

Review of measures for light-source color rendition and considerations for a two-measure system for characterizing color rendition

Kevin W. Houser,^{1,*} Minchen Wei,¹ Aurélien David,²
Michael R. Krames,² and Xiangyou Sharon Shen³

¹Department of Architectural Engineering, The Pennsylvania State University, University Park, PA, 16802, USA

²Soraa, Inc., Fremont, CA 94555, USA

³Inno-Solution Research LLC, 913 Ringneck Road, State College, PA 16801, USA
khouser@enr.psu.edu

Abstract: Twenty-two measures of color rendition have been reviewed and summarized. Each measure was computed for 401 illuminants comprising incandescent, light-emitting diode (LED) -phosphor, LED-mixed, fluorescent, high-intensity discharge (HID), and theoretical illuminants. A multidimensional scaling analysis (Matrix Stress = 0.0731, $R^2 = 0.976$) illustrates that the 22 measures cluster into three neighborhoods in a two-dimensional space, where the dimensions relate to color discrimination and color preference. When just two measures are used to characterize overall color rendition, the most information can be conveyed if one is a reference-based measure that is consistent with the concept of color fidelity or quality (e.g., Q_a) and the other is a measure of relative gamut (e.g., Q_g).

©2013 Optical Society of America

OCIS codes: (330.1690) Color; (330.1715) Color, rendering and metamerism; (230.3670) Light-emitting diodes.

References and links

1. CIE, "Methods of measuring and specifying colour rendering properties of light sources," in CIE 13 (CIE, Vienna, Austria, 1965).
2. W. Walter, "How meaningful is the CIE color rendering index?" *Light Design Appl.* **11**(2), 13–15 (1981).
3. T. Seim, "In search of an improved method for assessing the colour rendering properties of light sources," *Lighting Res. Tech.* **17**(1), 12–22 (1985).
4. K. W. Houser, "Lighting for quality," *Light Design Appl.* **32**(11), 4–7 (2002).
5. J. A. Worthey, "Color rendering: asking the question," *Color Res. Appl.* **28**(6), 403–412 (2003).
6. CIE, "TC 1-62: Color rendering of white LED light sources," in CIE 177:2007 (CIE, Vienna, Austria, 2007).
7. CIE, "TC 1-69: Color rendition by white light sources," (Accessed Nov 18, 2012) http://div1.cie.co.at/?i_ca_id=549&pubid=239.
8. CIE, *Division 1: vision and color, meeting minutes*, (Taipei, Taiwan, Sep. 26–27, 2012), 26–28.
9. R. Luo, "An update of the div. 1 meeting in Taipei," e-mail distributed to TC1–69 listserv. Oct 10, 2012.
10. X. Guo and K. W. Houser, "A review of color rendering indices and their application to commercial light sources," *Lighting Res. Tech.* **36**(3), 183–199 (2004).
11. M. P. Royer, K. W. Houser, and A. M. Wilkerson, "Color discrimination capability under highly structured spectra," *Color Res. Appl.* (Online) Nov 2011. 9 pgs. DOI:10.1002/col.20702.
12. D. B. Judd, "A flattery index for artificial illuminants," *Illum. Eng. (USA)* **62**, 593–598 (1967).
13. P. J. Bouma, "Physical aspects of colour; an introduction to the scientific study of colour stimuli and colour sensations," (Eindhoven: Philips Gloeilampenfabrieken (Philips Industries) Technical and Scientific Literature Dept. (1948).
14. W. A. Thornton, "The quality of white light," *Lighting Des. Appl.* **12**, 51–52 (1972).
15. J. A. Schanda, "A combined colour preference – colour rendering index," *Light. Res. Tech.* **17**(1), 31–34 (1985).
16. W. A. Thornton, "A validation of the color-preference index," *J. Illum. Eng. Soc.* **4**(1), 48–52 (1974).
17. W. Davis and Y. Ohno, "Color quality scale," *Opt. Eng.* **49**(3), 033602 (2010).
18. Y. Ohno and W. Davis, "NIST CQS version 7.5," Excel Software, Sep. 10, 2009, (personal communication, 2009).
19. Y. Ohno, NIST, "CQS 9.0.c (Win).xls," (personal communication, 2012).
20. CIE, "Colorimetry, 3rd edition," in CIE15:2004 (CIE, Vienna, Austria, 2004).
21. M. Wei and K. W. Houser, "Status of solid-state lighting based on entries to the 2010 US DOE Next Generation Luminaire competition," *Leukos.* **8**(4), 237–259 (2012).

22. K. Smet, University of British Columbia, and M. R. Luo, University of Leeds, “n-CRI v9.xlsm,” (personal communication, 2012)
23. K. Smet, University of British Columbia, “MemoryCRI.xls,” (personal communication, 2012)
24. M. Wei and K. W. Houser, “Colour discrimination of seniors with and without cataract surgery under illumination from two fluorescent lamp types,” in *Proceedings of CIE 2012 Lighting Quality and Energy Efficiency*, (Hangzhou, China, 2012), 359–368.
25. K. W. Houser and R. B. Gibbons, “Composite CRI,” *J. Illum. Eng. Soc.* **28**(1), 117–129 (1999).
26. M. Krames, Soraa Inc., 6500 Kaiser Drive, Fremont, CA, 94555, (personal communication, 2012).
27. F. Szabo, Virtual Environment and Imaging Technologies Laboratory, Department of Electrical Engineering and Information Systems, University of Pannonia, Veszprém, Hungary, “HRI.xls,” (personal communication, 2012).
28. G. B. Buck 2nd and H. C. Froelich, “Color characteristics of human complexions,” *Illum. Eng.* **43**(1), 27–49 (1948).
29. C. L. Sanders, “Color preferences for natural objects,” *Illum. Eng. (USA)* **54**, 452–456 (1959).
30. M. S. Rea and J. P. Freyssinier-Nova, “Color rendering: a tale of two metrics,” *Color Res. Appl.* **33**(3), 192–202 (2008).
31. M. S. Rea and J. P. Freyssinier, “Color rendering: beyond pride and prejudice,” *Color Res. Appl.* **35**(6), 401–409 (2010).
32. M. S. Rea, “A practical and predictive two-metric system for characterizing the color rendering properties of light sources used for architectural applications,” in *Proc. of SPIE-OSA Vol. 7652. International Optical Design Conference 2010*, 765206–1 – 765206–7.
33. M. S. Rea and J. P. Freyssinier, “The class A color designation for light sources,” in *Proc. Of Experiencing Light 2012: International Conference on the Effects of Light on Wellbeing*, (Eindhoven, The Netherlands, 2012).
34. A. Žukauskas, R. Vaicekauskas, F. Ivanauskas, H. Vaitkevicius, P. Vitta, and M. S. Shur, “Statistical approach to color quality of solid-state lamps,” *IEEE J. Sel. Top. Quantum Electron.* **15**(6), 1753–1762 (2009).
35. K. Smet, W. R. Ryckaert, M. R. Pointer, G. Deconinck, and P. Hanselaer, “Correlation between color quality metric predictions and visual appreciation of light sources,” *Opt. Express* **19**(9), 8151–8166 (2011).
36. J. P. Freyssinier and M. S. Rea, “Class A color classification for light sources used in general illumination,” in *Proc. of Light Sources 2012: The 13th International Symposium on the Science and Technology of Lighting*, June 24–29, 2012, Troy, New York, 337–338, (2012).
37. M. G. Figueiro, K. Appleman, J. D. Bullough, and M. S. Rea, “A discussion of recommended standards for lighting in the NICU,” *J. Perinatol.* **26**, S19–S26 (2006).
38. R. Dangol, M. Islam, M. Hyvarinen, P. Bhusal, M. Puolakka, and L. Halonen, “Subjective preferences and colour quality metrics of LED light sources,” *Lighting Res. Tech.* published online, (Jan 4, 2013), DOI: 10.1177/1477153512471520.
39. P. Bodrogi, S. Brückner, and T. Q. Khanh, “Ordinal scale based description of color rendering,” *Color Res. Appl.* **36**(4), 272–285 (2011).
40. K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, G. Deconinck, and P. Hanselaer, “A memory colour quality metric for white light sources,” *Energy Build.* **49**, 216–225 (2012).
41. K. A. G. Smet, J. Schanda, L. Whitehead, and M. R. Luo, “TC1-69 report: The CRI2012 colour rendering index,” unpublished report distributed to CIE TC1–69 in advance of publication in a journal (2012).
42. M. D. Fairchild, *Color appearance models, 2nd edition* (Wiley Interscience, 2005), 408 p.
43. P. R. Boyce and R. H. Simons, “Hue discrimination of light sources,” *Lighting Res. Tech.* **9**(3), 125–140 (1977).
44. ANSI, “Specifications for the chromaticity of solid state lighting products,” in ANSI/ANSI C78.377–2008, (National Electrical Manufacturers Association, Rosslyn, Virginia, 2008).
45. K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, G. Deconinck, and P. Hanselaer, “Memory colours and colour quality evaluation of conventional and solid-state lamps,” *Opt. Express* **18**(25), 26229–26244 (2010).
46. P. van der Burgt and J. van Kemenade, “About color rendition of light sources: The balance between simplicity and accuracy,” *Color Res. Appl.* **35**(2), 85–93 (2010).
47. CIE, “Method of measuring and specifying color rendering properties of light sources,” in CIE13.2–1995 (CIE, Vienna, Austria, 1995).
48. K. Hashimoto, T. Yano, M. Shimizu, and Y. Nayatani, “New method for specifying color-rendering properties of light sources based on feeling of contrast,” *Color Res. Appl.* **32**(5), 361–371 (2007).
49. W. A. Thornton, “Color-discrimination index,” *J. Opt. Soc. Am.* **62**(2), 191–194 (1972).
50. H. Xu, “Colour rendering capacity of illumination,” *J. Illum. Eng. Soc.* **13**(2), 270–276 (1984).
51. H. Xu, “Colour rendering capacity and luminous efficiency of a spectrum,” *Lighting Res. Tech.* **25**(3), 131–132 (1993).
52. S. A. Fotios, “The perception of light sources of different colour properties,” Doctor of Philosophy Thesis, UMIST, United Kingdom (1997).
53. M. R. Pointer, “Measuring colour rendering—a new approach,” *Lighting Res. Tech.* **18**(4), 175–184 (1986).
54. M. R. Pointer, “Measuring colour rendering—a new approach II,” NPL Report: DQL-OR 007 (2004).

