

Color Quality Metrics – Recent Progress and Future Perspective

Y. Ohno

Sensor Science Division, National Institute of Standards and Technology
100 Bureau Drive, Gaithersburg Maryland 20899 USA

ABSTRACT

The color quality of light sources for lighting includes the effects of white light chromaticity, as well as color rendition characteristics, including color fidelity and the aspects related to color preference and perception. In this paper, the current standards and the latest research on white light chromaticity and color saturation preferences are overviewed. The characteristics of TM-30 Fidelity Index and Gamut Index from the Illuminating Engineering Society (IES) are examined in comparison to the Color Rendering Index (CRI). Current activities in International Commission on Illumination (CIE) related to these and future perspective of color quality metrics are discussed.

1. INTRODUCTION

It is critical for light-emitting diode (LED) sources for general lighting to have acceptable color quality suitable for intended applications. There are two primary aspects of color quality of lighting sources: white light chromaticity and color rendition. These two aspects are interrelated, and both of these characteristics are critical for general lighting. The chromaticity ranges for general lighting products have been well standardized for many years [1, 2]. However, new research is raising questions about whether the current chromaticity ranges are most appropriate for all lighting applications. The current standards and recent studies on white light chromaticity are overviewed and discussed in section 2. Color rendition is considered largely in two aspects – color fidelity and an aspect related to color preferences. The current standardized metric, Color Rendering Index (CRI) [3], is designed to evaluate color fidelity. However, color fidelity does not always correlate well with visual evaluations of the color quality of lighting, and thus there is a need for another metric that evaluates these aspects of color preference and perception. This issue is discussed in section 3 with the results of recent studies conducted at National Institute of Standards and Technology (NIST). TM-30 [4] from the Illuminating Engineering Society of North America (IES) was recently published as a proposed new method for evaluating color rendition. This metric is discussed and compared to CRI in section 4, and the related CIE activities and future considerations are discussed.

2. WHITE LIGHT CHROMATICITY STANDARDS AND RECENT RESEARCH

The chromaticity of light sources for general lighting is normally chosen to be around the Planckian locus or the daylight locus, which are generally considered as the center locations for white light, as specified in the existing standards [1, 2]. Figure 1, for example, shows the American National Standards Institute (ANSI) chromaticity specifications for solid state lighting products [2]. While these standards are widely used, new research raises questions as to whether the current chromaticity ranges are most appropriate for all lighting applications. Shifts in chromaticity across the Planckian locus are expressed by the term, D_{uv} (symbol D_{uv}), defined in the standard [2] as shown in Fig. 2, and its use is discussed in elsewhere [5].

A recent study on a vision experiment [6] reported that perceived neutral white points determined by the subjects were $D_{uv} \approx -0.01$ at 2700 K to 3500 K, $D_{uv} \approx 0.00$ at

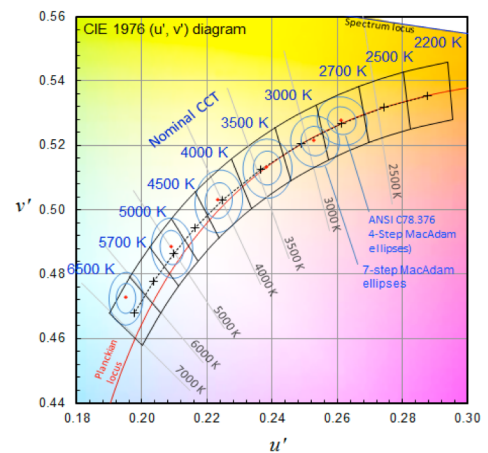


Fig. 1 The chromaticity specifications for solid-state lighting products by ANSI C78.377, 2015 revision [2] on the CIE (u' , v') diagram.

