the colours of ten well-known objects. After first computing their colours under the test source, the corresponding colour values under CIE illuminant D65 can be evaluated by use of a chromatic adaptation transform and then converted to the IPT colour space [11]. The degree of similarity of each object's chromaticity to the respective memory colour of each familiar object can then be calculated and combined to yield index Sa, then finally rescaling the values to give index Rm in the 0 … 100 range.

Table 1 is an illustration of the results achieved using several of the abovementioned systems to evaluate the two source SPDs defined in the earlier case study†. On average, the fidelity of the LED source is 20 units worse than the Fluorescent. In terms of gamut area, both sources score close to 100 and there is thus little to choose between them without further detail. Notice, however, that in terms of the memory colour index, there is only a 7 unit difference in favour of Fluorescent. This apparent discrepancy warrants further study, and this leads on to the investigation of Naturalness as a potential metric for the "*real*" appreciation of the differences between the sources in terms of their colour properties.

Table 1. Some of the colour properties of the two randomly selected sources used in the earlier case study

Light Sources (4000 K)	Fidelity Metrics			Gamut	Memory
	CIE Ra	NIST Qf	IES Rf	IES Rg	Smet Rm
Fluorescent	96	95	95	102	91
Combo LED	77	73	76	96	84

3. Evidence for Colour Naturalness

In the lighting world there does not appear to have been any great interest in the concept of Naturalness prior to about 2010. Such interest was clearly being driven during the preceding decade by the rapid adoption of LED light sources and studies of their colour properties.

3.1. Naturalness in Imaging

During the 1990s there had already been research undertaken into the naturalness of electronic images – specifically in respect of the capture and display of images of natural scenes [12, 13]. Experiments were carried out by Ridder et al. (1995) using colour monitors displaying natural-scene images which were rated by up to 21 observers. The scenes were shown in several different experiments with and without variations of chroma, hue, saturation or lightness; and the viewers were asked to rate the image quality, naturalness and colourfulness. The instructions given to the subjects included definitions of these three image attributes. The definition used for naturalness was "the degree of correspondence between the reproduced image and reality (*i.e.* the original scene as it is according to the viewer)". A significant finding was that "qualitatively optimal" images were more colourful than the images considered to be the most natural. Apparently, the subjective preference in quality was biased toward more colourful images even though the subjects realised that these images were less natural-looking. It will become clear below that generally similar

* For sources with a CCT of 4000 K or less, the reference is a Planckian radiator at the same CCT. At 5000 K or above, the reference illuminant is from the CIE D Series. Between 4000 K and 5000 K, the reference illuminant is a proportional blend of Planckian radiation and the D Series illuminant, each at the specified CCT.

† These results are specific to the particular SPDs selected. They have been included in order to give the reader an appreciation of the orders of magnitudes of the quantities of interest. They are not representative of either fluorescent sources or LED sources in general.

findings seem to apply when assessing coloured objects viewed under different light sources, and that care must be taken not to confuse "naturalness" with "qualitatively optimal".

3.2. Naturalness in Lighting

This study has been based on a search for published work on actual experimentation that has included assessments by human observers of colour Naturalness in various test situations, and Table 2 provides a concise summary of these works. At least ten research groups have been active over the past dozen or so years in researching the colour quality properties of light. Here (in *approximate* chronological sequence) are highlighted a number of those groups whose results are based on visual evaluations of quality factors including Naturalness.

Table 2. Published data on lighting Colour Quality: Visual Assessments including Naturalness