1. Introduction

Limitations of the general color rendering index (R_a) provided by the Commission Internationale de l’Éclairage (CIE), first introduced in 1965 [1], are well known and documented [2–5]. Nevertheless, R_a has worked well enough to remain in continuous use for nearly 50 years. R_a ’s limitations are especially pronounced when applied to highly-structured

spectral power distributions (SPDs)—that is, SPDs with sharp changes in slope, spikes, discontinuities, or some regions of smoothness and others that are spiky—including those produced by some solid state lighting (SSL) light sources such as light emitting diodes (LEDs). CIE TC1-62 *Colour Rendering of White LED Light Sources*, established in 2002, concluded that R_a cannot even correctly rank-order the color-rendering ability of light sources when LEDs are included [6], yet the committee did not recommend an alternative measure.

CIE TC1-69 *Colour Rendition by White Light Sources*, established in 2006, was formed “to investigate new methods for assessing the color rendition properties by white-light sources used for illumination, including solid-state light sources, with the goal of recommending new assessment procedures.” [7] At the CIE September 2012 meeting in Taipei, TC1-69 agreed to produce a technical report based on work to date, after which it will close. It is anticipated that the draft report will not make a single recommendation. It is expected to be a progress report that includes summaries from the different research groups that have developed their own approaches and measures for color rendition [8]. The membership of TC1-69 believe they need more time to verify the performance of the various models that have been developed before offering one solution for use by the lighting industry [9].

Two new TCs will continue where TC1-69 left off. TC1-90 *Colour Fidelity Index* was established “to evaluate available indices based on colour fidelity for assessing the colour quality of white-light sources with a goal of recommending a single colour fidelity index for industrial use”. TC1-91 *New Method for Evaluating the Colour Quality of White-Light Sources* was established “to evaluate available new methods for evaluating the colour quality of white-light sources with a goal of recommending new methods for industrial use. (Methods based on colour fidelity should not be included).” Both committees have been given four years to perform their work [8].

Meanwhile, industry and standards organizations have an exigent need for a measure (or measures) of color rendition that faithfully characterize the color quality of light sources, especially SSL light sources. The question is: Is there now enough information to recommend a solution that can faithfully and reliably serve the needs of the lighting industry? In this paper, we begin the first step of answering this question by reviewing the color measures that have already been put forth and to understand their similarities and differences as well as strengths and shortcomings.

This paper includes a summary of 22 indices for color rendition that now appear in the literature, including those considered by TC1-69 and that will continue to be evaluated by the new CIE TCs. This component of the paper is an extension and update of the 2004 review performed by Guo and Houser [10]. We then summarize our process for computing each of the 22 indices for a set of 401 SPDs, and present statistical analyses that quantify the similarities, differences, and interrelationships between the 22 indices, plus CCT.

Our objectives are: 1) Summarize the work that has already been done by others; 2) Quantitatively demonstrate that there are many commonalities among the existing measures, including evidence of how they cluster; 3) Contribute to the discussion regarding a two-measure system for characterizing color rendition.

2. Background

2.1 Existing measures of color rendition

The appendix provides a summary of the 22 indices that are considered in this paper. The reader is referred to the cited references for thorough treatment of the computational details. The indices fall into one of the three basic classes of color rendition: the accurate rendition of colors so that they appear as they would under familiar reference illuminants; rendition of objects such that they appear pleasant, vivid, or flattering; and the capability of an illuminant to allow an observer to distinguish between colors when viewed simultaneously. These dimensions of color rendition are respectively referred to as color *fidelity*, color *preference*, and color *discrimination* [11–15]. In column 3 of the appendix, we have classified the various

indices as “F”, “P”, and “D” based on our understanding of the authors’ original intent. The groupings are not based on numerical analyses and they should not prejudice one measure over another. Moreover, the classifications are not entirely independent. Gamut has sometimes been used as a proxy for both discrimination and preference, but gamut is an imperfect predictor of both. In the case of preference, for example, excessively large gamut can make object colors appear too saturated, which is neither natural nor preferred [12, 16]. In the case of either preference or discrimination, increases in saturation that lead to a larger gamut are almost always accompanied by hue shifts [17]. Thus, color discrimination and preference improvements associated with increased saturation may be offset by hue shifts. In the appendix, gamut-based indices have been denoted with either a “D” or “P” as per our understanding of the developer’s original intent. We also employed the letter “Q” as an abbreviation for “Quality” to denote that that index has been employed as something other than a pure measure of fidelity, preference, or discrimination.

As will be expanded upon in the discussion, there is little support to suggest that any single-number index can capture more than one dimension. It is not possible to simultaneously maximize fidelity, preference, and discrimination with any one illuminant because they have conflicting optimization criteria [10, 12–17].

2.2 SPDs and method of computing measures of color rendition

Table 1 provides a summary of the 401 SPDs employed in this study. These comprise 107 SPDs from the CQS programs provided by NIST [18, 19], 12 from CIE (i.e., F1 – F12) [20], 28 from Guo and Houser [10], 100 from Wei and Houser [21], 43 from the R_a 2012 program provided by Smet and Luo [22], 42 from the MCRI program provided by Smet [23], 4 from Wei and Houser [24], 4 from Royer, Houser, and Wilkerson [11], 20 from Houser and Gibbons [25], 4 from Soraa [26], 3 from the harmony rendering index (HRI) program provided by Szabó [27], and 20 additional SPDs that we digitized from other sources. We also computed and included 8 phases of blackbody radiation from 2,000 – 4,999 K and 6 CIE D illuminants from 5,000 – 8,000 K.

For the contemporary measures now being considered by CIE TC1-69, we contacted the respective authors and requested Excel spreadsheets that would allow us to compute their measures. Programs were generously provided, allowing us to compute CQS (Q_a , Q_f , Q_p , Q_g), MCRI, R_a 2012, RCRI, FCI94, and FCI02. An Excel spreadsheet previously developed by Guo and Houser [10] was employed to compute CCT, R_a , R_9 , R_aO , R_f , CPI, CDI, CRC84, CRC93, CSA, and PI. We wrote new code to compute FMG, FSCI, and GAI. All computations were performed using SPDs from 380 – 780 nm in 5 nm increments. If the SPD was not originally in that format, we employed derivative-constrained-spline interpolation. Extrapolation was never done; if the SPD did not extend to either 380 or 780 nm, then unreported values were set to zero. The computations yielded a data set from which we were able to perform statistical analyses of the interrelationships between the 22 indices. We also

Table 1. Summary of the SPDs employed in this study. Abbreviations are used in Figs. 2 and 6

Type of Illuminant	Counts and Abbreviations			
	Real Illuminants		Theoretical Models	
	Counts	Abbr.	Counts	Abbr.
LED Phosphor	130	LP-R	29	LP-T
LED Mixed	17	LM-R	51	LM-T
Fluorescent Broadband	30	FB-R	45	FL-T ^a
Fluorescent Narrowband	31	FN-R		
High-Intensity Discharge	31	HI-R	–	–
Tungsten Filament	17	TF-R	–	–
Blackbody Radiation	–	–	8	BB-T
D-Series Illuminants	–	–	6	DS-T
Other ^b	–	–	6	OT-T

^a Fluorescent models include broadband and narrowband

^b e.g., Equal-Energy, Clipped Incandescent, Ideal Prime Color

included CCT in our analyses since gamut area increases with CCT for CIE reference illuminants (i.e., blackbody radiation up to 5000 K and daylight phases at or above 5000 K).