4000 K, and $D_{uv} \approx 0.005$ at 6500 K (though the report did not use the Duv measure). This experiment was done with a lighting booth with white inner walls and no objects inside. Another experimental study [7] showed that subjects, viewing an office miniature, freely adjusted the chromaticity of illumination as they preferred, choosing points mostly below the Planckian locus, with an average Duv around -0.014.

In 2013, vision experiments were conducted at NIST for white chromaticity points perceived most natural in a real-size interior room setting using the NIST Spectrally Tunable Lighting Facility (STLF) [8], participated by 18 subjects. The average results are plotted with the ANSI specifications in Fig. 3. The results showed that the lights with average $D_{uv} \approx -0.015$, far below the Planckian locus, appeared most natural to the subjects [9]. The results from this experiment were verified with additional experiments conducted in 2015 [10] with 21 subjects, using modified spectra of lights, addressing the question raised by other researchers [11]. Further studies are desired to verify the applicability of these results in various real application conditions.

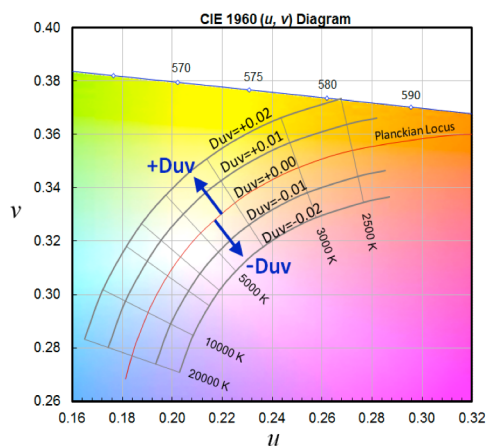


Figure 2. The scale of Duv shown on the CIE 1960 (u, v) diagram.

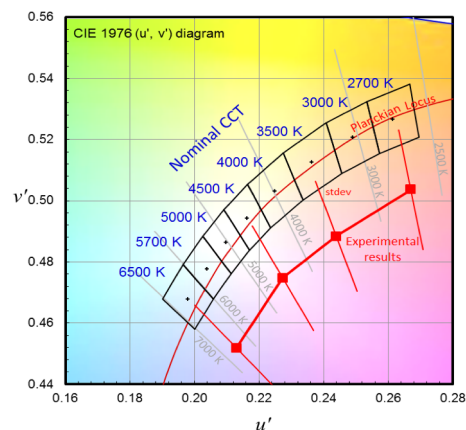


Figure 3. Average chromaticities judged to be most natural, plotted with ANSI C78.377 chromaticity specifications for SSL products.

3. RECENT RESEARCH ON COLOR SATURATION PREFERENCE

There have been many research articles about this topic, and it is generally well known that people prefer slightly over-saturated object colors in illuminated indoor scenes. Increased chroma, however, deviates from the fidelity, causing decreases in the CRI R_a score. To determine the level of chroma saturation that is generally most preferred, a large-scale vision experiment was conducted at NIST in 2014 using the NIST STLF [12]. The experiment was conducted with 20 subjects at correlated color temperatures (CCTs) of 2700 K, 3500 K, 5000 K (all $D_{uv} = 0$) and 3500 K with $D_{uv} = -0.015$. There were four different viewing targets: mixed fruits and vegetables and the entire room, skin tone, red fruits/vegetables only, and green fruits/vegetables only. Pairs of lights were presented, one of which was always the neutral saturation (the same level as the CRI reference illuminant). Figure 4 shows the STLF room setting (left) and the nine different levels of chroma saturation used in the experiment (right). The right figure shows the plots of the 15 test color samples used in Color Quality Scale (CQS) [13]. Figure 5 shows the average results for all subjects and all CCT/Duv conditions for each viewing target. The vertical axis is the percentage that subjects preferred the light with a given chroma saturation (horizontal axis) over the neutral light ($\Delta C_{ab}^* = 0$). The chroma saturation (ΔC_{ab}^*) is calculated for the red test sample in the CQS. The results were unexpectedly consistent across all CCT/Duv conditions and for all targets, showing that the subjects most preferred the chroma increase ΔC_{ab}^* between 4 to 8. The subjects' preference starts decreasing at larger saturations, as the objects start appearing unnatural. The subjects' preference at saturation level $\Delta C_{ab}^* \approx 16$ (most saturated condition) is still comparable to that of the neutral saturation ($\Delta C_{ab}^* = 0$).