Year	Ref. No.	Test sources	Test setup	Reference source(s)	CCTs (K)	Lux level	Observer Nos.	Observations recorded
2009	Jost-Boissard et al. [14]	12 LED	2 light booths	Tungsten-Halogen; Fluorescent	3050 3950	230 220	40	Attractiveness Naturalness Suitability
2015	Jost-Boissard et al. [15]	1 Fluorescent 7 LED 1 Halogen	3-booth setup using 2 at a time	Lighting pairs evaluated	3000 4000	230 210	45 36	Naturalness Colourfulness Visual preference Colour difference
2010	Smet et al. [16]	1 Halogen 1 Neodymium 1 Fluorescent 3 LED combos	2-booth setup, with sequential viewing	–	2750	250	92	Preference; Fidelity Vividness Naturalness; Attractiveness
2011	Smet et al. [17]	Ditto	Ditto	–	2700	Ditto	Ditto	Spearman correlation analysis
2012	Smet [4]	Ditto	Shared data with the above	–	Ditto	Ditto	Ditto	Re-analysis of the data from the two preceding papers
2012	Nascimento & Masuda [18]	D65 and simulated metamers	Video projection in dark room.	D65	6500	20 cd/m^2	6 4	Naturalness Preference Chromatic diversity
2013	Islam et al. [19]	21 LED 3 Fluorescent	Light booths	Fluorescent sources	2700 4000 6500	–	60	Naturalness Visual appearance Colourfulness
2013	Dangol et al. [20]	ditto	Shared data with the above	ditto	ditto	–	ditto	Shared observations with the above
2015	Dangol et al. [21]	6 LED SPDs 2 Fluorescent	Immersion using 2 identical office mock-ups	–	4000 6500	300 500	40	Preference Naturalness of objects Colourfulness Naturalness of hand
2017	Bhusal & Dangol [22]	–	Reviewed Islam et al. [19], Dangol et al. [20, 21]	–		–	–	Re-analysis of previous data from the 3 above papers
2017	Khanh et al. [23]	1 Halogen; 2 Fluorescent 2 Phosphor-converted LED	Immersion in office mock-up	None	2300 2700 4000 4100	470	38	Preference Naturalness Vividness
2017	Khanh et al. [24]	7 LED	Single light booth	None	3220	–	23	Preference; Naturalness; Vividness
2018	Khanh & Bodrogi [25]	7 selected LED SPDs (multi-LED source)	Cosmetic products in single viewing booth	–	3200	550	6 "Responsive observers"	Preference Naturalness Vividness
2018	Khanh et al. [26]	36 LED spectra at 4 CCTs	Immersion in a room	–	3100 4100 5000 5600	750	10 (or 9) "Responsive observers"	Preference Naturalness Vividness
2017	Royer et al. [27]	26 SPDs (mixtures of 7 LEDs)	Immersion in a room	Planckian	3500	215	28	Saturated; Dull; Normal; Shifted
2019	Esposito & Houser [28]	24 SPDs with various IES TM-30 properties	Single viewing booth	Prior adaptation to each scene	3500	650	40 total (20 for each spectrum)	Naturalness Vividness Preference
2020	Royer et al. [29]	90 SPDs	Immersion in a room	Prior adaptation to each scene	2700 3100 3500	310	25	Normalness Saturation Preference Acceptability

NOTE: "–" signifies feature(s) not specifically identified

Jost-Boissard et al. [14] reported on experiments with 12 different LED spectra at approx. 225 lux and CCTs of 3050 and 3950 kelvins, using simultaneous viewing of 2 light booths by a total of 40 observers (at different times). Naturalness and CQS were found to correlate with correlation a coefficient *r* of about 0.8 (CCT dependent). The authors comment that "One may not have a precise idea of what is naturalness and thus would not know what to expect." This appears to be a recurring theme in the majority of the work reported here, although many of the authors glided over this problem. In a later piece of work by Jost-Boissard et al. [15], using a range of different light sources at approx. 220 lux and CCTs of 3000 K and 4000 K, the results from 36+ observers indicated that both Ra and Qa correspond to visual Naturalness with $r^2 \geq 0.88$. This led to the comment that "A high fidelity score does not necessarily mean a natural rendition since it is based on comparison to a reference illuminant which itself may not be considered as the most natural". This acute observation has been echoed by many other workers in this field. Significant correlations with other metrics (including MCRI and CQS quantities) are also given in the paper.

Smet et al. [16] used six different sources (including three LED combinations) at 250 lux and 2750 K, in a 2-booth setup viewed by a total of 92 observers who were asked to evaluate the sources for Preference, Fidelity, Vividness, Naturalness and Attractiveness. Naturalness was found to correlate closely with Sa (or, MCRI), Ra, and Qa (all with Pearson correlation coefficients $r \geq 0.86$). As a justification for the use of Sa, they comment that "Colours often seem 'wrong' when they are not what we expect … them to be". A later paper [17] using a different statistical approach to examine thirteen metrics, led to Spearman correlation results for Naturalness: Ra and Qf ($r \geq 0.65$); Sa ($r = 0.45$).

Nascimento & Masuda (2012) [18] used a projection system onto a 3-m by 2-m screen viewed at a 3 m distance, with an average luminance of 20 cd/m^2, to display images of natural scenes. The scenes were lit by D65 and a number of D65 metamers. Observers had to select which illuminant was optimal in terms of naturalness, individual preference, and chromatic diversity. For the naturalness assessment, the observers were instructed to select which illuminant made the colours of objects appear as natural as possible; and the main finding was that " ... the illuminant selected was more spectrally structured than daylight and had a low color rendering index." This finding underlines the statement noted elsewhere that reference illuminants (of which D65 is one) do not necessarily provide the best perceived naturalness.