For the 10 gamut-area measures that are computed with reference to the gamut of a fixed reference illuminant (i.e., FMG, CDI, GAI, CRC84, CRC93, CSA, PI, FCI94, FCI02, and FSCI), we also computed modified versions using the gamut of a reference illuminant with the same CCT as that of the test source, either blackbody radiation or a CIE D illuminant. We edited the source code in the Excel spreadsheets and Visual Basic for Applications (VBA) macros in order to perform these computations.

3. Results and discussion

3.1 Data distribution

The distribution of all 22 indices and CCT were checked for normality using measures of skewness and kurtosis.

Skewness measures the extent to which a distribution of values deviates from symmetry around the mean. A value of zero means that the distribution is symmetric. Positive skewness indicates a greater number of smaller values and negative skewness indicates a greater number of larger values. Computed values of skewness were mostly within an acceptable range of ± 2 .

Kurtosis is a measure of the peakedness or flatness of a distribution. A kurtosis value near zero indicates a shape close to normal; positive kurtosis values indicate a shape flatter than normal and negative values indicate distributions which are more peaked than normal. Kurtosis values of ± 2 are usually considered acceptable for employing statistics that are based on the assumption of normality. Computed values of kurtosis for the 23 distributions ranged from 0.02 to 30.8 with most greater than 2.

In the following data analyses, we employed statistical methods that do not require normality of the data distribution.

3.2 Correlation

We employed Spearman's correlation method to analyze correlations between the 22 indices. Spearman's coefficient measures the rank order of points and makes no assumptions about data distribution. Table 2 is a matrix of Spearman correlation coefficients arranged to highlight three clusters of highly correlated measures.

3.3 Multidimensional scaling

Multidimensional scaling (MDS) was employed to further assist in identifying clusters of similarity and to identify the underlying dimensionality. MDS uncovers the structure and relationships in a set of data by finding a representation of the indicators (i.e., color measures) in a low-dimensional space such that the matrix of Euclidean distances among the indicators corresponds as closely as possible to some characteristics of the input matrix. We generated input dissimilarities (i.e., distances) between every pair of measures using the multivariate raw score data (i.e., values of the 23 computed measures for each of the 401 SPDs). Since the various measures employ different scales, they were converted to z-scores prior to performing the MDS analysis. The output was a spatial map of color measures, each represented as a point in a low-dimensional space. The greater the dissimilarity between a pair of color measures, the further apart the points lie on the spatial map.

The low-dimensional space was optimized to minimize stress, a criterion function indicating the lack of correspondence between the distances among points in the MDS map and the input matrix. Values of matrix stress smaller than 0.2 indicate a good fit. R^2 values are the proportion of variance that is accounted for by the MDS model, with a maximum value of 1.0. Our two-dimensional MDS solution has a matrix stress value of 0.0731 and R^2 of 0.976, indicating that this solution is an excellent fit to these data. We also computed one-dimensional and three-dimensional MDS solutions. The one-dimensional solution had poor

Table 2. Matrix of Spearman Rank correlation coefficients that also illustrates blocks of similarity from the MDS scaling solution (see next section). The upper left shading in orange identifies a cluster that can be called “fidelity-based” measures, the middle shading in green identifies a cluster that can be called “preference-based” measures, and the lower right shading in blue identifies a cluster that can be called “gamut-based” (discrimination) measures. The date that each index appears in the literature is also provided. ** indicates that the correlation is significant at the 0.01 level and * at the 0.05 level (2-tailed).

	R _a O 1965	R _f 1967	R _a 1974	R _o 1974	Q _a 2010	Q _f 2010	RCRI 2010	R _a 12 2012	CPI 1974	FCI94 2007	FCI02 2007	Q _g 2010	Q _p 2009
R _a O	1.000												
R _f	.901**	1.000											
R _a	.946**	.875**	1.000										
R _o	.792**	.766**	.843**	1.000									
Q _a	.921**	.930**	.937**	.838**	1.000								
Q _f	.950**	.894**	.952**	.826**	.979**	1.000							
RCRI	.875**	.842**	.891**	.813**	.906**	.925**	1.000						
R _a 12	.815**	.750**	.830**	.787**	.885**	.883**	.837**	1.000					
CPI	.588**	.799**	.607**	.701**	.714**	.610**	.614**	.497**	1.000				
FCI94	.266**	.434**	.326**	.480**	.405**	.315**	.347**	.272**	.655**	1.000			
FCI02	.300**	.453**	.372**	.541**	.461**	.374**	.412**	.376**	.643**	.977**	1.000		
Q _g	.716**	.824**	.757**	.803**	.872**	.794**	.789**	.785**	.841**	.623**	.685**	1.000	
Q _p	.332**	.551**	.378**	.483**	.468**	.345**	.356**	.284**	.821**	.802**	.790**	.739**	1.000
MCRI	.581**	.723**	.646**	.746**	.751**	.685**	.748**	.696**	.778**	.637**	.710**	.891**	.638**
CDI	.016	.105*	.024	.100*	.061	.013	.038	.079	.300**	-.153**	-.131**	.255**	.272**
FMG	-.003	.0816	.000	.075	.034	-.011	.015	.050	.272**	-.189**	-.169**	.216**	.237**
CRC84	-.060	.0537	-.059	-.005	-.039	-.081	-.044	-.099*	.266**	-.175**	-.189**	.134**	.242**
CRC93	.044	.084	.053	.123*	.094	.071	.097	.195**	.179**	-.236**	-.186**	.243**	.116*
CSA	-.032	.0089	-.041	.0057	-.023	-.046	-.028	.0259	.135**	-.351**	-.326**	.102*	.0636
GAI	.016	.105*	.023	.099*	.061	.013	.037	.078	.300**	-.154**	-.133**	.254**	.272**
PI	.579**	.539**	.570**	.494**	.553**	.573**	.572**	.537**	.444**	.012	.032	.539**	.244**
FSCI	.319**	.204**	.322**	.253**	.307**	.332**	.263**	.489**	.009	-.382**	-.308**	.221**	-.125*
CCT	-.081	-.126*	-.110*	-.102*	-.134**	-.111*	-.102*	-.057	-.104*	-.564**	-.548**	-.126*	-.240**

	MCRI 2010	CDI 1972	FMG 1977	CRC84 1984	CRC93 1993	CSA 1997	GAI 2004*	PI 1986	FSCI 2004	CCT
MCRI	1.000									
CDI	.240**	1.000								
FMG	.209**	.996**	1.000							
CRC84	.145**	.943**	.953**	1.000						
CRC93	.259**	.945**	.946**	.874**	1.000					
CSA	.0978	.966**	.977**	.927**	.951**	1.000				
GAI	.238**	1.000**	.996**	.944**	.944**	.966**	1.000			
PI	.495**	.591**	.575**	.514**	.602**	.553**	.591**	1.000		
FSCI	.0765	.485**	.487**	.341**	.594**	.548**	.485**	.460**	1.000	
CCT	-.102*	.790**	.816**	.777**	.818**	.908**	.791**	.471**	.543**	1.000

values for stress and R^2 and the three-dimensional solution only marginally improved the fit. The final MDS solution is presented as Fig. 1. We also computed MDS solutions for real light sources only ($N = 263$), theoretical illuminants only ($N = 147$), and LED sources only ($N = 233$). The MDS solutions for all three sub-sets had similar stress and R^2 values and spatial configurations comparable to that shown in Fig. 1, indicating the stability of the two-dimensional solution.

Figure 1 illustrates that most indices clustered into one of three distinct neighborhoods, highlighted by orange, green, and blue bubbles. The color scheme and the clusters correspond with the regions identified in the correlation matrix of Table 2. Based on the measures in each neighborhood, we refer to the three clusters as fidelity-based (orange), gamut-based (blue), and preference-based (green). Those labeled as gamut-based are generally intended to

characterize color discrimination. There is subjectivity in forming the neighborhoods. CCT, for example, has a positive and relatively high Spearman Rank correlation coefficient with the measures that we have called gamut-based ($0.78 < R < 0.91$). We did not include CCT in that group because it lies some distance away and it is not intended to be a measure of color rendition. What we have called the preference-based neighborhood is also larger than the other two, suggesting that that group of measures employ more dissimilar computational methodologies than the other two groups, leading to higher heterogeneity. Q_g may appear to be out of place in what we have called the preference-based neighborhood since it is a measure that is (relative) gamut-based; this is discussed below.

Orthogonal axes have been superimposed on the spatial map and are suggestive of two dimensions that underlie these data. The horizontal dimension, which we have labeled “Color Discrimination”, is bounded on the left by what we have called fidelity-based measures and on the right by what we have called gamut-based measures. Color discrimination is operationally similar to gamut area and to the magnitude of color-shift vectors in chromaticity space. Thus, the plot reveals a tradeoff between fidelity and discrimination. The vertical dimension, which we have labeled “Color Preference”, is bounded at the upper-end by what we have called preference-based measures. Ignoring FSCI for the moment, which we discuss below, the lower-end is bounded by what we have labeled fidelity-based measures and CCT. The spatial map suggests that what we have labeled as gamut-based measures are generally high in color discrimination, what we have labeled as preference-based measures are generally high in color preference, and what we have labeled as fidelity-based measures are relatively low on both.