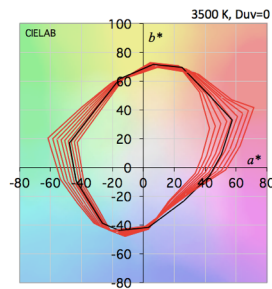
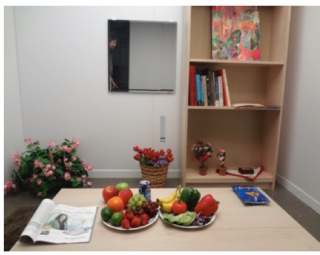


Figure 4. The experimental set up (left) and the CIELAB (a^* , b^*) plots of the 15 CQS samples for the nine different saturation levels used for the NIST experiment (3500 K, $D_{uv}=0$).

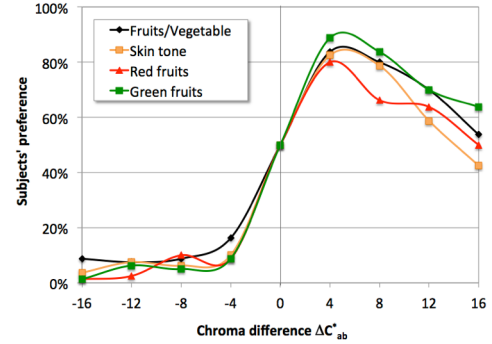


Figure 5. Subjects' preferences for all CCT/Duv conditions for the different viewing targets, as a function of the chroma difference from the reference illuminant.

Figure 6 shows the 3500 K results (average of all participants, all targets) plotted with the CRI R_a values and other metrics' values (discussed in section 4) for these lights. The R_a value sharply decreases as the chroma saturation increases from neutral. The CRI penalizes these preferred lights with increased chroma. For example, if $R_a \geq 80$ is required in a regulation, many preferred lights (shaded area) are disqualified. This sharp slope results from the heavy weights in the red region in the $W^*U^*V^*$ color space (see section 4). Further studies are desired on the effects for longer-term use of such preferred lights in real applications.

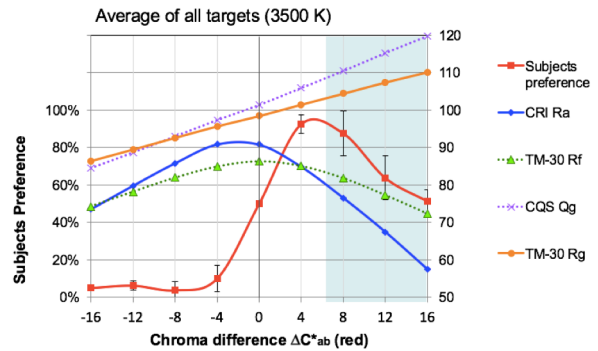


Figure 6. The 3500 K results plotted with the CRI R_a values and other metric values.

4. ANALYSES OF IES TM-30 COMPARED TO CRI

The IES TM-30 [4] was developed to address the shortcomings of CRI experienced with LED light sources for general lighting [14]. A major problem is that CRI is a color fidelity metric and it does not assess other aspects related to color preference and perception of illuminated scenes. The CRI R_a scores do not correlate well with visual evaluation for preferred light sources as discussed in the previous sections. The earlier research, Color Quality Scale [13], attempted to address this issue, but could not be agreed in CIE as an update of CRI, mainly due to the difficulty in defining what it measures scientifically. The CIE then decided to take the direction to develop new metrics separately, one for color fidelity (dealt with by TC1-90) and another for the aspects other than color fidelity (dealt with by TC1-91). IES TM-30 was developed as a proposal from IES in this direction, to provide an improved color fidelity metric and a means to assess the aspects related to color preference. Thus, TM-30 is a two-metric system, consisting of Fidelity Index R_f and Gamut Index R_g . The R_f is intended as an improvement of CRI R_a as a color fidelity metric. However, a color fidelity metric alone is not sufficient to assess all aspects of color rendition. Gamut Index is intended to provide preference-related information, though it does not directly indicate the degree of preference. A research article on the details of TM-30 is available [15].