Islam et al. [19] and Dangol et al. [20] describe an investigation based on 21 LED sources and 3 fluorescent lamps at 2700, 4000 and 6500 kelvins, to determine visual observations by 60 participants of "Naturalness of objects, Visual appearance of the lit environment, and Colourfulness of the Macbeth ColorChecker Chart". This writer's examination of their data showed that the average Naturalness response correlated with Qp and Qg (r^2 scores between 0.4 and 0.7 for the various sources at the different CCTs). Correlation (r^2 score) with the mean value of Qp and Qg has been calculated as 0.28 at 2700 K and ≥ 0.75 for 4000 K and 6500 K. We therefore offer the following simple prediction formula (Equation 1) which, based on this analysis, appears to be valid for CCT > 2700 K:

$$CQ_{Nat} = \frac{1}{2}(Q_p + Q_g) \tag{1}$$

where CQ_{Nat} represents the average Naturalness as defined in [19].

Dangol et al. (2015) [21] described studies of user acceptance of LED office lighting in terms of subjective evaluations of Preference, Naturalness and Colourfulness, undertaken by 40 observers. The method involved total immersion in mock-up offices lit in turn with six LED and two Fluorescent sources at either 4000 or 6500 K and either 300 or 500 lux. The subjective preferences for the lit environment were better explained by Qp plus Qg than by the other metrics used. Bhusal & Dangol [22] conducted a new analysis of the Islam et al. [19] and Dangol et al. [20] data, which showed a "moderate correlation" (Spearman $r = 0.48$) of Naturalness with TM-30 Rg.

Khanh et al. [23] used 5 different sources (including two LED types) viewed by 38 observers in a mock-up office, at 470 lux and CCTs between 2300 and 4100 kelvins, to scale Colour preference, Naturalness and Vividness for comparison against a total of 14 colour quality metrics. In one of the clearest instances of specific guidance for the observers, Naturalness was defined as "Subjective extent of how natural the colour appearance of an identified object (e.g. a rose) is under the current light source compared to the ideal colour appearance in your memory in the way you remember that object." This may help explain the relatively significant correlation of Naturalness with MCRI Rm ($r^2 = 0.72$) in their results.

Seven multi-LED spectra (all with $Ra \geq 96$) at 3220 K were evaluated by 23 observers using a single light booth [24]. Khanh et al. (2017 and 2019) analysed the merged data set, which made several proposals for a prediction formula for the Naturalness colour quality of any source [23, 24], of which one of the simplest is included here as our Equation 2 with r^2 of 0.64:

$$CQ_{Nat} = MCRI + 0.000019(CCT) + 0.087(Q_p) \tag{2}$$

In a study of reddish cosmetic products, presented in a single viewing booth, using 7 multi-LED spectra at 3200 K and 550 lux, Khanh & Bodrogi [25] showed a high correlation of Naturalness with stand-alone indices Qp and MCRI Rm ($r^2 = 0.88$ in both cases). Six "responsive observers" provided the subjective evaluations.

Khanh et al. [26] describe assessments of subjective colour preference, naturalness and vividness of two different colourful still life arrangements viewed in a real room. Coloured objects were illuminated by a four-channel LED light engine with 36 different spectra at four CCTs (3100–5600 K) and 750 lux. Analysis of the results showed a significant dependence of the subjective judgments on the correlated colour temperature. The authors propose a lighting design approach based on, first, the selection of the desired CCT, and then the application of a colour quality formula with specific coefficients applicable to that CCT, to find a measure of the Naturalness quality CQ_{Nat}. There are tables of coefficients to be used for the specific qualities (Preference, Naturalness, Vividness) at each of the four CCTs investigated. This paper goes a long way to demonstrating the complexity of finding a universal definition for a source colour naturalness metric.

Royer et al. [27] used 26 different LED spectra at 3500 K and 215 lux and a total of 28 observers who were asked to provide a range of subjective responses, including "*normalness*" (which is taken here to be analogous to Naturalness) and which is expressed on a scale of 1 to 8, with a "1" representing "most normal". A number of combinations of TM-30 metrics were investigated, and a proposed formula for Normalness (with $r^2 = 0.83$) is given in Equation 3:

$$Normalness = 9.37 – 0.069(Rf) – 3.76(Rcs, h1) \quad (3)$$

where, $Rcs, h1$ is the red chroma shift in the TM-30 system.

The authors also made the strong point that " *... guidance or thresholds derived from this experimental data should not be indiscriminately applied to other contexts*" [27]. Such a caveat almost certainly also applies to all the works cited in this section, and this needs to be borne in mind.