Consider NIST’s CQS indices Q_a , Q_p , Q_g , and Q_f as a way of interpreting the meaning of the dimensions. Q_f , a measure of fidelity, plots at nearly the same location as other fidelity-based test-color sample methods (e.g., R_a , R_aO , R_{a12}). Q_a plots above Q_f along the vertical dimension of color preference. This is appropriate because in the computation of Q_a , a test source that increases object chroma is not penalized (nor is it rewarded), whereas any shift is penalized in the computation of a pure fidelity measure such as Q_f . This was incorporated into the formulation of Q_a based on evidence that increases in object chroma are not detrimental to color quality as long as they are not excessive. Q_p , a measure of preference, plots further up the vertical dimension of color preference, and above Q_a . Q_p places additional weight on increases in object chroma and rewards a light source for that behavior. Note that Q_p plots at the lower portion of the preference neighborhood. This likely occurs because in the computation of Q_p , scores are rescaled so that the average score for the 12 reference fluorescent lamp spectra (i.e., CIE F1 – F12) is equivalent for Q_p and R_a [17]. Considering the color fidelity/discrimination dimension, Q_p plots further to the right of Q_a , and Q_g plots still further to the right. This reflects the fact that these indices progressively credit increases in gamut. Q_f plots slightly to the right of Q_a , though they are very close; this is reasonable since neither give credit for increasing gamut. Finally, Q_g , a measure of relative gamut area, plots well above Q_p on the color preference dimension and also falls within the color preference neighborhood. Q_g has a different scaling than Q_a , Q_f , and Q_p . It is normalized by the gamut area of a reference illuminant that is at the same CCT as the test illuminant. Values can be greater than 100, which can account for its position relative to Q_a , Q_f , and Q_p . Similar logic can be followed to understand the relationships between other measures on the spatial map.

To demonstrate the level of correlation between measures belonging to the same MDS cluster, Fig. 2 includes scatter plots that illustrates the correlation between R_a and Q_a and between R_a and R_{a12} . The greatest differences between R_a and Q_a appear when one of the two values is below 60. There does not appear to be marked differences between R_a and Q_a as a function of light source type. R_{a12} , in general, penalizes fluorescent light sources, especially those with R_a between 80 and 85. Nevertheless, the overall high correlations between measures within the same cluster suggest that the evaluation of spectra will be rather robust against the specific choice of a metric in a given MDS cluster. More pointedly, R_{a12} and Q_a show high correlation with R_a despite the many improvements and refinements in these newer formulations.

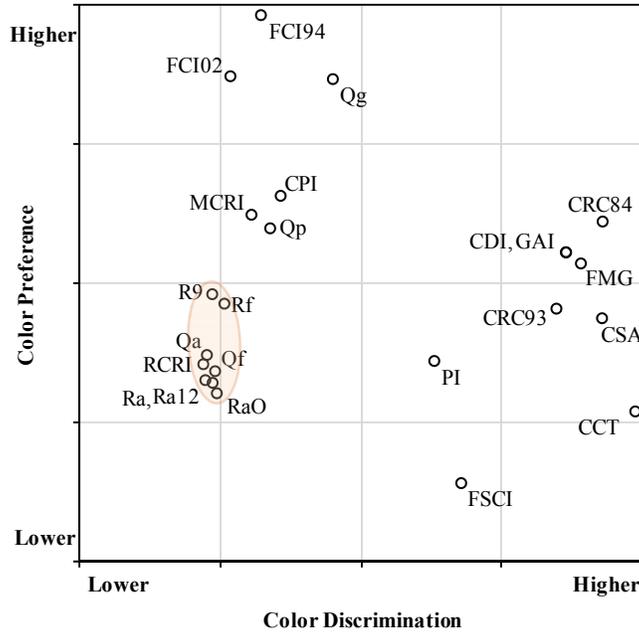


Fig. 1. Two-dimensional Euclidian-distance MDS solution for 22 color measures based on 401 SPDs.

Most gamut-based measures fall between the fidelity-based and preference-based measures on the color preference dimension. This is reasonable because increasing gamut, in comparison to the reference illuminants, frequently leads to an increase in preference [12, 16, 28–33]. However, the increases in saturation that are required to increase gamut are generally accompanied by hue shift that may not be preferable. Whereas larger gamut is always better for a gamut-based measure, a well constructed preference-based measure will account for the fact that oversaturation and hue shifts can make objects look unnatural and not-preferred, even when they increase gamut.

Not all indices may appear to cluster in the group that might have been expected based on the intent of the original developer. For example, Judd’s Flattery Index (R_f) was intended to characterize preference, but our data suggests that it actually resembles fidelity-based measures, raising questions about its efficacy in characterizing preference. In order to establish an R_f value of 90 for the reference illuminant and to employ the same constant as used in the computation of R_a (i.e., 4.6), Judd reduced the preferred color shifts to 1/5 of their experimental values. That decision effectively doomed the utility of the R_f to characterize preference and made it perform similarly to R_a . Thornton built upon Judd’s work when developing CPI and he preserved the full magnitude of preferred color shifts. CPI clusters in the neighborhood of preference-based measures.

Of the gamut measures plotted, Q_g may appear to be anomalous since it plots in the preference neighborhood. But Q_g (in version 9.0c of NIST’s CQS, which is what we have reported) is computed differently than all other gamut-based measures: it employs a variable reference at the same CCT as the test illuminant rather than employing the same reference for all test illuminants. Unlike the other gamut-based measures that cluster in the blue neighborhood at the right side of the spatial map, Q_g is not correlated with CCT (Spearman Rank correlation coefficient = -0.206). We discuss this further in the next section.

FSCI and PI do not cluster within any of the three identified groups. FSCI is computed differently than all other indices. It is based on a spectral-bands method that only considers the similarity of the test illuminant’s SPD to that of an equal-energy illuminant; it does not

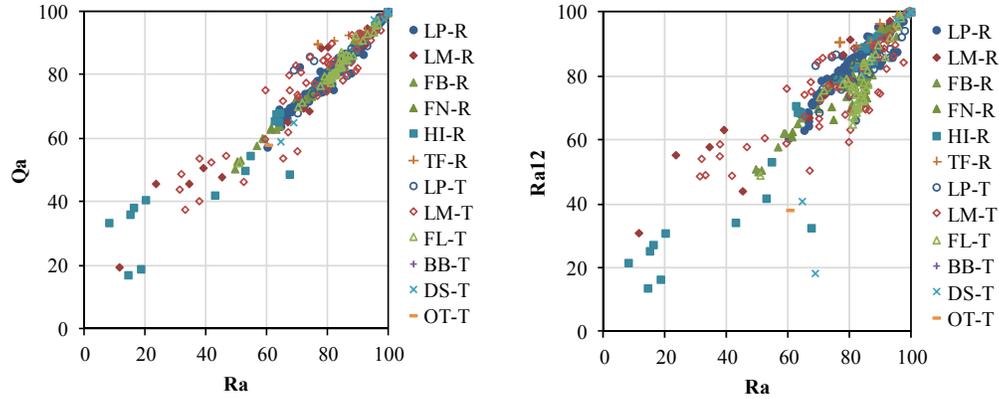


Fig. 2. Scatter plots of R_a vs. Q_a (left, Spearman $R^2 = 0.937$) and R_a vs. R_{a12} (right, Spearman $R^2 = 0.830$) for the 401 SPDs. Refer to the appendix for an explanation of the abbreviations.

consider test-sample colors or chromatic adaptation, nor does it have a reference that varies with CCT. It is not surprising that FSCI stands alone. In computing PI for the plot above, we always employed D65 as the reference illuminant, which could explain why it plots in the same general area as the measures that employ D65, CIE illuminant C, or equal-energy white as fixed references. Though not shown, we also computed PI for all 401 SPDs using a reference illuminant at the same CCT as the test illuminant. When that variable reference version of PI is plotted on the spatial map, it joins the fidelity cluster.

3.4 CCT correlations and renormalization

CCT plots far from the preference indices, suggesting that high CCT is not associated with high preference. Indeed, the correlation matrix of Table 2 shows a negative correlation between CCT and all of the indices that clustered into the preference neighborhood. CCT, however, is moderately correlated with the gamut-based measures, which is consistent with the fact that gamut tends to increase with CCT for CIE reference illuminants. As an example, Fig. 3 illustrates the correlation between CCT and GAI. GAI is favored by higher CCTs because gamut area increases with CCT [16]. Similar trends were observed between CCT and other fixed-reference gamut-based measures (e.g., CDI, CRC84, CRC93, CSA, FMG, PI, FSCI).