One of the significant problems of CRI as a color fidelity metric is the very non-uniform object color space $W^*U^*V^*$ used to evaluate color differences of test color samples. Figure 7 shows how distorted the $W^*U^*V^*$ color space is for a 2700 K Planckian source and standard daylight (D65). As shown, this distortion is less prominent at higher CCTs. This distortion causes the color differences in the red-green region (particularly red) to be weighted heavily and the yellow-green region very lightly

weighted in CRI. This feature of CRI might be suitable for preference aspects, but is a significant problem as a color fidelity metric. CIE 1976 $L^*a^*b^*$ (CIELAB) [16] is the current CIE standard object color space. CAM02UCS is the latest color space based on the color appearance model [17] and has been experimentally verified that the color differences calculated on this color space correlate well with the perceived color differences of objects of various colors, better than CIELAB.

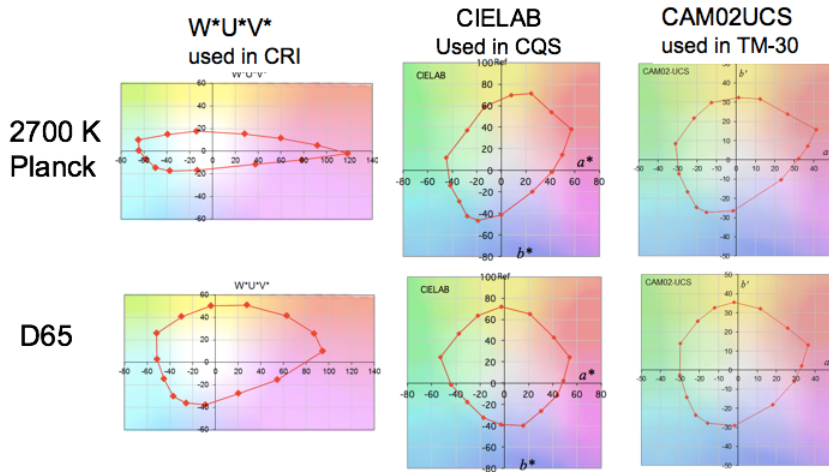


Figure 7. Two-dimensional plots of the 15 saturated color samples of CQS under 2700K Planckian radiation and under D65 illumination on $W^*U^*V^*$, CIELAB, and CAM02UCS object color spaces.

Due to the nonuniformity of the color space, there are two cases of score deviation in CRI; (1) light sources that create relatively large chroma increases in yellow-blue direction (e.g., triphosphor fluorescent lamps and some phosphor-type LED lamps) tend to get higher CRI R_a score than R_f , (2) light sources that create relatively large chroma increases in red-green directions (e.g., some RGB or RGBA LED sources) tend to get lower R_a scores than R_f . Such examples of light source are shown in Fig. 8, where (a) and (b) are examples of case (1) above, and (c) is an example of case (2) above. Case (2) is also demonstrated in Fig. 6 above. Compared to CRI, TM-30 R_f is designed to evaluate the color differences of all hue colors evenly and more accurately as perceived by human observers.