Esposito & Houser [28] investigated 24 SPDs with CCT of 3500 K and illuminance of 650 lux having various TM-30 Rf, Rg and gamut-shape properties. In each case the lighting was assessed by 20 observers viewing a coloured scene in a single display booth containing 12 familiar (natural and manufactured) products; and they were asked to evaluate the colour qualities of Naturalness, Vividness and Preference. In terms of Naturalness, a best fit with the viewer observations in this experiment (with $r^2 = 0.92$) was given by the model in Equation 4:

$$Naturalness = 1.464 + 0.02674(Rf) + 0.188(Rcs, h1) – 15.41(R^2cs, h1) – 0.05305(\psi) + 0.000602(Rf \times \psi) \quad (4)$$

where, ψ is a specified gamut-ellipse property, and the remaining terms are as defined in TM-30-18.

An experimental validation of colour rendition specification criteria based on ANSI/IES TM-30-18 was published by Royer et al. [29]. 25 observers were asked to rate 90 SPDs at three CCTs (2700 K, 3100 K, 3500 K) and 310 lux in terms of Normalness, Saturation, Preference and Acceptability. Viewing was carried out by full immersion in a test room. The primary purpose of the experiment was to explore the performance of a previously proposed 3-tier set of colour specification criteria intended for the guidance of lighting designers and users. Of interest in the present context is the data listed in Table 3 of that paper, which permits the subjective results to be compared against the calculations of TM-30-18 parameters for each SPD group; and a new analysis of the data by the present author showed that Normalness and Rf have a linear relationship ($r^2 = 0.605$). Their Table 4 shows $r^2 = 0.83$ for the prediction of mean Normalness, but it is not clear which model was used [29].

3.3. Spectral Similarity

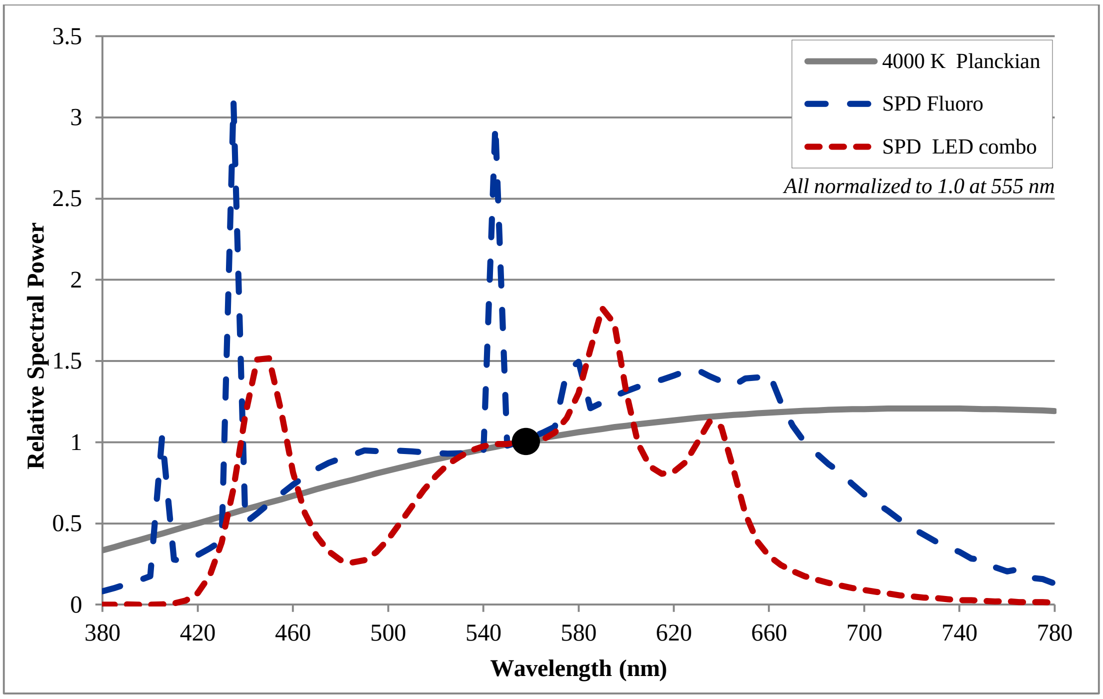
The idea of spectral similarity as a measure of Naturalness was mentioned at the beginning of this paper. The proposal, published by Bridgelux Inc. [1], defined the ASD (average spectral difference) in the following terms*: "The ASD value, expressed as a percentage, always compares a test source to a reference source at the same CCT". The reference source is determined by use of the TM-30 reference-source methodology. The two SPDs are then "Y-normalized so that they are comparable in the visible spectrum". The absolute values of the differences between the two SPDs are sampled at 1 nm intervals from 425 to 690 nm. These values are averaged (arithmetic mean) and the result converted to a percentage to give the ASD. It will be evident that the ASD should be zero to achieve the highest possible Naturalness in this approach.