From a practical standpoint, it is not ideal for a measure of color rendition to be correlated with CCT. In the practice of lighting design it is common to first select CCT to set the overall color-tone of the environment. After CCT is selected, other color quality characteristics can then be evaluated, such as fidelity, discrimination, and preference. Most measures of rendition accommodate this process by using a reference illuminant of variable CCT (matched to the test spectrum), but this is not the case for measures based on gamut: all measures of gamut studied, with the exception of Q_g , are computed with reference to a high CCT illuminant such as D65, CIE Illuminant C, or equal-energy.

To remove dependence upon CCT for all measures that are tied to a reference illuminant at a fixed CCT, we recomputed these measures using a reference illuminant at the same CCT as the test illuminant. For the reference, we employed blackbody radiation below 5000 K and a phase of daylight at or above 5000 K. These modified measures are abbreviated as CDI_{VR} , $CRC84_{VR}$, $CRC93_{VR}$, CSA_{VR} , PI_{VR} , $FCI94_{VR}$, $FCI02_{VR}$, FMG_{VR} , and $FSCI_{VR}$, where “VR” stands for “variable reference”. Note that because CDI and GAI have a correlation of 1.0 and are essentially equivalent (i.e., they measure the gamut area enclosed by the eight test-color samples that are used in the computation of R_a , but using different reference illuminants), we did not compute a variable reference version of GAI. In effect, this means that the renormalized measures from the blue cluster of Fig. 1 employ a variable reference at the same CCT as the test illuminant. This is the procedure already used in the definition of Q_g . (Note: In [17, 18] Davis and Ohno computed Q_g with reference to D65. The change to a variable

reference was introduced in version 9.0 of NIST's CQS formulation [19]. To our knowledge, a description of this new formulation of Q_g has not heretofore appeared in the refereed literature. We only discovered NIST's new formulation after developing our own VR versions of the other gamut-based measures.) Fig. 4 provides the MDS spatial map for these data (Matrix Stress = 0.107, $R^2 = 0.961$). When CCT dependence is removed, three neighborhoods appear on the spatial map. The neighborhoods can be classified as fidelity-based (orange oval, left), those based on target chromaticity coordinates (blue oval, middle), and those based on relative gamut (green oval, right). $FSCI_{VR}$ and CCT do not cluster with any of these three groups, likely for the reasons previously discussed.

The horizontal axis is much more strongly pronounced than the vertical axis. We have labeled it "Relative Saturation", an aspect of color perception that relates to preference, discrimination, and gamut. Measures in the leftmost cluster, which are fidelity-based measures, have comparatively low relative saturation and high fidelity. Relative saturation increases for the three measures in the cluster that are based on target appearance, where the target chromaticities that underlie these measures are intended to relate to an aspect of preference. MCRI has fixed target chromaticities that represent a presumed ideal archetype. Q_p has target chromaticities on constant hue lines that allow for chroma enhancement. CPI has target chromaticities that allow for chroma enhancement in comparison to a reference illuminant. Relative saturation increases still more for measures in the rightmost cluster, which contains measures that are based on relative gamut. Q_g is the only measure in the rightmost cluster that is an existing measure of color rendition. All of the others are based on our computational modification where we employed a reference of variable CCT.

Because of chromatic adaptation, and because CCT is selected to set the overall color tone of an environment as part of the lighting design process, we believe that variable-reference measures are especially relevant to applied lighting design. If the relative gamut is greater than that of the reference, and illuminance is lower than that provided by daylight, then an increase in preference and discrimination might be expected relative to the reference at that same CCT. If the relative gamut is smaller than that of the reference, then a decrease in preference and discrimination might be expected relative to the reference at the same CCT.

The collapse of the three MDS clusters to one dimension (relative saturation) is especially interesting in view of the relative information contained in the various measures of color rendition. As already discussed and presented in Table 3, measures pertaining to the same cluster have high correlation. The left (fidelity) and right (relative gamut) clusters are most differentiated along the horizontal dimension; that is, they are least correlated along this

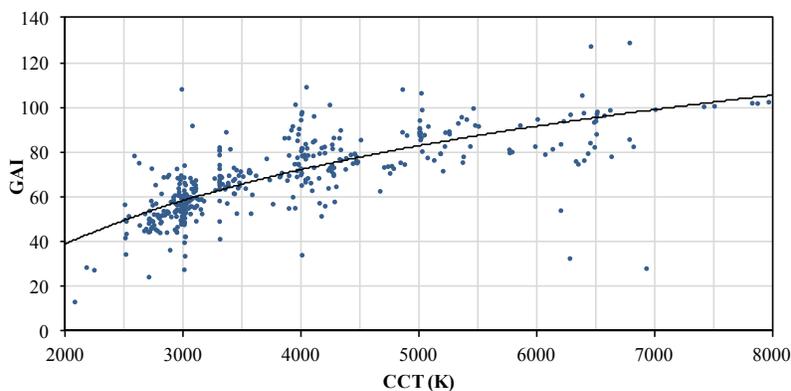


Fig. 3. Scatter plot of CCT vs. GAI illustrating that higher CCTs favor larger gamut areas. Spearman rank correlation coefficient between GAI and CCT, $R^2 = 0.791$. R^2 for the logarithmic fit shown as a black line is 0.574. The trend is similar for all measures of gamut area, including CDI, CRC84, CRC93, CSA, and FMG. Though not as pronounced as shown above, a positive trend with CCT also exists for $FSCI$ and PI.

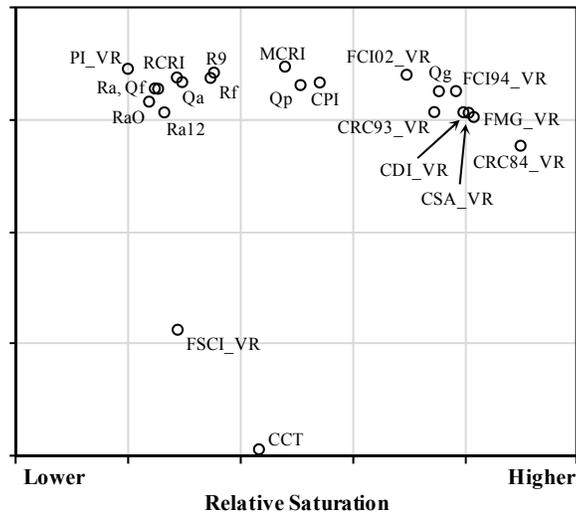


Fig. 4. Two-dimensional Euclidian-distance MDS solution for 22 measures based on 401 SPDs, where all measures have been computed using a reference illuminant at the same CCT as the test illuminant.

dimension. Thus, if a SPD is characterized with just two derived measures, a maximum amount of uncorrelated information will be retained if one measure comes from each of these two clusters.

4. Discussion of considerations for a two-dimensional color scale

Any proposal for a new measure of color rendition must be guided by practical and theoretical considerations. Like others [5, 10, 12, 17, 30, 34, 35], we believe that one number cannot fully encapsulate the multidimensional problem of color rendition. We also accept that the lighting industry needs a simple and readily interpretable tool for communicating color quality. To address these considerations, below we consider the concept of a two-dimensional scale that could be simplified into a one-word categorical rating. The framework builds upon the work of Guo and Houser [10], Rea and Freyssinier-Nova [30–32, 36], Figueiro and colleagues [37], and Dangol and others [38]. A one-word categorical rating has also been proposed by Bodrogi and his colleagues [39].

Guo and Houser performed a factor analysis based on 8 measures of color rendition for 34 illuminants, yielding two underlying factors that they labeled as reference-based (fidelity) and gamut-based. Guo and Houser suggested the use of one reference-based (fidelity) and one gamut-based measure when evaluating light sources for general illumination, with the caveat that they did not support using a blackbody reference below 5000 K [10]. Rea and Freyssinier-Nova created a set of 8 white-light illuminants that varied in R_a , GAI, and FSCI. They performed a series of psychophysical experiments related to color discrimination, vividness, and naturalness. Their central conclusion was that GAI should be used in conjunction with R_a and that neither measure is adequate on its own [30]. Rea and Freyssinier-Nova later made numerical recommendations, proposing that an illuminant should have both a R_a of between 80 and 100 and a GAI between 80 and 100 for spaces where color is important [31, 32]. Dangol and others performed psychophysical experiments at three different CCTs (2700, 4000, 6500 K) using lighting booths equipped with fluorescent and LED sources. They purposely adjusted the LED spectra to vary some measure of color rendition (i.e., Q_p , Q_g , FCI) while maintaining R_a at a value of 80. They concluded that people's judgments of naturalness and overall preference could not be predicted with a single measure, but required the joint use of a fidelity-based measure (e.g., Q_p) and a gamut-based measure (e.g., Q_g or GAI) [38]. Smet and his colleagues suggest that R_a and GAI can be combined into a single number that relates to naturalness [40].

In consideration of Rea and Freyssinier-Nova's proposal, Fig. 5 is a scatter plot of R_a versus GAI for the 401 SPDs in our data set, coded by CCT ranges. The plot illustrates that higher CCT light sources are strongly favored in simultaneously achieving high values of both R_a and GAI. For example, fourteen illuminants in our data set have values of R_a and GAI between 90 and 100; their range of CCT is 3921 to 7412 K with a mean of 5687 K.