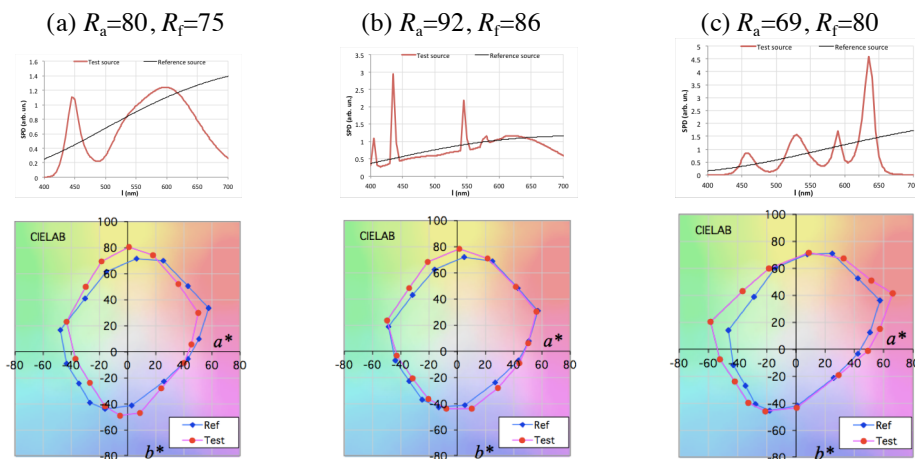


Figure 8. The spectral power distributions of three example light sources and the CIELAB plots of the CQS 15 samples for these sources – (a) a phosphor type LED lamp, (b) fluorescent lamp, both showing strong yellow-blue saturation, and (c) an RGBA LED source showing strong red-green saturation.

Another significant problem of CRI is the use of only eight medium-saturated test color samples, which can cause anomalies (unreasonably higher value than real performance) in the R_a score.

Figure 9 shows the spectral reflectance curves of the test color samples of (a) CRI, (b) CQS, and (c) TM-30, plotted with an RGB LED source spectrum. This light source has unacceptably poor rendering of red objects ($R_0 = -97$) though it produces a generally acceptable value of $R_a = 82$. It is observed that the CRI test sample curves are fairly flat in the red region, which means the R_a score is less sensitive to the location of the red peak, allowing a good R_a score in spite of poor red rendering. The TM-30 and CQS detect this problem and the effect is reflected in the score. The Q_f in the figure