There does not appear to have been any controlled visual experimentation to verify this quantity as a Naturalness metric. An online article has discussed several criticisms of the ASD, including the following [30]:

- "While spectral similarity is easily defined and calculated, that doesn't establish that it provides useful information, which is perhaps the reason no previous spectral similarity metric has gained traction in the lighting community."
- A number of past experiments have shown that "spectral similarity measures like ASD are not well correlated with people's descriptions of the naturalness of a light source."

Figure 4 is included here as a demonstration of the ASD method which we modified to permit the use of the two previously-defined test sources (from the earlier case study) with 5-nm increments in their SPDs. The spectra were all normalized to unity at 555 nm, although no explicit definition of the normalization wavelength was found in [1].

* Direct quotes in this paragraph are from [1] and the remainder is paraphrased from the same source.



Using a modified Bridgelux calculation (with 5 nm increments in SPDs), the Average Spectral Differences compared with the Planckian reference are: Fluorescent: ASD = 27.3 %, LED combination: ASD = 42.3%.

Figure 4. SPDs of the two 4000 K test sources discussed previously, together with the 4000 K Planckian reference to illustrate the spectral similarity approach

3.4. Summary of Metrics

This paper has presented a concise selection of relevant visual research data published over roughly the past decade. It is not claimed to be exhaustive, and the author acknowledges that contributions have been made by many research groups in this period. Where it has been possible to pinpoint specific standalone metrics that correlate with Naturalness, these are included in Table 3 along with their coefficients of determination.

Table 3. Summary of potential Colour Naturalness metrics with quoted (or deduced) values for r^2

Year	CIE *Ra*	MCRI *Sa*	MCRI *Rm*	CQS *Qa*	CQS *Qf*	CQS *Qp*	CQS *Qg*	TM-30-18 *Rf*
2010	0.77	0.76		0.74				
2013						≥ 0.50	≥ 0.40	
2015	≥ 0.88	≥ 0.36		≥ 0.88	≥ 0.84	≥ 0.64		
2017			0.72					
2018			0.88			0.88		
2020								0.61

Notes: Data are extracted from the references discussed in 3.2 and Table 2, and are identified by year so as to track the emergence of the different fidelity metrics.

r^2 = coefficient of determination (where r = Pearson correlation coefficient) correct to 2 decimal places.

MCRI Rm results have been interpreted from the relevant papers' contents.

Normalness (where used) has been interpreted as equivalent to Naturalness.

In addition to these metrics, there have been the proposals for Naturalness prediction formulae, as in Equations 1 to 4. Some performance data for the different suggested formulae are presented in Table 4.

Table 4. Summary of potential Colour Naturalness formulae with quoted (or deduced) values for r^2

Year	Eq. (1)	Eq. (2)	Eq. (3)	Eq. (4)
2013	0.61			
2017		0.64	0.83	
2019				0.92

Notes: Data are extracted from the references discussed in 3.2 and Table 2, and are identified by year so as to track the emergence of the different fidelity metrics.

r^2 = coefficient of determination (where r = Pearson correlation coefficient) correct to 2 decimal places.

Normalness (where used) has been interpreted as equivalent to Naturalness.

It should be noted that the correlation statistics given in Tables 3 and 4 have been extracted (or directly deduced) from the references listed in Table 2. As such, they are specific to the experimental conditions established by each research group, and are not directly comparable to one another. Thus, a major purpose of this paper is to make a plea for greater standardization of the research efforts in this field.

3.5. Application to Specific SPDs

Let us again, purely as an illustrative exercise, consider the two randomly selected SPDs introduced in the earlier case study, and view the results when the various metrics and formulae are applied to them. A number of these were given in Table 1, and Table 5 lists some additional results to explain the computation of the prediction formulae (Equations 1 to 4). It is notable that the relatively low fidelity of the LED source goes hand-in-hand with its poor performance in the rendition of colours with strong red content; *viz.* in the CIE system its *R9* = –16, and in TM-30-18 its red-bin value $Rcs,h1 = -15$.