There are five illuminants that have an R_a of between 90 and 100 and a GAI greater than 100: two phases of daylight at 7,500 and 8,000 K, two versions of the F40T12/C75 lamp (7821, 7867 K), and one theoretical fluorescent lamp (6378 K). We suggest that these illuminants could be excellent in some applications. Daylight, for example, is widely considered to be the standard by which other sources are judged for color rendition, and F40T12/C75 lamps have been successfully employed for decades in color critical applications. Yet, these lamps fail to meet Rea and Freyssinier-Nova's criteria because GAI exceeds 100. Recall that Rea and Freyssinier-Nova compute GAI with reference to an equal-energy spectrum, which has a CCT of 5455 K and a smaller gamut than that of these sources. These examples underscore a limitation of employing a fixed reference and simultaneously setting an upper limit on GAI, as has been done by Rea and Freyssinier-Nova. Some common phases of daylight with a CCT greater than that of EEW produce a gamut greater than EEW. Thus, those phases of daylight have GAI greater than 100 and would be considered unacceptable in the Rea and Freyssinier-Nova model. Even without specifying details—such as light level, the objects being illuminated, or the specific color rendering objectives—we suggest that daylight is an excellent illuminant for color rendition.

We support the development of a two-measure system. Below we offer considerations that we believe to be relevant when developing a system suitable for applied lighting.

1. Test-Color Samples: Davis and Ohno showed that light sources can perform poorly with saturated test-color samples even when they perform well with the 8 desaturated test-color samples employed in the computation of R_a and GAI. Computations by Davis and Ohno suggest that the inverse is never true [17]. Smet and his colleagues suggest a set of imaginary test-color samples that span the visible spectrum [41]. The fundamental rationale is that steep slopes and sharp changes in slope (within a test-color sample or an illuminant) are more likely to lead to color shifts than gentle slopes and gentle changes in slope (within a test-color sample or an illuminant). Therefore, it is more difficult for an illuminant to minimize color shifts when the test-color samples have steep slopes, and steep slopes in the spectral

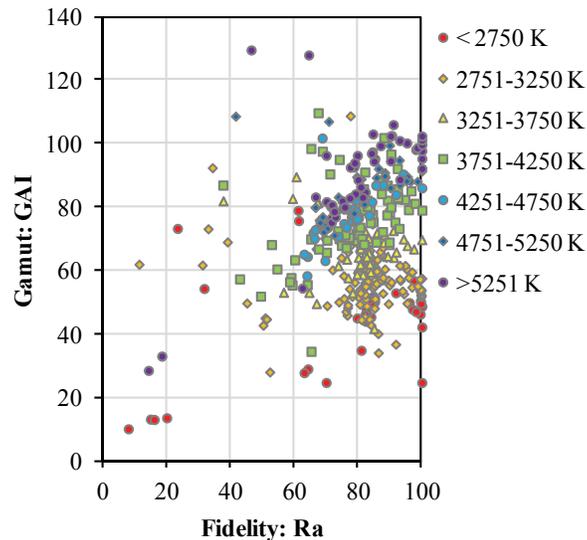


Fig. 5. Scatter plot of R_a vs. GAI for the 401 SPDs binned into ranges of CCT.

reflectance distribution of a test-sample color are characteristic of more saturated colors. It follows that more saturated test color samples provide a more difficult test of a light source's color rendition ability. Regardless of the specific approach adopted, there is evidence that test-color samples should not be desaturated.

2. Consideration of Illuminance: Daylight is an outstanding light source that renders colors naturally and with high fidelity. It has been employed as the reference for most gamut-based indices. Yet, a person's experience and reference for daylight is outdoors, where illuminance can be as high as 100,000 lx. Daytime outdoor illuminance is almost always greater than the 50 – 1000 lx that is typical of indoor illuminance from electric light sources. The magnitude of illuminance strongly affects the appearance of colored objects. Perceived hues are dependent upon illuminance (Bezold-Brucke effect) [42], colors appear more saturated under higher illuminance (Hunt effect) [42], and color discrimination performance is dependent upon illuminance level [30, 43].

If an electric light source increases object saturation relative to a reference illuminant at a typical indoor illuminance level, then the object may appear more like it would under daylight at a typical outdoor daytime illuminance level. Nevertheless, it may not be practical to consider illuminance directly within a color rendition system. Most measures of color rendition are based on relative colorimetry, which ignores illuminance level. In our computations, only MCRI considered illuminance level or the degree of adaptation. Yet, given that color quality measure will almost always be employed at illuminance levels much less than that provided by daylight, a source should not be harshly penalized (if at all) for increases in gamut. As a practical matter, no measure that is based only on chromaticity will be able to predict color appearance.

3. Consideration of CCT: Lower CCT light sources can have excellent color qualities, including excellent color-discrimination performance despite having a smaller gamut [11, 30]. This characteristic of human perception is not captured when one of the dimensions of a two-measure system is pegged to a particular CCT, as occurs when a single reference is employed for a gamut-based measure. A two-measure system should reflect quality in lighting applications, which includes a consideration of CCT to set the overall color tone of the environment. We do not believe that it is appropriate for a two-measure system that is intended to characterize color quality to strongly favor higher CCT illuminants. This occurs because most two-measure proposals include a measure of gamut area, where gamut area is normalized to a reference with a relatively high CCT and relatively large gamut. There are alternate approaches. One would be to use references of different gamut areas at a limited number of CCTs, perhaps aligning with ANSI bins [44]. A second approach would be to eliminate the use of reference illuminants and instead use target chromaticity coordinates [12, 16, 45], possibly using different targets for different CCTs. A third approach is for both dimensions of a two-measure system to be based on comparison to a reference illuminant at the same CCT. The later approach is what we explore below, employing "VR" normalized gamut indices in conjunction with a measure based on fidelity.

4. Categorical Definitions: Rea and Freyssinier-Nova provided only two categories in their two-dimensional plot of fidelity and gamut: acceptable and unacceptable. We build on their basic idea [32], and that offered by Bodrogi and his colleagues [39], by suggesting the use of word-categories to define regions of a two-dimensional plot. Numerical regions can be defined to represent excellent, good, fair, and poor color quality. We believe that a simple word scale has the potential to capture overall color quality, would be especially useful to end-users, and could be considered for consumer packaging.

5. Choice of Measures: When more than one measure is employed for evaluating color quality, they should reflect two salient and scalable aspects of color rendition that have relatively low correlation with each other. Table 1 shows high correlation among three groups of indices, Fig. 1 illustrates that most existing measures cluster into one of three neighborhoods, and Fig. 4 illustrates that when CCT dependence is removed the measures essentially collapse onto one dimension. When considering the paired use of two indices, the two indices should be selected from opposite ends of the horizontal axis shown in Fig. 4. We

developed a spreadsheet tool that plots the relationship between any two user-selected indices to assist in evaluating candidate measures. After studying all paired combinations, and considering the above criteria, we tentatively suggest Q_a to represent the fidelity neighborhood and Q_g to represent the preference neighborhood. A plot is provided as Fig. 6. We initially expected to select more pure measures of fidelity and gamut, essentially updating and confirming the proposal of Rea and Freyssinier-Nova [32]. van der Burgt and van Kemenade also suggest that a pure fidelity measure should be one component of a two-measure system [46]. A pure fidelity measure, however, penalizes all color shifts and may incorrectly penalize some illuminants for favorable increases in chroma. We believe that Q_a is more reflective of color quality in application because it does not penalize illuminants for chroma increases. Traditional gamut measures are inappropriate because of their dependence upon CCT. Q_g was selected because it is an existing measure of relative gamut and because it shares some of the same computational framework as Q_a , such as test-sample colors. We believe that each of the two dimensions should have individual meaning and predict a criterion of color rendition or a hybrid criterion. Figure 6 takes a hybrid-criterion approach by commingling Fidelity / Quality / Naturalness on the horizontal axis and Preference / Discrimination on the vertical axis. We believe that this is reflective of how light sources perform, how people perceive color, and how designers select light sources for applied lighting.