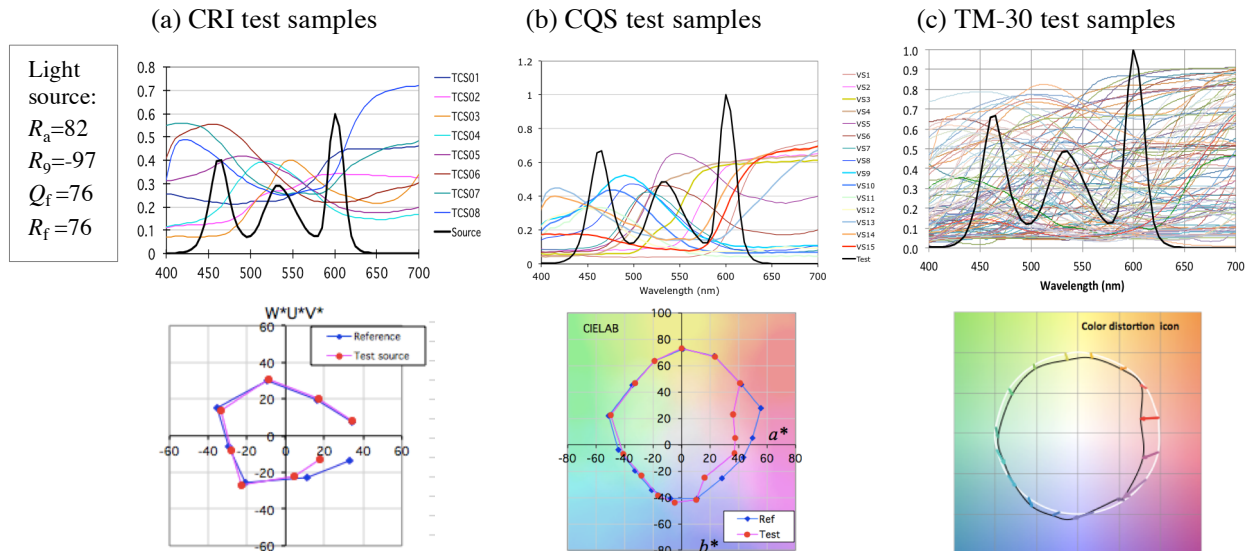


Figure 9. The spectral reflectance curves of the test color samples of (a) CRI, (b) CQS, and (c) TM-30, plotted with the spectral distribution of an RGB LED source, and color gamut presentations of the RGB source in each metric.

is the color fidelity scale in CQS [13].

Another example of an anomaly in CRI R_a is shown in Fig. 10, which shows a case of an RGB LED source optimized for high R_a score. The TM-30 Color Vector Graphics show significantly large hue shifts in the blue-green region, however, the CRI eight samples do not detect this problem.

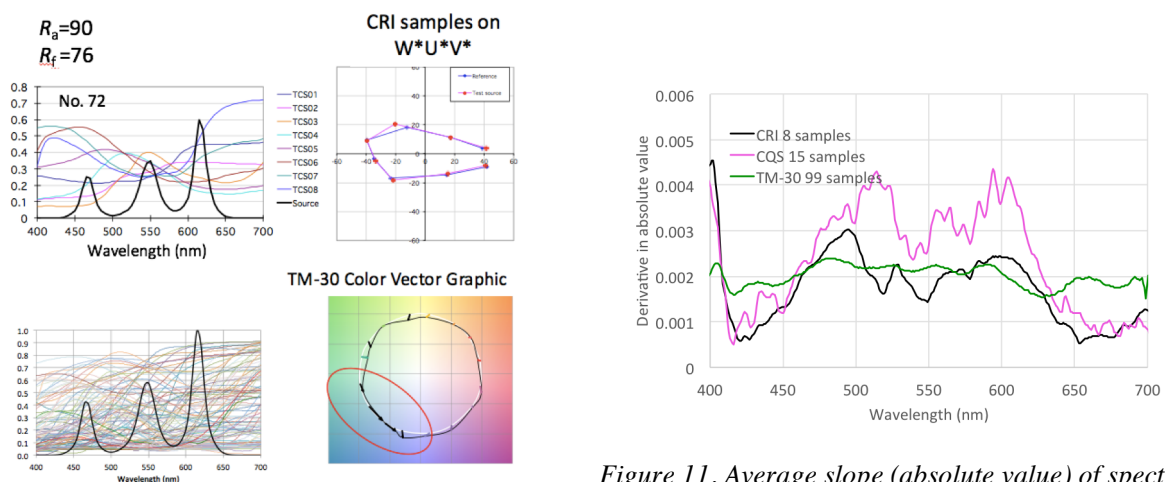


Figure 10. An example of RGB source showing a large difference in value of R_a and R_f .

Figure 11. Average slope (absolute value) of spectral reflectance curves of all test color samples of CRI, CQS, and TM-30

The slopes of the spectral reflectance curves in Fig. 9 imply how sensitively the spectral component of the light source at that wavelength affects the calculation of the metric, though the effects are compounded with integration by the color matching functions. Figure 11 shows the average slopes (in absolute value) of all the test color samples of CRI, CQS, and TM-30 at each wavelength. The curves of CRI and CQS go down at near both ends (deep blue and deep red) of the spectral region, while TM-30 curves are kept fairly flat. This feature of TM-30 allows the metric to evaluate all (possible) object colors uniformly, while it is considered that CRI may not be able to evaluate accurately the colors of some objects having spectral reflectance curves changing rapidly in the deep red and deep blue region. Such colors are typically purple. This is the main reason that the R_f scores for narrow-band spectra that do not cover enough wide spectral region (e.g., triphosphor fluorescent lamps and RGBA LED sources) are lower than the R_a scores.