Table 5. Additional results for the Fluorescent lamp and LED combination used in the case study

Category	Metric		Fluorescent Lamp	LED Combination
CIE – CRI	R9		95	–16
NIST – CQS v7.5	Qg		102	93
	Qp		98	77
IES – TM-30-18	Rcs,h1		–1	–15
	Esposito's Ψ* [28]		–80°	–70°
Prediction Formulae *Notes for Eq. 4: (i). It is nonlinear (ii). We estimated Ψ values (to ± 2°)	**Equation Number**	**Expected Range**		
	Eq. (1)	0 … 100	100	85
	Eq. (2)	0 … 100	99.6	90.8
	Eq. (3)	8 … 1	6.58	>> 8
	Eq. (4)*	--	–11.9	<< –100

We note that Equations 3 and 4 were derived for well-defined experimental conditions that included light sources of CCT = 3500 K in both cases, while the two sources of our case study both have a nominal CCT = 4000 K. The LED combination is a particularly poor performer for red surfaces, probably explaining the out-of-range results it gives for these two equations. The reader should note that the two example SPDs (as used above) were randomly selected from samples used in an earlier study [4]. They are not intended to be representative of their respective classes (either fluorescent or LED). They are included for the sole purpose of illustrating the uses (and possible merits) of the metrics under discussion.

4. Discussion

In terms of the colour properties of lighting, and the methods for their measurement and classification, we have been on a journey of gradual discovery and development over roughly the last seventy years. At each stage of the process the overarching influence was the marketing of "new and improved" light sources. Two further important factors were the power of the available computing resources at a given time, plus the continuously unfolding knowledge of colour science and colour modelling.

Post-WW II, in the period (the 1950s - 1960s) when fluorescent tubes were first being deployed in large numbers, with new CCTs and spectral compositions that were a complete change from the previously-dominant incandescent filament lamps, the necessity for a colour rendition metric became unavoidable. Calculations were kept as simple as possible, to permit the use of hand-operated calculators or the (later) alternative of batch processing on "main frame" computers using software such as Fortran.

After a number of abortive early attempts to design a colour rendition metric using spectral comparisons and spectral-band methods (e.g. [31]) the CIE colour rendering index (Ra) emerged in 1965 as the most rational and meaningful metric – being based on real psycho-physical data and the best available 3-dimensional colour model at the time. It was updated in 1974 to include the effects of chromatic adaptation, and in 1995 [2] to include a number of (mainly editorial) corrections.

The CRI continued its broad acceptance, with small modifications, through the period of the development of high-intensity discharge lighting and tri-phosphor fluorescent lamps (roughly the 1970s to 1990s). However, its credibility began to suffer later in that period as suspicions grew of so-called "gaming" in which the SPD design was focussed purely on Ra optimization – at the expense, at times, of the actual visual effects of the light source.

These events, together with advances in computing and in the science of colour perception, provided the catalysts for the development of the newer approaches to colour rendition [6-8] – as outlined earlier in this paper. The greatest impetus, however, was provided by the development of high-intensity LEDs and their marketing as high-efficiency, long-life light sources. Unfortunately, early spectral designs were able to meet the then-current CRI criteria while still giving unacceptably poor visual effects – hence the urgent need for alternative metrics.

The range of ideas contained within this paper leads one to suggest the concept of a "*Fidelity Continuum*" (Figure 5) in which we might express Naturalness as a type of "*Visual Fidelity*" – taking it beyond existing fidelity metrics such as Ra, Qf, Rf, etc.

Purely Physical	**...**	***Psycho-Physical (CIE colour matching & colour appearance modelling)***			***Colour Memory Prediction***	**...**	***Visually Based***
ASD etc.	*Spectral Bands*	Ra	Qf	Rf	Rm	...	$R(nat)?$

Figure 5. Illustrating the concept of the Fidelity Continuum

5. Conclusions

It seems safe to say that the preceding content clearly demonstrates a strong interest in the naturalness concept among the members of the lighting community. That is not to say that the findings (in terms of the metrics and predictive equations discussed) can be regarded as definitive. Rather, they are pointers towards possible future approaches, and are indicative of a need for a measure of standardization in research efforts. There would seem to be three essential steps to enable progress in this field, which needs to be underpinned by further visual research:

1. Agreement on a definition for Naturalness in lighting;
2. Creation of a database of SPDs with their corresponding colour Naturalness gradings.
3. Encouragement for collaboration and standardization wherever possible.

In terms of step 1, it has already been noted that Khanh et al. [23] provided a definition as an aid to their observer panel, and this would appear to have much to recommend it. Also, as noted earlier, De Ridder et al. [12] made a similar definition with respect to the assessment of colour naturalness in images. A further suggestion from outside the field of illumination has been put forward by Goodman [32] as part of a project on the feasibility of the measurement of naturalness, and which states that Naturalness is: "The probability that a material or object is perceived as being natural, i.e. perceived as being derived from nature." This paper has also made an attempt at a Naturalness definition – given in the first paragraph of the Introduction (the capacity of a light source to illuminate coloured surfaces in such a way as to provide the "most natural" view of the colours). It ought to be possible to use these concepts to find an appropriate definition to fit the needs of lighting.