Rather than indiscriminately computing color differences, the Q_a computation does not penalize (nor does it reward) an illuminant for increasing object chroma. Hue shifts are penalized, as are chroma shifts that desaturated object colors. For example, and as illustrated

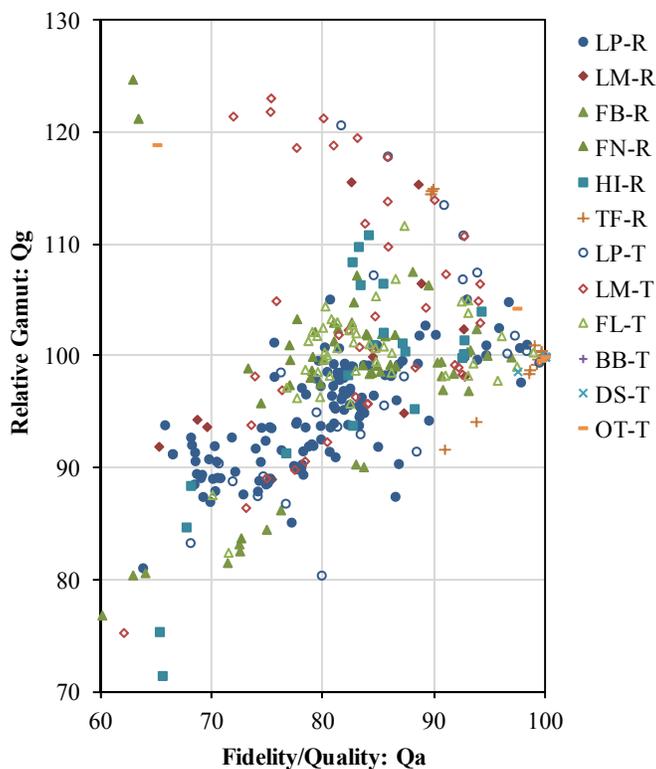


Fig. 6. Plot of Q_a vs. Q_g . The horizontal axis is related to fidelity and is a proxy for quality or naturalness when used for general illumination. The vertical axis is a measure of relative gamut and is a proxy for preference and discrimination. Refer to Table 2 for an explanation of the abbreviations.

in Fig. 6, neodymium lamps are rated highly (see the three orange + symbols with $Q_a \approx 90$ and $Q_g \approx 115$). They have an especially large gamut relative to a blackbody at the same CCT and that leads to a high score for Q_g . But the increase in gamut is accompanied by a hue shift, which lowers the score for Q_a .

The scatter plot of data points form a triangular pattern, converging near 100 for both Q_a and Q_g . This occurs because all sources with $Q_a = 100$ will have a value of Q_g near 100, while lower values of Q_a allow for a much wider range of Q_g . These measures provide useful information individually, are mutually complimentary, and limit each other when considered together.

As a next step, we believe that the lighting community should develop a two-measure system color rendition and we will enthusiastically participate in that work. In our opinion, the first goal of a two-measure system should be to communicate the maximal amount of information about a light-source's color rendition potential. It must be readily interpretable by design professionals and formulated in such a way that it can be simplified even further into grades, classes, or words that would be understood by the general public. We do not believe that it is necessary for a new system to incorporate existing measures, but we also see no reason to invent new measures if what already exists can be intelligently combined into a two-measure system. While we cannot predict the future, we can imagine SPDs becoming even more discontinuous and spiky, just as we can imagine paints, plastics, and other materials with highly structured spectral reflectance distributions. This is our rationale for moving away from measures that rely on CIE test-color samples 1 – 8. It is also desirable to employ the same set of test-color samples for both measures. We believe that each of the two dimensions should have individual meaning and predict a criterion of color rendition or a hybrid criterion. Figure 6 takes a hybrid-criterion approach by commingling Fidelity/Quality on the horizontal axis and Preference/Discrimination on the vertical axis. We believe that this is reflective of how light sources perform and of how people perceive color. The straw-man of Fig. 6 is readily interpretable, providing information that can be helpful to expert users. If regions of the two-dimensional space of Fig. 6 were to be defined with words such as excellent, good, fair, and poor, these words could have meaning to non-experts, including the general public.

Table 3 list details about some of the 401 illuminants in our data set, including numerical scores for CCT, R_a , Q_f , and Q_a .

5. Conclusions

While we appreciate and respect the considerable work that has gone in to developing new and improved measures of color rendition, especially as part of CIE TC 1-69, the above analyses suggest that the newer indices are not remarkably different from the older ones. Many of the newer measures have stronger theoretical underpinnings, for example by employing improved test-color samples and the latest CIE color appearance models, chromatic adaptation models, and/or colors spaces. Nevertheless, when the output of the computations is a single number, frequently on a scale of 0 – 100, these improved computational engines yield results that are highly similar to longstanding measures that were based on cruder underlying models. The basic problems of color rendition—fidelity, discrimination, and preference—are well established, not subjects of debate, and are clearly revealed in our correlation and MDS analyses. In our assessment, the newer measures that have been recently proposed do not represent reconceptualizations of the basic dimensions that define quality white light. The improvements are at the margins.

Unless new aspects are brought to bear, we believe that the work of the new CIE TCs is unlikely to lead to an index or indices that are unequivocally superior to those that already exist. We question the pragmatism of asking the lighting community to be patient while two new CIE committees wrestle with details that may be lost when complex computations are distilled to an integer scale. We hope that these new committees will reach consensus, and that their recommendations will be mutually supportive, but is it prudent to ask the lighting industry to continue to wait?

Table 3. Performance of some of the 401 illuminants studied.

Source Type	Description	CCT	R _a	Q _a	Q _g
Tungsten Filament	Sylvania Tru Aim MR16 Halogen	2776	100	100	100
Tungsten Filament	Philips 75W Halogena	2836	100	100	100
LED Phosphor Real	Xicato XSM Artist 3000 K	2940	97	98	101
LED Phosphor Real	Soraa Vivid MR16 2700 K	2724	96	95	100
LED Phosphor Real	Soraa Vivid MR16 3000 K	2969	96	95	101
Fluorescent Broadband	CIE F8	4997	96	97	100
Fluorescent Broadband	F32T8/TL930	2908	95	94	102
Fluorescent Broadband	F40/C75	7412	93	95	100
HID	MHC100UMP4K	4256	92	93	101
Tungsten Filament	Solux halogen	4144	90	91	92
LED Phosphor Real	Xicato XSM 80 3000 K	2496	88	81	105
Fluorescent Broadband	F40/CWX	4030	87	85	99
HID	White SON HPS	2760	87	85	107
Fluorescent Narrowband	F32T8/TL850	5072	86	86	102
Fluorescent Narrowband	F32T8/TL835	3480	86	85	102
LED Mixed Real	GE SoftWhite LED	2976	86	84	100
Fluorescent Narrowband	F32T8TL830	2940	85	84	102
Fluorescent Narrowband	F32T8TL835	3700	85	84	102
Fluorescent Narrowband	F32T8TL841	4194	83	84	99
LED Phosphor Real	Philips EnduraLED 10W MR16 2700K	2789	82	82	97
LED Phosphor Real	Soraa Premium MR 16 3000 K	3005	82	82	96
LED Phosphor Real	Philips EnduraLED 10W MR16 3000K	3167	81	82	96
LED Phosphor Real	Soraa Premium MR16 2700 K	2708	80	81	94
Tungsten Filament	Neodiumium Incandescent	2757	77	90	115
Fluorescent Narrowband	F32T8/TL741	4663	70	73	84
HID	H38JA100DX	4037	68	49	110
HID	C100S54C	2171	64	68	88
Fluorescent Broadband	Cool White FL	4290	63	63	80
HID	Topanga Plasma 5500 K	6197	63	66	72
HID	MH100W	3923	55	55	77
Fluorescent Broadband	F34T12/LW/RS/EW	4165	50	50	72

Against this backdrop, common sense suggests that it is not possible for any single index to fully encapsulate the multidimensional problem of color rendition. For example, if a light source is to be optimized for a measure of color preference, then it's necessary to make some colors appear more saturated than they would under a reference illuminant [12, 16, 28, 29]. Thus, there is an intrinsic tradeoff between measures of fidelity and preference. If a light source is to be optimized for a measure of gamut area, then the goal will be to create colors that are as saturated as possible. Maximizing a gamut-based measure will come at the expense of fidelity and preference since extreme color saturation is neither preferred nor natural.

As a single number index, R_a has fulfilled the practical need for a measure that is simple to understand and readily interpretable. Its utility is proven despite its limitations and shortcomings. Yet, as light sources have become more spectrally complex, R_a has been pushed beyond its limits. We believe that any single-number index that has been or will be proposed as a replacement for R_a will still suffer from a fundamental problem: one number is not enough to characterize all dimensions of color rendition and one number cannot faithfully summarize color quality.

While an expert user might be comfortable evaluating multiple indices, it is still necessary to condense information into a limited number of measures. When just two measures are used, the maximal amount of information that is relevant to applied lighting is retained if one of them is a fidelity-based measure that is consistent with the concept of color fidelity or quality (such as Q_a), and the other is a measure of relative-gamut (such as Q_g). To meet the needs of some users, such as residential homeowners, even multiple measures may be further compressed into a single scale, such as a word scale.