5. CIE ACTIVITIES AND FUTURE PROSPECT

The IES published the position statement of CRI (2015) [18], which indicates that TM-30 is not a required standard and is issued for evaluation by the international lighting community. The R_f metric was proposed for consideration in CIE TC1-90, which accepted it as the metric for the first TC draft and work is in progress. However, a color fidelity metric alone is not sufficient to evaluate all aspects of color rendition of light sources, and another metric to evaluate the aspect related to color preference is needed, and is being addressed in TC1-91. CIE also published a position statement on CRI and color quality metrics [19], which clarified these points. The TM-30 R_g (Gamut Index) is proposed for inclusion in CIE TC1-91 report, but the gamut area is not sufficient to address the aspects of general color preference. CIE is encouraging further research on this topic, and hopes to develop a recommendation of a more useful color preference-related metric in the future. Then, it is expected that two metrics (a color fidelity metric and a color preference-related metric) will be used to specify and evaluate light sources for general lighting.

References

- [1] IEC 60081-1997, Double-capped fluorescent lamps – Performance specifications, Annex D.
- [2] ANSI_NEMA_ANSLG, C78.377-2015 Specifications for the Chromaticity of Solid State Lighting Products.
- [3] CIE 13.3: 1995, Method of measuring and specifying colour rendering properties of light sources.
- [4] IES TM-30-15, IES Method for Evaluating Light Source Color Rendition.
- [5] Y. Ohno, Y. Practical Use and Calculation of CCT and Duv, *LEUKOS*, 10:1, 47-55 (2013).
- [6] M. S. Rea and J. P. Freyssinier, 2013, White Lighting, *Color Research and Application*, **38**- 2.
- [7] E. E. Dikel, G. J. Burns, J. A. Veitch, S. Mancini, and G. R. Newsham, Preferred Chromaticity of Color-Tunable LED Lighting, *LEUKOS*, 10:2, 101-115, 101-115, 2014.
- [8] C. Miller et al, NIST spectrally tunable lighting facility for color rendering and lighting experiments, in *Proc. CIE 2009: Light and Lighting Conference*. 5 pages, 2009.
- [9] Y. Ohno and M. Fein, Vision Experiment on Acceptable and Preferred White Light Chromaticity for Lighting, CIE x039:2014, pp. 192-199 (2014).
- [10] Y. Ohno and S. Oh, Vision Experiment II on White Light Chromaticity for Lighting, CIE x042:2016, pp. 175-184 (2016).
- [11] M. Wei and K. Houser, 2015. What is the cause of apparent preference for sources with chromaticity below the blackbody locus?, *Leukos*, DOI: 10.1080/15502724.2015.1029131.
- [12] Y. Ohno, M. Fein, and C. Miller, Vision Experiment on Chroma Saturation for Color Quality Preference, Proc. 28th Session of CIE, CIE 216: 2015, pp. 60-69 (2015).
- [13] W. Davis and Y. Ohno, 2010. Color Quality Scale,” *Optical Engineering*, 033602, **49**, 3, 033602-1 to 033602-16.
- [14] CIE 177: 2007, Colour rendering of white LED light sources.
- [15] A. David et al, Development of the IES method for evaluating the color rendition of light, *Optics Express* Vol. 23, Issue 12, pp. 15888-15906 (2015).
- [16] ISO 11664-4:2008(E)/CIE S 014-4/E:2007: Colorimetry — Part 4: CIE 1976 L*a*b* Colour Space
- [17] CIE 159:2004: A Colour Appearance Model for Colour Management Systems: CIECAM02.
- [18] IES PS-8-15 Color Rendering Index, <https://www.ies.org/PDF/PositionStatements/PS-8-15.pdf>.
- [19] CIE Position Statement on CRI and Color Quality Metrics,

http://www.cie.co.at/index.php?i_ca_id=981 (2015).