Steps 2 and 3 will depend on an agreement regarding a satisfactory definition for Naturalness (step 1) and, in particular, will depend on the ability of research groups to access the funding to carry out the necessary visual experimentation. As an interim step towards a research database, it may be possible for the authors mentioned in this paper to provide tables of SPDs for the sources they have investigated.

6. Declarations

6.1. Data Availability Statement

The data presented in this study are available on request from the corresponding author

6.2. Funding

The author received no financial support for the research, authorship, and/or publication of this article.

6.3. Acknowledgements

- The author is grateful for the facilities provided by the university.
- The author acknowledges the paper "A memory colour quality metric for white light sources" [4] as the source of the data used in the case study, as well as the calculation tool for the memory colour rendering index Rm.
- The author also wishes to acknowledge NIST and IES respectively for their CQS [6] and TM-30 [7] calculation tools that have considerably simplified the calculation work.
- Sincere thanks go to colleagues and co-workers who have contributed to the development of this script.

6.4. Institutional Review Board Statement

Not Applicable.

6.5. Declaration of Competing Interest

The author declares that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the author.

7. References

[1] Bridgelux Inc. (2020). Average Spectral Difference, a new method to make objective comparisons of naturalness between light sources. Bridgelux White Paper, 1-16, California, United Stated.

[2] CIE 13.3-1995. (1995) Method of measuring and specifying colour rendering properties of light sources. CIE (Commission Internationale de l´Eclairage), 1-16, Vienna, Austria.

[3] Philips Lighting. (2021). Dynamic Lighting. Available online: https://www.lighting.philips.co.nz/systems/themes/dynamic-lighting. (accessed on May 2022).

[4] Smet, K. A. G., Ryckaert, W. R., Pointer, M. R., Deconinck, G., & Hanselaer, P. (2012). A memory colour quality metric for white light sources. Energy and Buildings, 49, 216–225. doi:10.1016/j.enbuild.2012.02.008.

[5] Van Trigt, C. (1999). Color rendering, a reassessment. Color Research & Application., 24(3), 197-206. doi:10.1002/(SICI)1520-6378(199906)24:3%3C197::AID-COL6%3E3.0.CO;2-S

[6] Davis, W. & Ohno, Y (2010). Color quality scale. Optical Engineering, 49(3), 033602. doi:10.1117/1.3360335.

[7] IES TM-30-15. (2015). IES Method for Evaluating Light Source Color Rendition. Illuminating Engineering Society of North America (IES), New York, United States.

[8] CIE 224-2017. (2017). Colour Fidelity Index for accurate scientific use. CIE (Commission Internationale de l´Eclairage), 1-44 Vienna, Austria.

[9] Smet, K. (2018). IES TM-30. Face Seminar Session Boom, October 2-3, 2018, Light & Lighting Laboratory, Gent, Belgium. Available online: https://assets.website-files.com/57698e8d8a6044ef2ce2d17b/5bc842dfe3e80e99724e7b20_SMET_TM30.pdf. (accessed on April 2022).

[10] Smet, K. A. G., & Hanselaer, P. (2016). Memory and preferred colours and the colour rendition of white light sources. Lighting Research and Technology, 48(4), 393–411. doi:10.1177/1477153514568584.

[11] Ebner, F., & Fairchild, M. D. (1998). Development and testing of a color space (IPT) with improved hue uniformity. Final Program and Proceedings - IS and T/SID Color Imaging Conference, 8–13.

[12] De Ridder, H., Blommaert, F. J. J., & Fedorovskaya, E. A. (1995). Naturalness and image quality: Chroma and hue variation in color images of natural scenes. Human Vision, Visual Processing and Digital Display VI, San Jose. SPIE, 2411, 51–61.

[13] Yendrikhovskij, S. N., Blommaert, F. J. J., & De Ridder, H. (1999). Color reproduction and the naturalness constraint. Color Research and Application, 24(1), 52–67. doi:10.1002/(SICI)1520-6378(199902)24:1<52::AID-COL10>3.0.CO;2-4.

[14] Jost-Boissard, S., Fontoynont, M., & Blanc-Gonnet, J. (2009). Perceived lighting quality of LED sources for the presentation of fruit and vegetables. Journal of Modern Optics, 56(13), 1420–1432. doi:10.1080/09500340903056550.

[15] Jost-Boissard, S., Avouac, P., & Fontoynont, M. (2015). Assessing the colour quality of LED sources: Naturalness, attractiveness, colourfulness and colour difference. Lighting Research and Technology, 47(7), 769–794. doi:10.1177/1477153514555882.