Appendix: summary descriptions of measures for color rendition

Measure	Abbrev.	Type	Description	Ref.
Former CIE Test Color Method (1965)	R_aO	F	The 1965 version of R_a , now deprecated, based on a different chromatic adaptation model and somewhat different chromaticity-space computations. The range is the same as with the current version of R_a .	[1]
Flattery Index (1967)	R_f	P	Accounts for desirable shifts in hue and saturation by considering an 'ideal' configuration of chromaticity coordinates in the CIE 1960 UCS. Ten test-color samples are considered with unequal weighting. Like R_a , R_f has a reference illuminant at the same CCT. The maximum value is 100 and the reference has a value of 90.	[12]
CIE Test Colour Method (1974)	CRI, R_a	F	Scale that represents the mean resultant color shift of eight test-color samples under an illuminant in comparison to a reference illuminant of the same CCT in the CIE 1960 UCS. The maximum value is 100 and the reference always has a value of 100, irrespective of CCT. Negative values are possible.	[47]
Special Index 9 of CIE Test Colour Method (1974)	R_9	F, Q	A special index of CRI that characterizes the resultant color shift of saturated red. The maximum value is 100 and the value of R_9 for the reference illuminant is always 100, irrespective of CCT. Negative values are possible. It has sometimes been considered as a proxy for color quality, though there is not a strong theoretical basis for doing so.	[47]
Color Quality Scale version 9.0c (2012)	CQS, Q_a	F, Q	Q_a is a scale of 0 – 100 that maintains a similar computational structure as R_a while employing fundamental improvements: a better chromatic adaptation model (CMCCAT2000), 15 saturated test-color samples, illuminants are not penalized for increases in chroma (i.e., a saturation factor is incorporated), computations are performed in the CIELAB color space, color differences are combined with a root mean square, and sources with extremely low CCTs are penalized because they have smaller gamut areas. It is scaled so that 12 reference fluorescent lamp spectra have equivalent values of Q_a and R_a . Like R_a , all CQS indices (i.e., Q_a , Q_f , Q_p , Q_g) are based on comparison to a reference illuminant at the same CCT. Note that the reference cited is for an earlier version of CQS that employed a somewhat different formulation that described in this paragraph and employed in this paper (e.g., seven of the 15 test-color samples have been changed, the CCT factor was removed, and Q_g in v9.0 is scaled based on a reference illuminant with equal CCT).	[17, 19]
Color Fidelity Scale of CQS version 9.0c (2012)	Q_f	F	Computed the same way as Q_a except the saturation factor is excluded. It has a similar function to R_a . Scaled from 0 – 100 and so that 12 reference fluorescent lamp spectra have equivalent values of Q_f and R_a .	[17, 19]
Rank-Order Color Rendering Index (2010)	RCRI	F	This ordinal rating scale is computed using CIE CAM02 formulae and 17 test-color samples. A reference illuminant is defined in the same manner as employed for R_a . The RCRI scale is defined with reference to a predicted number of excellent and good ratings of the test-sample colors, which were empirically determined. RCRI has a range of 0 – 100.	[39]
CRI2012 Colour Rendering Index (2012)	R_a2012 , nCRI, R_a12	F	R_a2012 is a scale of 0 – 100 that maintains a similar computational structure as R_a while employing fundamental improvements: computations are done using CIE CAM02-UCS, an (imaginary) set of test-color samples was developed and optimized, and color differences are combined with a root mean square. Like Q_a , it is scaled so that 12 reference fluorescent lamp spectra have equivalent values of R_a2012 and R_a . Like Q_a , it also employs the same reference illuminants as R_a .	[41]

Color Preference Index (1974)	CPI	P	Conceptually similar to R_f in that it credits illuminants for rendering an array of test-color samples in desirable ways, as defined by an ‘ideal’ configuration of chromaticity coordinates in the CIE 1960 UCS. It equally weights the 8 test-colors that contribute to the index. Like R_a , CPI has a reference illuminant at the same CCT. The maximum value is 156 and the reference has a value of 100.	[16]
Feeling of Contrast Index: CIE CAT94 (2007)	FCI94	P	FCI is derived from a transformation of gamut area formed by a four-color combination of red, green, blue, and yellow in CIELAB. D65 is always employed as the reference illuminant and CIECAT94 is employed for chromatic adaptation. FCI corresponds to a ratio of D65 illumination to the test illuminant illumination for equal “feeling of contrast”, which is probably most closely related to color preference. D65 has a FCI94 of 100 and values higher and lower are possible; the practical range is about 20 – 200.	[48]
Feeling of Contrast Index: CIE CAM02 (2007)	FCI02	P	FCI02 is conceptually similar to FCI94 except that it is computed using the CIECAM02 color appearance model and CIECAT02 chromatic adaptation formulae. D65 has a value of 100; the practical range is about 20 – 150.	[48]
Relative Gamut Area Scale of CQS version 9.0c (2012)	Q_g	–	Computed as relative gamut area formed by the (a^* , b^*) coordinates of the 15 test-color samples in CIELAB normalized by the gamut area of a reference illuminant at the same CCT and multiplied by 100. Scaling is different from Q_a , Q_f , and Q_p and can be greater than 100. Q_g does not employ a chromatic adaptation transform. See comments above under Q_a regarding the difference between Q_g v7.5 and Q_g v9.0c.	[17, 19]
Color Preference Scale of CQS version 7.5 (2009)	Q_p	P	This index places additional weight on preference of object color appearance based on the idea that increases in chroma are generally preferred. It is scaled from 0 – 100 and so that 12 reference fluorescent lamp spectra have equivalent values of Q_p and R_a . Q_p was dropped from CQS version 9.0 with the belief that additional visual experiments are needed before Q_p can be placed into practice.	[17, 18]
Memory Color Rendering Index (2010)	MCRI	P, Q	MCRI is based on observers’ memory of the preferred color of 10 familiar objects (e.g., fruits, flowers, skin, neutral grey). There is no reference illuminant; the reference is color memory. Tristimulus values for the objects are transformed to corresponding colors under D65 illumination using CIECAT02 and then transformed to the IPT color space, where MCRI is computed. MCRI has a range of 0 – 100. The result is also affected by the degree of adaptation and illuminance.	[45]
Color Discrimination Index (1972)	CDI	D	A higher CDI is associated with a larger gamut in the CIE 1960 UCS chromaticity diagram. The gamut is normalized to 100 based on CIE illuminant C. The practical range is about 10 – 130.	[49]
Farnsworth-Munsell Gamut (1977)	FMG	D	The area enclosed by a line joining the positions of all 85 test-color samples of the Farnsworth-Munsell 100 Hue Test is computed in the CIE 1960 UCS. The index is then normalized to 100 based on CIE illuminant C. Values greater than 100 are possible.	[43]
Color Rendering Capacity (1984)	CRC84	D, Q	A scale of color rendering potential based on the number of object colors that an illuminant can theoretically render. This 1984 version is based on computation in the CIE 1960 (u , v , Y) space. CRC84 has a theoretical range of 0.0 – 1.0, though only a contrived source would fall outside the range of 0.15 – 0.40.	[50]
Color Rendering Capacity (1993)	CRC93	D, Q	This 1993 update yields an index related to the volume of a color solid computed in CIELUV. It is calculated as a ratio of the color solid volume obtained under a test illuminant to that obtained with an equal-energy spectrum. The minimum value is 0.0 and the maximum value can exceed 1.0.	[51]
Cone Surface Area (1997)	CSA	D, Q	The base of a cone is formed using the gamut of the eight CRI test-color samples within the CIE 1976 UCS diagram (u' , v') and the height (w') is determined from the chromaticity of the test illuminant. In this way, CSA combines a measure of gamut with source chromaticity. CSA is not reliant upon, or normalized to, a reference illuminant.	[52]

Gamut Area Index (2004)	GAI	D	Gamut area is defined as the area enclosed by the polygon created by the eight test-color samples used in the R_a calculation within the CIE 1976 UCS diagram (u' , v'). The gamut area for an equal-energy reference spectrum is assigned a GAI of 100. Other illuminants are normalized to this value, yielding either higher or lower values of GAI. The practical range is about 10 – 130. The authors have not suggested using this alone, and have only suggested using this in conjunction with R_a .	[30]
Full Spectrum Color Index (2004)	FSCI	F	FSCI is a measure of how much a light source's SPD deviates from an equal-energy spectrum. It is scaled so that the equal-energy reference receives a score of 100 and warm white fluorescent receives a score of 50. Negative values are set to zero. It is intended to be a measure of fidelity. It is a linear transformation from the Full-Spectrum Index (FSI).	[30]
Pointer's Index (1986)	PI	F, Q	Pointer's full index has 16 sub-indices related to hue, lightness, and chroma for red, yellow, green and blue. The sub-indices can be combined to produce an overall index, which is what we have included in this study. Computations are done using Hunt's 1982 color appearance model. The test-sample colors are the 18 color samples in the Macbeth color checker. Any illuminant can be employed as the reference; in this work we have always used D65. The scale is 0 – 100. Pointer revisited his method in 2004 [54] to take advantage of the latest CIE recommendations on color-appearance models, but we performed computations based on his 1986 method.	[53]

Acknowledgments

We gratefully acknowledge Soraa for funding this work. We thank our professional colleagues for sharing their SPDs and computational spreadsheets: Wendy Davis, Xin Guo, Ronnier Luo, Yoshi Ohno, Michael Royer, Kevin Smet, Ferenc Szabó, and Andrea Wilkerson.