[16] Smet, K. A. G., Ryckaert, W. R., Pointer, M. R., Deconinck, G., & Hanselaer, P. (2010). Memory colours and colour quality evaluation of conventional and solid-state lamps. Optics Express, 18(25), 26229–26244. doi:10.1364/oe.18.026229.

[17] Smet, K., Ryckaert, W. R., Pointer, M. R., Deconinck, G., & Hanselaer, P. (2011). Correlation between color quality metric predictions and visual appreciation of light sources. Optics Express, 19(9), 8151–8166. doi:10.1364/oe.19.008151.

[18] Nascimento, S. M. C., & Masuda, O. (2012). Psychophysical optimization of lighting spectra for naturalness, preference, and chromatic diversity. Journal of the Optical Society of America A, 29(2), A144–A151. doi:10.1364/josaa.29.00a144.

[19] Islam, M. S., Dangol, R., Hyvärinen, M., Bhusal, P., Puolakka, M., & Halonen, L. (2013). User preferences for LED lighting in terms of light spectrum. Lighting Research and Technology, 45(6), 641–665. doi:10.1177/1477153513475913.

[20] Dangol, R., Islam, M., Lisc, M. H., Bhusal, P., Puolakka, M., & Halonen, L. (2013). Subjective preferences and colour quality metrics of LED light sources. Lighting Research and Technology, 45(6), 666–688. doi:10.1177/1477153512471520.

[21] Dangol, R., Islam, M. S., Hyvärinen, M., Bhushal, P., Puolakka, M., & Halonen, L. (2015). User acceptance studies for LED office lighting: Preference, naturalness and colourfulness. Lighting Research and Technology, 47(1), 36–53. doi:10.1177/1477153513514424.

[22] Bhusal, P., & Dangol, R. (2017). Performance of different metrics proposed to CIE TC 1-91. International Journal of Sustainable Lighting, 19(2), 91–103. doi:10.26607/ijsl.v19i2.36.

[23] Khanh, T. Q., Bodrogi, P., Vinh, Q. T., & Stojanovic, D. (2017). Colour preference, naturalness, vividness and colour quality metrics, Part 1: Experiments in a room. Lighting Research and Technology, 49(6), 697–713. doi:10.1177/1477153516643359.

[24] Khanh, T. Q., Bodrogi, P., Vinh, Q. T., & Stojanovic, D. (2017). Colour preference, naturalness, vividness and colour quality metrics, Part 2: Experiments in a viewing booth and analysis of the combined dataset. Lighting Research & Technology, 49(6), 714-726. doi:10.1177/1477153516643570.

[25] Khanh, T. Q., & Bodrogi, P. (2018). Colour preference, naturalness, vividness and colour quality metrics, Part 3: Experiments with makeup products and analysis of the complete warm white dataset. Lighting Research and Technology, 50(2), 218–236. doi:10.1177/1477153516669558.

[26] Khanh, T. Q., Bodrogi, P., Vinh, Q. T., Guo, X., & Anh, T. T. (2018). Colour preference, naturalness, vividness and colour quality metrics, Part 4: Experiments with still life arrangements at different correlated colour temperatures. Lighting Research and Technology, 50(6), 862–879. doi:10.1177/1477153517700705.

[27] Royer, M. P., Wilkerson, A., Wei, M., Houser, K., & Davis, R. (2017). Human perceptions of colour rendition vary with average fidelity, average gamut, and gamut shape. Lighting Research and Technology, 49(8), 966–991. doi:10.1177/1477153516663615.

[28] Esposito, T., & Houser, K. (2019). Models of colour quality over a wide range of spectral power distributions. Lighting Research and Technology, 51(3), 331–352. doi:10.1177/1477153518765953.

[29] Royer, M. P., Wei, M., Wilkerson, A., & Safranek, S. (2020). Experimental validation of colour rendition specification criteria based on ANSI/IES TM-30-18. Lighting Research and Technology, 52(3), 323–349. doi:10.1177/1477153519857625.

[30] Livingston, J., Royer, M., Whitehead, L. (2020). Spectral Similarity as a Measure of Naturalness? FIRES Forum 11 September 2020, Illuminating Engineering Society. Available online: https://www.ies.org/fires/spectral-similarity-as-a-measure-of-naturalness/ (accessed on April 2022).

[31] Crawford, B. H. (1963). The colour rendering properties of illuminants: The application of psychophysical measurements to their evaluation. British Journal of Applied Physics, 14(6), 319–328. doi:10.1088/0508-3443/14/6/302.

[32] Goodman, T. (2008). Measurement of Naturalness. 12th IMEKO Man, Science and Measurement Symposium, September 3-5, Annecy, France.